AMÉLIORER LA SÉCURITÉ DE L'APPRENTISSAGE FÉDÉRÉ AVEC LE CHIFFREMENT POLYMORPHIQUE et HOMOMORPHIQUE

Thèse présentée

dans le cadre du programme de doctorat en ingénierie de l'UQAC

offert par extension à l'UQAR

en vue de l'obtention du grade de Philosophiae doctor (Ph.D)

PAR

© Mohammad Moshawrab

Juillet 2024

Composition du jury:

Pr. Chan-Wang Park, président du jury, Université du Québec à Rimouski (UQAR)
Pr. Mehdi Adda, directeur de recherche, Université du Québec à Rimouski (UQAR)
Pr. Abdenour Bouzouane, codirecteur de recherche, Université du Québec à Chicoutimi (UQAC)
Dr. Nacim Ihaddadene, membre externe, JUNIA, école d'ingénieurs
Pr. Djamal Rebaine, membre interne, Université du Québec à Chicoutimi (UQAC)

Dépôt initial le 28 Mars 2024 Dépôt final le 16 Juillet 2024

À vous, ceux qui sont toujours de mon côté,
même si je suis loin parfois

REMERCIEMENTS

En ce moment de réflexion, je souhaite exprimer ma sincère gratitude aux personnes qui ont contribué à façonner ma carrière universitaire. Si cet espace n'est pas dédié à l'expression de l'amour, il est certainement un lieu de profonde gratitude.

J'exprime ma profonde gratitude à mon tuteur académique et superviseur, le professeur Mehdi Adda, pour ses conseils constants et sa confiance inconditionnelle en mes capacités. Ses efforts dévoués ont non seulement rendu ce moment possible, mais aussi vraiment exceptionnel.

Je tiens également à souligner les précieuses contributions de mon co-superviseur, le professeur Abdenour Bouzouanne, et le soutien permanent du Dr. Hussein Ibrahim, qui a été un excellent collaborateur, depuis le début de ce voyage.

Je tiens à remercier tout particulièrement le Dr Ali Raad, doyen de la Faculté des arts et des sciences de l'Université islamique du Liban, dont le dévouement constant tout au long de mes études doctorales a été vraiment appréciable.

Tous ceux qui ont joué un rôle dans cette réalisation et qui ont offert leur soutien, j'adresse mes salutations les plus chaleureuses, mes remerciements les plus sincères et ma profonde gratitude. Vos contributions collectives ont été le moteur de ma réussite.

Je suis profondément redevable à mes parents, dont la générosité, matérielle qu'émotionnelle, m'a propulsé dans la position dans laquelle je me trouve aujourd'hui. Leurs sacrifices et leur amour ont été la fondation de mon parcours jusqu'à présent. À mes chers frères, Kassim et Haidar, je dois une grande reconnaissance pour leur soutien constant à chaque instant. À mes chères sœurs, Enass et Zainab, vous tenez une place spéciale dans mon cœur, et votre présence me sera toujours précieuse. Aux membres de ma famille bien-aimée, Evan, Abbas, Ali, Ahmad et Fatima, j'exprime mon amour et mon appréciation les plus sincères.

Finalement, mon épouse bien-aimée, Aline, a été mon soutien indéfectible, partageant mes rêves, ma vie et mes aspirations. Sa résilience et son amour ont été mon guide, et pour cela, elle a toute ma gratitude.

RÉSUMÉ

Depuis sa création, l'Intelligence Artificielle (IA) a toujours été un point focal de la recherche, équipant les machines, y compris les ordinateurs et autres, de la capacité d'agir de manière autonome, en émulant ainsi l'intelligence humaine. La sous-domaines de l'IA, plus particulièrement l'apprentissage machine (ML) et l'apprentissage profond (DL), ont fait preuve d'une efficacité remarquable dans la résolution des tâches d'analyse des données. L'apprentissage machine, caractérisé par un ensemble d'algorithmes permettant aux ordinateurs d'apprendre à partir de données d'entraînement sans programmation explicite, a été largement adopté dans l'industrie, les soins de santé, les transports, l'éducation, le commerce électronique et divers autres secteurs. Cette adoption a été catalysée par son aptitude à découvrir des motifs dans les données, à apprendre d'eux, et à faire des prédictions en conséquence.

La croissance de l'apprentissage machine a été facilitée par les progrès des techniques informatiques. Ces progrès ont permis aux spécialistes de l'apprentissage automatique d'analyser des ensembles de données plus importants et de s'attaquer à des problèmes plus complexes, repoussant ainsi les limites de ce qui peut être réalisé. Néanmoins, le domaine reste confronté à une série de défis que l'on peut classer en quatre grandes catégories : les défis liés aux données, les défis liés aux modèles, les défis liés à la mise en œuvre et les défis généraux. Parmi ceux-ci, les questions primordiales de la vie privée et de la sécurité entrent dans la catégorie des défis généraux. D'une part, les modèles de ML restent vulnérables à un spectre d'attaques et de menaces, quels que soient les mécanismes de sécurité mis en œuvre. D'autre part, les préoccupations en matière de protection de la vie privée ont donné naissance à un cadre réglementaire qui restreint l'accès aux données, limitant ainsi les performances des modèles de ML. Il s'agit là d'un obstacle important, car l'efficacité des modèles intelligents est souvent proportionnelle à leur capacité à accéder à des ensembles de données diversifiés et complets. En réponse à ces défis de sécurité et de confidentialité, Google a introduit l'apprentissage machine fédéré, également connu sous le nom de Federated Learning (FL). Le FL a été initialement conçu comme une approche de ML préservant la vie privée, car il élimine le

besoin de centraliser les données des utilisateurs pour l'apprentissage des modèles. À la place, les modèles sont distribués aux clients participants, formés localement, puis agrégés par le serveur pour générer un modèle global. Dans ce contexte, un algorithme d'agrégation d'apprentissage fédéré est défini comme le mécanisme utilisé par le serveur pour envoyer des modèles aux clients impliqués dans le cycle d'FL, recevoir les modèles entraînés de leur part et les combiner en un seul modèle global entraîné. Bien que le FL soit prometteur en matière de protection de la vie privée, il reste sensible aux menaces de sécurité. Les chercheurs étudient activement des méthodes pour sécuriser le FL contre diverses attaques, notamment les attaques byzantines, les attaques par inversion, les attaques par porte dérobée, etc., en mettant en œuvre des mécanismes tels que le chiffrement homomorphe, le calcul multipartite sécurisé, la méthode des multiplicateurs à sens alternatif et bien d'autres encore. Bien que des progrès considérables aient été réalisés dans le renforcement des algorithmes FL contre certaines attaques, des vulnérabilités telles que l'attaque par inversion persistent, permettant à des entités malveillantes de discerner les données des utilisateurs contenues dans les modèles entraînés. Cela souligne le besoin pressant de faire progresser les mesures de sécurité dans le domaine du FL.

Motivée par l'impératif de renforcer l'apprentissage fédéré contre une multitude d'attaques et reconnaissant le potentiel du chiffrement polymorphe et homomorphe dans l'amélioration de la sécurité, cette recherche présente quatre nouveaux frameworks d'agrégation de l'apprentissage fédéré : PolyFLAG_SVM, PolyFLAM, PolyFLAP et HP_FLAP. Les modèles proposés intègrent le chiffrement polymorphe et homomorphe dans leur architecture, ce qui garantit que les messages échangés entre le serveur et les clients restent protégés contre les entités malveillantes. Les frameworks prennent en charge la formation de plusieurs modèles d'apprentissage machine, permettant aux utilisateurs de sélectionner le modèle le mieux adapté à leur problème spécifique. Ce qui distingue ces frameworks, c'est l'intégration du chiffrement homomorphe et polymorphe, qui renforce leur résilience face aux menaces. Le chiffrement homomorphe permet au serveur d'agréger les paramètres échangés sans déchiffrement, tandis que le chiffrement polymorphe garantit que chaque message échangé entre le serveur et les clients FL est chiffré avec une clé de chiffrement distincte, réduisant ainsi le risque de compromis des clés à pratiquement zéro, puisque ces clés ne sont jamais réutilisées dans le cycle du FL. Cette double couche de sécurité renforce la sécurité globale du FL, en contrant diverses menaces, y compris les attaques par inversion, faisant ainsi progresser le domaine en question. En outre, les frameworks proposés intègrent des techniques de réduction des coûts de communication pour améliorer leur efficacité.

Pour valider l'efficacité de ces frameworks, une évaluation complète a été menée, englobant les garanties théoriques, l'analyse de la complexité temporelle et spatiale, les évaluations de l'utilisation des ressources et les évaluations de la qualité de l'apprentissage. Des tests approfondis ont été effectués sur trois ensembles de données distincts, dont un ensemble de données simulées et des données réelles liées à la santé provenant de SHAREEDB et des ensembles de données binaires de Surgical Deepnet. Les résultats empiriques soulignent l'amélioration substantielle de la sécurité, car même dans le cas rare d'une clé de chiffrement compromise ou ayant fait l'objet d'une fuite, le risque pour l'ensemble du système est minime, étant donné la non-réutilisation des clés entre les différentes sessions des clients. Bien que les frameworks proposés offrent effectivement des approches d'agrégation de FL sécurisées et efficaces en termes de communication, ils constituent une base sur laquelle d'autres avancées et intégrations avec les approches existantes peuvent être construites. Une telle intégration peut améliorer la fiabilité et la crédibilité des frameworks proposés et de l'environnement d'apprentissage fédéré dans son ensemble.

mots-clés: Confidentialité, sécurité, attaques par inversion, apprentissage machine fédéré, apprentissage fédéré, chiffrement polymorphe, chiffrement homomorphe, PolyFLAG_SVM, PolyFLAM, PolyFLAP, HP_FLAP, communication efficace.

ABSTRACT

Since its inception, Artificial Intelligence (AI) has remained a focal point of research, equipping machines, including computers and more with the capacity for acting autonomously, by emulating human intelligence. The subfields of AI, most notably Machine Learning (ML) and Deep Learning (DL), have demonstrated remarkable efficiency in solving data analysis tasks. Machine Learning, characterized by a set of algorithms enabling computers to learn from training data without explicit programming, has gained widespread adoption across industry, healthcare, transportation, education, e-commerce, and various other sectors. This adoption has been catalyzed by its aptitude for uncovering patterns in data, learning from them, and ability to making predictions accordingly.

The growth of Machine Learning has been facilitated by advancements in computing techniques. These advancements have empowered ML practitioners to analyze larger datasets and tackle more intricate problems, thus extending the boundaries of what can be achieved. Nevertheless, the domain still grapples with a range of challenges that can be categorized into four broad areas: data-related, model-related, implementation-related, and general challenges. Among these, the overarching issues of privacy and security fall within the category of general challenges. On one front, ML models remain vulnerable to a spectrum of attacks and threats, regardless of the security mechanisms implemented. On the other front, privacy concerns have given rise to a regulatory landscape that restricts access to data, thereby constraining the performance of ML models. This has emerged as a significant hurdle, as the effectiveness of smart models is often commensurate with their ability to access diverse and comprehensive datasets.

In response to these security and privacy challenges, Google introduced Federated Learning, also known as Federated Learning (FL). FL was initially conceived as a privacy-preserving ML approach, as it eliminates the need to centralize user data for model training. Instead, models are distributed to participating clients, trained locally, and later aggregated by the server to generate a global model. In this context, a Federated Learning aggregation algo-

rithm is defined as the mechanism used by the server to send models to clients involved in FL cycle, receive trained models from them and merge the them into a single trained global model. While FL has shown promise in preserving privacy, it remains susceptible to security threats. Researchers are actively exploring methods to secure FL against various attacks, including Byzantine attacks, inversion attacks, backdoor attacks, and more, by implementing mechanisms like Homomorphic Encryption, Secure Multi-Party Computation, the Alternating Direction Method of Multipliers and much more. While substantial progress has been made in strengthening FL algorithms against certain attacks, vulnerabilities like the inversion attack persist, enabling malicious entities to discern users' data contained within the trained models. This underscores the pressing need to advance security measures within the FL domain.

Motivated by the imperative to fortify Federated Learning against a multitude of attacks and recognizing the potential of Polymorphic and Homomorphic Encryption in enhancing security, this research introduces four novel Federated Learning aggregation frameworks: PolyFLAG_SVM, PolyFLAM, PolyFLAP, and HP_FLAP. These proposed models embed both Polymorphic and Homomorphic Encryption in their architecture, ensuring that messages exchanged between the server and clients remain safeguarded against malicious entities. The models support multiple smart models in training, providing flexibility for users to select the most suitable model for their specific problem. What sets these frameworks apart is the integration of both Homomorphic and Polymorphic encryption, bolstering their resilience against threats. Homomorphic Encryption allows the server to aggregate exchanged parameters without decryption, while Polymorphic Encryption guarantees that each message exchanged between the server and the FL clients is encrypted with a distinct encryption key, thus reducing the risk of key compromise to virtually zero, as these keys are never reused in the FL cycle. This dual-layered security enhances the overall security of FL, countering various threats, including inversion attacks, thereby advancing the FL domain. Moreover, the proposed frameworks incorporate communication cost reduction techniques to enhance their efficiency.

To validate the efficacy of these proposed frameworks, a comprehensive evaluation was conducted, encompassing theoretical guarantees, analysis of time and space complexity, resource utilization assessments, and assessments of learning quality. Extensive testing was performed across three distinct datasets, including a simulated dataset and real-life health-related data from SHAREEDB and the Surgical Deepnet Binary datasets. The empirical results unequivocally underscore the substantial enhancement in security, as even in the rare event of a

compromised or leaked encryption key, it poses minimal risk to the overall system, given the non-reuse of keys across different client sessions. While these proposed frameworks indeed offer secure, communication-efficient FL aggregation approaches, they present a foundation upon which further advancements and integrations with existing approaches can be built. Such integration can enhance the reliability and trustworthiness of the proposed frameworks and the Federated Learning environment as a whole.

keywords: Privacy, Security, Inversion Attacks, Federated Machine Learning, Federated Learning, Polymorphic Encryption, Homomorphic Encryption, PolyFLAG_SVM, PolyFLAM, PolyFLAP, HP_FLAP, communication-efficient.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI: Artificial Intelligence
ML: Machine Learning
DL: Deep Learning
FL: Federated Learning
GDPR: European Union's General Data Protection Regulation
FedAvg: Google's Federated Averaging Framework
SMC: Secure Multiparty Computation
IID: Independent and identically distributed
non-IID: non Independent and identically distributed
FDBS: Federated Database Systems
AUC: Area Under Curve
FLAG: Federated Learning with Gradient Aggregation
RFA: Robust Federated Aggregation
LAQ: Lazily Aggregated Quantized Gradient
FedHQ: Federated Learning with Heterogeneous Quantization
FAIR: Federated Learning with quality awareness
FedPSO: Federated Particle Swarm Optimization
LEGATO: Layerwise Gradient Aggregattion
MHAT: Model-Heterogenous Aggregation Training
SEAR: Secure and Efficient Aggregation framework
SGX: Software Guard Extensions
EPPDA: Efficient Privacy-Preserving Data Aggregation
FedBuff: Federated Buffered asynchronous aggregation
PE: Polymorphic Encryption
GM: Geometric Median
ADMM: Alternating Direction Method of Multiplier

GAN: Generative Adversarial Network
HE: Homomorphic Encryption
PRM: Processor Reserved Memory range
HeteroSAg: Secure Aggregation with Heterogeneous Quantization
BCW: Breast Cancer Wisconsin dataset
ESR: Epileptic Seizure Recognition dataset
CREDIT: Default of credit card clients dataset
SVHN: Street View House Numbers
PEFL: Privacy-friendly FL architecture
DTAHE: Federated Learning security mechanism based on additive homomorphic encryption
UKBB: UK Biobank neuroimaging dataset
HAM10000: Human versus machine data with 10,000 training images
CCFL: Communication-based Federated Learning
FLN: Federated Learning Network
AES: Advanced Encryption Standard Algorithm
RAS: Rivest-Shamir-Adleman Algorithm
PHE: Partially Homomorphic Encryption
FHE: Fully Homomorphic Encryption
ToKs: Table of Encryption Keys
HEToKs: Homomorphic Encryption Table of Keys
SVM: Support Vector Machines
LR: Logistic Regression
NB: Gaussian Naïve Bayes
SGD: Stochastic Gradient Descent
MLP: Multi Layer Perceptron
AC: Accuracy
PR: Precision
RE: Recall
F1: F1 Score
SP: Specificity
NPV: Negative Predictive Value

CHAPTER 1

General Introduction

Artificial Intelligence (AI) has experienced rapid growth over the past two decades. The concept of AI has been around since 1950, and the term itself was coined in 1965 at the Dartmouth Summer Workshop, which is considered the founding event of AI as a field [1]. However, the growth in Information and Communication Technologies (ICTs) and the increasing power of computers have contributed significantly to the increasing feasibility and adoption of AI [2]. AI technologies are becoming more advanced and are capable of analyzing enormous amounts of data, learning from past experiences, and making predictions based on patterns and trends [3]. Despite the different definitions provided for AI in [4–6], they all agree that it is the technology that enable machines to mimic human intelligence. A lot of researches, including [4–6] showed the efficiency of Artificial Intelligence, and discussed the difference in its applications in our daily lives.
Machine Learning (ML) [7] and Federated Learning (FL) [8] are popular sub-fields of the AI as depicted in Figure 1. Machine Learning is defined as a field of study that focuses on the development of algorithms that enable computer systems to learn from data and make predictions or decisions without being explicitly programmed. It involves the application of various approaches that allow computers to automatically improve their performance on a given task through experience [7].

Machine Learning (ML), allows computers to "learn" from training data and expand their knowledge over time without being explicitly programmed. Machine Learning algorithms attempt to find patterns in data and learn from them to make their own predictions. Traditionally, a computer program is developed by engineers and given a set of instructions that enable it to turn incoming data into its intended output. ML, by contrast, designs the program to learn with little or no human interaction and to expand its knowledge over time. The remark-

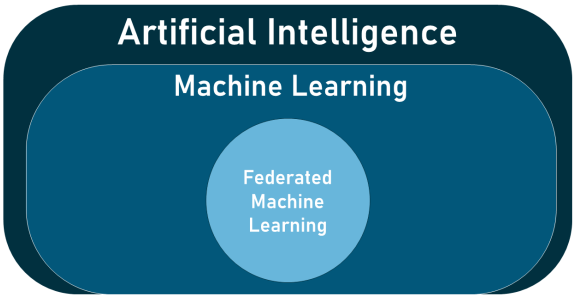

Figure 1: Artificial Intelligence Branches.

able success of ML, as well as its enormous potential in data analysis, have made it attractive to researchers in many fields. Later studies revealed the variety of applications of ML that can be observed in different fields such as: E-commerce and product recommendations, Image, speech and pattern recognition, user behavior analytics and context-aware smartphone applications [9,10], healthcare services [11–13], traffic prediction and transportation [11,14], Internet of Things (IoT) and smart cities [14], Cybersecurity and threat intelligence [15], Natural Language Processing and sentiment analysis [16], sustainable agriculture [17], industrial applications [18] and more.

1.1 Machine Learning Techniques Taxonomy

Artificial Intelligence and its descendant, Machine Learning, are used in a wide variety of real-world applications. Plenty implementations are available in the areas mentioned in the previous section. Moreover, the algorithms of ML can be classified into different groups depending on their classification perspective. These algorithms are traditionally classified into supervised, unsupervised, semi-supervised, and reinforcement learning [19–21]. However, this classification only considers the data analyzed by the model or the so-called learning style and ignores other possible classification bases. In this context, the function or goal of the algorithm as well as the architecture can serve as classification factors and provide an extended taxonomy for ML algorithms. Figure 2 below presents the proposed taxonomy for machine learning algorithms.

1.1.1 Classification per Learning Style

Machine Learning workflows specify what steps are performed in an ML project. Data acquisition, data preprocessing, model training and fine-tuning, evaluation, and production deployment are generally the common processes. Consequently, the type of data obtained determines the Machine Learning algorithm. From this point of view, the four categories listed below can be defined [22–24]:

- **Supervised Learning:** This refers to the types of ML where machines are trained with labeled input and then predict output based on that data. Labeled data means that the input data have been labeled with the corresponding output. The training data serve as a supervisor that teaches the computers how to correctly predict the output. Then it can be described as a process of providing the model ML with appropriate input and output data so that it can identify a function to map the input and output variables;
- **Unsupervised Learning:** An algorithm that operates only on input data and has no outputs or target variables. Consequently, unlike supervised learning, there is no teacher to correct the model. In other words, it is a collection of problems where a model is used to explain or extract relationships in data;
- **Semi-Supervised Learning:** This is a form of supervised learning in which the training data includes a small number of labeled instances and a large number of unlabeled examples. It attempts to use all available data, not just the labeled data as in supervised learning;
- **Reinforcement Learning:** This defines a class of problems where the intelligent model operates in a given environment and must learn how to act based on inputs. This means that there is no given training dataset, but rather a goal or collection of goals for the model to achieve, actions it can take, and feedback on its progress toward the goal. In other words, the goal is to learn what to do, how to map events to actions in order to maximize a numerical reward signal, not dictating to the model what actions to perform, but figuring out through trial and error which activities yield the greatest reward.

1.1.2 Classification per Function

Machine Learning algorithms, on the other hand, can be categorized by the goal of the model. The goal, also referred to as the function, is the output of the model and determines the type of model to be used. The different types of ML can be defined as follows [22–24]:

- **Classification:** the process by which a ML algorithm predicts a discrete output or so-called class. Depending on the type of class to be predicted, this class can be divided into the following groups:
  - **Binary Classification:** refers to algorithms that can predict only one of two labels, e.g., classifying emails as spam or not;
  - **Multi-Class Classification:** refers to algorithms with more than two class labels, where there are no normal and abnormal results. Instead, the examples are classified into one of several known classes;
  - **Multi-Label Classification:** the set of algorithms that predict the output of a label class, with no limit to how many classes the instance can be assigned to.
- Regression: the process by which a ML algorithm can predict a continuous output or a so-called numerical value;
- Clustering: the process of categorizing a set of data instances or points so that those in the same group are more similar and different from data points in other groups. It is essentially a collection of instances based on their similarity and dissimilarity;

1.1.3 Classification per Architecture

Another approach to classifying Machine Learning algorithms can be based on the underlying architecture of the system. In this context, two main categories can be defined [25, 26]:

- **Centralized Architecture:** the traditional ML architecture, where data is collected on a machine running the model;
- **Distributed Machine Learning:** the ML paradigm that benefits from a decentralized and distributed computing architecture where the ML process is split across different nodes, resulting in a multi-node algorithm and system that provides better scalability for larger input data.

1.2 Machine Learning Under Scope: Challenges

Accurate results in classification or regression are increasingly encouraging the incorporation of these techniques into areas of daily life [9]. The feasibility of using AI tools, and in

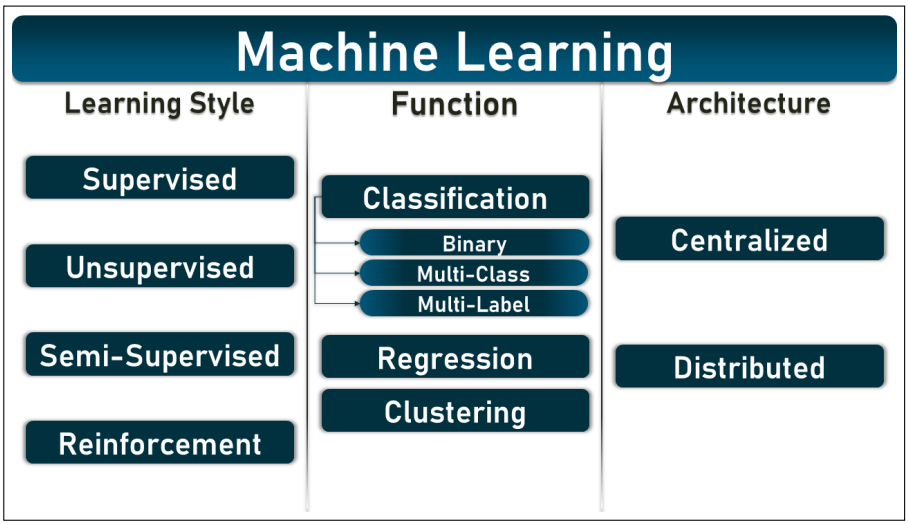

Figure 2: Machine Learning algorithms taxonomy.

particular ML, has been demonstrated by the high performance they offer and the possibility of integrating them in different domains. However, ML still suffers from several challenges that are extensively described and discussed in the literature [8]. However, these challenges are not classified into a single taxonomy, but grouped according to different aspects, including technical, ethical, social, and other factors. In this section, the common challenges are presented under a proposed taxonomy based on data-related, model-related, implementation-related, and other general aspects. In addition, these challenges are illustrated and summarized in Table 1 below.

1.2.1 Data Related Challenges

Machine Learning algorithms are typically implemented in a known pipeline consisting of data collection, preprocessing, exploration, model selection, training, evaluation, and deployment. Data, which constitute the main component of these algorithms, can present various challenges, such as [27, 28]:

- **Data availability and accessibility:** to train a model, one must have the necessary data, which may not be available on the spot or may be available but inaccessible for various reasons;
- **Data locality (Data Islands):** in the real world, data are scattered in different and unrelated entities called "data islands". Due to different regulations and laws, data

related to the same subject and available on different data islands cannot be accessed for use and analysis;

- **Data readiness:** even if data are available and accessible, several aspects should be considered, such as:
  - **Data Heterogeneity:** available data may have different characteristics or be composed of different forms. For example, health data for the same patient may be available in different forms, such as medical images, reports, videos, and structured data. The ability to deal with such heterogeneity is a challenging task;
  - **Noise and Signal Artifacts:** due to the interaction between data acquisition instruments and other electrical devices, data can be poisoned by noisy attributes that affect the overall results of ML models;
  - **Missing Data:** data collected by measuring devices may be incomplete for various reasons;
  - **Classes Imbalance:** in classification problems, the data collected for one group may dominate the data collected for other groups, affecting the learning of the smart model.
- **Data Volume:** is the amount, size, and scope of the data. In the context of ML, size can be defined either vertically by the number of records or samples in a dataset or horizontally by the number of features or attributes it contains. Data volume presents several challenges, such as:
  - **Course of Dimensionality:** dimensionality describes the number of features or attributes that are present in a dataset. Increasing dimensionality can have a negative impact on model performance.
- **Feature Representation and Selection:** the performance of ML models heavily depends on the choice of data representation or features, so selecting the optimal features will definitely improve the overall model performance.

1.2.2 Models Related Challenges:

In addition to the challenges posed by the data, some challenges are related to models such as [29, 30]:

- **Practicality and Performance:** achieving the most practical and feasible ML models remains the main goal for researchers from various fields, and the highest feasibility will lead to the highest adoption and integration of this technology;

- **Model Evaluation:** evaluating an ML model can be challenging, especially when traditional performance metrics may not, necessarily, reflect a model's feasibility In addition, ML models are susceptible to variance and bias that can affect their performance, results, and confidence. Given that variance is the variability of the model prediction for a given data point or a value indicating the spread of our data, and bias is the difference between the average prediction of our model and the correct value we are trying to predict;

- **Model Selection:** different models can produce different results even for the same problems. For example, support vector machines (SVM) and logistic regression (LR) can lead to different results, even when working with the same data at the same point in time. Thus, selecting the optimal model and tuning its parameters are not easy tasks;

- **Explainability:** some of the ML models are known by their black box identity. The failure to allow human users to comprehend and understand the results and output, or how they were extracted by a Machine Learning model can have a negative impact on trust in these models, even when high accuracies are achieved.

1.2.3 Implementation Related Challenges:

Assuming that the obstacles in the data and models have been overcome, implementing the models of ML can be a challenging task due to various obstacles such as [31, 32]:

- **Real-time Processing:** ML models are created and trained with available data. However, fitting these models to real-time problems presents several challenges;

- **Execution Time and Complexity:** due to many reasons, including but not limited to, the complexity of the data or models, multiple preprocessing steps and computation resources, ML models can require enormous computing power and long execution times. On the other hand, the structure of some ML models, Neural Networks for instance, are known to pose more complexity than others such as Linear Regression. A model complexity can also raise execution time challenges as well.

1.2.4 General Challenges:

Finally, other challenges besides technical aspects can be mentioned in this section, such as [29, 30]:

- **User Data Privacy and Confidentiality:** which is one of the most critical issues in the field of ML. Users tend not to share their data for various reasons, which affects the availability of the data and jeopardizes the entire ML cycle;
- **User Technology Adoption and Engagement:** due to privacy issues, unclear results, lack of explanation, and other reasons, users may not accept ML being integrated into their daily routine, or even accept its results;
- **Ethical Constraints:** various ethical constraints posed by ML have been widely discussed in the literature, such as control and morality, model ownership, environmental impact, and many others.

Table 1: Machine Learning domain common challenges.

Group	**Challenges**	
Data-Related Challenges	Data Availability and Accessibility Data Locality	
	Data Readiness	Data Heterogeneity Noise and Signal Artifacts Missing Data Classes Imbalance
	Data Volume	Course of Dimensionality
Models Related Challenges	Accuracy and Performance Model Evaluation Model Selection Explainability	
Implementation-Related Challenges	Real-Time Processing Execution Time and Complexity	
General Challenges	User Data Privacy and Confidentiality User Technology Adoption and Engagement Ethical Constraints	

1.3 Data Privacy:A Machine Learning Impediment

Privacy is a fundamental right, and protecting sensitive personal information is critical in today's digital age. Privacy issues can arise when collecting, storing, and analyzing data in the context of Machine Learning, as algorithms may rely heavily on personal data to train models and make predictions. For example, ML models applied in health, finance or some specific fields of life, may deal with highly sensitive personal data. The necessity to preserve privacy of sensitive data stems not only from their confidentiality, but also due to the increasing number of data breaches, which require more and more solutions as their negative impact grows. Consequently, not only individuals, but also society, governments, and organizations are strengthening the protection of data privacy and security. In this regard, several regulations and laws were enacted, such as the European Union's General Data Protection Regulation (GDPR) [33], China's Cyber Security Law of the People's Republic of China [34], the General Principles of the Civil Law of the People's Republic of China [35], the PDPA in Singapore [36], and hundreds of principles legislated around the world. While these regulations help protect private information, they pose new challenges to the ML field by making it more difficult to collect data to train models, which in turn makes it more difficult to improve the accuracy of model performance and to personalize those models. Consequently, data privacy and confidentiality are not a stand-alone challenges, but also trigger other challenges for ML, such as data availability, performance, personalization, and thus trust and acceptance.

1.4 Federated Learning: Privacy-Preserving ML Concept

Recent advancements in data collection and analysis have been significant, driven by the development of communication tools and AI techniques. However, data are often gathered in isolated "data islands", composed of entities such as foundations, institutions, individuals, or other organizations where data are collected and stored. To enhance AI model performance, a centralized approach is often sought, involving the collection of data into a central repository for unified processing, cleaning, and modeling. For instance, analyzing a patient's health data from various hospitals, clinics, or health centers can be most effective when done collectively. However, privacy regulations and data heterogeneity present challenges to this centralized data collection and analysis. As a result, researchers worldwide have focused on finding solutions to the issues posed by data islands and privacy.

1.4.1 An Overview of FL

In the context of privacy criticality, Google introduced a new notion in ML field, which was named Federated Machine Learning or Federated Learning [8]. This concept allowed training ML models, without the need to collect data on a central server, as it is the case in the traditional ML concept. Alternatively, a global model is sent from the central sever to the entities participating in the FL system. Each entity trains the received model on its own data, and send the trained model to the server where all models are aggregated into a trained global one. This mechanism eliminates the need to share private data, thus promoting the privacy, and keeping confidential data secured. Even it was introduced in 2016, FL is considered as a promising concept that can improve the entire field of Machine Learning. Despite this rise, Federated Learning is still in its infancy and still struggling with various challenges and issues. These challenges include convergence issues, algorithm complexity, communication and computational overhead, and more [8]. Most importantly, Federated Learning is still vulnerable to the well-known digital attacks, which will be discussed in more detail later. This vulnerability to attacks threatens the main concept of FL, which is privacy preservation.

1.4.2 Aggregation in Federated Learning

In Federated Learning, aggregation is the process of collecting and combining model updates from multiple clients that have trained their models locally on their respective datasets. This process is typically orchestrated by a central server or coordinator. The aggregated updates are then used to update the global model, which is subsequently redistributed to the clients for further local training. The aggregation step is crucial for maintaining model performance while preserving data privacy, as it allows the global model to learn from distributed data without requiring raw data to be shared between clients or with the central server.

1.4.3 Securing Federated Learning Against Attacks

FL is vulnerable to different threats and attacks including poisoning and inference attacks [37]. Those two types of attacks affect the FL system differently. Poisoning attacks affect the learning quality of the global model, by feeding the main server with models trained by false and useless data. Therefore, the global model's learning quality will be affected. On the other hand, inference attacks can allow malicious entities to capture the models exchanged between the server and clients, and therefore crack the private data embedded in those mod-

els through different mechanisms. Considering their severity, those threats and attacks can undermine the privacy-preserving identity of FL.

In this context several attempts were carried out to secure FL algorithms against such attacks. For example, authors in [38] presented a secured FL algorithm by the means of safe vector summing and cryptographic primitives in multiple stages. Their proposed model succeeded in securing FL against poisoning attacks, but it also shows some limits in its ability to withstand active attacks, ensure the use of well-formed input, and manage communication overhead. Similarly, Authors in [39] proposed RFA, an FL algorithm secured with geometric median-based aggregation, that succeeded to withstand poisoning attacks, but not inference attacks. Likewise the proposed algorithm LEGATO [40], where authors developed the algorithm to withstand poisoning attacks only. Beside its limitation to this kind of attacks, LEGATO was limited to ML models composed in layers such as Neural Networks. In addition, authors in [41] used the Alternating Direction Method of Multiplier (ADMM) to control the communication mechanism among the FL entities. Their proposed algorithm, named SecureD-FL withstand malicious entities attempts to crack other entities data by limiting connections among them. However, it do not take into consideration withstanding inference attacks as well. Moreover, authors of [42] uses trusted execution environment (TEE) hardware to secure the FL method. The proposed algorithm, named SEAR, proved efficiency against different attacks, but it struggles with the limited memory capacity in those hardware environments. Furthermore, authors in [43] proposed EPPDA, that uses homomorphic encryption for secret sharing among FL entities. EPPDA proved to withstand among different types of attacks, including inference attacks.

Additionally, authors in [44] addresses communication efficiency and resilience to Byzantine attacks. Despite its success against such attacks, but inference attacks were not taken into consideration in the proposed algorithm. Also, authors in [45], proposed FLDetector focuses on identifying potentially malicious clients, by the means of assessing the consistency of model updates received from those entities. Although it reduces the risk of malicious entities, but it did not take withstanding inference attacks into consideration in its mechanism. Also, the algorithm FLCert, proposed in [46] was built to resist poisoning attacks by classifying customers into groups and utilizing majority voting among global models to resist poisoning attacks. This algorithm didn't take inference attacks into their security strategy as well as most of the previous mentioned algorithm. Likewise, authors in [47], proposed ELSA that distributes trust among two servers instead of a central manager. Therefore, the proposed algorithm withstands poisoning attacks and malicious servers, but didn't take into considera-

tion the severity of inference attacks that can crack private data, even if exchanged between a trusted server and a legit entity. Moreover, authors in [48] proposed Multi-RoundSecAgg that excels in long-term privacy preservation, structured user selection, and fairness considerations. However, it introduces complexity in terms of multi-round confidentiality guarantees and structured user selection strategies, potentially increasing computational and operational complexity.
On the other hand, different solutions used Homomorphic Encryption to secure the models exchanged between the central server and the FL entities. For instance, authors in [49–55] embedded this encryption technique in their mechanisms. However, such implementations are still prone to the risk of cracking encryption keys, thus threatening the FL privacy identity.

1.5 Polymorphic & Homomorphic Encryption

In the context of securing data exchange against attacks, different techniques can be defined. For instance, Polymorphic and Homomorphic Encryption are two techniques that can be defined as below.

1.5.1 Polymorphic Encryption

To define Polymorphic Encryption, it is helpful to define both polymorphism and encryption, which are in fact, two different concepts:

- **Polymorphism:** denotes the capacity of an object or function to assume various forms or behaviors;
- **Encryption:** refers to the process of transforming regular data into an incomprehensible format to deter unauthorized access or usage. Furthermore, encryption techniques are plenty, and the well-known encompass AES (Advanced Encryption Standard), RSA (Rivest-Shamir-Adleman), and others [57].

Therefore, Polymorphic Encryption [56] can be characterized as an encryption approach that alters the algorithm or encryption keys to bolster security. With PE, it becomes significantly more challenging for attackers to decipher the encryption, even if they manage to obtain the ciphertext. Unlike conventional encryption methods that rely on fixed algorithms and static keys, PE employs dynamic algorithms and continually changing keys. This dynamic nature ensures that the encryption pattern is constantly evolving, making it extremely

difficult for attackers to predict or decode the encrypted data.
In traditional encryption, once an attacker discovers the key or understands the algorithm, they can decrypt the data. However, polymorphic encryption adds layers of complexity by frequently altering the cryptographic parameters. This continuous transformation of the encryption schema means that even if one set of keys is compromised, the subsequent data remains secure due to the ever-changing nature of the encryption process.
This method provides robust protection against a wide range of attacks, including brute force, man-in-the-middle, and quantum computing attacks [56]. The adaptability and resilience of Polymorphic Encryption make it a formidable defense mechanism in safeguarding sensitive data in an era where cyber threats are becoming increasingly sophisticated.

1.5.2 Homomorphic Encryption

Homomorphic Encryption [58], a recent innovation in encryption techniques, permits computations to be conducted on encrypted data without requiring decryption beforehand. In contrast to conventional encryption methods, where decryption is necessary for data utilization, homomorphic encryption allows mathematical operations to be directly applied to encrypted data while maintaining its secrecy. This breakthrough encompasses various forms, including partial homomorphic encryption (PHE) and fully homomorphic encryption (FHE), which are defined in the below list:

- Partial Homomorphic Encryption: supports a limited set of operations, such as addition or multiplication, on encrypted data. This is useful for specific applications where only certain computations are needed;
- Fully Homomorphic Encryption, on the other hand, enables arbitrary computations on encrypted data, providing the most flexibility and security. FHE has been a particularly challenging goal to achieve due to its complexity and computational overhead, but recent advancements have made it more feasible.

The implications of this advancement are significant for privacy and security, particularly in contexts like cloud computing and data sharing. By employing homomorphic encryption, sensitive data can remain safeguarded during processing, facilitating secure data analytics, private machine learning, and confidential data processing outsourcing. This means that organizations can outsource data processing tasks to third-party service providers, such as cloud computing platforms, without exposing the underlying data [58].

For instance, in the field of healthcare, homomorphic encryption allows for the analysis of encrypted patient data to gain insights without compromising patient privacy. In finance, it enables secure computations on encrypted financial records, allowing for private auditing and fraud detection. Overall, homomorphic encryption offers a powerful tool for maintaining data confidentiality and integrity in an increasingly data-driven world, where privacy concerns are paramount [58].

1.6 Problematic

Privacy and confidentiality concerns lead to additional ML challenges for ML, such as data availability, performance, personalization, and thus trust and acceptance. Consequently, the concept of Federated Learning was introduced as a solution to privacy issues.
However, FL is still prone to different attacks and threats. This issue has been of great interest in the literature, and numerous solutions have been proposed to make the algorithms and implementation of FL more robust to attacks [25]. Despite this, inference attacks have not been considered in these solutions, which in turn threatens user privacy and underpin the preserving privacy identity of Federated Learning, given the fact that it enables cracking private data out of the captured model.

1.7 Objectives and Research Questions

The vulnerability of current Federated Learning algorithms to various types of attacks, especially inference attacks, undermines the true identity of FL domain. The ability of a malicious entity to discover users' private data is certainly the opposite of the privacy-preserving idea that was initially proposed in Federated Learning notion. Therefore, this research investigates and analyzes the security aspects of Federated Learning, and proposes a set of frameworks with enhanced security levels to withstand various attacks, including inference attacks. Therefore the objective for this research can be summarized as:

Enhance the security of Federated Learning aggregation algorithms to stand against inference attacks by the means of Polymorphic and Homomorphic Encryption

Within this context, the research delves into Federated Learning, encryption methodologies, and the introduction of innovative aggregation frameworks. The exploration of these

facets aims to provide comprehensive answers to crucial research questions, ultimately advancing the understanding and fortification of the security landscape in Federated Learning. The research questions can be found in the below list (RQ in the below list is the acronym of 'Research Question' and RSQ is the acronym of Research Sub Question)

- **RQ1:** How to enhance privacy of Federated Learning aggregation algorithms and secure them against malicious attacks, especially inference attacks?
- **RQ2:** What is the impact of integrating Polymorphic & Homomorphic encryption in FL in terms of robustness against inference attacks and in terms of additional complexity, computation and communication cost?

1.8 Research Methodology

In this research project, four Federated Learning Frameworks with advanced and enhanced security are proposed. Those frameworks, named PolyFLAG_SVM, PolyFLAM, PolyFLAP, and HP_FLAP, compete with the existing Federated Learning frameworks by embedding both Polymorphic and Homomorphic Encryption technologies to secure the messages exchanged between FL server and clients. In this section, the research methodology followed to obtain the objectives of the project is explained. This methodology was conducted in four phases over a period of three years as detailed below and summarized in Table 2 below:

- **First Phase – Review:** formed of the examination and contextualization of the research topic, where the literature was reviewed and the existing Federated Learning frameworks were analyzed to find out where they stand in the vulnerability against attacks, specifically inference attacks. During this phase, the problem was framed, and a possible solution was proposed;

- **Second Phase – Design:** includes the design of the frameworks and setting the technical benchmarks of the implementation. For this purpose, primarily implementations of Federated Learning, Homomorphic, and Polymorphic encryption were performed. This stage helped in gaining knowledge about the tools needed to build the frameworks. The results obtained at this stage were recorded to be used later;

- **Third Phase – Implementation:** First, Polymorphic Encryption layer was added to the FL framework built in the second phase. After that, Homomorphic Encryption layer was added so that additional security could be obtained. This stage resulted in

four different frameworks that support up to five different smart models (as will be discussed later);

- **Fourth Phase – Evaluation & Use-Case Study:** included evaluation of the implemented frameworks with different datasets. For this purpose, a simulated Federated Learning environment was built, and a simulated dataset was used for primary evaluation. Later, two real-world datasets were used to conduct a use case study that helps in testing the proposed frameworks in real-world scenarios.

1.8.1 First Phase: Review & Critical Analysis of FL Frameworks/algorithms

In the literature review phase, an extensive exploration of existing research and scholarly works was undertaken to establish the theoretical foundation and identify gaps within the realm of Federated Learning.

- **Literature Review:** composed of a comprehensive exploration of existing scholarly works, research articles, and relevant publications in the field of study. The insights gained from this literature review not only informed subsequent phases but also played a pivotal role in shaping the overall research methodology. The literature review went into three main tasks:
  - **Task 1:** reviewing smart health, with a focus on the use of smart wearables in the detection of Cardiovascular Diseases. Due to the fact that smart health is relatively a wide and broad topic of research, Cardiovascular Diseases was selected as a starting point due to the severity of this diseases and to the wide interest it gains in the research domain. The objective was to gain a nuanced understanding of the current state of knowledge, theoretical frameworks, and methodologies employed by previous researchers in smart health. To fulfill an adequate review, systematic strategies were followed to include all relevant studies, analyze them, and extract emerging trends and gaps;
  - **Task 2:** During the review, challenges in the Machine Learning domain were set under focus, with Multimodal and Federated Learning being two trending domains studied. The first acts as a solution for heterogeneity, and the latter acts as a solution for privacy;
  - **Task 3:** Perform a deep review of Federated Learning, with a focus on the challenges and deficiencies hindering the domain;

- **Problem Definition:** The problem definition phase was instrumental in precisely articulating the challenges within the domain of Federated Learning that necessitated intervention. Central to this phase was the identification and elucidation of the vulnerability of existing Federated Learning frameworks to inference attacks. The scope was delineated by a nuanced understanding of the risks posed by these attacks, emphasizing the potential compromise of sensitive information. By explicitly defining the problem, this phase provided a clear foundation for subsequent research endeavours, guiding the formulation of a targeted and effective solution. The insights garnered during this phase not only shaped the trajectory of the research methodology but also contributed to the development of frameworks resilient to the identified vulnerabilities;

- **The Scope of Solution:** Within the literature review phase, attention was given to framing the scope of the solution, which addressed the vulnerability of Federated Learning frameworks to inference attacks. The proposed solution aimed at enhancing Federated Learning by introducing novel frameworks that incorporated Polymorphic and Homomorphic Encryption techniques, thereby bolstering the security measures. The delineation of the scope was intricately tied to countering the identified vulnerability, and it was strategically defined by incorporating the use of Polymorphic and Homomorphic Encryption. This strategic choice not only aimed at mitigating the susceptibility to inference attacks but also paved the way for a more secure and robust Federated Learning environment. The thorough literature review informed and justified the selection of these encryption techniques, providing a solid foundation for subsequent phases in the research methodology.

1.8.2 Second Phase: Design

The design phase marked a crucial transition from the theoretical groundwork laid in the literature review to the concrete formulation of a solution tailored to address the identified issues. This phase encompassed several key components:

- **Conceptual Framework:** Building upon the insights derived from the literature review, a robust conceptual framework was developed. This framework served as the blueprint for integrating Polymorphic and Homomorphic Encryption into the architecture. Each element of the framework was meticulously designed to ensure com-

patibility with existing infrastructure while addressing the specific security concerns identified in the problem definition phase;

- **Algorithm Development:** In order to deepen the technical understanding of the proposed solution, three fundamental algorithms were developed. The first algorithm encapsulated the core functionality of a FL approach, elucidating the decentralized model training process. Simultaneously, a Polymorphic Encryption algorithm and a Homomorphic Encryption algorithm were crafted to comprehend the intricacies of these cryptographic techniques. These algorithmic implementations provided invaluable insights into the technical nuances and potential challenges, laying a solid foundation for the subsequent design and implementation phases;

- **Frameworks Architecture:** The design phase also involved the creation of the architecture, outlining the structural components of the proposed Federated Learning framework. This encompassed defining the roles of different entities, specifying data flow, and elucidating the integration of encryption techniques at various stages of the learning process;

- **Encryption Integration:** A key focus of the design phase was the seamless integration of Polymorphic and Homomorphic Encryption. This involved developing protocols and algorithms to embed these encryption methods into the FL workflow without compromising computational efficiency or model accuracy. The design ensured that the encryption mechanisms not only bolstered security but also facilitated collaborative model training across decentralized devices.

1.8.3 Third Phase: Implementation

The implementation phase was a pivotal stage where the theoretical design was translated into functional frameworks. Four distinct frameworks were developed, each incorporating advanced cryptographic techniques to enhance the security of Federated Learning:

- **PolyFLAG_SVM:** The first framework was designed as a Federated Learning solution embedding polymorphic encryption. This framework specifically offered a Support Vector Machine (SVM) model. An SVM model is known with different advantages such as effective in high-dimensional spaces, robustness to overfitting, versatility in kernel selection, global optimization, memory efficiency, effective in cases of small

sample size, versatility in solving different types of problems including both classification and regression and many more. The architecture facilitated secure model training, ensuring that server and clients could exchange gradients while preserving the privacy of local datasets. To secure the exchange of gradients, each client an SVM model on its local data, generate its own gradients, wrap them within messages that are encrypted polymorphically and exchange them with the server. At each round, the server receives the gradients from all involved clients, and aggregate them to have a global model trained federally without the need to collect users' private data (source code available at [59]);

- **PolyFLAM:** Building upon the PolyFLAG_SVM foundation, the second framework, PolyFLAM, expanded the scope by incorporating polymorphic encryption to support five different models. Unlike the first framework, PolyFLAM enabled the exchange of entire models between the server and clients, enhancing the versatility of FL across various model architectures (source code available at [60]);

- **PolyFLAP:** The third framework represented an enhanced version of PolyFLAM, introducing polymorphic encryption and supporting five different models. However, PolyFLAP innovatively shifted the paradigm by exchanging model parameters between the server and clients. This approach aimed to further optimize communication efficiency while preserving the confidentiality of sensitive information during the Federated Learning process (source code available at [61]);

- **HP_FLAP:** In the final iteration, the HP_FLAP framework was developed, embedding both polymorphic and homomorphic encryption. This advanced framework supported four distinct models, providing a comprehensive approach to secure FL. In HP_FLAP, the exchange of model parameters between the server and clients was facilitated through a combination of polymorphic and homomorphic encryption, adding an extra layer of security to the collaborative learning process (source code available at [62]);

These implementations not only demonstrated the technical feasibility of integrating cryptographic techniques into Federated Learning but also served as experimental platforms for evaluating the efficacy of the proposed security-enhanced frameworks.

1.8.4 Fourth Phase: Evaluation and Use Case Study

The Evaluation and Use Case Study phase marked the culmination of the research, focusing on assessing the performance and applicability of the developed frameworks in diverse scenarios. Three distinct evaluations were conducted, each shedding light on the frameworks' capabilities:

- **Evaluation with Simulated Dataset:** The first evaluation utilized a simulated dataset to gauge the frameworks' effectiveness in controlled environments. The dataset was generated in tabular form, with 20 features and 9000 records and formed of real value numbers. Metrics such as complexity, data consumption, and accuracy were recorded. This simulated environment allowed for a thorough understanding of the frameworks' behavior under ideal conditions, serving as a baseline for subsequent evaluations;

- **Evaluation with SHAREEDB Dataset:** Moving beyond simulated scenarios, the frameworks were evaluated with real-world data using the SHAREEDB dataset. This evaluation aimed to assess the frameworks' performance in a more dynamic and complex setting, providing insights into their adaptability and robustness with genuine data challenges;

- **Evaluation with Surgical-Binary Dataset:** In the final evaluation, the frameworks were subjected to the unique characteristics of the Surgical-Binary dataset. This use case study delved into the frameworks' performance in a specialized domain, capturing the nuances of medical data. The assessment considered metrics related to complexity, data consumption, accuracy, and additional domain-specific parameters.

Throughout these evaluations, the frameworks were meticulously monitored, and comprehensive metrics were recorded to analyze their performance. This included assessing the computational complexity of the frameworks, evaluating their efficiency in terms of data consumption, and measuring the accuracy of collaborative learning. A rigorous comparison among the results obtained by each framework provided a comprehensive understanding of their relative strengths and weaknesses, informing potential areas for refinement and improvement.

1.8.4.1 Robustness Against Inference Attacks

On the other hand, testing the proposed frameworks against inference attacks included monitoring and recording the flow of encrypted exchanged messages among FL server and clients. It worth mentioning that parties who perform Inference attacks tends to eavesdrop or intercept exchanged model(s) between FL Server and client, in the attempt to crack users' private data out of this model. However, the main goal of the proposed models was to secure those messages by:

- Encrypting each message exchanged among FL server and clients by the means of the robust AES encryption algorithm;
- Using a unique encryption key for each of the exchanged messages, which guarantees that a leaked or cracked key do not risk any of the future messages as it will not be used again in the FL cycle;
- Additionally securing the exchanged parameters (in case of HP_FLAP), by adding the Homomorphic Encryption layer within the polymorphically encrypted messages. In such a way, if a malicious entity succeeded by cracking an exchanged model, it will not be able to gain access for the embedded parameters that are also encrypted. Furthermore, the FL server itself, will not gain access to the exchanged parameters in this case, where it aggregate all the collected parameters while exchanged, being homomorphically encrypted. This also grants the HP_FLAP framework additional security, even against the server itself.

Consequently, monitoring and evaluating the security level of the proposed frameworks is obtained by tracking the exchanged messages among FL server and clients, and recording the encryption keys used in each round with each client or at the server side. The polymorphism of the keys in use, beside the exchange of encrypted parameters, are the the guarantees of robustness of the proposed frameworks against inference attacks. During the evaluation phase, polymorphism and homomorphism were tracked and recorded in the system as will be explained later in Chapter 5.

To fulfill the research goal, and to execute the methodology as planned, a Gantt-Diagram was initially created with the above explained tasks. The diagram was followed, and updates were performed where needed, reaching to the results explained in this report. The diagram

is illustrated in Figure 3 below. In this chart, the scientific contributions named as P1, P2, .. are listed in Table 4 mentioned in the Appendix A.

Table 2: Research Plan

Problem	*Federated Learning is **prone to inference attacks**, and the available aggregation algorithms **lack robustness** against such attacks which **threatens** the privacy-preserving concept of FL*	
Research Objective	Enhance the security of Federated Learning aggregation algorithms to stand against inference attacks by the means of Polymorphic and Homomorphic Encryption	
Research Plan	**First Phase**	Literature Review, Analysis and Contextualization
		Problem Definition
		Defining the scope of the proposed solution
	Second Phase	Conceptual Framework
		Algorithm Development
		Frameworks Architecture
		Encryption Integration
	Third Phase	Integration of Polymorphic Encryption and Building PolyFLAG_SVM
		Extending the supported smart models and building PolyFLAM
		Enhancing Communication cost by building PolyFLAP
		Integrating Homomorphic Encryption and building HP_FLAP
	Fourth Phase	Evaluation with Simulated Dataset
		Evaluation with SHAREEDB dataset
		Evaluation with Surgical-Binary dataset
		Comparing performance results obtained from the proposed frameworks

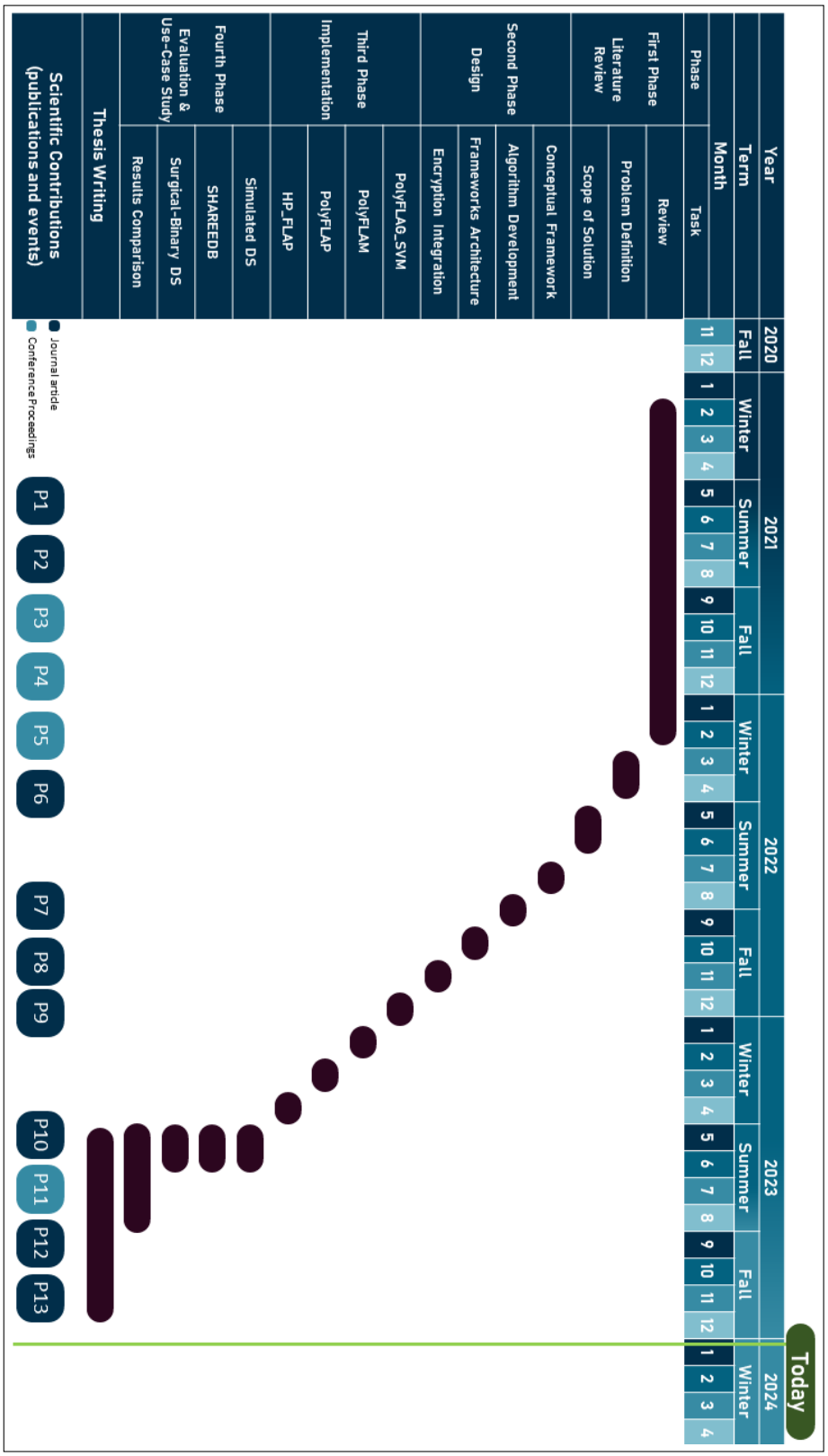

Figure 3: Research Gantt Diagram.

1.9 Novelty and Contribution

In this research, four frameworks are proposed to improve security and privacy in Federated Learning. The uniqueness of the proposed frameworks stems from their integration of polymorphic and homomorphic encryption, as will be shown later, and their ability to train different Machine Learning models suitable for different real-world analysis problems. This integration not only distinguishes the proposed systems from the current state of the art, but also creates new opportunities for strengthening the overall security of FL.
The contributions of this research include the development and rigorous evaluation of several frameworks that represent significant advances in the field of privacy-friendly Machine Learning. These frameworks integrate state-of-the-art cryptographic techniques to strengthen the security and privacy of the Federated Learning environment. Through the use of polymorphic encryption, each exchanged message is encrypted with a unique key, making any potential compromise of the key insignificant to the integrity of the overall system. Furthermore, the use of homomorphic encryption during model aggregation introduces an additional layer of security. In addition to the security improvements, the proposed frameworks also provide practical benefits by reducing communication costs through parameter and gradient exchange options. Furthermore, the inclusion of five different smart models: Support Vector Machines, Linear Regression, Naive Bayes, Stochastic Gradient Descent Classifiers, and Multi-Layer Perceptrons — makes these frameworks versatile and applicable to a wide range of data analysis challenges. The contributions of the proposed frameworks, and thus this thesis, can be summarized as follows:

- Improved security through the implementation of polymorphic encryption, which ensures that any potential compromise of the key does not pose a threat to the overall system;
- Improved privacy through homomorphic encryption during model aggregation, introducing an additional layer of security;
- Practical reduction in communication costs through the ability to exchange parameters and gradients;
- Enhanced versatility through support for five different smart models: Support Vector Machines, Linear Regression, Naive Bayes, Stochastic Gradient Descent Classifiers and Multi-Layer Perceptrons.

1.9.1 Scientific Publications

During the research journey, six articles were published in different journals, and other four were published as conferences proceedings. In addition, three more articles are currently under review in different scientific journals. All details related to those publications can be found in Appendix A.

1.10 Thesis Structure

This thesis presents a comprehensive survey of the field of federated learning along with proposed frameworks that integrate advanced cryptographic techniques to improve privacy, security, and efficiency. The remainder of the thesis is organized into several chapters, each addressing specific aspects and facets of the research journey and presented in the form of a journal article. Through this structured approach, a coherent presentation of the proposed work, its context, and its importance in the broader landscape of Machine Learning and privacy is discussed.

Paper presented in Chapter 2 explores the concept of Federated Learning in depth, first addressing privacy concerns in Machine Learning and demonstrating how Federated Learning contributes to privacy protection. The remainder of the chapter explains the technical underpinnings of Federated Learning, introduces a taxonomy for Federated Learning, distinguishes its key characteristics from those of other Machine Learning methods, and provides comprehensive insights into its various applications and challenges.

Paper presented in Chapter 3 is dedicated to the exploration of aggregation algorithms in Federated Learning. It begins with an explanation of the architecture of this concept and then delves into the different types of messages that are exchanged between the server and the clients in a Federated Learning system. The chapter further deals with a comprehensive analysis of different aggregation approaches and culminates with an examination of the current state-of-the-art implementations of Federated Learning aggregations, accompanied by a thorough presentation of the challenges involved.

Paper presented in Chapter 4 takes an in-depth look at the security and privacy issues that arise when implementing aggregation algorithms in FL. The chapter begins with an explanation of the various threats and attacks in Federated Learning. It then provides a thorough and comprehensive analysis of aggregation algorithms that have been developed with security and privacy considerations in mind. Furthermore, this chapter puts these secured aggregation

algorithms under the microscope, highlighting shortcomings and vulnerabilities and providing empirical evidence that underscores the central problem addressed in this research.

Papers presented in Chapter 5 is devoted to a detailed examination of the frameworks presented in this thesis. In first section, a paper discussing PolyFLAG_SVM is presented, while the second paper discusses PolyFLAM and PolyFLAP and the third discusses HP_FLAP. The four novel frameworks are explained in detail and their structure and workflows are explained. A comprehensive evaluation is then conducted, considering various facets and contrasting these frameworks with the current state of the art. Finally, the chapter culminates in the practical application of the proposed frameworks through real-world data testing, illustrating their utility and effectiveness through use case studies.

Chapter 6 concludes with a comprehensive executive summary that summarizes the entire body of work and provides key findings and reflections on the contributions and implications of this research. Also, it explores challenges and future prospects related to the proposed frameworks, highlighting potential barriers and opportunities for further research and development. Finally, the chapter presents a comparison between the proposed frameworks and the existing frameworks.

CHAPTER 2

Reviewing Federated Machine Learning: A General Overview of the Domain

Published in Sensors Journal 2023
Under the special issue ***Smart Environments for Health and Well-Being***
Volume 23; Issue 4; doi: 10.3390/s23042112

Résumé: Dans le monde réel, la collecte de données constitue un défi majeur, sinon le plus grand, dans le développement de modèles ML pour plusieurs raisons, dont la plus importante est la confidentialité. Ce chapitre examine en détail l'apprentissage fédéré en abordant d'abord les problèmes de confidentialité dans l'apprentissage automatique et en démontrant son rôle dans la protection de la confidentialité. Il se penche ensuite sur les aspects techniques de l'apprentissage fédéré, introduit une taxonomie pour le concept, met en évidence ses caractéristiques distinctives par rapport à d'autres méthodes d'apprentissage machine, et fournit une exploration complète de ses diverses applications et défis. Par conséquent, l'article présenté dans ce chapitre contribue de manière significative à l'état de l'art dans le domaine de l'apprentissage fédéré. Il propose des définitions complètes de l'apprentissage fédéré, clarifiant ses principes fondamentaux et ses applications. De plus, l'article établit une taxonomie de l'apprentissage fédéré, catégorisant ses différentes techniques et approches, et délimite les frontières entre l'apprentissage fédéré et d'autres technologies d'apprentissage automatique, mettant en avant les caractéristiques uniques et les avantages de l'apprentissage fédéré.

sensors

MDPI

Review

Reviewing Federated Machine Learning and Its Use in Diseases Prediction

Mohammad Moshawrab [1], Mehdi Adda [1,*], Abdenour Bouzouane [2], Hussein Ibrahim [3] and Ali Raad [4]

[1] Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC G5L 3A1, Canada
[2] Département d'Informatique et de Mathématique, Université du Québec à Chicoutimi, 555 Boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada
[3] Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, QC G4R 5B7, Canada
[4] Faculty of Arts & Sciences, Islamic University of Lebanon, Wardaniyeh P.O. Box 30014, Lebanon
* Correspondence: mehdi_adda@uqar.ca; Tel.: +1-(418)833-8800 (ext. 3197)

Abstract: Machine learning (ML) has succeeded in improving our daily routines by enabling automation and improved decision making in a variety of industries such as healthcare, finance, and transportation, resulting in increased efficiency and production. However, the development and widespread use of this technology has been significantly hampered by concerns about data privacy, confidentiality, and sensitivity, particularly in healthcare and finance. The "data hunger" of ML describes how additional data can increase performance and accuracy, which is why this question arises. Federated learning (FL) has emerged as a technology that helps solve the privacy problem by eliminating the need to send data to a primary server and collect it where it is processed and the model is trained. To maintain privacy and improve model performance, FL shares parameters rather than data during training, in contrast to the typical ML practice of sending user data during model development. Although FL is still in its infancy, there are already applications in various industries such as healthcare, finance, transportation, and others. In addition, 32% of companies have implemented or plan to implement federated learning in the next 12–24 months, according to the latest figures from KPMG, which forecasts an increase in investment in this area from USD 107 million in 2020 to USD 538 million in 2025. In this context, this article reviews federated learning, describes it technically, differentiates it from other technologies, and discusses current FL aggregation algorithms. It also discusses the use of FL in the diagnosis of cardiovascular disease, diabetes, and cancer. Finally, the problems hindering progress in this area and future strategies to overcome these limitations are discussed in detail.

Keywords: federated machine learning; federated learning; privacy preservation; aggregation algorithms; diseases prediction; cardiovascular diseases; diabetes; cancer; smart wearables; smart health

Citation: Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Reviewing Federated Machine Learning and Its Use in Diseases Prediction. *Sensors* **2023**, *23*, 2112. https://doi.org/10.3390/s23042112

Academic Editor: Alessandro Bevilacqua

Received: 20 January 2023
Revised: 4 February 2023
Accepted: 9 February 2023
Published: 13 February 2023

1. Introduction

Artificial intelligence (AI) is a rapidly advancing technology that is increasingly being integrated into various industries and aspects of daily life, leading to significant changes and advancements in the way we live and work. This truth is obvious and can be seen with one's own eyes; no evidence is needed to prove it. Ever since Alan Turing, considered the father of theoretical computer science and artificial intelligence, asked their famous question, "Can computers think?" [1], artificial intelligence has become a broad field of research. Despite the fact that AI has been researched for a long time, there is no single definition for this field. The authors in [2] defined it as a set of tools and techniques that use principles and devices from various fields such as computation, mathematics, logic, and biology to address the problems of realizing, modeling, and mimicking human intelligence and cognitive processes, while the authors in [3] defined it as programs that, in an arbitrary world, will cope no worse than a human.

Machine learning (ML), a derivative of AI, allows computers to "learn" from training data and expand their knowledge over time without being explicitly programmed. Machine learning algorithms attempt to find patterns in data and learn from them to make their own predictions. In short, machine learning algorithms and models learn through experience. Traditionally, a computer program is developed by engineers and given a set of instructions that enable it to turn incoming data into its intended output. ML, by contrast, designs the program to learn with little or no human interaction and to expand its knowledge over time. The remarkable success of ML, as well as its enormous potential in classification and regression problems and its ability to use both supervised and unsupervised learning techniques, have made it attractive to researchers in many fields. Later studies revealed the variety of applications of ML that can be observed in the field such as:

- E-commerce and product recommendations [4,5];
- Image, speech and pattern recognition [4,5];
- User behavior analytics and context-aware smartphone applications [4,5];
- Healthcare services [6–8];
- Traffic prediction and transportation [4,9];
- Internet of Things (IoT) and smart cities [9];
- Cybersecurity and threat intelligence [10];
- Natural language processing and sentiment analysis [11];
- Sustainable agriculture [12];
- Industrial applications [13].

1.1. Machine Learning under The Scope: Challenges

Accurate results in classification or regression are increasingly encouraging the incorporation of these techniques into areas of daily life. The feasibility of using AI tools, and in particular ML, has been demonstrated by the high performance they offer and the possibility of implementing them in different domains. However, ML still suffers from several challenges that are extensively described and discussed in the literature. However, these challenges are not classified into a single taxonomy, but grouped according to different aspects. In this section, the common challenges are presented under a proposed taxonomy based on data-related, model-related, implementation-related, and other general aspects. In addition, these challenges are illustrated and summarized in Figure 1 below.

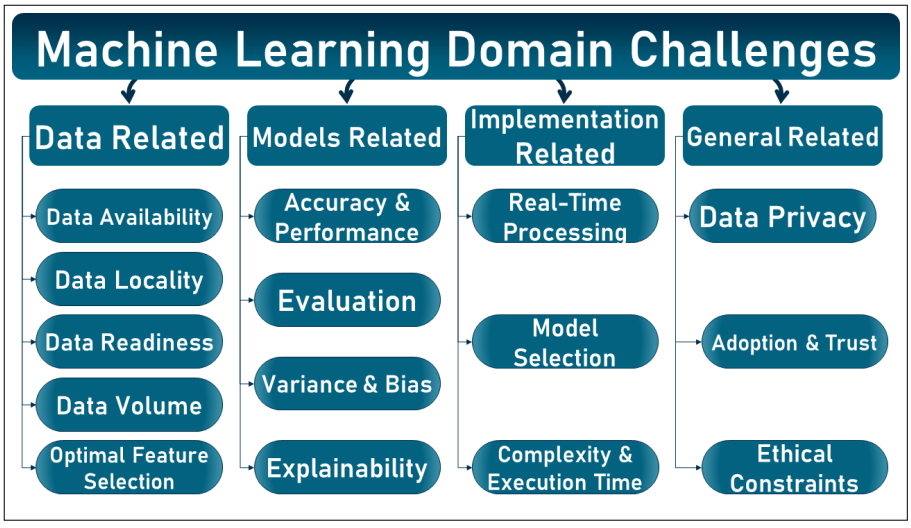

Figure 1. Machine Learning Domain Challenges.

1.1.1. Data Related Challenges

Machine learning algorithms are typically implemented in a known pipeline consisting of data collection, preprocessing, exploration, model selection, training, evaluation, and

deployment. Data, which constitute the main component of these algorithms, can present various challenges, such as [14,15]:

- Data availability and accessibility: to train a model, one must have the necessary data, which may not be available on the spot or may be available but inaccessible for various reasons;
- Data locality (data islands): in the real world, data are scattered in different and unrelated entities called "data islands." Due to different regulations and laws, data related to the same subject and available on different data islands cannot be accessed for use and analysis;
- Data readiness: even if data are available and accessible, several aspects should be considered, such as:
  - Data heterogeneity: available data may have different characteristics or be composed of different forms. For example, health data for the same patient may be available in different forms, such as medical images, reports, videos, and structured data. The ability to deal with such heterogeneity is a challenging task;
  - Noise and signal artifacts: due to the interaction between data acquisition instruments and other electrical devices, data can be poisoned by noisy attributes that affect the overall results of ML models;
  - Missing data: data collected by measuring devices may be incomplete for various reasons;
  - Classes imbalance: in classification problems, the data collected for one group may dominate the data collected for other groups, affecting the learning of the smart model.
- Data volume: is the amount, size, and scope of the data. In the context of ML, size can be defined either vertically by the number of records or samples in a dataset or horizontally by the number of features or attributes it contains. Data volume presents several challenges, such as:
  - Course of dimensionality: dimensionality describes the number of features or attributes that are present in a dataset. Increasing dimensionality can have a negative impact on model performance;
  - Bonferroni principle [16]: the Bonferroni principle states that when searching for a particular type of event in a given set of data, the probability of finding that event is high. Therefore, the accuracy of a ML model subject to the Bonferroni principle may be compromised.
- Feature representation and selection: the performance of ML models heavily depends on the choice of data representation or features, so selecting the optimal features will definitely improve the overall model performance.

1.1.2. Models Related Challenges:

In addition to the challenges posed by the data, the models themselves can present researchers with various problems, such as [17,18]:

- Accuracy and performance: achieving the highest accuracy for ML models remains the main goal for researchers from various fields, and the highest accuracy will lead to the highest adoption and integration of this technology;
- Model evaluation: evaluating an ML model can be challenging, especially when traditional performance metrics such as accuracy, precision, and recall do not reflect a model's feasibility;
- Variance and bias: where variance is the variability of the model prediction for a given data point or a value indicating the spread of our data, and bias is the difference between the average prediction of our model and the correct value we are trying to predict. ML models are susceptible to variance and bias, which can affect their performance, results, and confidence;

- Explainability: some of the ML models, especially deep learning models, are known by their black box identity. The lack of explanations of how they work can have a negative impact on trust in these models, even when high accuracies are achieved.

1.1.3. Implementation Related Challenges:

Assuming that the obstacles in the data and models have been overcome, implementing the models of ML can be a challenging task due to various obstacles such as [19,20]:

- Real-time processing: ML models are created and trained with available data. However, fitting these models to real-time problems presents several challenges;
- Model selection: different models can produce different results even for the same problems. For example, support vector machines (SVM) and logistic regression (LR) can lead to different results, even when working with the same data at the same point in time. Thus, selecting the optimal model and tuning its parameters are not easy tasks;
- Execution time and complexity: due to the complexity of the data or models, multiple preprocessing steps, and many other reasons, ML models can require enormous computing power and long execution times.

1.1.4. General Challenges:

Finally, other challenges besides technical aspects can be mentioned in this section, such as [17,18]:

- User data privacy and confidentiality: which is one of the most critical issues in the field of ML. Users tend not to share their data for various reasons, which affects the availability of the data and jeopardizes the entire ML cycle;
- User technology adoption and engagement: due to privacy issues, unclear results, lack of explanation, and other reasons, users may not accept ML being integrated into their daily routine, or even accept its results;
- Ethical constraints: various ethical constraints posed by ML have been widely discussed in the literature, such as control and morality, model ownership, environmental impact, and many others.

1.2. Privacy Challenge: Federated Machine Learning Motivation

The challenges in machine learning and its derivatives have been thoroughly studied, and researchers are trying to find answers to all of them without focusing on just one. Nevertheless, the workflow of ML mainly consists of data acquisition and preprocessing, feature engineering, model training, model evaluation, and model deployment. The structure of the workflow reflects the importance of data in ML. The performance of ML models heavily depends on the availability of data. Although achieving highly accurate models depends on the technical structure of the models themselves, the cleanliness and readiness of the data, the optimal feature selection, and many other aspects, it is well known that the availability of more data to train the models increases their accuracy [14,15]. However, in the real world, data collection is a big challenge, if not the biggest, in developing ML models for several reasons, most importantly privacy and confidentiality.

Not only individuals, but also society, governments, and organizations are strengthening the protection of data privacy and security. In this regard, several regulations and laws were enacted, such as the European Union's General Data Protection Regulation (GDPR) [21], China's Cyber Security Law of the People's Republic of China [22], the General Principles of the Civil Law of the People's Republic of China [23], the PDPA in Singapore [24], and hundreds of principles legislated around the world. While these regulations help protect private information, they pose new challenges to the ML field by making it more difficult to collect data to train models, which in turn makes it more difficult to improve the accuracy of model performance and to personalize those models. Consequently, data privacy and confidentiality are not a stand-alone challenges, but also trigger other challenges for ML, such as data availability, performance, personalization, and thus trust and acceptance.

Overcoming Privacy Challenges

The criticality of privacy has been a hot research topic for years, pushing to find different solutions to protect the information exchanged by subjects. To this end, various privacy algorithms were proposed, such as encrypting data before exchange through various algorithms such as differential privacy [25], k-order anonymity [26], homomorphic encryption [27], and other methods. However, these methods were not able to provide definitive and unbreakable solutions, as several attacks have been observed in ML such as the model inversion attack [28] and the membership inference attack [29], which are able to derive raw data by accessing the model.

Recently, Google proposed a new concept in the machine learning domain known as "federated machine learning" or "federated learning" [30]. The main concept behind FL is to eliminate the exchange of user data between peripherals. FL is a type of collaborative distributed/decentralized ML privacy-preserving technology where a model is trained without the need to transfer data from the edges to a central server, but models are sent to peripherals to be trained on local data, and then sent back to a central aggregation server to generate the global model without knowing the embedded data.

Federated learning has proven to be a great solution to user privacy issues, opening the door to collecting more data to train ML models and improve their accuracy and efficiency. Moreover, FL enables training models with data from different entities known as data islands and merging the knowledge into a global trained model, which increases the efficiency of the models. In addition, FL enabled the handling of heterogeneous data scattered in different data spaces with different characteristics, and facilitated the so-called "learning transfer" where models can share their knowledge without transferring users' private data. Nevertheless, FL is still in its infancy and is still vulnerable to various challenges.

1.3. Machine Learning and Healthcare

The development of information and communication tools, in parallel with the emergence of artificial intelligence and its branches such as ML and DL has produced effective solutions to health challenges. Moreover, AI is considered the most promising technology for improving healthcare services, as it can be applied to almost all areas of medicine and will revolutionize healthcare delivery to patients and populations. This tremendous contribution is not due to magic, but to AI's data processing capabilities that surpass those of humans, especially in terms of its ability to perform large calculations in a short period of time. Given the promise, initiatives to use AI as a solution to healthcare problems have recently significantly expanded, with the number of AI healthcare applications exceeding thousands in the last decade [31,32].

AI is playing an increasingly important role in healthcare and has the potential to revolutionize the way healthcare professionals diagnose, treat, and monitor patients. One of the most important ways in which AI can be used in healthcare is to analyze large amounts of medical data. By using machine learning algorithms to identify patterns and trends in these data, AI can help medical professionals make more accurate diagnoses, predict which patients are at risk of developing certain diseases, and develop more personalized treatment plans [33]. AI can also be used to monitor patients' health and vital signs in real-time, and to alert medical professionals to potential problems. This can be particularly useful for patients with chronic conditions who need close monitoring to avoid complications. For example, using AI in smart wearables, a person's heart rate and blood pressure can be continuously monitored and the data analyzed to detect the early signs of cardiovascular diseases (CVDs), as shown in [34]. In addition, smart wearables equipped with sensors and machine learning algorithms can play a critical role in detecting and monitoring diabetes by continuously tracking and analyzing biometric data such as blood glucose levels, heart rate, and activity levels, enabling early detection and intervention [35]. In addition, the potential of smart wearables and machine learning models in detecting fatigue in the workplace has been shown to be highly feasible, contributing to disease prevention [36]. Overall, AI has the potential to significantly improve the quality of healthcare for patients and make healthcare more efficient and cost-effective. However, it also

raises ethical and legal issues that need to be addressed for the successful implementation of AI in healthcare.

With healthcare being of critical importance, the performance of ML in healthcare needs to be enhanced. Increasing this performance requires using the latest techniques and overcoming any barriers that may impede progress. The barriers to the development of the use of ML in healthcare are the same for all ML implementations in all diseases and correspond to the previously described problems. Therefore, potential solutions that can help promote the use of ML will lead to improved applications in these areas.

1.4. Outline and Main Contributions of This Article

In this article, FL and its use in disease prediction and diagnosis have been studied. To achieve this goal, this article explores this topic in depth in the following sections. In Section 2, FL is discussed from various perspectives, including technical perspectives, aggregation algorithms, and others. Then, in Section 3, the use of Federated learning technology in detecting and predicting various diseases is presented by listing the state-of-the-art in each area and discussing the implementations mentioned in the literature. Later, in Section 4, the challenges that hinder the progress in this field are discussed and therefore some future perspectives that could help in overcoming these challenges are proposed. In this context, this article attempts to answer the following questions:

- What is federated machine learning?
- What are the motivations for this technology?
- What are the technical perspectives on which FL is based?
- What taxonomy can be used to classify FL algorithms and techniques?
- What are the differences between FL, traditional ML (including deep learning), distributed and decentralized ML, and federated database systems?
- What are the existing FL aggregation algorithms and what is the contribution of each?
- What are the available FL frameworks?

The topic of federated learning has been a hot and trending topic in recent years. As a result, dozens if not hundreds of studies have already addressed this topic, with a large number of these studies reviewing federated learning. However, none of the articles proposed an inclusive and full taxonomy for FL, or even compared FL to classical ML, decentralized ML and federated database systems. Furthermore, the federated aggregation algorithms were not reviewed with any of the previous studies. Furthermore, the use of FL in diseases prediction such as CVDs and diabetes were not reviewed. Consequently, this article proposes several new ideas, contributing to the body of FL knowledge by:

- Proposing a novel and comprehensive taxonomy that classifies FL into the maximum number of possible categories;
- Establishing clear and precise boundaries to distinguish between FL, traditional ML, distributed and decentralized ML, and federated database systems;
- Discussing existing aggregation algorithms in FL and evaluation of the contributions of each to the field;
- Reviewing and discussing the state of the art of FL in diagnosing:
  - Cardiovascular disease;
  - Diabetes;
  - Cancer.
- Presenting the challenges faced by FL and the possible future perspectives that can be pursued to increase the efficiency of the technology.

2. Federated Learning

Artificial intelligence and its derivatives, such as machine learning and deep learning, are gaining attraction and confidence in a variety of fields. For example, deep learning surpassed human performance in the game of Go, where AlphaGo and AlphaGo Zero achieved superhuman feats by beating the world champions of the game. However,

the high accuracy achieved by these models required that they be trained on data that spanned 29 million records [37,38]. This underscores the description of such technologies as "data hungry," with the need to improve the accuracy of the models requiring larger datasets. This is undoubtedly the case not only in gaming, but also in other sectors such as education, industry, healthcare, and others. Moreover, this is not the only problem that hinders the development of ML and DL. With the development of ICT tools and especially mobile networks, data collection has become easier and larger datasets are being obtained. However, an urgent problem that requires effective solutions is the privacy and security of data, with the disclosure of information about individuals never being a minor issue and recently attracting the attention of both governments and researchers [39,40].

2.1. Overview and Definition(s)

The increasing efficiency of artificial intelligence tools is leading them to be used in various areas of life. However, the challenges faced by these technologies lead researchers to always look for appropriate solutions, which is why federated learning, or what is sometimes called federated machine learning, was found.

2.1.1. Data Islands and Privacy Dilemma: Concept behind FL

The ability to collect and analyze large amounts of data has recently made great strides, especially with the development of communication tools and AI methods. However, data are collected in what are known as "data islands." Data islands are defined as foundations, institutions, individuals, or other entities where data are collected and stored [41,42]. To improve the performance of AI models, the idea of the centralized server is pursued, with the common method being to collect data in a centralized repository and perform unified processing, cleaning, and modeling. For example, a patient's health data scattered across different hospitals, clinics, or health centers have the greatest potential when analyzed together [43]. However, privacy regulations and restrictions, as well as data heterogeneity, limit the ability to collect and simultaneously analyze such data. Consequently, the search for solutions to the data islands and privacy dilemma has attracted the attention of researchers worldwide and was the motivation for the concept of federated learning [44]. In Figure 2, the concept of data islands is illustrated by showing how medical data may be stored in different institutions and cannot be shared due to the sensitivity of the health data. Instead, the parameters are shared with the FL server as shown in the figure.

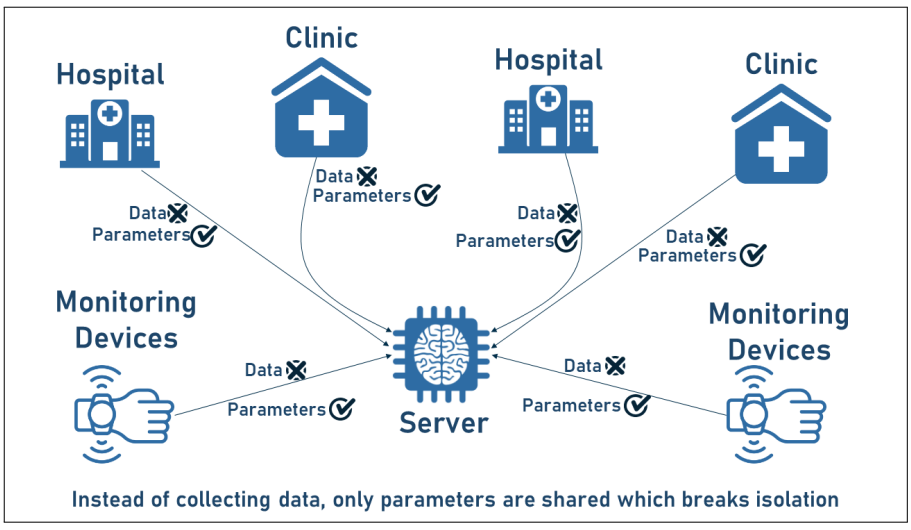

Figure 2. Data islands concept illustrated by medical entities.

2.1.2. Motivations behind FL Concept

The critical importance of data protection has led to the development of various protection algorithms aimed at protecting data through encryption or other methods, but they have failed to provide an inevitable solution against attacks. Moreover, the data annotation in some fields, such as the medical industry, relies on the knowledge of professionals, resulting in a rarity of valid data that are detrimental to industrial development. Accordingly, the need to deal with private data or data scattered in islands while maintaining their privacy is the main motivation behind the concept of federated learning [45]. The fact that private and confidential data available in scattered sources are more usable for ML models than those centralized on the server provides FL with the opportunity for collaboration between these data sources to improve the accuracy of ML models [46]. Because the data can be analyzed without having to be transferred to a central server, FL helps address the challenges mentioned earlier. The FL architecture, communication methods, security mechanisms, and other factors allow the model to be trained on edge devices, the data islands, by sending them the model itself, rather than collecting and aggregating the data in a centralized space [47]. In other words, instead of aggregating training data from different sources, FL enables the training of the shared global model using a central server, while keeping the data at their main sources of origin [48]. This not only preserves data privacy, but also reduces data transfer costs by limiting the transfer to only the necessary parameters rather than the entire datasets. This also allows dealing with a scalable number of devices, ranging from ten to ten million [49,50]. All in all, FL is an emerging and promising technology that helps one solve the ML challenges by preserving data privacy, increasing the model performance, reducing the data transfer costs, improving scalability, and more. Therefore, it has the potential to challenge the prevailing ML paradigm [51,52].

2.1.3. FL Definition(s)

Federated learning was originally introduced by Google in 2016, where it was used in Google Keyboard to predict users' text input on tens of thousands of Android devices without transferring data from the devices to central servers [30]. However, the authors in [43] claim that the term FL was introduced before and that its core idea is distributed deep learning, such as the privacy-protected deep learning system proposed in [52]. Although it is still considered a new concept, it is increasingly attracting researchers' attention, and its definition can be found in various places in the literature. For example, the authors in [42,45] define it by explaining how it works, mentioning that federated learning is a type of collaborative distributed/decentralized machine learning technology where privacy is maintained and a model is trained without the need to transfer data from the edges to a central server, but instead weight updates are sent to a central aggregation server to build the global model. A statistical definition is given in both [41,44], where FL is defined as follows:

> "Define N data owners {F1, ...FN}, all of whom wish to train a machine learning model by consolidating their respective data {D1, ...DN}. A conventional method is to put all data together and use D = Di ∪ ... ∪ DN to train a model MSUM. A federated learning system is a learning process in which the data owners collaboratively train a model MFED, in which process any data owner Fi does not expose its data Di to others. In addition, the accuracy of MFED, denoted as VFED should be very close to the performance of MSUM, VSUM. Formally, let δ be a non-negative real number, if $|VFED - VSUM| < \delta$. We say the federated learning algorithm has δ-accuracy loss."

2.2. FL Technical Inspection

The potential for federated learning lies in the architecture upon which it is built. To understand this structure, it is necessary to study the various aspects of this technology and its various parts, which will be presented in this section.

2.2.1. Underlying Architecture

Federated learning is a collaborative decentralized approach of machine learning where data are analyzed by the model without being transmitted from the edges to the central server, which acts as an aggregator. This is made possible by the architecture behind this technology. The technical architecture of FL consists of the three main components: the parties, the manager, and the communication framework, which are discussed below [41,42,44]:

- Parties: are also referred to as customers, users, or individuals, and are the data owners and beneficiaries of FL. They are indicated by:
  - Hardware specifications such as storage, processing, and power capacities;
  - Scalability and stability;
  - Data distribution.
- Manager: known as a server or aggregator, is the high-performance central server that acts as a model aggregator rather than a data collector;
- Communication–computation framework: the computation handles the model training and the communication handles the exchange of model parameters between the parties and the manager. Several frameworks were developed to manage the relationships between different FL entities, which are discussed in detail later;

In the various available frameworks for communication computation, the steps taken in the application of FL differ but they share a common basic concept which is:

- The parties federally train their own model using their local data without sharing it;
- The global model is updated by the locally trained models;
- The global model is then shared with the different parties/data owners;
- The above steps are repeated until the global model achieves the desired performance.

2.2.2. FL Communication-Computation Frameworks

The different FL communication–computation frameworks are due to the different centralized concepts. Currently, there are two FL concepts: centralized managers and decentralized managers. Each of these concepts manages communication between parties differently, where [46]:

- Centralized design (client-server architecture): in this approach, data flow is often asymmetric, with the manager aggregating information from the parties and sending them back to the updated model. In addition, communication between the manager and the parties can be synchronous or asynchronous;
- Decentralized design (peer-to-peer architecture): In this approach, communication is performed between the parties themselves without the need for a central manager. This allows each party to directly update the global parameters.

The above concepts are currently implemented in various FL frameworks which will be discussed later. Two popular FL architectures are mentioned below: the centralized federated average (FedAvg) [30] and the decentralized FL framework [53], which are discussed and explained below as well as shown in Figure 3:

- Federated average learning, which is the basis of FL and is determined in the following steps:
  - The manager sends the model to the parties involved;
  - The parties train the received model with their local data;
  - The updated models are sent back to the manager;
  - The above steps are repeated until the model achieves the desired performance.
- Decentralized federated learning SimFL, where no central manager/server is required. In this framework, the following steps are applied:
  - The parties first update the gradients of their local data;
  - Then, the gradients are sent to a selected party;
  - Next, the selected party uses its local data and the gradients to update the model;

– Then, the model is sent to all other parties;
– To ensure fairness and to use the data from the different parties, each party is selected to update the model for approximately the same number of rounds and the above steps are therefore repeated until the final model is reached.

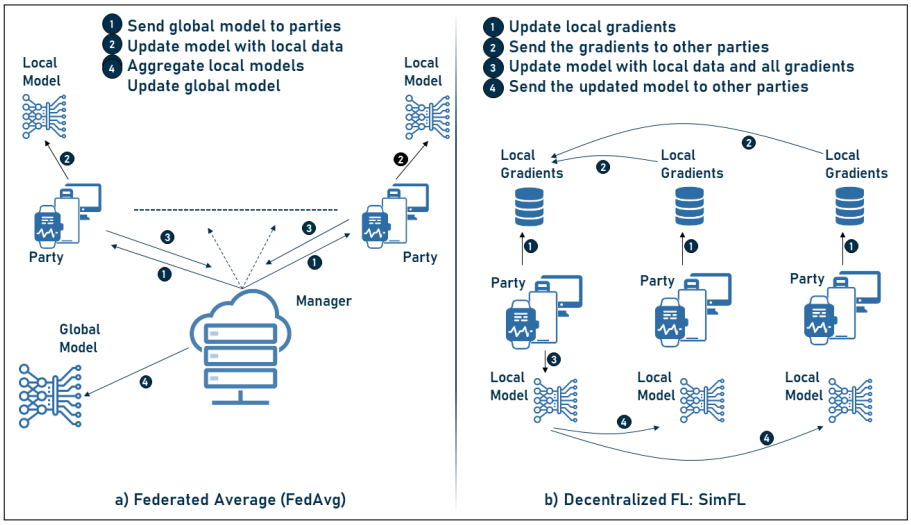

Figure 3. Communication–computation frameworks.

2.3. Federated Learning Taxonomy

The different ways of applying federated learning have contributed to the creation of numerous classifications within this technology, which can be considered differently according to the different subdivision bases or points of view. In light of this, the study of the literature in FL concludes to subdivide it based on six approaches, which are listed below and explained in this section:

- Data partitioning;
- Machine learning model;
- Privacy mechanism;
- Communication architecture;
- Scale of federation;
- Motivation of federation.

2.3.1. Data Partitioning

Federated learning provides the ability to train models without the need to collect data from edge devices. In addition, in the FL environment, a device's local storage of data samples (pictures, documents, etc.) is considered its sample space. On the other hand, the feature space is the collection of characteristics used to characterize the data points, often expressed as a vector with a large number of dimensions. This set of characteristics may be put to use in a wide range of classification and regression applications. FL is able to develop a model that can efficiently aggregate information from the various sample and feature spaces, which are typically dispersed throughout the parties (clients, users, etc.). Depending on the data structure and point of view, the samples and features in federated machine learning (FL) may be seen as rows or columns. Traditional machine learning uses a table-like data structure with rows representing samples and columns representing features; in FL, however, the samples are generally dispersed over numerous devices or locations, leading to a lack of unified data structure. If this is the case, we may think of the samples as columns and the features as rows, with each feature being shared across all devices. Finally, the representation is determined by the nature of the issue and the FL

technique used. Figure 4 below describes the difference between features and samples in both traditional ML and federated ML.

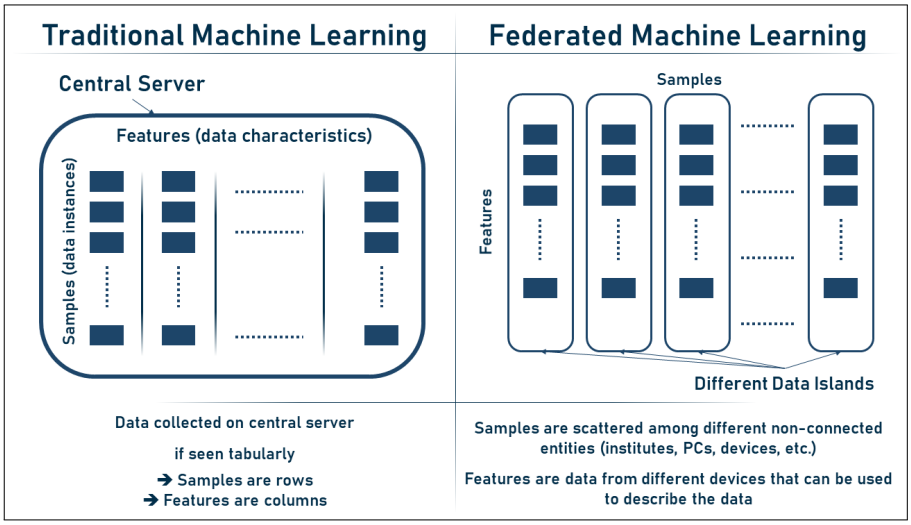

Figure 4. Samples vs. features in traditional and federated ML.

In this context, the different forms of data partitioning in federated learning environments form three categories that are described below [41,42,44].

- Horizontal FL: also known as sample-based federated learning, and is the case when the data on the parties share the same feature space but differ in the samples. In other words, in horizontal FL partitioning, the datasets are partitioned horizontally (by parties), and then the part of the data that have the same features but the parties are not exactly the same is taken out for training. It is therefore characterized by the following:
  - Is the most commonly used data partitioning strategy in implementations of FL;
  - Is suitable to increase the sample size;
  - Can train the local models using their local data with the same architecture, since these data share the same feature space;
  - Simplifies the update of the global model by averaging all local models.
- Vertical FL: also known as feature-based learning, when the data share the same or similar sample space (parties) but differ in the feature space (data). In other words, in vertical FL partitioning, the dataset is split vertically (by features), then part of the data where the parties are the same but the features are not exactly the same are taken out for training:
  - Which is challenging in terms of implementation;
  - Which makes it more complex to update a global model by averaging because the data may not be similar between parties;
  - Which has much more room for improvement to be applied in more complicated ML approaches.
- Federated transfer learning: this is the case when the datasets scattered between the parties differ not only in the samples but also in the feature space. In this partitioning method, the data are not segmented, but the learning is transferred to overcome the lack of data or tags. Therefore, it is characterized by:
  - Being an effective way to protect both data security and user privacy while breaking the boundaries of data islands;
  - Enabling the transfer of knowledge from one domain to another for better learning outcomes;

– Offering plenty of room for growth to make it more flexible with different data structures;
– Triggering the issue of communication efficiency.

Furthermore, Figure 5 below illustrates the three categories of federated machine learning divided by the type of data.

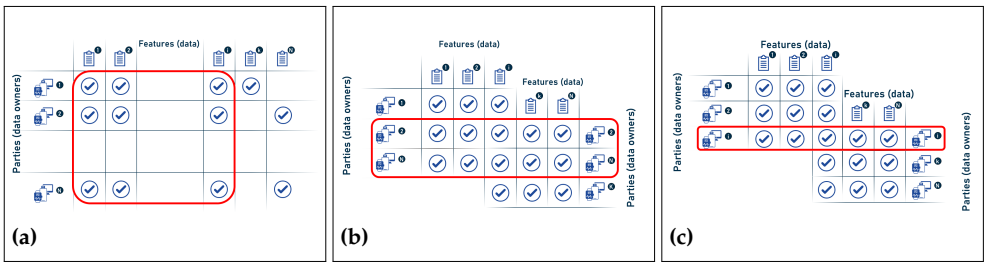

Figure 5. (**a**) Horizontal FL; (**b**) vertical FL; and (**c**) federated transfer learning.

In Table 1, the differences between the alternate groups of FL, classified based on the type of data, are summarized.

Table 1. Differences among FL groups divided by type of data.

	Horizontal Transfer Learning	**Vertical Transfer Learning**	**Federated Transfer Learning**
Data Distribution Similarity	Same	Different	Different
Output/Label Space Similarity	Different	Same	Same
Type of Task	Single task	Single ask	Federated task

2.3.2. Machine Learning Models

Federated learning was created to overcome problems with machine learning algorithms. Therefore, it is of great interest to train a modern ML model for a specific task. Researchers have worked diligently to develop new models or reinvent existing models to fit the federated learning architecture. For example, the ML models used in FL include but are not limited to: [41,42,44]:

- Linear models: support vector machines, linear regression, ridge regression, lasso regression, among others;
- Decision tree: gradient boosting, decision trees, random forests, among others;
- Neural networks: convolutional neural networks, multi-layer perceptron, deep neural networks, and others.

2.3.3. Privacy Mechanism

It is clear that the main goal behind the development of FL technology is to protect the privacy of the data of individuals, organizations, and companies participating in the machine learning process. The main concept to preserve this privacy is that the parties involved do not share their data with other entities, but only exchange some model parameters. However, these parameters may still reveal sensitive information about the data. FL was exposed to several attacks that may occur at any stage of the process of FL, including the inputs, the learning process, and the learned model [54]. In the list below, several attacks are discussed and detailed based on the model stage targeted by the Machine Learning attack [46]:

- Inputs: During this phase, malicious parties can perform "data poisoning attacks" [55–57], in which the labels of the training samples with a particular class are changed so that the final model performs poorly on that class;

- Learning process: during this process, parties can perform "model poisoning attacks" [58,59] or Byzantine fault [60,61] to upload some designed model parameters at the local model level. Such attacks can negatively affect the accuracy of the learning process due to the poisoned local updates;
- The learned model: once the learned model is published, it is exposed to attacks such as model inversion attack [28] and membership inference attack [29] and others. Such attacks can potentially infer raw data by accessing the model. For example, they can determine whether a particular dataset was used in the training process. Finally, inference attacks can also be performed in the FL manager learning process, where the server has access to the parties' local updates.

To overcome such problems and achieve the goals, various approaches such as model aggregation, cryptographic methods, and differential privacy are used in Federated Learning systems. These techniques help avoid the risk of attacks and backdoors and are described below [41–43]:

- Model aggregation: is one of the most common privacy preserving mechanisms in FL systems and the main concept behind the FL technique, where the global model is trained by aggregating the model parameters of all parties without sharing the original data in the training process;
- Cryptographic methods: In this approach, the parties must encrypt their messages before sending them to the manager or other parties, work with the encrypted messages, and decrypt the encrypted output to obtain the final result. In this context, various algorithms have been used in FL systems, such as:
  - Homomorphic encryption [39]: Users can compute and process the encrypted data without revealing the original data, and at the same time the user decrypted the processed data with the key, which is exactly the expected result. However, due to the additional encryption/decryption operations, homomorphic encryption incurs extremely high computational overhead;
  - Secure multiparty computation (SMC) [62]: in this algorithm, the server is guaranteed to learn the parties' inputs only in their entirety. However, SMC does not provide any confidentiality guarantee for the final model, which is still vulnerable to inference and model inversion attacks and can also be a reason for additional computational overhead.
- Differential privacy [63]: is a new definition of privacy in which the final results of the model are insensitive to the changes of a particular dataset by minimizing the impact of a single dataset on the computation of the results. This method has been proven successful for data poisoning attacks, but may not be usable for model poisoning attacks.

2.3.4. Methods for Resolving Heterogeneity

The different equipment of the parties involved in the FL system and the diversity of the data stored in them can have a negative impact on the efficiency of the overall learning process. To solve the problems caused by this heterogeneity, four types of distractions are used in FL implementations [41]:

- Asynchronous communication: the synchronous scheme can be easily disrupted by the diversity of devices. Therefore, asynchronous communication can help resolve this diversity;
- Device sampling: limiting the use of a party/device to only the necessary iterations, not necessarily participating in every single iteration;
- Fault-tolerant mechanism: in an environment with multiple working participants, the failure of one participant can affect the performance of the entire environment. A fault-tolerant mechanism helps prevent the entire system from collapsing if one of the parties fails;
- Model heterogeneity: is used to resolve data heterogeneity and includes three strategies:

- Each individual party has its own model;
- A global model that is suitable for all parties;
- Relevant learning models for tasks.

2.3.5. Communication Architecture

Following the various client/server approaches taken in FL systems, there are two main categories in communication architecture, which are [46]:

- Centralized design: this assumes the existence of a central server that aggregates the local models trained by the parties and sends them back for updating. Communication between the manager and the local parties can be synchronous or asynchronous;
- Decentralized design: in this approach, communication is between the parties, and each can directly update the global model without the need for a central aggregation manager.

2.3.6. Scale of Federation

Federated learning can be classified into two groups based on the scale of federation, namely: cross-silo FLS and cross-device FLS [42,46]. These two approaches differ in the number of parties and the amount of data stored in each party, where [64]:

- Cross-silo FL: this approach is used when the participating parties are fewer in number, have relatively large amounts of data, have relatively high computational power, and are available for all rounds of learning. This approach is best suited when the participants are organizations or computers;
- Cross-Device FL: in contrast, the number of parties involved in the learning process is relatively large, they have a small amount of data, and are equipped with relatively low computing power. This approach is best suited when the participants are mobile devices.

2.3.7. Federation Motivation

Finally, the reasons for using FL systems can be categorized as follows [46]:

- Regulations: where laws restrict the sharing of private information between different companies, such as the GDPR, Chinese laws, or PDPA or other laws;
- Incentives: where FL is motivated by a desire to develop services.

The various categories of federated learning that may be obtained from grouping various points of view are outlined below in Table 2 along with a summary of the advantages associated with each category.

2.4. Federated Learning: Borderlines

Federated learning is the result of the accumulation of technological improvements in machine learning. Motivated by privacy preservation, inspired by the concept of distributed computing, and executed by advanced communication technologies, FL has become an efficient and feasible technology. In this section, we highlight the limitations of FL systems to differentiate them from traditional and previous ML technologies.

2.4.1. FL vs. Classic ML

Both FL and classical ML aim to optimize the learning goal. However, they differ in the architecture of their models. Since the classical ML can be implemented in both centralized and distributed approaches, this section compares FL only with the centralized classical ML, while the comparison with the distributed ML is performed in the next section. Centralized classical ML is the concept where data characterized by the same features are collected from different users on a central server where they are then processed and analyzed. In this context, the two concepts are compared using [47]:

- Motivation: classical ML focuses on the learning goal, while FL focuses on both the learning goal and privacy;

- Data identity: in classical ML, user data are described as independently and identically distributed (IID), while in FL, it is possible to deal with unbalanced non-IID data coming from different parties, be it individuals or institutions;
- Centralization: in the classical ML, all data and computations are centralized around one server, while FL provides both centralized and distributed server architecture;
- Data access: in the classic ML, the central server has full access to the user data, while this is not the case in FL;
- Communication and data transfer: in classic ML, all the user data are fully transmitted to the central server, while in FL, only minimal parameters or trained models are exchanged.

Table 2. Summarized Taxonomy for Federated Learning Systems.

Taxonomy	Category	Structure	Advantages
Data partitioning	Horizontal FL	Different parties and similar data features	Holds larger variety of parties
	Vertical FL	Similar parties and different data features	Holds larger variety of data features
	Federated transfer learning	Different parties and different data features	Holds larger variety of parties and data features
Machine learning models	Linear models	Linear regression, ride regression, lasso regression	Ease of implementation
	Decision tree	gradient boosting, decision trees, random forests	Accurate, stable, and can map non-linear relationships
	Neural networks	-	Learning capabilities, highly robust and fault-tolerant
Privacy mechanisms	Model aggregation	Central manager learns by aggregating the locally trained model	Avoid transmitting original data
	Cryptographic methods	Using encryption algorithms such as homomorphic encryption and secure multi party computation (SMC) to encrypt the messages exchanged among parties	Enables the calculation and processing of encrypted data
	Differential privacy	Reducing the impact of a single data record on the calculation of the global model	Reduce the effect of data poisoning attacks
Methods for solving heterogeneity	Asynchronous communication sampling	To resolve the heterogeneity of parties	Solve the problem of communication delays and avoid simultaneous training with heterogeneity of parties
	Fault-tolerant mechanism	To resolve the failure of parties	Prevent whole system from collapsing if one of the parties failed
	Heterogeneous model	To resolve the heterogeneity of data	Resolve the issue of models diversity
Communication architecture	Centralized design	Architecture controlled by a central aggregation manager/server	Reduces communication cost
	Decentralized design	Communication performed among parties without the existence of a central manager/server	Reduces the risk of backdoor attacks
Scale of federation	Cross-silo FL	Parties are less in number, hold large amounts of data and equipped with high computation power	Fits for FL among institutions
	Cross-device FL	Parties are high in number, hold less amount of data and equipped with less computation power	Fits for FL among individuals
Motivation of federation	Regulations	Motivated by laws such as GDPR and others	Enhancing ML services
	Incentives	Motivated by desire of updating some services	

2.4.2. FL vs. Distributed and Decentralized ML

The architecture of the FL system is based on the concept of distributed computing. Therefore, FL is considered a collaborative distributed learning technology. On the other hand, distributed classical ML is the concept that collects data characterized by the same features from different users on more than one central server where they are processed and analyzed. Thus, the concept of distributed classical ML is to distribute the data analysis tasks to multiple servers instead of just one. Thus, it can be said that distributed classical

ML models are trained using the same methodology as centralized ML models, except that they are trained separately on multiple servers. In this context, the two concepts are compared using [41–43]:

- Motivation: in distributed classical ML, the main goal is to accelerate the processing phase, while in FL, both privacy and processing phases are targeted;
- Data identity: in the distributed classical ML, the data are described as IID records, while in FL, it is unbalanced non-IID records due to heterogeneity;
- Centralization: in the distributed classical ML, no central server is included in the architecture, while in FL, both centralization and distribution are provided;
- Data access: in the distributed classic ML, the data are distributed among several servers, but the global model still has access to the user data and, moreover, some servers can have access to all the data of a user at a given time;
- Communication and data transmission: in distributed classical ML, all user data are transmitted to the network of servers, while in FL, only minimal parameters or trained models are exchanged.

2.4.3. FL vs. Federated Database System

Federated database systems (FDSs) [65] are systems that are able to combine multiple database entities and manage them as one overall system. This concept was proposed to achieve integration between multiple independent databases. Moreover, it can manage heterogeneous databases distributed among different storage units. Moreover, FDS focuses on basic operations such as insert, delete, update, and other database operations. In this context, the two concepts are compared using [44,65]:

- Motivation: in FDBS, the main goal is to perform database operations over diverse and independent databases, while the main goal of FL is to process heterogeneous and independent databases to learn from data;
- Data identity: both can support non-IID databases;
- Centralization: both support the decentralization of database storage, but in FDBS, the processing is handled by a central server;
- Data access: in FDBS, unlike FL, the processing server has access to all data;
- Communication and data transfer: in FDBS all data are transferred in contrast to FL.

The boundaries between federated ML and classical machine learning, distributed and decentralized machine learning, and the federated database are shown in Figure 6 below.

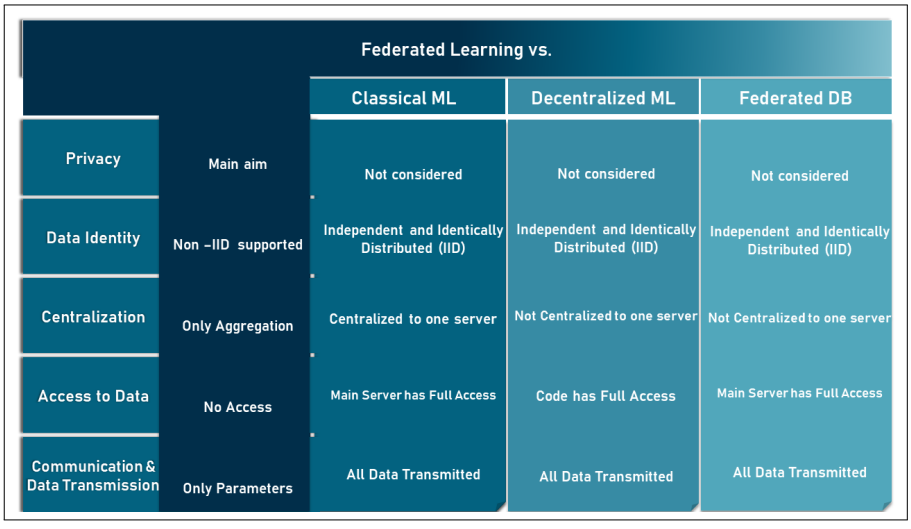

Figure 6. Borderlines between FL, ML, decentralized ML and federated DB.

2.5. FL Aggregation Algorithms: State of the Art

The first implementation of federated learning was proposed by Google to train Android keyboards to predict text input [30]. Despite its success in training machine learning models without the need to collect user data, the performance of FedAVG is poorly understood and encounters a number of problems and drawbacks, as discussed in [66]. These drawbacks can be summarized below:

- Performance issues:
  - Suffering from 'client-drift' and convergence;
  - Tuning difficulty;
  - High communication and computation cost;
  - Significant variability in systems characteristics on each network device;
  - Existence of non-identically distributed data across the network;
  - Heterogeneity of devices, users and network channels;
  - Sensitivity to local models;
  - Scalability issues.
- Security and privacy issues: FL is still under the risk of several breaching attacks such as:
  - Poisoning attacks;
  - Inference attacks;
  - Backdoor attacks.

Therefore, there was a great need to improve the performance of the federated learning FedAvg aggregation algorithm to overcome its drawbacks. In this context, several implementations have been carried out in the last 5 years. Given the diversity of challenges in this area, researchers are continuously investing in developing or improving FL aggregation algorithms. To this end, there are twenty-seven aggregation algorithms in the literature to date. These algorithms are listed in Table 3 below. An in-depth analysis of these algorithms can summarize the areas to which they contribute in the following list, which is also detailed in the table:

- Improving model aggregation;
- Reducing convergence;
- Handling heterogeneity;
- Enhancing security;
- Reducing communication and computations cost;
- Handling users' failures (fault tolerance);
- Boosting learning quality;
- Supporting scalability, personalization, and generalization.

Table 3. Contributions of existing FL aggregation algorithms.

Ref#	Year	Given Name	Model Aggregation	Convergence Reduction	Heterogeneity	Security	Communication Cost	Computation Cost	Fault Tolerance	Learning Quality	Scalability	Personalization
[30]	2017	FedAVG	✓			✓	✓					
[66]	2017	-				✓						
[67]	2019	RFA	✓				✓					
[68]	2020	SCAFFOLD	✓	✓			✓					
[69]	2020	FedOPT FedADAGAR FedYOGI FedADAM	✓	✓	✓							
[70]	2020	FedBoost					✓					

Table 3. *Cont.*

Ref#	Year	Given Name	Model Aggregation	Convergence Reduction	Heterogeneity	Security	Communication Cost	Computation Cost	Fault Tolerance	Learning Quality	Scalability	Personalization
[71]	2020	FedProx		✓	✓							
[72]	2020	FedMA	✓				✓					
[73]	2020	-	✓	✓			✓					
[74]	2020	-					✓	✓				
[75]	2020	-	✓				✓					
[76]	2020	LAQ					✓					
[77]	2020	SAFA	✓				✓	✓	✓			
[78]	2021	FedDist			✓							✓
[79]	2021	FEDHQ	✓	✓								
[80]	2021	FAIR	✓							✓		
[81]	2021	FedPSO					✓			✓		
[82]	2021	LEGATO				✓	✓	✓			✓	
[83]	2021	MHAT	✓		✓			✓				
[84]	2021	-				✓						
[85]	2021	-	✓		✓					✓		
[86]	2021	SEAR				✓		✓				
[87]	2021	Turbo-Aggregate	✓					✓			✓	
[88]	2022	EPPDA				✓			✓			
[89]	2022	FedBuff	✓									
[90]	2022	HeteroSAg	✓			✓	✓					
[91]	2022	LightSecAgg						✓	✓			

However, the achievements of previous federated learning aggregation algorithms have mainly focused on the aggregation itself or on reducing communication costs. The other contribution areas have been less studied. For example, among the 27 algorithms mentioned, 15 targeted global model aggregation and 12 targeted communication cost reduction, while only three targeted learning quality improvement and only one targeted personalization. This distribution is shown in the diagram in Figure 7 below.

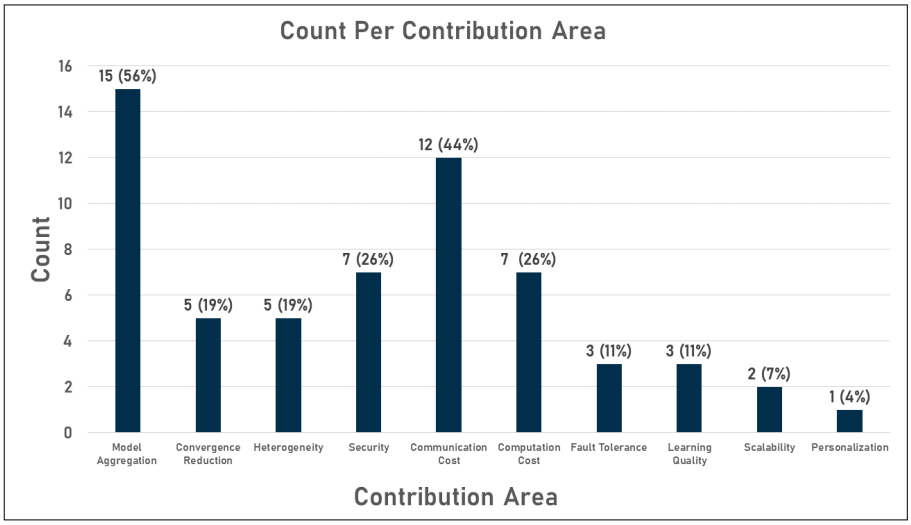

Figure 7. Aggregation algorithms count per contribution area.

Analysis of the distribution of implementations per contribution domain shows that the state of the art in federated learning algorithms has produced a number of robust aggregation

algorithms that are also acceptable from the point of view of reduced communication costs. However, from a security point of view, all the presented implementations focused on only one type of attack, namely the Byzantine attack. Other attacks have not been extensively covered in the literature, raising the question of how robust the available methods are against attacks such as reversal attacks, which are the main concern of FL, where attackers can detect users' private data based on the local trained model exchanged within the network. In addition, few efforts have been made to improve the learning quality of the models from FL, which in turn raises the question of the extent to which the accuracy of the traditional algorithms from ML is comparable to that of the models from FL. Finally, personalization has only been investigated in a single study, as shown in the table and chart.

2.6. FL Available Frameworks/Platforms

Despite its novelty, federated learning has been a popular topic in the research community. The increasing interest in this field assisted in having several frameworks or platforms that implement FL. Some of those frameworks are [65,92,93]:

- Tensorflow federated (TFF) algorithm [94]: an open source framework for experimenting with FL that enables developers to experiment with novel FL algorithms as well as simulating existing ones on their data;
- Federated AI technology enabler (FATE) [95]: relies on homomorphic encryption and supports a range of FL architectures and secure computation algorithms including logistic regression, tree-based algorithms, neural networks and transfer learning;
- PySyft [96]: developed by OpenMined and decouples private data from model training using federated learning, differential privacy and multiparty computation;
- Tensor/IO [97]: a lightweight cross-platform library for on-device machine learning, bringing the power of TensorFlow and TensorFlow Lite to iOS, Android, and React native applications;
- Tensorflow encrypted: provides an interface similar to that of TensorFlow and aims to make the technology readily available without requiring the user to be an expert in ML, cryptography, distributed systems, and high-performance computing;
- CoMind: built on top of TensorFlow and provides high-level APIs for implementing FL and FedAvg specifically;
- Horovod: based on the open message passing interface (MPI) and works on top of popular deep learning frameworks, such as TensorFlow and PyTorch;
- LEAF benchmark: is a modular benchmarking framework for machine learning in federated settings, with applications in FL, multi-task learning, meta-learning, and on-device learning aiming to capture the reality, obstacles, and intricacies of practical FL environments.

2.7. Training and Evaluation of Federated Learning Algorithms

FL is known as a privacy-preserving technology, where the data are not transferred to nor collected at a central server to allow model training. However, when training a federated machine learning model, updates are aggregated from multiple decentralized nodes: each node trains a local model on its own data and then shares the model updates with other nodes, allowing the global model to converge towards a stable solution while protecting the privacy and security of the individual data points. Additionally, there exist, in fact, norms and standards that may be used to evaluate federated machine learning algorithms. However, due to the fact that federated machine learning is still a relatively new field, these norms and standards are still in the process of developing. These norms include, but are not limited to [30,94–97]:

- Model accuracy: in the case of FL, model accuracy is a frequent parameter used to assess performance. Precision, recall, F1-score, and area under the curve (AUC) are various ways in which a model's efficacy may be evaluated;
- Communication overhead: since communication delays might have a negative effect on the efficiency of a federated machine learning system, it is crucial to keep this

in mind. The length of time spent communicating, the number of times messages need to be sent back and forth, and the overall quantity of data communicated are all indicators of communication overhead;
- Convergence speed: the speed with which a model reaches a stable solution is known as its convergence speed. Since the models in federated machine learning need to be trained across numerous participants, this is a crucial factor to take into account;
- Privacy: since the data are being shared across several parties, privacy and security are crucial concerns in federated machine learning. Examples of privacy and security standards include data encryption, differential privacy, and safe multiparty computing.

These are some of the norms and standards that are used to assess federated machine learning algorithms. However, given that the area of study is still developing, new norms and standards may appear as the technology progresses.

3. Federated Learning in Action

Federated machine learning is emerging as a privacy-friendly technology that is expected to boost the performance of machine learning algorithms by enabling more data analytics. The ability to analyze more data or even instances with heterogeneous architectures will help increase the accuracy of smart models and thus increase their adoption in various domains. This is already demonstrated in the literature where FL is already being used in various domains such as healthcare, transportation, Internet of Things, and others [43,50,51].

3.1. FL: Areas of Implementation

Federated learning was initially used to improve the text prediction service for Android Google keyboards. However, its success and efficiency motivated its implementation in other domains. As an innovative modeling mechanism that allows training a global model with heterogeneous data from different parties without compromising user data privacy and security, FL has demonstrated its feasibility for training models that classic ML models do not allow due to factors such as intellectual property rights, privacy regulations, data confidentiality, statistical heterogeneity, and others. In addition, several FL implementations have been performed in different domains such as:

- Smart healthcare: due to the sensitive nature of healthcare data, FL is a promising solution to improve the ML healthcare service while maintaining privacy [51,98];
- Smart retail: the ability to gather knowledge from different institutions enables the smart retail sector to thrive by analyzing data scattered on different islands [43];
- Transportation: FL helps improve autonomous driving decisions by training vehicles with data from different geographic locations that enable accurate learning [43,99];
- Natural language processing (NLP): with the ability to handle heterogeneous data, FL is a good choice to improve the performance of NLP models [43,100];
- Finance: the banking sector is one of the biggest beneficiaries of FL, where the data of customers scattered in different institutions can be analyzed to assess credit risk [43,50].

3.2. Federated Learning and Disease Prediction

In addition, federated learning has the potential to play an important role in healthcare by enabling the training of models using distributed and decentralized health data [51,93,98]. This can help protect patient privacy while enabling the creation of more accurate and personalized models and the analysis of more data, as long as privacy is maintained. Federated ML can also enable the training of models with data that are difficult to obtain and consolidate, such as data from under-served or rural areas. In addition, ML can help eliminate healthcare data islands by enabling data sharing and analysis across multiple organizations. In addition, FL has increased its efficiency in learning from data that are distributed across multiple sites and cannot be combined into a single dataset, or when data reside in multiple clinical systems [101]. In summary, FL can significantly improve the quality of healthcare by making it more data-driven and personalized [93,98,101].

3.2.1. Federated ML and Cardiovascular Diseases: State-of-the-Art

Cardiovascular diseases, which comprise the deadliest diseases, claimed 18.6 million lives worldwide in 2019, accounting for 32% of global mortality. For this reason, researchers in the field of machine learning have been addressing the issue of cardiovascular diseases and trying to find more feasible solutions that can help in predicting these diseases to reduce their deadly impact. Several implementations have been performed in the literature to predict CVDs or heart-related information, whether using smart wearables equipped with smart machine learning models [34] or using only machine learning models as shown in [102].

However, with the advent of federated learning, it became possible to analyze data from diverse and heterogeneous sources, supporting the accuracy and feasibility of applying FL algorithms in cardiology. Consequently, FL has been considered in several implementations in the treatment of heart disease. For example, the authors in [103] were the first to apply FL in the field of cardiovascular disease. They analyzed various electronic health records (EHRs) to predict the hospitalizations of patients with heart disease in a given year based on their medical history described in the EHRs. To this end, they developed a federated optimization scheme (cPDS) to solve the sparse support vector machine (sSVM) problem and used the Boston Medical Center electronic health records to train and test their model. In addition to maintaining privacy, their model proved to scale well, and its performance was measured by the area under curve (AUC), which reached as high as 0.78.

In addition, the authors implemented a regression model in [104] to predict heart rate using federated learning. They used Polar smartwatches to collect their own data, which were analyzed using FL sequential Bayesian and empirical Bayes-based hierarchical Bayesian models. The former model was proposed to work based on a centralized FL architecture, while the latter provides an alternative decentralized but more scalable method from the perspective of a hierarchical Bayes model. They succeeded in creating a privacy-friendly and scalable model that predicted heart rate with high accuracy. Similarly, in [105], the authors implemented a time-series-to-time-series generative adversarial network (T2T-GAN), which is a centralized FL model based on LSTM, to predict blood pressure. Their study was performed using the "Cuff-Less Blood Pressure" estimation, an open source dataset available in the Kaggle datastore [106] for training and the "College of Queensland vital signs dataset" [107] for testing. In addition to the novelty of their model, they were able to maintain privacy and predict blood pressure with high accuracy.

In addition, the study [108] was performed to predict the presence of cardiovascular disease. With the goal of developing a personalized privacy-preserving model and reducing the difference between global and local data, a novel feature alignment model was developed to predict the presence of various cardiac arrhythmias. They analyzed electrocardiography (ECG) recordings from their privately collected data and their classification model achieved 87.85% accuracy. Similarly, in [109], the authors created a classification model to predict the cardiovascular risks. They analyzed the Nursing Electronic Learning Laboratory (NeLL) EHR data using a sequential pattern mining (SPM)-based framework. They created both centralized and decentralized models that could predict risk with high accuracy while protecting patient privacy.

In the same context, [110] proposed a cardiovascular arrhythmia prediction model based on federated learning. The authors built a centralized federated transfer learning and explainable 1D convolutional neural network (CNN) trained with the MIT-BIH arrhythmia database [111]. They succeeded in preserving privacy, increasing explainability, reducing communication costs, and creating a personalized model with up to 98.9% arrhythmia prediction accuracy.

Finally, in [112], the authors developed a 3D CNN for predicting hypertrophic cardiomyopathy with FL. Their centralized FL model was trained with the M&M [113] and ACDC challenges [114] datasets consisting of cardiovascular magnetic resonance images. Their

model preserved privacy and achieved a performance of 0.89 AUC. The following Table 4 summarizes and presents the federated learning implementations performed with FL.

Table 4. Federated machine learning implementations in CVDs prediction.

Ref	Year	Type	Parameter Studied	Predicted outcome	Model	FL Architecture	Contribution	Dataset Used	Performance
[103]	2018	Classification	Electronic health records	Hospitalization for CVD patients	Federated optimization scheme (cPDS) for solving sparse support vector machine		Scalability Privacy	Electronic heart records from the Boston Medical Center	Best 0.78 AUC
[104]	2020	Regression	Heart rate	Heart rate	Federated ;earning based on sequential Bayesian method (FD Seq Bayes)	Centralized Decentralized	Privacy Scalability	Private	-
					Empirical Bayes-based hierarchical Bayesian method (FD HBayes-EB)				
[105]	2021	Regression	Blood pressure	Blood pressure	Time-series-to-time-series generative adversarial network (T2T-GAN) (based on LSTM)	Centralized	Novelty Privacy	Cuff-Less blood pressure estimation [106] University of Queensland vital signs dataset [107]	Mean error of 2.95 mmHg and a standard deviation of 19.33 mmHg
[108]	2021	Classification	ECG	Arrythmias	Customized alignment Model	Centralized	Personalization Privacy	Private	Accuracy: 87.85%
[109]	2021	Classification	Electronic health records	Cardiovascular risk	Sequential pattern mining (SPM) Based Framework	Centralized Decentralized	Privacy	Nursing Electronic Learning Laboratory (NeLL)	-
[110]	2022	Classification	ECG	Arrythmias	1D-convolutional neural Networks	Centralized	Privacy Explainability Communication cost reduction Personalization	MIT-BIH arrhythmia Database [111]	Accuracy: 98:9%
[112]	2022	Classification	Cardiovascular magnetic resonance images	Hypertrophic cardiomyopathy	3D-convolutional neural networks	Centralized	Privacy	M&M challenge [113] ACDC challenge [114]	Best 0.89 AUC

3.2.2. Federated ML and Diabetes: State-of-the-Art

In addition to its role in predicting cardiovascular diseases, federated learning has also been used in diabetes detection. According to recent figures from the World Health Organization (WHO), diabetes affects approximately 422 million people worldwide, most of whom live in low and middle-income countries, and 1.5 million deaths are directly attributable to it each year. Most frustrating, however, is the fact that both the incidence and prevalence of diabetes have substantially increased in recent decades [115]. The criticality of these diseases and the increase in their numbers require innovative solutions to help manage these situations. In this context, several implementations of federated learning have already been carried out.

Additionally, in [116], the authors evaluated the effectiveness of federated neural network-based retinal microvasculature segmentation and classification of referable diabetic retinopathy (RDR) using optical coherence tomography (OCT) and OCT angiography (OCTA). For this purpose, several datasets were used, including SFU prototype swept-source OCTA, RTVue XR Avanti (OptoVue, Inc.), Angioplex (Carl Zeiss Meditec), and PLEX Elite 9000 (Carl Zeiss Meditec). The obtained results show that FL achieves comparable performance to conventional DL models while maintaining data confidentiality.

In addition, the authors of [117] developed a decentralized, privacy-protected, FL algorithm to identify individuals at high risk of developing diabetes-related problems. In their experiments, they trained and evaluated models using the "Health Facts EMR Data" dataset from Cerner. The results showed that FL can be used not only to maintain privacy but also to address issues such as class-imbalance when using real-world clinical data. In addition, FL showed similar performance to the gold standard of centralized learning, and the use of class-balancing strategies improved performance across all cohorts. In addition, in [118], the authors proposed the use of deep learning models for the diagnosis of diabetes, also known as the Diabetes Management Control System (DMCS). The system can predict patients' glucose levels at each evaluation time point, while the classification model was designed to identify anomalous data points using a convolutional neural network (CNN) and a multilayer

perceptron model (MLP). Considering the sensitive nature of patient physiological data contained in the datasets, the authors developed independent learning (IL) and federated learning to protect the privacy of user data. However, the dataset used to train and evaluate the proposed models was generated by a simulator. The results of their study show that the FL method has a higher retrieval rate ($\geq$98.69%) than the IL method ($\leq$97.87%). In addition, the FL-CNN model performed better than the MLP model with a recall value of 99.24% compared to 98.69% for the former and the latter, respectively.

Furthermore, in [119], the authors investigated the privacy threat of gradient inversion attacks to reconstruct identifiable retinal fundus images during diabetic retinopathy classification training with federated learning. Despite the fact that the primary goal of the research is privacy-related, the authors conducted their evaluation using the fine-grained annotated diabetic retinopathy (FGADR) dataset [120], which allows for the advanced exploration of DR diagnosis. The results show that the reconstructed images matched the respective baseline images with an accuracy level of 72.0%. In addition, the authors proposed an FL-based model for predicting diabetes in [121]. The experimental results showed that federated learning helps to overcome data isolation phenomenon, also known as data islands, between healthcare institutes, and successfully collects patient data from different facilities, which can not only improve the accuracy of the trained model but also successfully protect patient privacy. Furthermore, in [122], the authors investigated the use of federated learning to detect diabetic retinopathy and non-DR images. To this end, they created three models, including standard, FedAVG, and FedProx, and evaluated their models with five publicly available diabetic retinopathy datasets, including EyePACS [123], Messidor [124], IDRID [125], APTOS [126], and College of Auckland (UoA) [127]. The three models achieved accuracies of 92.19%, 90.07%, and 85.81%, respectively.

The aforementioned implementations of federated learning in the detection of diabetes. In FL, the model can be developed using data from different healthcare facilities without requiring a facility to provide its entire dataset, improving the generalizability of the model while maintaining data confidentiality. The state of the art in the use of federated learning in diabetes discussed in this section is summarized in Table 5:

Table 5. Federated machine learning implementations in diabetes prediction.

Ref	Model	Data Used	Performance
[116]	FL deep neural network	SFU prototype swept-source OCTA RTVue XR Avanti (OptoVue, Inc.) Angioplex (Carl Zeiss Meditec) PLEX Elite 9000 (Carl Zeiss Meditec)	Performance is comparable to conventional DL models
[117]	Not identified	Health Facts EMR Data dataset from Cerner	Performance is similar to the gold standard of centralized learning
[118]	FL convolutional neural network (CNN) FL multilayer perceptron (MLP)	Generated by simulator	FL-CNN recall: 99.24% FL-MLP recall: 98.69% performed better than traditional DL
[119]	Not identified	Fine-Grained Annotated Diabetic Retinopathy (FGADR) dataset [120]	Accuracy: 72%
[121]	Not identified	Private data collected from different healthcare facilities	-
[122]	Standard FL FedAVG FedProx	EyePACS [123] Messidor [124] IDRID [125] APTOS [126] University of Auckland (UoA) [127]	Standard FL Accuracy: 92.19% FedAVG Accuracy: 90.07% FedProx Accuracy: 85.81%

3.2.3. Federated ML and Cancer: State-of-the-Art

Differently speaking, cancer, which is the disease characterized by the uncontrolled multiplication and spread of aberrant cells throughout the body, is of particular interest to federated learning researchers. This disease is known to be a leading cause of death worldwide, responsible for approximately 10 million deaths in 2020, accounting for 16% of total mortality [128] that year. Therefore, there is an increasing interest in finding technological assistance solutions for the diagnosis and prediction of cancer.

In this context, Alexander Chowdhury et al. [129] conducted a comprehensive literature review to identify the latest applications of federated learning for cancer research and clinical oncology analysis. Their study came up with several positive results that contribute

to the understanding of the use of federated learning in cancer diagnosis. Their results showed that many studies have been conducted in this area, but only 56% of them were focused on cancer research, while the others used cancer datasets to benchmark a general method. The studies dedicated to cancer research are listed in Table 6 below:

Table 6. Federated machine learning implementations in cancer prediction.

Ref	Disease	Data Used	Performance
[130]	Brain tumor	Brain MRI Segmentation Kaggle dataset [131]	FL results outperform the baseline but classical ML models competed with their results
[132]	Brain tumor	BraTS dataset [133]	Dice = 0.86 for both FL and ML scenarios
[134]	Brain tumor	BraTS dataset [133]	FL performance is similar to ML models
[135]	Brain tumor	Private data	Dice=0.86 for both FL and ML scenarios
[136]	Skin cancer	ISIC 2018 dataset [137]	Accuracy = 91% for both FL and ML scenarios
[138]	Skin cancer	ISIC 2019 Dermoscopy dataset [137]	Accuracy: 89% which outperformed previous implementations
[139]	Breast cancer	Private data from 7 different institutions	FL perform 6.3% on average better than classical ML
[140]	Breast cancer	Obtained from Netherlands Cancer Registry (NCR)	Not available
[141]	Prostate cancer	Private data	FL model exhibited superior performance and generalizability to the ML models
[142]	Lung cancer	Private data from 8 institutes across 5 countries	Not available
[143]	Pancreatic cancer	Data from hospitals in Japan and Taiwan	FL models have higher generalizability than ML models
[144]	Thyroid cancer	Private data from 6 institutions	DL models outperformed FL models
[145]	Anal cancer	Private data from 3 institutions	Not available

3.3. Discussion

Federated learning is a method for training ML models using decentralized data residing on different devices or systems as opposed to a central server. In the field of disease diagnosis, FL could be used to train models on a huge, distributed dataset of patient data from different hospitals or clinics. This method allows information and knowledge to be shared between facilities while protecting the privacy and security of patient data. Using a larger, more diverse dataset also allows for more accurate and robust models. However, implementations of federated learning for disease prediction, particularly cardiovascular disease, diabetes, and cancer, can be discussed from several perspectives, which are discussed in more detail in this section.

3.3.1. Models Performance: Competition between FL and ML

In classical ML, data collection is the first step in the execution of the known pipeline. It is also known that the accuracy of a trained ML model can be improved by collecting additional data. Therefore, it is agreed in theory that the accuracy of FL models will surpass that of traditional ML models because FL can access more data due to its nature.

In this context, the prediction results presented in Table 4 using FL show the high feasibility and accuracy. For example, the models in [110] achieved 98.9% accuracy in detecting cardiac arrhythmias, whereas the models in [108] had 87.85% accuracy. In addition, both models in [103,112] had area under the curve values of 0.78 and 0.89, respectively. However, these results are not better than any classical ML models used to predict CVDs. Even though the results of [110] are relatively high, a comparison between other implementations and classical implementations shows that the accuracy of the classical ML is higher. For example, the machine learning models proposed in [102] achieved over 91% accuracy in predicting CVDs 12 months before their onset. These results outperform all FL implementations in Table 4 except [110].

On the other hand, the FL implementations in diabetes diagnosis showed relatively high performance values, with the authors in [118] recording an accuracy of 99.24%, which is better than the traditional ML models used in this field, as explained by the authors. Moreover, in [116,117], the authors stated that the results obtained were comparable to those obtained with traditional DL models. However, the results in [119] are not as high as those obtained with other implementations, with an accuracy of 72%, which is lower than the results obtained with conventional ML models, as shown in [35].

Furthermore, the results presented in Table 6 were inconsistent in comparing the performance between FL and the classical ML and DL models. In this regard, the results obtained in [132,134,136,136] proved that the FL and ML models (including the classical ML and DL) have the same performance. However, the results obtained in [130,138,139,141] proved that the FL models outperform the earlier implementations of ML. In contrast, the authors of the results in [144] clearly stated that the models of DL outperform the models of FL, in contrast to the results in [143] where the authors stated that the models of FL have higher generalizability than the models of ML, but not higher accuracy.

In summary, although FL may theoretically have higher performance in machine learning, the results obtained are not yet sufficient to prove this hypothesis in the field of disease prediction. The FL implementations in this field are very accurate and feasible, but in some cases, the models of ML are still able to provide higher accuracy even if privacy is not preserved.

3.3.2. Real World vs. Research Implementations

Federated ML was proposed by Google in 2016 [30]. Although FL is still in its infancy, it has found widespread application in research, particularly in disease prediction.

However, most of the implementations performed, whether these were for cardiovascular diseases, diabetes, or cancer prediction, have been implemented as research studies rather than production methods. Moreover, most of these implementations are performed with publicly available data rather than using clinical or real-world data. For example, in the case of cardiovascular disease prediction, only [103] used real-world data from healthcare institutions and in the study in [104], real-world data from 10 individuals were used, whereas the others used either publicly available datasets or unspecified private data. In addition, none of these implementations were carried through to production readiness, but were conducted only as research studies.

In addition, the models for diabetes detection based on FL only used [121] data from a laboratory, whereas [118] used a dataset generated from a simulator and used other publicly available datasets. In addition, none of these implementations were taken to production maturity; all were conducted as research studies only. In contrast, for cancer detection, the studies in [139,142–145] used data from laboratories, whereas others used publicly available datasets, with the exception of [135,141], which used their own data without explaining their source. Similarly to the cardiovascular disease and diabetes cases, all studies were only research studies that were not production projects and were not made commercially available for further use. These findings support the fact that FL is still in its infancy and further efforts are needed to move into production phases with FL.

3.3.3. Dedication to Disease Diagnosis

The implementations of federated machine learning that have been performed in the field of predicting diseases such as cardiovascular disease, diabetes, and cancer have not all directly been for diagnosing diseases. For example, in the prediction of cardiovascular diseases, all of the studies listed in Table 4 were aimed at proving privacy-preservation concepts. In addition, the studies in [103,104] attempted to solve scalability problems using CVDs, while [108] attempted to solve personalization nodes using FL, and [110] addressed explainability, where reducing communication costs contributed to both privacy and personalization. In this context, only [109] addressed the disease itself, without targeting other FL-related topics, because it used a dataset from a clinical laboratory.

In contrast, the diabetes FL-based implementations summarized in 3 were all devoted to the disease itself, without targeting other FL-related topics. The same is true for the studies listed in Table 4, as this table only includes FL-based models dealing with cancer, whereas the authors in [129] mentioned dozens of articles proposing some FL-based models trained with cancers but focusing on FL-related topics.

FL-based models are therefore able to analyze data from different institutions that are not connected or related in the real world, using specific disease datasets while targeting

other FL-related ideas such as scalability, communication costs, personalization, and so on. This may potentially help increase the efficiency and accuracy of intelligent models in predicting disease by giving them access to more data, while also helping to advance the field itself, clearly a win–win scenario for machine learning and health scientists.

3.3.4. Use of Smart Wearables

Smart wearables are known to provide people with continuous, long-term, and real-time monitoring. For example, fitness trackers and smartwatches have the potential to play an important role in the early detection and management of various diseases such as cardiovascular disease [34], diabetes [35], or even fatigue detection in the workplace, as shown in [36]. These tools can continuously monitor health data, such as the heart rate, and provide data that can help identify potential health problems. They also allow data to be collected outside of traditional healthcare settings, such as doctors' offices and hospitals, so that a larger number of patients can be cared for over longer periods of time. Overall, the use of smart wearables can lead to the earlier diagnosis and treatment of diseases, improving outcomes and reducing healthcare costs.

The importance of smart wearables stems from their specifications, which have resulted from improvements in information and communication technologies (ICTs), the Internet of Things (IoT), and artificial intelligence. Smart wearables, as seen in [34–36], can be known as:

- Non-invasive: do not penetrate the skin to collect data;
- Compact: should not be bulky or large so as not to interfere with life activities;
- Affordable: to increase its acceptance;
- Rugged: to withstand harsh operating conditions such as light scratches or shocks;
- Easy to use: should have an intuitive interface;
- Durable power source: able to operate for a long period of time.

Despite the potential and usefulness of using smart wearables for disease detection using federated machine learning models, only one study ([104]) has employed a smart wearable to predict the onset of cardiovascular disease using data collected from a smartwatch for continuous, long-term, and real-time monitoring. In the other studies on cardiovascular disease, diabetes, or cancer, the use of smartwatches was not considered in the research. Therefore, there is still a great opportunity to merge smart wearables with the field of federated machine learning to enable private and secure model training without sharing confidential data.

3.3.5. Limitations in the Use of FL for Disease Prediction

In this sense, the use of federated machine learning in the detection and prediction of CVDs, diabetes and cancer is still in its early stages. In addition to the fact that not all FL implementations beat classical ML models, very rare real-world examples in this context can be obtained. In addition, it is also rarely seen that FL researchers used smart wearables in their experiments. All these details are mentioned in Table 7 below, which summarizes the results discussed in this section to provide a complete overview of how implementations based on FL have contributed to different concepts. Moreover, other limitations and challenges that are obtained in the field of FL and its implementations in disease prediction are mentioned in Section 4.1 below.

Table 7. Federated machine learning implementations in CVDs prediction.

Ref	Disease	FL Beats ML *(Performance)*	Real-World Implementation	Disease-Oriented	Wearable
[103]	CVDs	No	No	No	No
[104]		No	Yes	No	Yes
[105]		No	No	No	No
[108]		No	No	No	No
[109]		No	No	Yes	No
[110]		Yes	No	No	No
[112]		No	No	No	No
[116]	Diabetes	Yes	No	Yes	No
[117]		Same	No	Yes	No
[118]		Yes	No	Yes	No
[119]		No	No	Yes	No
[121]		Not available	No	Yes	No
[122]		Yes	No	Yes	No
[130]	Cancer	Yes	No	Yes	No
[132]		Same	No	Yes	No
[134]		Same	No	Yes	No
[135]		Same	No	Yes	No
[136]		Same	No	Yes	No
[138]		Yes	No	Yes	No
[139]		Yes	No	Yes	No
[140]		Not available	No	Yes	No
[141]		Yes	No	Yes	No
[142]		Not available	No	Yes	No
[143]		No	No	Yes	No
[144]		No	No	Yes	No
[145]		Not available	No	Yes	No

4. FL in Disease Prediction: Challenges and Future Perspectives

Federated learning, the new and emerging technology, is promising and has already proven its efficiency in improving ML algorithms without compromising privacy. However, this technology still faces many challenges that require further research, which requires further development and improvement in this technology so that it can be further implemented in real-world scenarios. These challenges require further future work to bring this technology to a higher level so that it becomes more flexible and useful, contributing to its adoption in different areas of life. This section discusses these challenges and identifies the corresponding future perspectives needed to overcome obstacles and develop FL. These challenges demand further future work to bring this technology to a higher level to make it more flexible and useful, contributing to its adoption in different areas of life. This section discusses these challenges and identifies the corresponding future perspectives needed to overcome obstacles and develop FL.

4.1. Challenges

Federated learning is still in its early stages and still faces some obstacles. However, there is no unified classification of these challenges in the literature, and they can be considered differently depending on their nature, causes, and possible solutions. In this section, the challenges have been studied in detail and classified into three main categories [41,43,45,46,48,49,65,146,147]:

- Data source-related challenges (parties embedded in FL):
  - Structural heterogeneity;
  - Statistical heterogeneity;
  - Data specifications—amount and readiness.

- Learning process-related challenges:
  - Privacy;
  - High communication cost;
  - Aggregation techniques;
  - Personalization techniques;
  - Evaluation complexity.
- Other vulnerability-related challenges:
  - Federated fairness;
  - Application areas.

4.1.1. Data Source-Related Challenges

⋆ **Structural heterogeneity:** This is also referred to as system heterogeneity. Since federated learning mainly aims to deal with data scattered in different islands, called parties, these parties may differ in terms of network state, storage space, performance, and the processing capabilities of the devices containing the parties' data. Therefore, due to network failures, not all devices may be ready and online at each processing iteration, which is known as device failure. On the other hand, devices with better-processing capabilities train faster than other devices, resulting in unbalanced training times. Therefore, device failure and unbalanced training times can cause some devices to lag behind the global model if they are still training with outdated parameters, with these devices being referred to as laggards.

⋆ **Statistical heterogeneity:** Due to the differences between FL embedded parties, the data generated and collected are generally not independently and identically distributed (non-IID). Moreover, the data sizes of the different parties can be very different, resulting in an unbalanced distribution. This definitely increases the complexity in terms of analysis, modeling and evaluation.

⋆ **Data specifications—amount and readiness:** In classical machine learning and deep learning, the amount of training data is one of the factors affecting the performance of the models, where large amounts of data can increase the accuracy of the learned model. However, in a distributed environment, the amount of data on each party is not the same, and it may be insufficient for local training on some parties, which therefore affects the accuracy. In addition, heterogeneous data on the parties may require different preprocessing steps, where some parties can process some missing data while others do not.

4.1.2. Learning Process-Related Challenges

⋆ **Privacy:** Despite the fact that federated learning aims to building smart models that do not collect user data, it is still vulnerable to data leakage caused by attacks. This is possible because of the transmission of gradients and partial parameters, whether this is between parties and manager in the centralized architecture or between parties themselves in the decentralized architecture. Those parameters are under the risk of cracking on three levels: the inputs, learning process, or learned model, as previously discussed. Usually, attacks are performed by adversaries ranging from malicious clients in a party to a malicious party which only has black-box access to the model. The types of attacks can be summarized into the following groups [54]:

- Poisoning attacks: these are conducted by injecting noise into the FL system, and are also split into two categories:
  - Data poisoning attacks: these are the most common attacks against ML models and can be either targeted toward a specific class or non-targeted. In a targeted attack, the noisy records of a specific target class are injected into local data so that the learned model will act badly on this class;
  - Model poisoning attacks: these are similar to data poisoning attacks, where the adversary tries to poison the local models instead of the local data.

- Inference attacks: in some scenarios, it is possible to infer, conclude, or restore the party local data from the model updates during the learning process;
- Backdoor attacks: secure averaging allows parties to be anonymous during the model update process. Using the same functionality, a party or group of parties can introduce backdoor functionality in in FL global model. Then, a malicious entity can use the backdoor to mislabel certain tasks such as choosing a specific label for a data instance with specific characteristics. For sure, the proportion of the compromised devices and FL model capacity affects in the intensity of such attacks.

⋆ **High communication cost:** this is induced by the huge number of involved devices, encryption and privacy preserving computations, local models and parameter-exchange batches. In addition, it is known that the life cycle of modern data is short and that the speed of iterative updating of data is fast, because the most important advantage is timeliness. Therefore, the cost of communication is a difficult topic that is worth studying;

⋆ **Aggregation techniques:** in centralized federated learning, the local models are aggregated into a global model at the central server. Due to the variety of amounts of data at each party, different results of local models, communication bottlenecks and other challenges, the method behind aggregating the global model is a challenging topic. In addition, most of the existing aggregation algorithms target the aggregation itself, communication/computation cost reduction or heterogeneity the most, while other topics such as personalization and scalability are less investigated;

⋆ **Personalization:** According to [148], there is a gap between the accuracy of local and global models, which impose personalization as a challenging topic in FL. However, there are no clear metrics to evaluate the performance of personalization techniques, which should be a hot topic for further research;

⋆ **Evaluation complexity:** In classical ML and DL, the models are evaluated by defined metrics such as accuracy, communication cost, computation speed, among others. In contrast, the evaluation of an FL system will add more parameters to be evaluated such as privacy, additional communication cost, and robustness against attacks.

4.1.3. Other Vulnerabilities

⋆ **Federated fairness:** fairness is an emerging area of ML, investigating how to confirm that the results of a model do not depend on sensitive attributes in a way that is considered unfair. FL creates new problems for researchers regarding fairness and requires a greater focus on improving the fairness of existing algorithms. At present, it is unclear whether existing fairness methods and frameworks that have been shown to be effective in ML will also be effective in FL;

⋆ **Application areas:** federated learning has mainly been applied to supervised learning algorithms. Therefore, when using FL in domains that require data exploration, such as reinforcement learning, unsupervised learning, semi-supervised learning, and others, some challenges may arise;

⋆ **User adoption:** one of the main obstacles to integrating federated machine learning into disease diagnosis is user acceptance, adoption, and participation. Although FL is known as a privacy-friendly technology, FL is still new and has mixed user adoption due to privacy concerns, discomfort, ethics, and other contextual factors.

Therefore, these difficulties give rise to the study questions below. In addition, these questions are illustrated in Figure 8 below (the initialism RQ in the list below and in the figure refers to the term "research question"):

- **RQ1:** Heterogeneity has a negative impact on the performance of a federated learning system. What are the solutions to deal with diversity?
- **RQ2:** Real-world data are noisy and usually not suitable for analysis by intelligent models. How can peripheral data be processed before these are used for model training?
- **RQ3:** Federated machine learning is vulnerable to security breaches and attacks. What mechanisms are in place to strengthen these algorithms against malicious entities?

- **RQ4:** The additional computations and sharing of models incur additional communication and computational costs in the FL system. What techniques can be used to manage the increasing costs?
- **RQ5:** The available aggregation algorithms consider aggregation, reduction in communication and computational costs, and privacy the most, while other issues such as personalization and scalability are the least considered. What further steps need to be taken to improve the performance of FL's aggregation algorithms?

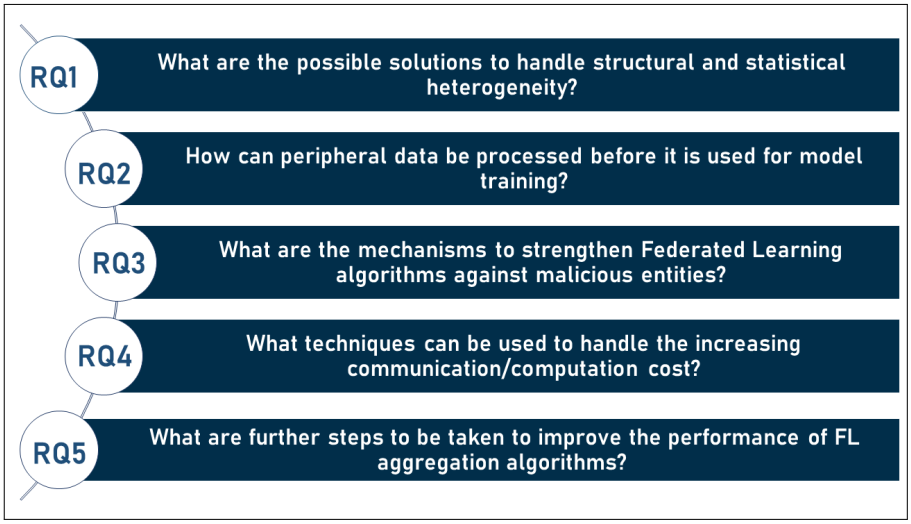

Figure 8. Research questions arising from analyzing the usage of FL in disease prediction.

4.2. Future Perspectives

Federated learning technology is still in its infancy, and there is much room for improvement and enhancement that can increase its efficiency and feasibility. Based on the literature review and investigation of the major challenges in this area, the following future prospects can be identified in FL [41,43,45,46,48,49,65,146,149]:

⋆ **Managing heterogeneity:** Heterogeneity in federated learning systems can result from both data and hardware, which is known as statistical or structural heterogeneity. To overcome heterogeneity, federated learning researchers may consider the following:

- Structural heterogeneity:
  - Fault tolerance: FL considers the impact of low participation in the training process to resist device failures by storing user updates in a trusted cache architecture to mitigate their unreliable impact on the global model;
  - Resource allocation: to solve resource scarcity, most of the previous work is devoted to properly allocating resources to heterogeneous devices.
- Statistical heterogeneity:
  - Data clustering: separating independent data into multiple clusters, then processing FL on each cluster, which is not suitable for training bulk data due to conversion overhead;
  - Modify local training mode: put cross-entropy loss into the transfer process and assign different local update times to each party in each processing iteration;
  - Meta learning [150]: Improve training on non-IID data by creating a small subset of data that are shared among all edge devices.

⋆ **Privacy preservation enhancement:** even though the main goal of FL is to preserve privacy by sharing the trained model between entities instead of raw data, the privacy preservation concept needs further enhancement, especially towards:

- Enhancing security mechanisms: by proposing new robust and feasible security mechanisms that are protected against data attacks and cracking;
- Verifying the returned model: most privacy preserving methods (FL) assume that the clients are reasonably honest. Although this is in line with training rules, curiosity in acquiring private data remains. Therefore, the returned model should be checked to determine whether it can be considered non-malicious.

⋆ **Communication optimization:** due to the system and structural heterogeneity, as well as the decentralized nature of FL, the research area of the communication cost reduction is a hot topic to attend to. There are plenty things to be considered in this area, such as:

- Gradient aggregation: it is worthwhile to introduce adaptive weighting for each party or an ML method to learn how to aggregate these gradients in an efficient way;
- Handle heterogeneity: efficiently handling heterogeneous data and devices will definitely reduce communication rounds;
- Novel models of asynchrony: in the environment of FL, there is a large variety of devices where the synchronous scheme can be easily disrupted. Therefore, it is better to use an asynchronous scheme that can handle this diversity, solve the communication delay problem, and avoid concurrent training with heterogeneous devices; Therefore, the development of asynchronous FL platforms is a possible area of study;
- One-/few-shot learning: to minimize communication costs, reducing the number of learning rounds could be a viable solution. Some researchers are exploring the possibility of training the local models with only one iteration and updating the global model accordingly.

⋆ **Performance optimization:** The trade-off between communication, performance, and privacy is an active research area in FL. Performance optimization can be achieved using various approaches, such as:

- Incentive mechanism: to encourage parties' participation in the training process in a feasible way, it is important to encourage high-quality users to contribute to the process by granting them some rewards, while neglecting or rejecting untrustworthy users because the inconsistent quality of data provided by users;
- Handle party dropouts: as one of the biggest challenges in networks with a large number of devices, handling dropouts will reduce communication costs, especially related to delayed parties;
- Personalization: improving FL personalization is much needed by users and has far-reaching applications. Many involved data holders will prefer to receive more personalized models to better meet their needs.

⋆ **Toward unsupervised learning:** unsupervised data are a large part of the data available in real life, and unsupervised learning is an area of great interest around the world. Therefore, it is of great efficiency to move towards unsupervised learning models with FL;

⋆ **Production of FL:** due to its novelty and lack of popularity, FL still needs to be put into production so that it can gain trust and be used in more areas of life;

⋆ **Benchmarks:** since the technology is still in its infancy, there is a large window of opportunity for benchmarking to define its future by ensuring that it is based on real-world circumstances, assumptions and datasets.

For this reason, we can summarize the prospects on the following emerging research topics. Moreover, these research topics are shown in Figure 9 as follows (the symbol TR in the list below and in Figure refers to the term "trending research topic"):

- **TR1:** Fault tolerance, resource allocation, data clustering, modifying local training models, and meta learning help handle heterogeneity;
- **TR2:** Preprocessing of data at peripherals to enhance their readiness may boost the overall model accuracy;

- **TR3:** More security perspectives are needed to strengthen FL against attacks;
- **TR4:** More communication/computation cost reduction is needed to boost the performance of FL algorithms;
- **TR5:** more perspectives are needed to be taken into consideration in aggregation algorithms such as privacy, personalization, and scalability.

Figure 9. Research topics that may serve as solutions to the challenges in the domain.

Figure 10 below summarizes the challenges–future solutions relationship and illustrates how future views may act as potential solutions in the domain, all of which can assist in enhancing research on the use of federated machine learning in disease diagnosis and prediction.

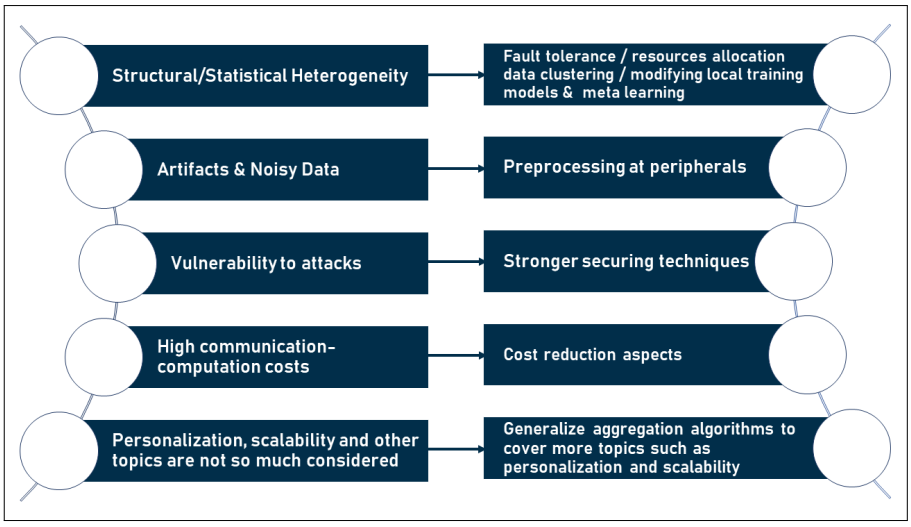

Figure 10. Challenges–future solutions chart.

5. Conclusions

It is hoped that the federated ML will solve the privacy problems of ML. It is attractive because it allows models to be trained without revealing sensitive information. Several aggregation strategies for FL knowledge have been proposed, although the field is still in its infancy. There are several examples of the application of this technology in various industries, including healthcare, banking, and others. As explained in this article, it has been used in healthcare as a diagnostic tool for a number of diseases, including cardiovascular

disease, diabetes, and cancer. Federated machine learning has achieved some successes so far, but still faces challenges such as the diversity of data and devices in the FL network, the possibility of security breaches and attacks, and the high cost of computation and communication. To help future researchers understand where we are now with this technology and what they need to take the following steps, this article presents a number of future directions that could be pursued to address these obstacles and improve the efficiency of this technology.

Author Contributions: Conceptualization: M.M. and M.A.; formal analysis: M.M.; investigation: M.M.; methodology: M.M. and M.A.; supervision: M.A., A.B., H.I. and A.R.; visualization: M.M.; writing—original draft: M.M.; writing—review and editing: M.A., A.B., H.I. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number 06351.

Acknowledgments: Acknowledgments: We acknowledge the support of Centre d'Entrepreneuriat et de Valorisation des Innovations (CEVI).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Conflicts of Interest: Not applicable.

References

1. Turing, A.M. Computing machinery and intelligence. In *Parsing the Turing Test*; Springer: Dordrecht, The Netherlands, 2009; pp. 23–65.
2. Frankish, K.; Ramsey, W.M. (Eds.) *The Cambridge Handbook of Artificial Intelligence*; Cambridge University Press: Cambridge, UK, 2014.
3. Hernández-Orallo, J.; Minaya-Collado, N. A formal definition of intelligence based on an intensional variant of algorithmic complexity. In Proceedings of International Symposium of Engineering of Intelligent Systems (EIS98), Tenerife, Spain, 11–13 February 1998; pp. 146–163.
4. Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef] [PubMed]
5. Sharma, N.; Sharma, R.; Jindal, N. Machine learning and deep learning applications-a vision. *Glob. Transit. Proc.* **2021**, *2*, 24–28. [CrossRef]
6. Pallathadka, H.; Mustafa, M.; Sanchez, D.T.; Sajja, G.S.; Gour, S.; Naved, M. Impact of machine learning on management, healthcare and agriculture. *Mater. Today Proc.* 2021, *in press*. [CrossRef] [CrossRef]
7. Ghazal, T.M.; Hasan, M.K.; Alshurideh, M.T.; Alzoubi, H.M.; Ahmad, M.; Akbar, S.S.; Al Kurdi, B.; Akour, I.A. IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet* **2021**, *13*, 218. [CrossRef]
8. Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.L. Machine learning for medical imaging. *Radiographics* **2017**, *37*, 505. [CrossRef]
9. Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. *Future Internet* **2019**, *11*, 94. [CrossRef]
10. Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine learning and deep learning methods for cybersecurity. *IEEE Access* **2018**, *6*, 35365–35381. [CrossRef]
11. Nagarhalli, T.P.; Vaze, V.; Rana, N.K. Impact of machine learning in natural language processing: A review. In Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, Tirunelveli, India, 4–6 February 2021; pp. 1529–1534.
12. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef]
13. Larrañaga, P.; Atienza, D.; Diaz-Rozo, J.; Ogbechie, A.; Puerto-Santana, C.; Bielza, C. *Industrial Applications of Machine Learning*; CRC Press: Boca Raton, FL, USA, 2018.
14. L'heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A. Machine learning with big data: Challenges and approaches. *IEEE Access* **2017**, *5*, 7776–7797. [CrossRef]
15. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. *Neurocomputing* **2017**, *237*, 350–361. [CrossRef]
16. Leskovec, J.; Rajaraman, A.; Ullman, J.D. *Mining of Massive Data Sets*; Cambridge University Press: Cambridge, UK, 2020
17. Paleyes, A.; Urma, R.G.; Lawrence, N.D. Challenges in deploying machine learning: A survey of case studies. *ACM Comput. Surv. (CSUR)* **2020**, *55*, 1–29. [CrossRef]

18. Char, D.S.; Shah, N.H.; Magnus, D. Implementing machine learning in health care—Addressing ethical challenges. *N. Engl. J. Med.* **2018**, *378*, 981. [CrossRef]
19. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [CrossRef]
20. Injadat, M.; Moubayed, A.; Nassif, A.B.; Shami, A. Machine learning towards intelligent systems: Applications, challenges, and opportunities. *Artif. Intell. Rev.* **2021**, *54*, 3299–3348. [CrossRef]
21. Albrecht, J.P. How the GDPR will change the world. *Eur. Data Prot. L. Rev.* **2016**, *2*, 287. [CrossRef]
22. Parasol, M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams. *Comput. Law Secur. Rev.* **2018**, *34*, 67–98. [CrossRef]
23. Gray, W.; Zheng, H.R. General Principles of Civil Law of the People's Republic of China. *Am. J. Comp. Law* **1986**, *34*, 715–743. [CrossRef]
24. Chik, W.B. The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform. *Comput. Law Secur. Rev.* **2013**, *29*, 554–575 [CrossRef]
25. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, 4–7 March 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 265–284.
26. El Emam, K.; Dankar, F.K. Protecting privacy using k-anonymity. *J. Am. Med. Inform. Assoc.* **2008**, *15*, 627–637. [CrossRef]
27. Li, P.; Li, J.; Huang, Z.; Li, T.; Gao, C.Z.; Yiu, S.M.; Chen, K. Multi-key privacy-preserving deep learning in cloud computing. *Future Gener. Comput. Syst.* **2017**, *74*, 76–85. [CrossRef]
28. Fredrikson, M.; Jha, S.; Ristenpart, T. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; (pp. 1322–1333).
29. Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership inference attacks against machine learning models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), IEEE, San Jose, CA, USA, 22–26 May 2017; pp. 3–18.
30. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics PMLR, Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
31. Ramesh, A.N.; Kambhampati, C.; Monson, J.R.L., Drew, P.J. Artificial intelligence in medicine. *Ann. R. Coll. Surg. Engl.* **2004**, *86*, 334. [CrossRef] [PubMed]
32. Maddox, T.M.; Rumsfeld, J.S.; Payne, P.R. Questions for artificial intelligence in health care. *JAMA* **2019**, *321*, 31–32. [CrossRef] [PubMed]
33. Nayyar, A.; Gadhavi, L.; Zaman, N. Machine learning in healthcare: Review, opportunities and challenges. In *Machine Learning and the Internet of Medical Things in Healthcare*; Academic Press: Cambridge, MA, USA, 2021; pp. 23–45.
34. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review. *Sensors* **2023**, *23*, 828. [CrossRef] [PubMed]
35. Makroum, M.A.; Adda, M.; Bouzouane, A.; Ibrahim, H. Machine learning and smart devices for diabetes management: Systematic review. *Sensors* **2022**, *22*, 1843. [CrossRef]
36. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review. *Sensors* **2022**, *22*, 7472. [CrossRef]
37. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Hassabis, D.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [CrossRef]
38. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Hassabis, D.; et al. Mastering the game of go without human knowledge. *Nature* **2017**, *550*, 354–359. [CrossRef]
39. Aono, Y.; Hayashi, T.; Wang, L.; Moriai, S. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 1333–1345.
40. Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; Vladu, A. Towards deep learning models resistant to adversarial attacks. *arXiv* **2017**, arXiv:1706.06083.
41. Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li,W.; Gao, Y. A survey on federated learning. *Knowl.-Based Syst.* **2021**, *216*, 106775. [CrossRef]
42. Li, Q.; Wen, Z.; Wu, Z.; Hu, S.; Wang, N.; Li, Y.; Liu, X.; He, B. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Trans. Knowl. Data Eng.* **2021**. [CrossRef]
43. Li, L.; Fan, Y.; Tse, M.; Lin, K.Y. A review of applications in federated learning. *Comput. Ind. Eng.* **2020**, *149*, 106854. [CrossRef]
44. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol. (TIST)* **2019**, *10*, 1–19. [CrossRef]
45. Mammen, P.M. Federated learning: Opportunities and challenges. *arXiv* **2021**, arXiv:2101.05428.
46. Zhang, K.; Song, X.; Zhang, C.; Yu, S. Challenges and future directions of secure federated learning: a survey. *Front. Comput. Sci.* **2022**, *16*, 165817. [CrossRef]
47. Asad, M.; Moustafa, A.; Ito, T. Federated Learning Versus Classical Machine Learning: A Convergence Comparison. *arXiv* **2021**, arXiv:2107.10976.

48. Mahlool, D.H.; Abed, M.H. A Comprehensive Survey on Federated Learning: Concept and Applications. *arXiv* **2022**, arXiv:2201.09384.
49. Zhang, H.; Bosch, J.; Holmström, Olsson, H. Engineering Federated Learning Systems: A Literature Review. In Proceedings of the 11th International Conference, ICSOB 2020, Karlskrona, Sweden, 16–18 November 2020; Springer, Cham, Switzerland, 2020; pp. 210–218.
50. Lim, W.Y.B.; Luong, N.C.; Hoang, D.T.; Jiao, Y.; Liang, Y.C.; Yang, Q.; Niyato, D.; Miao, C. Federated learning in mobile edge networks: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2031–2063. [CrossRef]
51. Xu, J.; Glicksberg, B.S.; Su, C.; Walker, P.; Bian, J.; Wang, F. Federated learning for healthcare informatics. *J. Healthc. Inform. Res.* **2021**, *5*, 1–19. [CrossRef]
52. Shokri, R.; Shmatikov, V. Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; pp. 1310–1321.
53. Li, T.; Sahu, A.K.; Talwalkar, A.; Smith, V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. [CrossRef]
54. Lyu, L.; Yu, H.; Yang, Q. Threats to federated learning: A survey. *arXiv* **2020**, arXiv:2003.02133.
55. Chen, X.; Liu, C.; Li, B.; Lu, K.; Song, D. Targeted backdoor attacks on deep learning systems using data poisoning. *arXiv* **2017**, arXiv:1712.05526.
56. Li, B.; Wang, Y.; Singh, A.; Vorobeychik, Y. Data poisoning attacks on factorization-based collaborative filtering. In Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 5–10 December 2016; Volume 29.
57. Alfeld, S.; Zhu, X.; Barford, P. Data poisoning attacks against autoregressive models. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30, Number 1.
58. Bagdasaryan, E.; Veit, A.; Hua, Y.; Estrin, D.; Shmatikov, V. How to backdoor federated learning. In Proceedings of the International Conference on Artificial Intelligence and Statistics, PMLR, Sicily, Italy, 3–5 June 2020; pp. 2938–2948.
59. Xie, C.; Huang, K.; Chen, P.Y.; Li, B. Dba: Distributed backdoor attacks against federated learning. In Proceedings of the International Conference on Learning Representations, Jakarta, Indonesia, 18–20 September 2019.
60. Castro, M.; Liskov, B. Practical Byzantine fault tolerance and proactive recovery. *ACM Trans. Comput. Syst. (TOCS)* **2002**, *20*, 398–461. [CrossRef]
61. Blanchard, P.; El Mhamdi, E.M.; Guerraoui, R.; Stainer, J. Machine learning with adversaries: Byzantine tolerant gradient descent. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017.
62. Bayatbabolghani, Fattaneh, and Marina Blanton. "Secure multi-party computation." In Proceedings of the 2018 ACM SIGSAC conference on computer and communications security 2018, Toronto, Canada, 15–19 October 2018
63. Dwork, C. Differential privacy: A survey of results. In Proceedings of the International Conference on Theory and Applications of Models of Computation, Xi'an, China, 25–19 April 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–19.
64. Kairouz, P.; McMahan, H.B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A.N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Zhao, S.; et al. Advances and open problems in federated learning. *Found. Trends® Mach. Learn.* **2021**, *14*, 1–210. [CrossRef]
65. Rahman, K.J.; Ahmed, F.; Akhter, N.; Hasan, M.; Amin, R.; Aziz, K.E.; Muzahidul Islam, A.K.M.; Hossain Mukta, S.; Islam, A.N. Challenges, applications and design aspects of federated learning: A survey. *IEEE Access* **2021**, *9*, 124682–124700. [CrossRef]
66. Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H.B.; Patel, S.; Ramage, D.; Segal, A.; Seth, K. Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017; pp. 1175–1191.
67. Pillutla, K.; Kakade, S.M.; Harchaoui, Z. Robust aggregation for federated learning. *IEEE Trans. Signal Process.* **2022**, *70*, 1142–1154. [CrossRef]
68. Karimireddy, S.P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; Suresh, A.T. Scaffold: Stochastic controlled averaging for federated learning. In Proceedings of the International Conference on Machine Learning, PMLR, Bangkok, Thailand, 18–22 November 2020; pp. 5132–5143.
69. Reddi, S.; Charles, Z.; Zaheer, M.; Garrett, Z.; Rush, K.; Konečný, J.; Kumar, S.; McMahan, H.B. Adaptive federated optimization. *arXiv* **2020**, arXiv:2003.00295.
70. Hamer, J.; Mohri, M.; Suresh, A.T. Fedboost: A communication-efficient algorithm for federated learning. In Proceedings of the International Conference on Machine Learning PMLR, Bangkok, Thailand, 18–22 November 2020; pp. 3973–3983.
71. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated optimization in heterogeneous networks. In Proceedings of the Machine Learning and Systems, Austin, TX, USA, 2–4 March 2020; Volume 2, pp. 429–450.
72. Wang, H.; Yurochkin, M.; Sun, Y.; Papailiopoulos, D.; Khazaeni, Y. Federated learning with matched averaging. *arXiv* **2020**, arXiv:2002.06440.
73. Guo, H.; Liu, A.; Lau, V.K. Analog gradient aggregation for federated learning over wireless networks: Customized design and convergence analysis. *IEEE Internet Things J.* **2020**, *8*, 197–210. [CrossRef]
74. Choi, B.; Sohn, J.Y.; Han, D.J.; Moon, J. Communication-computation efficient secure aggregation for federated learning. *arXiv* **2020**, arXiv:2012.05433.
75. Ye, D.; Yu, R.; Pan, M.; Han, Z. Federated learning in vehicular edge computing: A selective model aggregation approach. *IEEE Access* **2020**, *8*, 23920–23935. [CrossRef]

76. Sun, J.; Chen, T.; Giannakis, G.B.; Yang, Q.; Yang, Z. Lazily aggregated quantized gradient innovation for communication-efficient federated learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2031–2044. [CrossRef] [PubMed]
77. Wu, W.; He, L.; Lin, W.; Mao, R.; Maple, C.; Jarvis, S. SAFA: A semi-asynchronous protocol for fast federated learning with low overhead. *IEEE Trans. Comput.* **2020**, *70*, 655–668. [CrossRef]
78. Sannara, E.K.; Portet, F.; Lalanda, P.; German, V.E.G.A. A federated learning aggregation algorithm for pervasive computing: Evaluation and comparison. In Proceedings of the 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), IEEE, Kassel, Germany, 22–26 March 2021; pp. 1–10
79. Chen, S.; Shen, C.; Zhang, L.; Tang, Y. Dynamic aggregation for heterogeneous quantization in federated learning. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 6804–6819. [CrossRef]
80. Deng, Y.; Lyu, F.; Ren, J.; Chen, Y.C.; Yang, P.; Zhou, Y.; Zhang, Y. Fair: Quality-aware federated learning with precise user incentive and model aggregation. In Proceedings of the IEEE INFOCOM 2021—IEEE Conference on Computer Communications, IEEE, Vancouver, BC, Canada, 10–13 May 2021; pp. 1–10.
81. Park, S.; Suh, Y.; Lee, J. FedPSO: Federated learning using particle swarm optimization to reduce communication costs. *Sensors* **2021**, *21*, 600. [CrossRef]
82. Varma, K.; Zhou, Y.; Baracaldo, N.; Anwar, A. LEGATO: A LayerwisE Gradient AggregaTiOn Algorithm for Mitigating Byzantine Attacks in Federated Learning. In Proceedings of the 2021 IEEE 14th International Conference on Cloud Computing (CLOUD), IEEE, Chicago, IL, USA, 5–10 September 2021; pp. 272–277.
83. Hu, L.; Yan, H.; Li, L.; Pan, Z.; Liu, X.; Zhang, Z. MHAT: An efficient model-heterogenous aggregation training scheme for federated learning. *Inf. Sci.* **2021**, *560*, 493–503. [CrossRef]
84. Jeon, B.; Ferdous, S.M.; Rahman, M.R.; Walid, A. Privacy-preserving decentralized aggregation for federated learning. In Proceedings of the IEEE INFOCOM 2021—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), IEEE, Vancouver, BC, Canada, 10–13 May 2021; pp. 1–6.
85. Wang, Y.; Kantarci, B. Reputation-enabled federated learning model aggregation in mobile platforms. In Proceedings of the ICC 2021—IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6.
86. Zhao, L.; Jiang, J.; Feng, B.; Wang, Q.; Shen, C.; Li, Q. Sear: Secure and efficient aggregation for byzantine-robust federated learning. *IEEE Trans. Dependable Secur. Comput.* **2021**, *19*, 2239–3342. [CrossRef]
87. So, J.; Güler, B.; Avestimehr, A.S. Turbo-aggregate: Breaking the quadratic aggregation barrier in secure federated learning. *IEEE J. Sel. Areas Inf. Theory* **2021**, *2*, 479-489. [CrossRef]
88. Song, J.; Wang, W.; Gadekallu, T.R.; Cao, J.; Liu, Y. Eppda: An efficient privacy-preserving data aggregation federated learning scheme. *IEEE Trans. Netw. Sci. Eng.* **2022**, 1. [CrossRef]
89. Nguyen, J.; Malik, K.; Zhan, H.; Yousefpour, A.; Rabbat, M.; Malek, M.; Huba, D. Federated learning with buffered asynchronous aggregation. In Proceedings of the International Conference on Artificial Intelligence and Statistics PMLR, Virtual Conference, 28–30 March 2022; pp. 3581–3607.
90. Elkordy, A.R.; Avestimehr, A.S. Heterosag: Secure aggregation with heterogeneous quantization in federated learning. *IEEE Trans. Commun.* **2022**, *70*, 2372–2386. [CrossRef]
91. So, J.; Nolet, C.J.; Yang, C.S.; Li, S.; Yu, Q.; E Ali, R.; Guler, B.; Avestimehr, S. Lightsecagg: A lightweight and versatile design for secure aggregation in federated learning. In Proceedings of the Machine Learning and Systems, Santa Clara, CA, USA, 29 August–1 September 2022; Volume 4, pp. 694–720.
92. Sheth, A.P.; Larson, J.A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Comput. Surv. (CSUR)* **1990**, *22*, 183–236. [CrossRef]
93. Kumar. Y.; Singla, R. Federated learning systems for healthcare: Perspective and recent progress. In *Federated Learning Systems*; Springer: Cham, Switzerland, 2021; pp. 141–156.
94. Google. 2019. TensorFlow Federated. Retrieved 1 July 2022. Available online: https://www.tensorflow.org/federated (accessed on 1 July 2022).
95. Liu, Y.; Fan, T.; Chen, T.; Xu, Q.; Yang, Q. FATE: An Industrial Grade Platform for Collaborative Learning With Data Protection. *J. Mach. Learn. Res.* **2021**, *22*, 10320–10325.
96. Ryffel, T.; Trask, A.; Dahl, M.; Wagner, B.; Mancuso, J.; Rueckert, D.; Passerat-Palmbach, J. A generic framework for privacy preserving deep learning. *arXiv* **2018**, arXiv:1811.04017.
97. GitHub—doc-ai/tensorio: Declarative, On-Device Machine Learning for iOS, Android, and React Native. Deploy. Predict. Train. GitHub. . Available online: https://github.com/doc-ai/tensorio (accessed on 1 July 2022).
98. Antunes, R.S.; André da Costa, C.; Küderle, A.; Yari, I.A.; Eskofier, B. Federated Learning for Healthcare: Systematic Review and Architecture Proposal. *ACM Trans. Intell. Syst. Technol. (TIST)* **2022**, *13*, 1–23. [CrossRef]
99. Tan, K.; Bremner, D.; Le Kernec, J.; Imran, M. Federated machine learning in vehicular networks: A summary of recent applications. In Proceedings of the 2020 International Conference on UK-China Emerging Technologies (UCET), IEEE, Glasgow, UK, 20–21 August 2020; pp. 1–4.
100. Liu, M.; Ho, S.; Wang, M.; Gao, L.; Jin, Y.; Zhang, H. Federated learning meets natural language processing: A survey. *arXiv* **2021**, arXiv:2107.12603.
101. Goecks, J.; Jalili, V.; Heiser, L.M.; Gray, J.W. How machine learning will transform biomedicine. *Cell* **2020**, *181*, 92–101. [CrossRef]

102. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability. *Procedia Comput. Sci.* **2022**, *203*, 231–238. [CrossRef]
103. Brisimi, T.S.; Chen, R.; Mela, T.; Olshevsky, A.; Paschalidis, I.C.; Shi, W. Federated learning of predictive models from federated electronic health records. *Int. J. Med. Inform.* **2018**, *112*, 59–67. [CrossRef]
104. Fang, L.; Liu, X.; Su, X.; Ye, J.; Dobson, S.; Hui, P.; Tarkoma, S. Bayesian inference federated learning for heart rate prediction. In Proceedings of the International Conference on Wireless Mobile Communication and Healthcare, Virtual Event, 19 November 2020; Springer: Cham, Switzerland, 2020; pp. 116–130.
105. Brophy, E.; De Vos, M.; Boylan, G.; Ward, T. Estimation of continuous blood pressure from ppg via a federated learning approach. *Sensors* **2021**, *21*, 6311. [CrossRef]
106. Cuff-Less Blood Pressure Estimation. (4 June 2017). Kaggle. Retrieved 1 July 2022. Available online: https://www.kaggle.com/datasets/mkachuee/BloodPressureDataset (accessed on 1 July 2022).
107. Liu, D.; Görges, M.; Jenkins, S.A. University of Queensland vital signs dataset: Development of an accessible repository of anesthesia patient monitoring data for research. *Anesth. Analg.* **2012**, *114*, 584–589. [CrossRef]
108. Tang, R.; Luo, J.; Qian, J.; Jin, J. Personalized Federated Learning for ECG Classification Based on Feature Alignment. *Secur. Commun. Netw.* **2021**, *2021*, 6217601. [CrossRef]
109. Lee, E.W.; Xiong, L.; Hertzberg, V.S.; Simpson, R.L.; Ho, J.C. Privacy-preserving Sequential Pattern Mining in distributed EHRs for Predicting Cardiovascular Disease. In Proceedings of the AMIA Summits on Translational Science Proceedings, Bethesda, MD, USA, 17 May 2021; pp. 384–393.
110. Raza, A.; Tran, K.P.; Koehl, L.; Li, S. Designing ecg monitoring healthcare system with federated transfer learning and explainable AI. *Knowl.-Based Syst.* **2022**, *236*, 107763. [CrossRef]
111. MIT-BIH Arrhythmia Database v1.0.0. (24 February 2005). PhysioNet. Available online: https://physionet.org/content/mitdb/1.0.0/ (accessed on 1 July 2022.)
112. Linardos, A.; Kushibar, K.; Walsh, S.; Gkontra, P.; Lekadir, K. Federated learning for multi-center imaging diagnostics: A simulation study in cardiovascular disease. *Sci. Rep.* **2022**, *12*, 3551. [CrossRef] [PubMed]
113. Campello, V.M.; Gkontra, P.; Izquierdo, C.; Martin-Isla, C.; Sojoudi, A.; Full, P. M.; Maier-Hein, K.; Zhang, Y.; He, Z.; Lekadir, K.; et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge. *IEEE Trans. Med. Imaging* **2021**, *40*, 3543–3554. [PubMed]
114. Bernard, O.; Lalande, A.; Zotti, C.; Cervenansky, F.; Yang, X.; Heng, P.A.; Cetin, I.; Lekadir, I.; Camara, O.; Jodoin, P.M.; et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Trans. Med. Imaging* **2018**, *37*, 2514–2525. [CrossRef] [PubMed]
115. Diabetes. 2 December 2022. Available online: https://www.who.int/health-topics/diabetes#tab=tab_1. (accessed on 1 January 2023).
116. Lo, J.; Timothy, T.Y.; Ma, D.; Zang, P.; Owen, J.P.; Zhang, Q.; Wang, R.K.; Beg, M.F.; Lee, A.Y.; Sarunic, M.V.; et al. Federated learning for microvasculature segmentation and diabetic retinopathy classification of OCT data. *Ophthalmol. Sci.* **2021**, *1*, 100069. [CrossRef] [PubMed]
117. Islam, H.; Mosa, A. A Federated Mining Approach on Predicting Diabetes-Related Complications: Demonstration Using Real-World Clinical Data. In Proceedings of the AMIA Annual Symposium San Diego, CA, USA, 30 October–3 November 2021; American Medical Informatics Association: Bethesda, MA, USA, 2021; Volume 2021, p. 556.
118. Astillo, P.V.; Duguma, D.G.; Park, H.; Kim, J.; Kim, B.; You, I. Federated intelligence of anomaly detection agent in IoTMD-enabled Diabetes Management Control System. *Future Gener. Comput. Syst.* **2022**, *128*, 395–405. [CrossRef]
119. Nielsen, C.; Tuladhar, A.; Forkert, N.D. Investigating the Vulnerability of Federated Learning-Based Diabetic Retinopathy Grade Classification to Gradient Inversion Attacks. In Proceedings of the International Workshop on Ophthalmic Medical Image Analysis, Singapore, 22 September 2022; Springer: Cham, Switzerland, 2022; pp. 183–192.
120. "FGADR Dataset—Look Deeper Into Eyes." FGADR Dataset—Look Deeper Into Eyes. | FGADR. Available online: csyizhou.github.io/FGADR/blob/NateBYWang-patch-1//FGADR (accessed on 12 January 2023).
121. Liu, J.; Lu, X.; Yang, H.; Zhuang, L. A Diabetes Prediction System Based on Federated Learning. In Proceedings of the 2022 International Conference on Big Data, Information and Computer Network (BDICN), IEEE, Sanya, China, 20–22 January 2022; pp. 486-491.
122. Nasajpour, M.; Karakaya, M.; Pouriyeh, S.; Parizi, R.M. Federated Transfer Learning For Diabetic Retinopathy Detection Using CNN Architectures. In Proceedings of the SoutheastCon 2022, IEEE, Mobile, AL, USA, 26 March–3 April 2022; pp. 655–660.
123. Cuadros, J.; Sim, I. EyePACS: An open source clinical communication system for eye care. *Stud Health Technol Inform.* **2004**; 207–211.
124. Decencière, E.; Zhang, X.; Cazuguel, G.; Lay, B.; Cochener, B.; Trone, C.; Gain, P.; Ordonez, R.; Massin, P.; Klein, J.C.; et al. Feedback on a publicly distributed image database: The Messidor database. *Image Anal. Stereol.* **2014**, *33*, 231–234. [CrossRef]
125. Porwal, P.; Pachade, S.; Kamble, R.; Kokare, M.; Deshmukh, G.; Sahasrabuddhe, V.; Meriaudeau, F. Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research. *Data* **2018**, *3*, 25. [CrossRef]
126. APTOS 2019 Blindness Detection | Kaggle. APTOS 2019 Blindness Detection | Kaggle. Available online: https://www.kaggle.com/c/aptos2019-blindness-detection (accessed on 12 January 2023).

127. Chalakkal, R.J.; Abdulla, W.H.; Sinumol, S. Comparative analysis of university of Auckland diabetic retinopathy database. In Proceedings of the 9th International Conference on Signal Processing Systems, Auckland, New Zealand, 27–30 November 2017; pp. 235–239.
128. Cancer. 2022. Available online: https://www.who.int/news-room/fact-sheets/detail/cancer (accessed on 13 January 2023).
129. Chowdhury, A.; Kassem, H.; Padoy, N.; Umeton, R.; Karargyris, A. A Review of Medical Federated Learning: Applications in Oncology and Cancer Research. In Proceedings of the International MICCAI Brainlesion Workshop, Virtual Event, 27 September 2021; Springer: Cham, Switzerland, 2022; pp. 3–24.
130. Yi, L.; Zhang, J.; Zhang, R.; Shi, J.; Wang, G.; Liu, X. SU-Net: An efficient encoder-decoder model of federated learning for brain tumor segmentation. In Proceedings of the International Conference on Artificial Neural Networks, Bratislava, Slovakia, 15–18 September 2020; Springer: Cham, Switzerland, 2020; pp. 761–773.
131. Mazurowski, M.A.; Clark, K.; Czarnek, N.M.; Shamsesfandabadi, P.; Peters, K.B.; Saha, A. Radiogenomics of lower-grade glioma: Algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with The Cancer Genome Atlas data. *J.-Neuro-Oncol.* **2017**, *133*, 27–35. [CrossRef] [PubMed]
132. Sheller, M.J.; Reina, G.A.; Edwards, B.; Martin, J.; Bakas, S. Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In Proceedings of the International MICCAI Brainlesion Workshop; Springer: Cham, Switzerland, Granada, Spain, 16 September 2018; pp. 92–104.
133. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Van Leemput, K.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2014**, *34*, 1993–2024. [CrossRef] [PubMed]
134. Sheller, M.J.; Edwards, B.; Reina, G.A.; Martin, J.; Pati, S.; Kotrotsou, A.; Milchenko, M.; Xu, W.; Marcus, D.; Bakas, S.; et al. Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Sci. Rep.* **2020**, *10*, 1–12. [CrossRef] [PubMed]
135. Sheller, M.; Edwards, B.; Reina, G.A.; Martin, J.; Bakas, S. NIMG-68. Federated Learning in Neuro-Oncology for Multi-Institutional Collaborations without Sharing Patient Data. *Neuro-Oncol.* **2019**, *21* (Suppl. S6), vi176–vi177. [CrossRef]
136. Cai, X.; Lan, Y.; Zhang, Z.; Wen, J.; Cui, Z.; Zhang, W. A many-objective optimization based federal deep generation model for enhancing data processing capability in IoT. *IEEE Trans. Ind. Inform.* **2021**, *19*, 561–569. [CrossRef]
137. Codella, N.C.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.; Halpern, A. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, Washington, DC, USA, 4–7 April 2018; pp. 168–172.
138. Hashmani, M.A.; Jameel, S.M.; Rizvi, S.S.H.; Shukla, S. An adaptive federated machine learning-based intelligent system for skin disease detection: A step toward an intelligent dermoscopy device. *Appl. Sci.* **2021**, *11*, 2145. [CrossRef]
139. Roth, H.R.; Chang, K.; Singh, P.; Neumark, N.; Li, W.; Gupta, V.; Gupta, S.; Qu, L.; Ihsani, A.; Kalpathy-Cramer, J.; et al. Federated learning for breast density classification: A real-world implementation. In Proceedings of the Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning, Lima, Peru, 4–8 October 2020; Springer: Cham, Switzerland, 2020; pp. 181–191.
140. Rooijakkers, T. *CONVINCED—Enabling Privacy-Preserving Survival Analyses Using Multi-Party Computation*; TNO: Hague, The Netherlands, 2020.
141. Sarma, K.V.; Harmon, S.; Sanford, T.; Roth, H.R.; Xu, Z.; Tetreault, J.; Xu, D.; Flores, M.G.; Ramas, A.G.; Arnold, C.W.; et al. Federated learning improves site performance in multicenter deep learning without data sharing. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 1259–1264. [CrossRef] [PubMed]
142. Deist, T.M.; Dankers, F.J.; Ojha, P.; Marshall, M.S.; Janssen, T.; Faivre-Finn, C.; Masciocchi, C.; Vincenzo, V.; Wang, J.; Dekker, A.; et al. Distributed learning on 20 000+ lung cancer patients—The Personal Health Train. *Radiother. Oncol.* **2020**, *144*, 189–200. [CrossRef] [PubMed]
143. Wang, P.; Shen, C.; Roth, H.R.; Yang, D.; Xu, D.; Oda, M.; Misawa, K.; Chen, P.-T.; Liu, K.-L.; Mori, K.; et al. Automated pancreas segmentation using multi-institutional collaborative deep learning. In Proceedings of the Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning, MICCAI 2020, Lima, Peru, 4–8 October 2020; Springer: Cham, Switzerland, 2020; pp. 192–200.
144. Lee, H.; Chai, Y.J.; Joo, H.; Lee, K.; Hwang, J.Y.; Kim, S.M.; Kim, S.-M.; Kim, K.; Nam, I.-C.; Kong, H.J.; et al. Federated learning for thyroid ultrasound image analysis to protect personal information: Validation study in a real health care environment. *JMIR Med. Inform.* **2021**, *9*, e25869. [CrossRef] [PubMed]
145. Choudhury, A.; Theophanous, S.; Lønne, P.I.; Samuel, R.; Guren, M.G.; Berbee, M.; Brown, P.; Lilley, J.; van Soest, J.; Appelt, A.L.; et al. Predicting outcomes in anal cancer patients using multi-centre data and distributed learning—A proof-of-concept study. *Radiother. Oncol.* **2021**, *159*, 183–189. [CrossRef] [PubMed]
146. Bharati, S.; Mondal, M.R.H.; Podder, P.; Prasath, V.S. Federated learning: Applications, challenges and future directions. *Int. J. Hybrid Intell. Syst.* **2022**, *18*, 19–35. [CrossRef]
147. Nguyen, D.C.; Pham, Q.V.; Pathirana, P.N.; Ding, M.; Seneviratne, A.; Lin, Z.; Dobre, O.; Hwang, W.J. Federated learning for smart healthcare: A survey. *ACM Comput. Surv. (CSUR)* **2022**, *55*, 1–37. [CrossRef]

148. Lo, S.K.; Lu, Q.; Wang, C.; Paik, H.Y.; Zhu, L. A systematic literature review on federated machine learning: From a software engineering perspective. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–39. [CrossRef]
149. Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; Chandra V. Federated learning with non-iid data. *arXiv* **2018**. arXiv:1806.00582
150. Jiang, Y.; Konečný, J.; Rush, K.; Kannan, S. Improving federated learning personalization via model agnostic meta learning. *arXiv* **2019**, arXiv:1909.12488.

CHAPTER 3

Reviewing Federated Learning Aggregation Algorithms; Strategies, Contributions, Limitations and Future Perspectives

Published in Electronics Journal 2023
Under the special issue ***Collaborative Artificial Systems***
Volume 12; Issue 10; doi: 10.3390/electronics12102287

Résumé: Les algorithmes d'agrégation dans l'apprentissage fédéré sont discutés et examinés en détail dans ce chapitre. Ce chapitre se concentre exclusivement sur les algorithmes d'agrégation dans le cadre de l'apprentissage fédéré, en commençant par une présentation de son architecture, suivie d'une exploration des différents types de messages échangés entre le serveur et les clients dans un système d'apprentissage fédéré. Le chapitre procède à une évaluation approfondie de diverses méthodes d'agrégation et conclut en examinant minutieusement les implémentations les plus avancées d'agrégations d'apprentissage fédéré, ainsi qu'un examen approfondi des défis associés.

electronics

Review

Reviewing Federated Learning Aggregation Algorithms; Strategies, Contributions, Limitations and Future Perspectives

Mohammad Moshawrab [1,*], Mehdi Adda [1,*], Abdenour Bouzouane [2], Hussein Ibrahim [3] and Ali Raad [4]

[1] Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC G5L 3A1, Canada
[2] Département d'Informatique et de Mathématique, Université du Québec à Chicoutimi, 555 Boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada
[3] Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, QC G4R 5B7, Canada
[4] Faculty of Arts & Sciences, Islamic University of Lebanon, Wardaniyeh P.O. Box 30014, Lebanon
* Correspondence: mohammad.moshawrab@uqar.ca (M.M.); mehdi_adda@uqar.ca (M.A.); Tel.: +1-(581)-624-9394 (M.M.)

Abstract: The success of machine learning (ML) techniques in the formerly difficult areas of data analysis and pattern extraction has led to their widespread incorporation into various aspects of human life. This success is due in part to the increasing computational power of computers and in part to the improved ability of ML algorithms to process large amounts of data in various forms. Despite these improvements, certain issues, such as privacy, continue to hinder the development of this field. In this context, a privacy-preserving, distributed, and collaborative machine learning technique called federated learning (FL) has emerged. The core idea of this technique is that, unlike traditional machine learning, user data is not collected on a central server. Nevertheless, models are sent to clients to be trained locally, and then only the models themselves, without associated data, are sent back to the server to combine the different locally trained models into a single global model. In this respect, the aggregation algorithms play a crucial role in the federated learning process, as they are responsible for integrating the knowledge of the participating clients, by integrating the locally trained models to train a global one. To this end, this paper explores and investigates several federated learning aggregation strategies and algorithms. At the beginning, a brief summary of federated learning is given so that the context of an aggregation algorithm within a FL system can be understood. This is followed by an explanation of aggregation strategies and a discussion of current aggregation algorithms implementations, highlighting the unique value that each brings to the knowledge. Finally, limitations and possible future directions are described to help future researchers determine the best place to begin their own investigations.

Keywords: federated machine learning; federated learning; collaborative artificial systems; distributed machine learning; decentralized machine learning; distributed intelligent systems; aggregation algorithms; privacy-preserving technology

Citation: Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Reviewing Federated Learning Aggregation Algorithms; Strategies, Contributions, Limitations and Future Perspectives. *Electronics* **2023**, *12*, 2287. https://doi.org/10.3390/electronics12102287

Academic Editors: Jesús Ángel Román Gallego, María-Luisa Pérez-Delgado, María Concepción Vega Hernández, Alfonso Jose Lopez Rivero and Daniel Hernández De la Iglesia

Received: 24 March 2023
Revised: 16 May 2023
Accepted: 16 May 2023
Published: 18 May 2023

1. Introduction

The industrial revolutions, from the first to the fourth, marked significant turning points in human history, as they brought about a fundamental shift from manual labor to machine production. This led to an increase in production efficiency that enabled faster production at lower costs [1]. With the advent of information and communication technologies (ICTs) such as computers and later the Internet, the pace of technological progress has increased even further. These tools have revolutionized the way people communicate, work, and access information. They enable real-time global communication and interaction and facilitate access to vast amounts of data at a glance. The development of computers and machines has led to unprecedented levels of automation, making many tasks faster and more efficient than ever before [2].

In the 1960s, the introduction of artificial intelligence (AI) as a branch of computer science played an important role in revolutionizing computer and machine technology worldwide [3]. AI, known as the algorithms that enable computers to perform tasks that normally require basic intelligence, and to autonomously interpret and analyze large amounts of data, make predictions, act independently, interact with the environment, and even perform difficult tasks [4]. Moreover, the field of AI has been a hot research topic since its invention, which has led to several AI branches and offshoots, such as machine learning, deep learning, and others [5]. In this context, machine learning is defined as a set of algorithms that allow computers to "self-learn" from training data and improve their knowledge over time without being explicitly programmed. Machine learning algorithms aim to detect patterns in data and learn from them to make their own predictions [6]. In short, machine learning algorithms and models learn through experience. Technically, a computer program is written by engineers and given a set of instructions that enable it to convert input data into a desired output. In contrast, machine learning algorithms are designed to learn with minimal or no human intervention and improve their knowledge over time. The great success of ML and its great potential in classification and regression problems, as well as its ability to handle both supervised and unsupervised learning approaches, have attracted researchers from various fields [7]. Later reviews show the variety of applications of ML, which can be found in almost all areas of our lives, especially in the areas listed in Table 1 below.

Table 1. Machine learning common fields of implementation.

Field of Implementation
E-commerce and product recommendations [8,9]
Image, speech, and pattern recognition [8,9]
User behavior analytics and context-aware smartphone applications [8,9]
Healthcare services [10–12]
Traffic prediction and transportation [8,13]
Internet of things (IoT) and smart cities [13]
Cybersecurity and threat intelligence [14]
Natural language processing and sentiment analysis [15]
Sustainable agriculture [16]
Industrial applications [17]

1.1. Machine Learning Techniques: A Taxonomy

Artificial intelligence and its descendant, machine learning, are used in a wide variety of real-world applications. Thousands, if not millions, of implementations are available in the areas mentioned in the previous section. Moreover, the algorithms of ML can be classified into different groups depending on their classification perspective. These algorithms are traditionally classified into supervised, unsupervised, semi-supervised, and reinforcement learning [5–7]. However, this classification only considers the data analyzed by the model or the so-called learning style and ignores other possible classification bases. In this context, the function or goal of the algorithm as well as the architecture can serve as classification factors and provide an extended taxonomy for ML algorithms.

1.1.1. Classification per Learning Style

Machine learning workflows specify what steps are performed in an ML project. Data acquisition, data preprocessing, model training and fine-tuning, evaluation, and production deployment are generally the common processes. Consequently, the type of data obtained determines the machine learning algorithm. From this point of view, the four categories listed below can be defined [8–10]:

- Supervised Learning: This refers to the types of ML where machines are trained with labeled input and then predict output based on that data. Labeled data means that the input data have been labeled with the corresponding output. The training data serve

as a supervisor that teaches the computers how to correctly predict the output. Then it can be described as a process of providing the model ML with appropriate input and output data so that it can identify a function to map the input and output variables;

- Unsupervised Learning: An algorithm that operates only on input data and has no outputs or target variables. Consequently, unlike supervised learning, there is no teacher to correct the model. In other words, it is a collection of problems where a model is used to explain or extract relationships in data;
- Semi-Supervised Learning: This is a form of supervised learning in which the training data includes a small number of labeled instances and a large number of unlabeled examples. It attempts to use all available data, not just the labeled data as in supervised learning;
- Reinforcement Learning: This defines a class of problems where the intelligent model operates in a given environment and must learn how to act based on inputs. This means that there is no given training dataset, but rather a goal or collection of goals for the model to achieve, actions it can take, and feedback on its progress toward the goal. In other words, the goal is to learn what to do, how to map events to actions in order to maximize a numerical reward signal, not dictating to the model what actions to perform, but figuring out through trial and error which activities yield the greatest reward.

1.1.2. Classification per Function

Machine learning algorithms, on the other hand, can be categorized by the goal of the model. The goal, also referred to as the function, is the output of the model and determines the type of model to be used. The different types of ML can be defined as follows [8–10]:

- Classification: the process by which a ML algorithm predicts a discrete output or so-called class. Depending on the type of class to be predicted, this class can be divided into the following groups:
  - Binary Classification: refers to algorithms that can predict only one of two labels, e.g., classifying emails as spam or not;
  - Multi-Class Classification: refers to algorithms with more than two class labels, where there are no normal and abnormal results. Instead, the examples are classified into one of several known classes;
  - Multi-Label Classification: the collection of algorithms that predict the output of a label class, with no limit to how many classes the instance can be assigned to.
- Regression: the process by which a ML algorithm can predict a continuous output or a so-called numerical value;
- Clustering: the process of categorizing a set of data instances or points so that those in the same group are more similar and different from data points in other groups. It is essentially a collection of instances based on their similarity and dissimilarity;
- Dimensionality Reduction: the process of minimizing the number of variables in the supplied data, either by selecting only relevant variables (feature selection) or by creating new variables that reflect several others (feature extraction);
- Representation Learning: the process of determining appropriate representations for input data, which often involves dimensionality reduction.

1.1.3. Classification per Architecture

Another approach to classifying machine learning algorithms can be based on the underlying architecture of the system. In this context, two main categories can be defined [18,19]:

- Centralized Architecture: the traditional ML architecture, where data is collected on a machine running the model;
- Distributed Machine Learning: the ML paradigm that benefits from a decentralized and distributed computing architecture where the ML process is split across different

nodes, resulting in a multi-node algorithm and system that provides better performance, higher accuracy, and better scalability for larger input data.

That being said, federated machine learning, also known as federated learning, which is the core topic of this paper, is a decentralized ML strategy that will be discussed in detail in later sections. Figure 1 depicts the proposed taxonomy for machine learning algorithms.

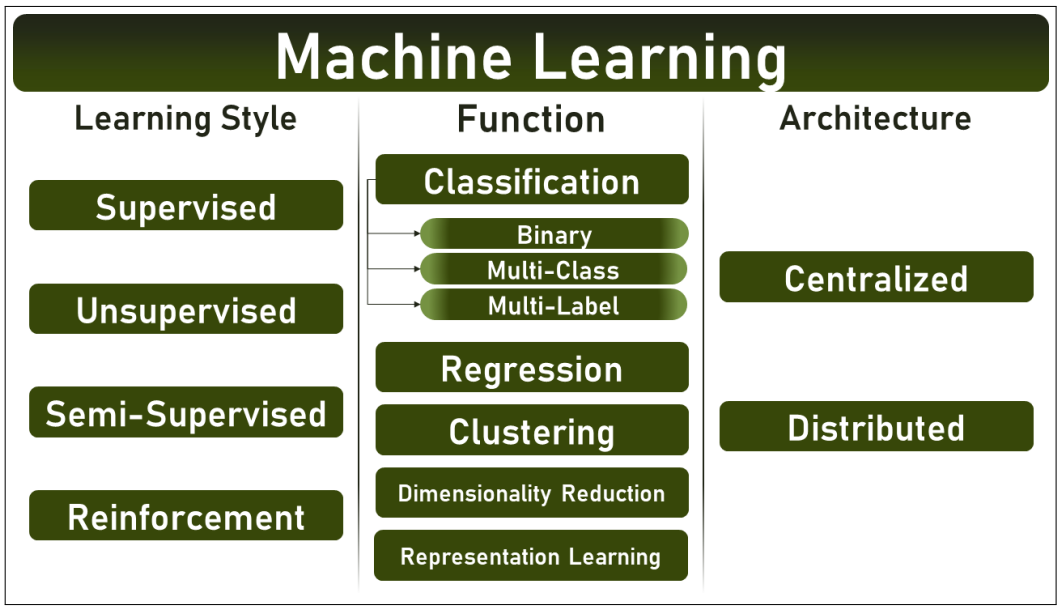

Figure 1. Machine learning algorithms taxonomy.

1.2. Machine Learning Challenges

Machine learning has become an important part of modern technology, enabling computers to perform complicated tasks with increasing efficiency and accuracy. However, despite its obvious benefits, there are several problems in the field of machine learning, ranging from technical issues of data quality and algorithm development to ethical concerns about privacy, fairness, and transparency. Therefore, there is also a great need to address these difficulties and ensure that machine learning benefits society in a responsible and sustainable manner. The known challenges in ML are discussed below.

1.2.1. Data Related Challenges

Machine learning algorithms are trained with datasets to determine relationships between them, discover trends and patterns, and predict future outcomes. However, since the workflow of the ML algorithms begins with data acquisition, as described earlier, data plays a critical role in shaping the quality and efficiency of a machine learning algorithm. The following describes the most common challenges in ML related to data [20,21]:

- Data Availability/Accessibility: data may not be available or accessible;
- Data Locality (Data Islands): data are scattered into different and non-related entities;
- Data Readiness: data may be heterogeneous, noisy or imbalanced;
- Data Volume: difficulty of working with datasets that are too large or too small;
- Feature Representation and Selection: selecting the optimal features for best results.

1.2.2. Models Related Challenges

After preparing the data for the ML algorithms, selecting the most appropriate model for the problem at hand is another problem that experts usually grapple with. The challenges associated with the ML models themselves are in the following list [22–24]:

- Accuracy and Performance: increasing the accuracy of the models;

- Model Evaluation: correct evaluation of the performance of the models;
- Variance and Bias: affects the results and confidence;
- Explainability: resolving back-box identity of ML models.

1.2.3. Implementation-Related Challenges

In addition, the implementation phase, which refers to the training of the model, evaluation and results, and other steps, is a big area of interest. The implementation phase is associated with several challenges, which can be summarized as follows [22–24]:

- Real-Time Processing: adapting models to operate in real time;
- Model Selection: selecting the best model suitable for the problem under study;
- Execution Time and Complexity: ML models may require high computational power.

1.2.4. General Challenges

On the other hand, there are a number of challenges that do not fall into any of the previously mentioned categories. More attention needs to be paid to these challenges, which are identified below, in order to increase the efficiency of the ML domain and improve its usability [20,23,25]:

- Users' Data Privacy and Confidentiality: data are protected by numerous regulations;
- User Technology Adoption and Engagement;
- Ethical Constraints.

1.3. Privacy Criticality: Federated Learning as a Solution

Privacy is a fundamental right, and protecting sensitive personal information is critical in today's digital age. Privacy issues can arise when collecting, storing, and analyzing data in the context of machine learning, as algorithms rely heavily on personal data to train models and make predictions. These challenges stem from the increasing number of data breaches, which require more and more solutions as their negative impact grows. According to a survey conducted by IBM in 2020, 56% of data breaches are due to malicious attacks, while 32% are due to system glitches or human error, as mentioned in the IBM Cost of a Data Breach Report [26]. This report is a global study that examines data breaches in different countries and industries. The report analyzes data breaches in different regions, including North America, Europe, Asia-Pacific, and the Middle East, and covers various industries, including healthcare, financial services, retail, and manufacturing. According to this report, the average total cost of a data breach in 2020 was $3.86 million USD, with an average cost of $150 USD per record, highlighting the economic burden of these illegal acts in addition to ethical violations and privacy [26]. Therefore, it is more than necessary to reduce the impact of malicious attacks or system disruptions on ML.

In this context, Google proposed federated learning in 2016, which later proved to be a solution to privacy issues [27]. Federated learning is thus defined as a machine learning method that allows numerous devices or organizations to train a model collaboratively without sharing raw data. Alternatively, the model is trained on local data, and only the model updates are shared with a central server, which enables privacy-preserving and decentralized model training [27,28]. By decentralizing the ML process, and reducing the amount of data transferred between devices and servers, federated learning was able to minimize the risk of data leaks from malicious attacks and system failures. These results were confirmed by different studies, for instance [29], where the authors proved that their federated learning framework preserves up to 99% of bandwidth and 99% of energy for clients during communication.

After that, in the federated learning domain, an aggregation algorithm is defined as the technique that aggregates the result of training multiple smart models on the clients' side using their local data. This algorithm is the part that handles the fusion of the results obtained from the local clients training, and updating the global model. The aggregation

algorithms in federated learning are discussed and reviewed in detail in this paper. In addition, Figure 2 illustrates the aggregation algorithm in the federated learning domain.

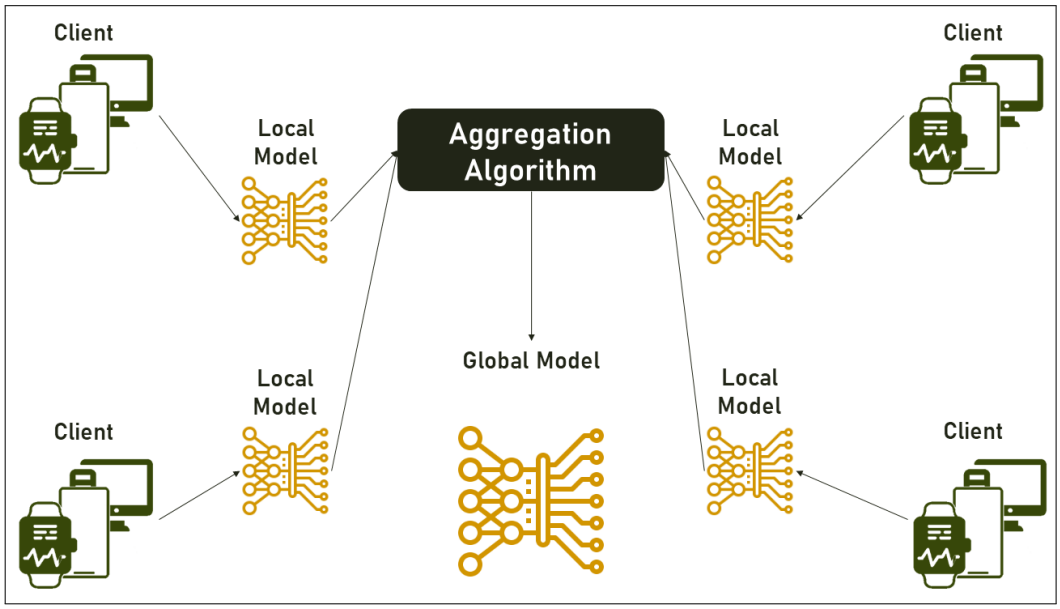

Figure 2. Federated learning aggregation algorithms.

1.4. Article Outline and Contributions

In this article, aggregation algorithms in the field of federated machine learning are discussed in detail. To achieve this goal, the topic is discussed in detail in the following sections. In Section 2, federated learning is discussed from different perspectives, including both definitions and technical perspectives. In addition, aggregation is defined and its various approaches are explored. In Section 3, the state of the art of aggregation algorithms in federated learning is presented and a taxonomy for the available algorithms is proposed. In Section 4, an overview of these algorithms is given, exploring the contribution of each algorithm along with its limitations and future prospects. Finally, the aforementioned sections are followed by a conclusion that summarizes the entire work. In this context, this article attempts to answer the following research questions:

- What is federated machine learning?
- What is aggregation in federated learning?
- What are the different aggregation strategies?
- What is the state of the art in FL aggregation algorithms?
- What possible taxonomy can be established for these algorithms?
- What are the area(s) of contribution for each proposed aggregation algorithms?
- What are the limitations to date in this area?
- What future perspectives can be pursued to improve aggregation in the field FL?

The topic of federated machine learning has been a hot and timely topic lately. Although it was first used in 2016, FL has become the focus of interest among computer science researchers because it is expected to play a role in advancing machine learning as a privacy-preserving technology that will help overcome the increasing conflicts associated with it. Dozens, if not hundreds, of studies have already been published in this regard. However, to our knowledge, none of these studies have addressed, inclusively and completely, the issue of aggregation algorithms in FL, as is the case in this study. For example, the authors of [30] discussed privacy and security in FL aggregation algorithms, but did not mention other aggregation approaches that address other goals, such as reducing communication and computational overhead, scalability, or other issues, as is the case here. Consequently, there is a great need to study this area in order to direct future efforts to

the crucial work that best contributes to the advancement of the field FL. Therefore, this article attempts to fill the gap in this area by providing a complete overview of the currently available federated learning aggregation algorithms, discussing their contributions and limitations, and providing future perspectives that researchers can pursue in their future studies. The contributions of this article can be summarized as follows:

- Differentiating between exchanging model updates, parameters or gradients in FL;
- Explaining aggregation algorithms in federated learning domain;
- Describing aggregation and its different approaches;
- Presenting the state of the art of aggregation algorithms in federated learning;
- Proposing a taxonomy that can be followed in categorizing FL aggregation algorithms;
- Discussing the contributions of each of the available FL aggregation algorithms;
- Studying the limitations of the aforementioned algorithms;
- Examining possible future prospects, which can be used as a starting point for further studies to improve aggregation algorithms in the field of FL.

2. Materials and Methods: Studying Federated Learning and Aggregation

Privacy and security are paramount in the age of big data. The more data that are collected and shared, the greater the risk of data breaches. Federated machine learning offers a compelling answer to these problems by allowing data to be analyzed and shared without ever leaving the device on which it was collected. Federated machine learning can realize the full potential of big data while protecting privacy and security by leveraging advanced algorithms and unique aggregation approaches. This section presents the technological foundations of federated machine learning and the various aggregation strategies that can be used to harness the potential of scattered data. Different approaches to secure privacy-preserving methods are explored, ranging from simple averaging to more advanced methods such as secure multi-party computing and differential privacy.

2.1. Federated Learning: An Overview

In federated machine learning, many parties collaborate to train a single model without exposing their own data to the other entities or a central server. The term "federated machine learning" can also refer to distributed learning with multiple participants. In this technique, each participant trains a model using only the data specific to that participant, and then shares the refined model parameters with a central repository. After receiving updates to the model from all participants, the aggregator merges them into a single, updated version of the model. This process is iteratively repeated until the accuracy of the combined model reaches the target level. Federated machine learning makes it possible to ensure privacy in machine learning, where sensitive data remain under the control of its original owners by ensuring that the data are stored locally and that data transfer between parties is kept to a minimum [27,28,31].

2.2. Federated Learning: Technical Perspectives

Federated learning is emerging as a privacy-preserving machine learning technology that is not only capable of protecting private data, but also improving the quality of models by facilitating access to more data. This potential stems from the underlying architecture and technical perspectives considered in this context.

2.2.1. Underlying Architecture

Typically, a federated machine learning environment consists mainly of a set of four groups of entities, namely, the main server, the parties, the communication framework, and the aggregation algorithm [31–33]. Each of these entities assumes a specific role in the federated learning process. These entities can be defined as follows:

- Central Server: the entity responsible for managing the connections between the entities in the FL environment and for aggregating the knowledge acquired by the FL clients;

- Parties (Clients): all computing devices with data that can be used for training the global model, including but not limited to: personal computers, servers, smartphones, smartwatches, computerized sensor devices, and many more;
- Communication Framework: consists of the tools and devices used to connect servers and parties and can vary between an internal network, an intranet, or even the Internet;
- Aggregation Algorithm: the entity responsible for aggregating the knowledge obtained by the parties after training with their local data and using the aggregated knowledge to update the global model.

Following this, the classical approach of the learning process is achieved in the environment of FL by repeating the following steps:

1. Central server receives connection from clients and sends them initial global model;
2. Parties receive initial copy of model, train it with their local data, and send results back to central server;
3. The central server receives the locally trained models, which are aggregated with the correct algorithm;
4. The central server updates the global model based on the aggregation results and sends the updated version to the clients;
5. Repeat the above steps until the model converges or until the server decides to stop.

In Figure 3 below, the underlying architecture, entities, and process steps are illustrated for a better description of the FL environment.

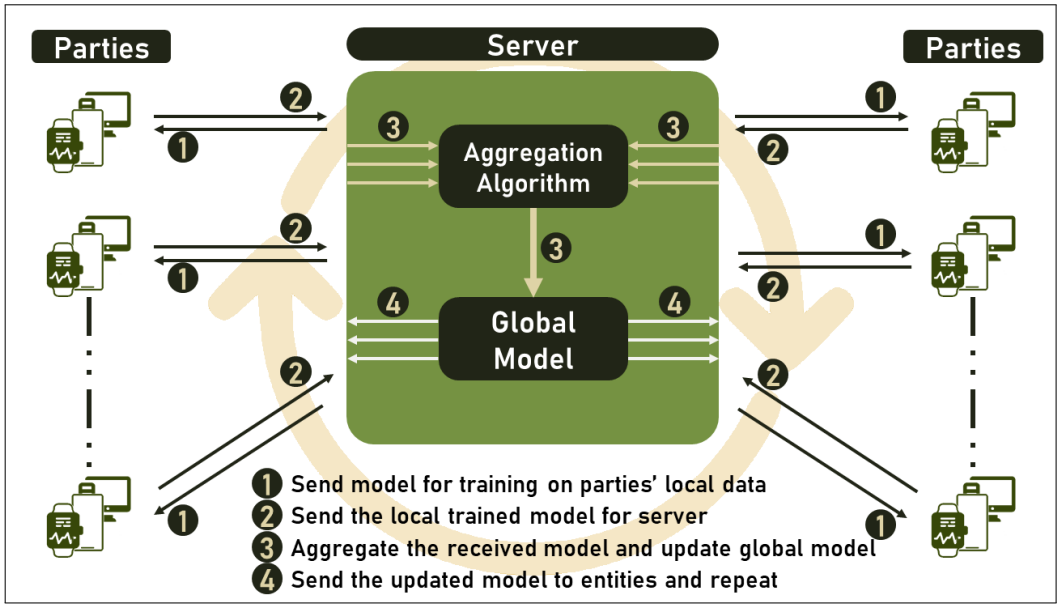

Figure 3. Federated learning process and environment.

2.2.2. Exchanging Models, Parameters, or Gradients

In classical machine learning, data are collected on the server so that the models can be trained directly, building their ability to predict future instances. In contrast, in the federated learning environment, the data are not collected on the server, but the models are shared between the server and the clients so that training can be performed on the local data, which helps to maintain privacy. The term "exchange of models" is often used in federated learning research, but it does not describe the different approaches to message exchange between the central server and the clients. For example, there are other alternatives for sending and receiving models, such as exchanging gradients or model parameters instead of the model itself. In this context, the different approaches to message exchange in the FL environment can be described as follows:

- Exchanging Models: This is the classical approach, where models are exchanged between server and clients. This approach is very costly in terms of communication and also poses security problems, since the models can be intercepted with malicious intent to extract the data used for training;
- Exchanging Gradients [34,35]: Instead of submitting the entire model to the server, clients in this method submit only the gradients they compute locally. Federated learning with gradient aggregation (FLAG) is another name for this strategy. Each client computes the gradients using its own local data and then submits them to the server, which indicates the direction in which the parameters of the model should be updated to minimize the loss function. After the server collects the gradients, it applies them to the global model. This method has the advantage of both maintaining privacy and reducing communication costs. The divergence of local models is one of the challenges that can arise with this strategy when clients use different learning rates and optimization strategies;
- Exchanging Model Parameters [36,37]: This concept is mainly tied to neural networks where model parameters and weights are usually used interchangeably. Parameters, sometimes called weights, are the values assigned to connections between neurons in a neural network where the input from one layer of neurons is used by the next layer to produce an output, which is then weighted. During training, the weights are adjusted to reduce the discrepancy between the expected and actual output. This method has the potential to reduce the burden of communication costs in an FL environment while maintaining the confidentiality features of the FL approach. However, this method assumes that all clients have the same model architecture, which may not be the case for all implementations, leading to numerous problems. There is also the possibility that the method will not be effective if the client data are too large or if the data are not balanced on the client side;
- Hybrid Approaches: Two or more of the above methods can be combined to form a hybrid strategy that is particularly suited to a particular application or environment. For example, the server can broadcast the initial parameters for the clients to all nodes and then receive updated models from the nodes, which it then combines with its own to create a global model.

In Table 2 below, the different types of messages exchanged between server and clients in federated learning environment are summarized, along with their advantages and disadvantages.

Table 2. FL exchanged messages: models updates vs. parameters vs. gradients.

Type	Concept	Advantages	Disadvantages
Exchanging Models	Models are sent between server and clients	Ease of implementation	High communication cost Less secure
Exchanging Gradients	Only gradients are exchanged between entities	Lower communication cost Higher security	Local models divergence
Exchanging Model Parameters	Only weight and parameters are exchanged	Lower communication cost Higher security	Limitation to neural networks Unified client model architecture Not effective with big or imbalanced data
Hybrid Approach	Merging two or more of the above approaches	Fit for specific cases	Generated frameworks may not be re-used

2.3. What Is Aggregation in FL?

Federated learning is a collaborative, decentralized machine learning technology where entities within the network collaborate in training a global model without sacrificing the security of private data. To make the process of integrating the obtained results

efficient, an aggregation approach is essential, whether the messages exchanged are the models themselves, some or all of their parameters, or gradients. Each client trains its own model on its own data and then transmits these results to the server, where an aggregation approach uses these results to generate the group's collaborative relationship. Then, this information is used to update the global model. The central server can leverage the diversity of the training data without actually having access to the raw data by aggregating the model updates sent from each device. The various aggregation methods available for use in federated machine learning each have their own advantages and disadvantages. However, aggregation in federated machine learning goes beyond simply merging model updates. In addition to tracking model performance across devices, additional statistical indicators such as loss functions or accuracy measurements can also be aggregated. Furthermore, aggregation can be carried out in a hierarchical manner, aggregating local models on intermediate servers before sending them to the central server, enabling large-scale federated learning systems. This is why the aggregation algorithm is such a fundamental concept in federated learning; it ultimately determines the success of model training and whether or not the resulting model is practical to use [28,29,31,32].

2.4. Different Approaches of Aggregation

Aggregation algorithms in federated learning are important because of their role in updating global models. There are many aggregation approaches that can be followed in building the aggregation algorithm in a federated learning environment. In federated learning, a variety of aggregation algorithms are used depending on the goals to be achieved, such as protecting user privacy, increasing the convergence rate, and reducing the damage caused by fraudulent customers. Each of these approaches has its advantages and disadvantages, and some are better suited to certain contexts of federated learning than others. In this section, the best- known aggregation algorithms are mentioned, since there may be approaches other than those presented here.

2.4.1. Average Aggregation

This is the initial approach and the most commonly known. In this approach, the server summarizes the received messages, whether they are model updates, parameters, or gradients, by determining the average value of the received updates. Since the set of participating clients is denoted by "N" and their updates are denoted by "w_i", the aggregate update "w" is calculated as follows [38]:

$$w = (1/N) * \sum_{i=1}^{N} w_i \tag{1}$$

2.4.2. Clipped Average Aggregation

This method is similar to average aggregation, where the average of received messages is calculated, but with an additional step of clipping the model updates to a predefined range before averaging. This approach helps reduce the impact of outliers and malicious clients that may transmit large and malicious updates [39]. Since "N" denotes the set of participating clients and "w_i" denotes their relative weights, "$clip(x, c)$" is a function, which clips the values of "x" to a range of "$[-c, c]$", and "c" is the clipping threshold, the total clipped aggregate update "w" is calculated as [39]

$$w = (1/N) * \sum_{i=1}^{N} clip(w_i, c) \tag{2}$$

2.4.3. Secure Aggregation

Techniques such as homomorphic encryption, secure multiparty computation, and secure enclaves make the aggregation process more secure and private in this way. These methods can ensure that client data remain confidential during the aggregation process,

which is critical in environments where data privacy is a high priority [40]. Secure aggregation is the result of integrating security techniques, such as those mentioned earlier, with one of the available aggregation algorithms to create a new secure algorithm. However, one of the most popular secure aggregation algorithms is the differential privacy aggregation algorithm, which proposes a different technique for integrating clients results. This technique is detailed in the next section.

2.4.4. Differential Privacy Average Aggregation

This approach adds a layer of differential privacy to the aggregation process to ensure confidentiality of client data. Each client adds random noise to its model update before sending it to the server, and the server compiles the final model by aggregating the updates with the random noise. The amount of noise in each update is carefully tuned to compromise between privacy and model correctness. If "N" denotes the set of participating clients and "w_i" denotes their relative weights, "n_i" is a random noise vector, drawn from a Laplace distribution with a scale parameter "b", and "b" is a privacy budget parameter, the differentially private aggregate update "w" is calculated as follows [41]:

$$w = \frac{1}{N} \sum_{i=1}^{N} (w_i + b \cdot n_i) \tag{3}$$

2.4.5. Momentum Aggregation

This strategy should help solve the slow convergence problem in federated learning. Each client stores a "momentum" term that describes the direction of model changes in the past. Before a new update is sent to the server, the momentum term is appended to the update. The server collects the updates enriched with the momentum term to build the final model, which can speed up convergence [42].

2.4.6. Weighted Aggregation

In this method, the server weights each client's contribution to the final model update depending on client performance or other parameters such as the client's device type, the quality of the network connection, or the similarity of the data to the global data distribution. This can help give more weight to consumers that are more reliable or representative, improving the overall accuracy of the model. Given that "N" denotes the set of participating clients and "w_i" their relative weights, and their corresponding weights "a_i", the weighted aggregate update "w" is computed as follows [43]:

$$w = (\sum_{i=1}^{N} a_i * w_i) / (\sum_{i=1}^{N} a_i) \tag{4}$$

2.4.7. Bayesian Aggregation

In this approach, the server aggregates model updates from multiple clients using Bayesian inference, which allows for uncertainty in model parameters. This can help reduce overfitting and improve the generalizability of the model [44].

2.4.8. Adversarial Aggregation

In this method, the server applies a number of techniques to detect and mitigate the impact of customers submitting fraudulent model changes. This may include methods such as outlier rejection, model-based anomaly detection, and secure enclaves [45].

2.4.9. Quantization Aggregation

In this approach, model updates are quantized into a lower bit form before being delivered to the server for aggregation. This reduces the amount of data to be transmitted and improves communication efficiency [46].

2.4.10. Hierarchical Aggregation

In this way, the aggregation process is carried out at multiple levels of a hierarchical structure, such as a federal hierarchy. This can help reduce the communication overhead by performing local aggregations at lower levels of the hierarchy before passing the results on to higher levels [47].

2.4.11. Personalized Aggregation

During the aggregation process, this approach considers the unique characteristics of each client's data. In this way, the global model can be updated in the most appropriate way for each client's data, while ensuring data privacy [48].

2.4.12. Ensemble Bases Aggregation

The model is trained on different subsets of clients, called ensembles, and the resulting models are integrated to produce the final model. Each ensemble may have a specific subset of clients and models trained on that customer. The models from each ensemble are then merged to create a final model. This method can help reduce the impact of non-IID data while improving model accuracy. To increase model accuracy, ensemble-based aggregation can be combined with other aggregation approaches, such as weighted aggregation [49].

In Table 3, these aggregation algorithms are summarized, showing at the same time their main concept as well as their advantages and disadvantages.

Table 3. FL aggregation approaches: concepts, advantages, and disadvantages.

Approach	Main Concept	Advantages	Disadvantages
Average Aggregation	Average the clients updates	Simple and easy to implement Can improve model accuracy	Sensitive to outliers and malicious clients May not perform well in cases of non-IID data
Clipped Average Aggregation	Clip the model updates to a predefined range before taking the average	Reduces the effect of outliers and malicious clients Can improve model accuracy	More computationally intensive
Secure Aggregation	Use techniques such as homomorphic encryption or secure multi-party computation to ensure privacy	Provides a high level of privacy protection Can still achieve good model accuracy	May be computationally expensive and slower than other aggregation methods Requires careful implementation and management of security protocols
Differential Privacy Average Aggregation	Add random noise to the model updates before aggregation to ensure privacy	Provides a high level of privacy protection	May be slower and less efficient than other aggregation methods The level of noise added can impact model accuracy
Momentum Aggregation	Add a momentum term to the model updates before aggregation to improve convergence speed	Can improve convergence speed and reduce the impact of noisy or slow clients	May be sensitive to the choice of momentum term and the level of noise in the updates
Weighted Aggregation	Weight the contributions of different clients based on performance or other factors	Can improve model accuracy by giving more weight to more reliable or representative clients	Requires careful calibration of weights and may be sensitive to bias or noise in performance metrics

Table 3. *Cont.*

Approach	Main Concept	Advantages	Disadvantages
Bayesian Aggregation	Use Bayesian inference to aggregate model updates and take uncertainty into account	Can improve model generalization and reduce overfitting	Can be computationally expensive and require large amounts of data The Bayesian model assumptions may not hold for all types of data
Adversarial Aggregation	Detect and mitigate the impact of malicious clients or outlier model updates	Can improve model accuracy and reduce the impact of malicious clients	May be computationally expensive and require sophisticated detection and mitigation techniques
Quantization	Reduce the bit representation of model updates before transmission	Can improve communication efficiency and reduce bandwidth requirements	May introduce quantization error that can impact model accuracy The level of quantization needs to be carefully chosen
Hierarchical Aggregation	Perform aggregation at different levels of a hierarchical structure	Can improve communication efficiency by performing local aggregation at lower levels	Requires a well-defined hierarchical structure and careful management of data and aggregation protocols
Personalized Aggregation	Takes clients' unique characteristics into account	Improves model performance by adapting to individual client data	Maintains privacy, but may require additional communication and computational overhead
Ensemble-based Aggregation	Aggregate the model updates of multiple models trained on different subsets of the data	Can improve model accuracy by leveraging the diversity of the models	May be computationally expensive and require careful management of the ensemble models

In this section, federated machine learning has been explained and discussed in detail. Federated learning is a machine learning-based technology that allows smart models to be trained without the need to collect users' private data on central servers. Alternatively, and because of the technical architecture on which FL is built, models are sent to users to be trained on their data, preserving privacy. Another approach in FL involves the exchange of model parameters or gradients. In this context, the messages exchanged between the central server and the clients of FL must be aggregated to produce the final global model. Consequently, the aggregation algorithms in FL are the mechanisms used to integrate knowledge from local models into a global model. Originally, the average aggregation was proposed by Google in their FL aggregation algorithm called FedAvg [38]. Later, several aggregation concepts were proposed in different studies, as explained earlier. In the next section, various FL aggregation algorithms are discussed and the state of the art is also presented.

3. Results: FL Aggregation Algorithm Implementations

In federated machine learning, both clients and server collaborate in training a smart model. The different approaches taken in aggregating locally trained models have led to several aggregation algorithms in recent years. Although it was first proposed by Google in 2016 [38], federated learning emerged as a trending topic that attracted researchers and led to dozens of studies in this area. In this context, several implementations for FL aggregation algorithms can be found in the literature and will be discussed in this section.

3.1. State of the Art

Aggregation algorithms for federated learning are being studied extensively, and researchers are making great efforts to advance this field. In the last six years, twenty-seven implementations were carried out in this context. These implementations are described below. A graphical summary of these implementations is shown in Figure 4 below.

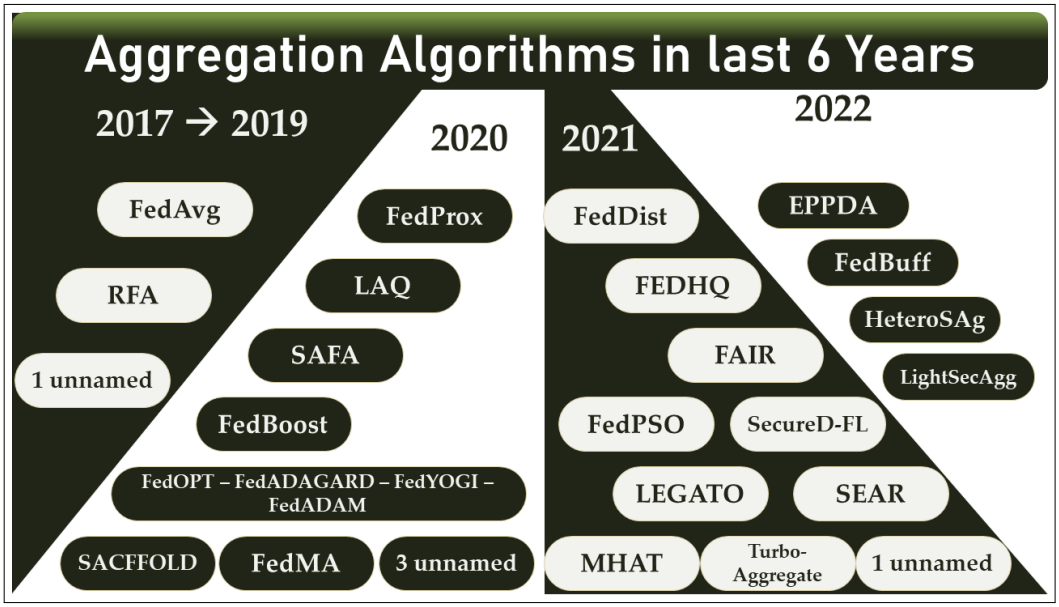

Figure 4. Implementations of FL aggregation algorithms.

3.1.1. "2017–2019": Introducing FL to the Market

The first implementation of a federated learning framework, called FedAvg, was proposed by Google [38], and they were the first to propose training smart models without collecting user data. This article provided the first practical method for FL of deep networks based on iterative model averaging. The authors of this article used five different model architectures and four datasets to evaluate their model. In the same year, the authors of [50] developed a novel communication-efficient, failure-robust protocol for secure aggregation of multiple and high-dimensional data. The proposed protocol allows a server to compute the sum of large data held by the user in a secure manner. The obtained results prove the security of their protocol in the "honest but curious" and "active adversary" settings, maintaining this security even if an arbitrarily chosen subset of users drops out at any point in time. Furthermore, in [51], the authors presented a new approach that is robust to possible poisoning of local data or model parameters. Their model, called robust federated aggregation (RFA), aggregates local updates using the geometric median, which can be efficiently computed using a Weiszfeld-type algorithm. The authors also offered two variants of RFA: a faster variant with robust one-step aggregation and another with intra-device personalization. They tested their model with three tasks from computer vision and natural language processing and their results competed with classical aggregation.

3.1.2. "2020": A Big Step

The year 2020 brought a boost in the development of FL aggregation algorithms. For instance, authors of [52] proposed SCAFFOLD, a new algorithm that uses control variance to correct for 'client drift' in its local updates. The obtained results showed that SCAFFOLD requires fewer rounds of communication and is not affected by data heterogeneity or client sampling. Moreover, SCAFFOLD proved that exploiting client data similarity leads to faster convergence. Moreover, in [53], the authors proposed different versions of federated learning models using different adaptive optimizations, including ADAGRAD, ADAM, and

YOGI, and analyzed their convergence in the presence of heterogeneous data for general nonconvex settings. The obtained results proved the feasibility of these models in reducing convergence in FL. Moreover, in [54], the authors presented an alternative approach called FedBoost, which uses an ensemble of pre-trained base predictors. This method can be used to train a model that can overcome the limitations of communication bandwidth and client memory capacity. With their proposed model, the cost of communication between server and clients could be reduced.

In addition, the authors of [55] proposed FedProx, which is able to deal with heterogeneity in federated learning networks. FedProx is a generalization and reparametrization of FedAvg, and they proved that their model provides more robust convergence than FedAvg over a range of real-world heterogeneous datasets. Moreover, the authors of [56] proposed a federated matching average (FedMA) algorithm, which constructs the joint global model layer-by-layer by matching and averaging hidden elements with similar feature extraction signatures. Their results show that FedMA outperforms the classical algorithms of FL in processing real-world datasets and also reduces the overall communication overhead. In the same context, the authors of [57] investigated the analog gradient aggregation (AGA) solution to overcome the communication resource constraints in FL applications. They proposed both new communication and learning approaches to improve the quality of gradient aggregation and accelerate the convergence speed. In addition, in article [58], the authors proposed a low-complexity approach that preserves user privacy and uses significantly fewer computational and communication resources.

Furthermore, in [59], the authors proposed a new approach to selective model aggregation based on a two-dimensional contract theory as a distributed framework to facilitate the interaction between FL entities. They tested their approach with two datasets, MNIST and BelgiumTSC. The obtained results showed that their model outperformed the original FL model, i.e., FedAvg. Moreover, the authors of [60] developed a new model that is characterized by adaptive communication of quantized gradients. The key idea of their model is quantization of gradients as well as skipping of less informative quantized gradient communications by reusing previous gradients. Quantization and skipping lead to 'lazy' worker–server communication, which explains the name of their model as the lazily aggregated quantized (LAQ) gradient.Their model showed a significant reduction in communication compared to other FL approaches. In addition, the authors of [61] proposed a semi-synchronous FL protocol, referred to as SAFA, to improve the convergence rate in heterogeneous FL networks. The authors introduced new designs for model distribution, client selection, and global aggregation to reduce the negative effects of stragglers, crashes, and model staleness. The obtained results demonstrate that the proposed model efficiently shortens the duration of interconnection rounds, reduces the waste of local resources, and improves the accuracy of the global model at an acceptable communication cost.

3.1.3. "2021": FL toward More Enhancements

In addition, the authors proposed FedDist in [62], a novel approach for FL aggregation, which is able to change its architecture by detecting dissimilarities between clients. This approach improves the personalization and specificity of the model without compromising generalization. In addition, the authors of [46] proposed federated learning with heterogeneous quantization (FedHQ), which accelerates convergence by computing and piggybacking the instantaneous quantization error as each client uploads the local model update, and the server dynamically computes the appropriate weight for the current aggregation. Their results show that the performance of FedHQ outperforms FedAvg with an accelerated convergence rate. Similarly, in [63], the authors proposed a novel system known as federated learning with quality awareness (FAIR), which consists of three main components. The first component is learning quality estimation, which uses historical learning records to estimate the user's learning quality. The second component is the quality-aware incentive mechanism, which reverses the auction problem to encourage the participation of users with high learning quality. The third component is the model

aggregation, where only ideal models are aggregated in this process to optimize the global model. Their conducted experiments demonstrated the effectiveness of FAIR.

Similarly, in [64], the authors proposed the new model called federated particle swarm optimization (FedPSO), which has increased robustness to unstable network environments. This is achieved by modifying the data that clients send to servers by transmitting score values instead of large weights of local models. In addition, FedPSO has improved network communication performance. Tests conducted by the authors have shown that their model has an improved communication approach where data transmission has been minimized, and that it has improved accuracy even in unstable networks. The authors of [65] also presented their model, called layerwise gradient aggregattion (LEGATO). LEGATO is a scalable and generalizable aggregation approach. Their model uses a dynamic gradient weighting scheme that processes gradients based on layer-specific robustness. Experiments conducted by the authors showed that LEGATO is computationally more efficient than previous models of FL. Moreover, LEGATO proved its efficiency against attacks such as the Byzantine attack. In addition, the authors proposed a new model in [66] called model-heterogenous aggregation training (MHAT) FL. The model relies on knowledge distillation to extract update information from the heterogeneous model of all clients and then train a supporting model on the server to understand the information aggregation. By relieving clients from using a unified model, computational resources are significantly reduced, and the convergence accuracy of the model also remains acceptable. The efficiency and applicability of this model has been demonstrated through several tests in this paper.

In response, the authors of [67] proposed a new federated learning model with an improved communication protocol to minimize privacy leakage. Unlike previous work that used differential privacy or homomorphic encryption, the proposed protocol controls the communication between participants in each round of aggregation. This communication pattern was inspired by combinatorial block design theory. The authors evaluated their model using tests with nine datasets distributed over fifteen sites. The obtained results demonstrate the efficiency of this model in minimizing privacy leakage. In addition, the authors of [68] proposed a new FL model based on a reputation-based aggregation methodology. The methodology scales the aggregation weights of users according to their reputation value, which is calculated using the performance metrics of their trained local model in each training round. This reputation value can therefore be considered as a metric for evaluating the direct contributions of each trained local model. The tests conducted by the authors have shown that their model outperforms previous implementations, especially in non-independent and identically distributed (non-IID) FL scenarios. In addition, in [69], the authors proposed a new model FL called the secure and efficient aggregation framework (SEAR). SEAR is a Byzantine-robust model for federated learning. The model relies on intel software guard extensions (SGX) to protect clients' locally trained models from Byzantine attacks. Considering the memory limitation in their concurrent trusted Intel SGX memory, the authors proposed to use two data storage modes to efficiently implement aggregation algorithms. Experiments conducted by the authors showed that SEAR is computationally efficient and robust against attacks. Furthermore, in [70], the authors proposed a secure aggregation framework FL called turbo-aggregate. This framework uses a circular multi-group strategy to efficiently aggregate locally trained models. Moreover, the framework uses additive secret sharing to incorporate aggregation redundancy to deal with user failures while maintaining the privacy of all users. The framework was tested and the results showed that, first, it provides an increase in aggregation speed of up to 40 times compared to previous implementations and, second, the total runtime grows almost linearly with the number of users, which increases scalability.

3.1.4. "2022": The Journey Continues

Recently, in [71], the authors proposed an efficient privacy-preserving data aggregation (EPPDA) mechanism. EPPDA is based on secret sharing and has an efficient fault-tolerance method to deal with user disconnection. The authors tested their model to show that it

is robust against reverse attacks and user connection disruption. In addition, the authors of [72] proposed a new FL model called federated buffered asynchronous aggregation (FedBuff). FedBuff is independent of the optimizer choice and combines the best features of synchronous and asynchronous FL. FedBuff was found to be 3.3 times more efficient than synchronous FL and up to 2.5 times more efficient than asynchronous FL. In addition, the authors of [73] proposed HeteroSAg, that enables secure aggregation with heterogeneous quantization. Their strategy was based on a grouping scheme that divides the network into groups and divides local model updates from users into segments. Therefore, aggregation is applied to segments with specific coordination between users instead of being applied to the local model. This strategy allows the edge users to adapt to their available communication resources, thus achieving a better trade-off between training accuracy and communication time. The tests conducted by the authors also show that HeteroSAg is robust against Byzantine attacks. Finally, in [74], LightSecAgg was proposed, which is based on reconstructing the aggregate mask of active users using " mask coding/decoding" instead of "random-seed reconstruction of the dropped users ". LightSecAgg shows a reduction in overhead for resilience against lost users. In addition, it provides a modular system design and optimized parallelization on the device for a scalable implementation that improves the speed of concurrent data exchange. The authors tested their model with four datasets to show its resilience to dropouts and significant reduction in training time.

3.2. FL Aggregation Algorithm Implementations Taxonomy

The growing interest in federated learning aggregation approaches promises to energise the field and encourage the adoption of this emerging technology in real-world applications. The available aggregation algorithms can be classified under different aspects besides the year of introduction, as mentioned earlier.

3.2.1. Classification by Area of Contribution

The analysis of the previously mentioned implementations leads to a summary of their contribution areas in the list below. In addition, Table 4 below shows a summary of the contribution of each implementation:

- Improving model aggregation;
- Reducing convergence;
- Handling heterogeneity;
- Enhancing security;
- Reducing communication and computation cost;
- Handling users' failures (fault tolerance);
- Boosting learning quality;
- Supporting scalability, personalization and generalization.

However, the achievements of the federated learning aggregation algorithms mentioned earlier focused mainly on the aggregation itself or on reducing communication costs. The other contribution areas were less explored. For example, of the twenty-seven algorithms mentioned, fifteen targeted global model aggregation and twelve targeted communication cost reduction, while only three targeted learning quality improvement, and only one targeted personalization. This distribution is shown in the graph provided in Figure 5 below (in the pie chart, total will not add up to 100% since one study may contribute to more than one area).

Table 4. Contributions of FL aggregation algorithm implementations.

Ref.	Year	Given Name	Model Aggregation	Convergence Reduction	Heterogeneity	Security	Communication Cost	Computation Cost	Fault Tolerance	Learning Quality	Scalability	Personalization
[38]	2017	FedAVG	✓									
[50]	2017	-				✓	✓					
[51]	2019	RFA	✓			✓						
[52]	2020	SCAFFOLD	✓	✓			✓					
[53]	2020	FedOPT FedADAGAR FedYOGI FedADAM	✓	✓	✓							
[54]	2020	FedBoost					✓					
[55]	2020	FedProx		✓	✓							
[56]	2020	FedMA	✓				✓					
[57]	2020	-	✓	✓			✓					
[58]	2020	-					✓	✓				
[59]	2020	-	✓				✓					
[60]	2020	LAQ					✓					
[61]	2020	SAFA	✓				✓	✓	✓			
[62]	2021	FedDist			✓							✓
[46]	2021	FEDHQ	✓	✓								
[63]	2021	FAIR	✓							✓		
[64]	2021	FedPSO					✓			✓		
[65]	2021	LEGATO				✓	✓	✓			✓	
[66]	2021	MHAT	✓		✓			✓				
[67]	2021	-				✓						
[68]	2021	-	✓		✓					✓		
[69]	2021	SEAR				✓		✓				
[70]	2021	Turbo-Aggregate	✓					✓			✓	
[71]	2022	EPPDA				✓			✓			
[72]	2022	FedBuff	✓									
[73]	2022	HeteroSAg	✓			✓	✓					
[74]	2022	LightSecAgg						✓	✓			

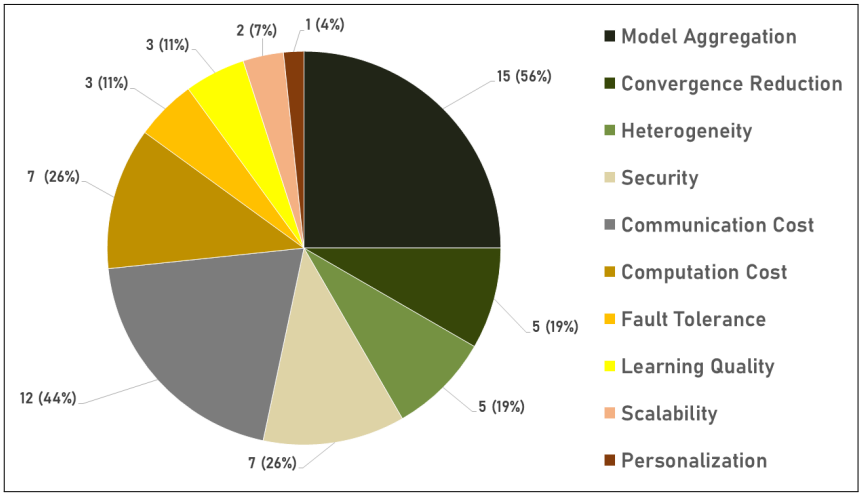

Figure 5. Count per contribution area.

3.2.2. Classification by the Aggregation Approach

On the other hand, considering the aggregation approaches followed in the algorithms, we can classify these implementations into the mapping shown in Table 5 below.

Table 5. Aggregation approaches followed in state of the art of FL aggregation algorithms.

Ref.	Year	Given Name	Aggregation Approach
[38]	2017	FedAVG	Averaging Aggregation
[50]	2017	-	Secure Aggregation
[51]	2019	RFA	Averaging Aggregation
[52]	2020	SCAFFOLD	Secure Aggregation
[53]	2020	FedOPT	Weighted Aggregation
		FedADAGAR	Differential Privacy Average Aggregation
		FedYOGI	Personalized Aggregation
		FedADAM	-
[54]	2020	FedBoost	Ensemble-Based Aggregation
[55]	2020	FedProx	Weighted Aggregation
[56]	2020	FedMA	Personalized Aggregation
[57]	2020	-	Personalized Aggregation
[58]	2020	-	Secure Aggregation
[59]	2020	-	Personalized Aggregation
[60]	2020	LAQ	Quantization Aggregation
[61]	2020	SAFA	Secure Aggregation
[62]	2021	FedDist	Weighted Aggregation
[46]	2021	FEDHQ	Quantization Aggregation
[63]	2021	FAIR	Personalized Aggregation
[64]	2021	FedPSO	Ensemble-Based Aggregation
[65]	2021	LEGATO	Personalized Aggregation
[66]	2021	MHAT	Personalized Aggregation
[67]	2021	-	Secure Aggregation
[68]	2021	-	Weighted Aggregation
[69]	2021	SEAR	Secure Aggregation
[70]	2021	Turbo-Aggregate	Secure Aggregation
			Personalized Aggregation
[71]	2022	EPPDA	Secure Aggregation
[72]	2022	FedBuff	Ensemble-Based Aggregation
[73]	2022	HeteroSAg	Secure Aggregation
			Quantized Aggregation
[74]	2022	LightSecAgg	Secure Aggregation

As shown in the table above, most implementations focus on the secure aggregation approach, which was implemented in 10 of the 27 available studies. Figure 6 below illustrates the distribution of implementations per the aggregation approach followed, with the approaches isted in Section 2.4 and summarized in Table 3.

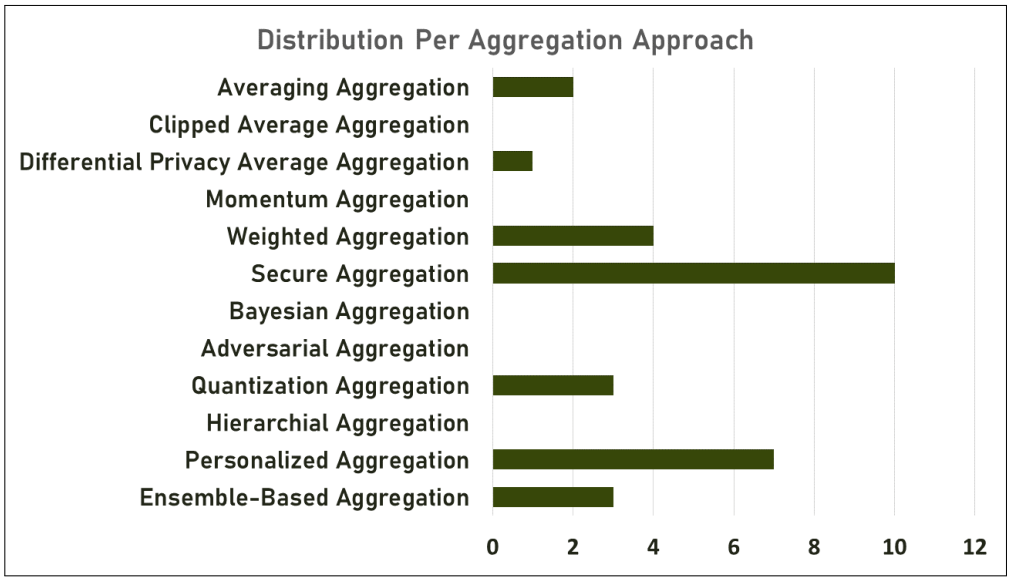

Figure 6. Count per aggregation approach.

Federated learning is growing rapidly as it is expected to play a critical role in revolutionizing the field of machine learning. Since the first FL aggregation algorithm, called FedAvg [38], dozens of aggregation algorithms have been proposed. In this context, FedAvg was fraught with some challenges and shortcomings, which was the main goal of the later studies. Therefore, each of the proposed algorithms contributed to the body of knowledge in FL with a different topic. For example, some were focused on reducing convergence costs, some on reducing computation and communication costs, some on security, and so on. Consequently, the proposed techniques can be classified from the perspective of their contribution domain, or they can even be classified according to the aggregation approach they follow. All these details have been mentioned in this section, and in the next section, the areas to which they contribute will be discussed in detail and, finally, challenges and future perspectives will be identified.

4. Discussion

Federated machine learning introduced a new concept to the field of artificial intelligence. It offers the possibility of improving the accuracy of intelligent models while preserving privacy, since user data are not collected on a central server as in classical machine learning. Instead, model updates, parameters, or gradients are shared between the server and FL clients, which are then aggregated to train or update the global model. In this context, different aggregation strategies can be followed, which also leads to a plethora of aggregation algorithms. Consequently, each aggregation follows one or more strategies and is characterized by one or more contributions. Moreover, there are some limitations in these implementations. All these details are discussed in this section.

4.1. Contributions of Aggregation Algorithms

Analysis of the distribution of implementations per area of contribution shows that research in federated learning aggregation algorithms has produced a number of robust algorithms that are also acceptable from the point of view of reduced communication costs. However, from a security point of view, all the implementations carried out focused

on only one type of attack, namely the Byzantine attack. Other attacks have not been extensively covered in the literature, which raises the question of how robust the available methods are against attacks such as reverse attacks, which are the main concern of FL, where attackers can determine users' private data from the local trained model exchanged within the network. In addition, few efforts have been made to improve the learning quality of FL's models, which in turn raises questions about the extent to which the accuracy of ML's traditional algorithms is comparable to that of FL's models. Finally, personalization has only been investigated in a single study, as shown in the table and the graph.

4.1.1. Aggregation

Advances in aggregation strategies in federated learning have been substantial in recent years. Originally, the focus was on simple averaging methods such as federated averaging, which takes the average of local model updates from each client and then updates the global model using the averaged aggregation. This strategy was introduced by Google in 2016, and their proposed framework became known as FedAvg [38]. However, later studies such as [75,76] have shown that FedAvg has several challenges in terms of performance, such as the following:

- Suffering from 'client-drift' and convergence;
- Tuning difficulty;
- High communication and computation cost;
- Significant variability in clients' system characteristics;
- Non-identically distributed data across the network;
- Heterogeneity of devices, users and network channels;
- Sensitivity to local models;
- Scalability issues.

In this regard, the successor aggregation algorithms have tried to solve the above problems, investigating the communication and computation costs in more than ten algorithms such as SCAFFOLD [52], FedBoost [54], SAFA [61] and others. In addition, issues related to heterogeneity, such as the diversity in clients' data and devices, sensitivity to local models, and others have been cited by aggregation algorithms such as FedYOGI [53], FedMA [56], FAIR [63], LEGATO [65], and others, where these algorithms succeeded in creating personalized aggregation algorithms that demonstrated their feasibility in different scenarios, such as clients' data and device heterogeneity, and more. As a result, aggregation itself has grown beyond the initial average integration to gain the ability to address more complex problems. For example, the introduction of the secure aggregation algorithm in [50] opened the door to improving the security of aggregation algorithms. In addition, the weighted and differential aggregation with average privacy introduced in [53] enabled more advanced aggregation algorithms where both security and communication cost are considered in these strategies. The later aggregation algorithms introduced many more aggregation concepts, showing the progress in this area mentioned earlier.

4.1.2. Convergence Reduction

In federated machine learning setup, the term convergence is used to describe the point at which the parameters of the model reach a stable and accurate state on all clients that contribute to the FL process. FedAvg suffered from client drift and convergence problems, as mentioned earlier. However, later implementations of FL aggregation algorithms included several mechanisms to address this problem.

For example, the developers of SCAFFOLD [52] attempted to reduce the communication rounds required for convergence by introducing an adaptive sampling strategy [77]. Then, SCAFFOLD dynamically selects a subset of clients in each communication round based on their similarity to the current global model. The selected clients are then used to update the global model, reducing the diversity between the global model and the selected clients, reducing the required communication rounds, and increasing the convergence speed.

In addition, FedOPT [53] improves convergence by applying optimization steps to both local and global models that allow for more accurate client updates and better alignment with the server's optimization goal, thereby accelerating convergence and improving the overall accuracy of the global model. In addition, the FedProx [55] aggregation algorithm includes a proximal term [78] in the optimization objective to increase the similarity between client updates, making the global model generalizable and able to represent the data of all clients. In other words, the proximal term encourages the FL client updates to be more similar compared to the global model, which increases the convergence speed. Overall, it can be said that the convergence speed problem has been intensively studied by researchers and many solutions have been proposed, including but not limited to those previously mentioned.

4.1.3. Heterogeneity

Traditionally, federated learning aggregation algorithms followed average aggregation to calculate the mathematical median of the received updates before updating the global model based on this average. However, this approach did not seem to be suitable for scenarios where the participating clients have heterogeneous data or so-called non-informally, identically distributed (Non-IID) data [79]. To address this issue, FedOPT [53] proposed to perform local optimization of the clients' dataset using the current global model parameters as a starting point, allowing clients to fit their models to their data, resulting in improved accuracy and generalizability. In contrast, FedMA [56] proposed a matched averaging approach based on finding clients with comparable data distributions and then taking the average of their model updates. The new global model parameters are based on the calculated weighted average. A similar approach, called distribution matching, is also included in FedHQ [46]. Overall, the handling of heterogeneity has been improved by the aggregation algorithms for federated learning developed after the introduction of FedAvg.

4.1.4. Security

On the other hand, security has been an active area of study in aggregation algorithms for federated learning. Due to the fact that FL is vulnerable to various types of attacks and threats including, but not limited to, poisoning attacks such as Byzantine attacks, inference attacks such as backdoor attacks, and more [80]. However, security enhancements were later introduced to include various security aspects. For example, in [50], the authors proposed a secure vector summation strategy that uses a protocol with a fixed number of rounds, lower processing cost, high fault tolerance, and only a single server that can be trusted with a small amount of information. In this architecture, the server has a dual role: it must both transmit messages between the different participants and perform the necessary computations. The authors also offer two variants of their protocol; the first is more efficient and has a better chance of being secure in the simplest model against honest but curious adversaries. Nevertheless, the alternative has been shown to be secure in the random oracle paradigm and guarantees anonymity even when faced with active adversaries, such as a hostile server.

In addition, to make the aggregation process more resilient to poisoning local data or model parameters of participating devices, the authors of [51] proposed robust federated aggregation (RFA). The authors contributed to the aggregation step and presented a better aggregation technique for federated learning, since compromised devices can only affect the global model through updates. The proposed technique aggregates model updates without revealing the unique contribution of each device and is based on the geometric median, which can be easily estimated using a Weiszfeld-type algorithm [81]. The experiments conducted by the authors show that RFA can compete with traditional aggregation at a low level of corruption and has greater resilience at a high level of corruption.

In addition, the authors of [67] have developed a decentralized aggregation protocol for federated learning that protects user privacy, called SecureD-FL. The proposed approach to data aggregation is based on a refined form of the alternating direction multiplier

(ADMM) [82]. This communication pattern is inspired by combinatorial block design theory and is used by the proposed method to minimize privacy loss and ensure privacy from honest but curious adversaries in each aggregation round. To reduce the amount of personal information leaked during the aggregation process, the algorithm selects which subset of users (called a group) should have a conversation during each iteration.

In addition, the authors of [69] proposed SEAR, a secure aggregation algorithm that uses a hardware-based trusted execution environment instead of time-consuming cryptographic tools. For example, they used Intel SGX [83] trusted execution environment (TEE) to aggregate the locally trained models in a secure and trusted hardware environment. This is a secure area of the central processor where the confidentiality and integrity of the loaded code and data can be well protected.

Furthermore, the authors of [71] proposed efficient privacy-preserving data aggregation (EPPDA), which exploits the homomorphism of secret exchange [84]. In this context, secret sharing is able to protect the clients' secret data and thus reduce the influence of some malicious clients, which makes this algorithm a private, fault-tolerant algorithm. The cryptographic primitives can be summarized in the following steps: secret sharing, key exchange protocol, authenticated encryption, and the signature method.

Finally, in [73], the HeteroSAg aggregation algorithm protects the privacy of each user's local model updates by masking each user's model update such that the mutual information between the masked model and the unique model is zero. The efficiency of HeteroSAg and its robustness against Byzantine attacks lie in the FL system cycle, which executes a segment grouping strategy based on dividing edge users into groups and segmenting local model updates for those users. In summary, security has been studied and improved in FL aggregation algorithms, with several attempts in this area, as explained. Furthermore, the security mechanisms used in the implemented aggregation algorithms are summarized in Table 6.

Table 6. Security mechanisms followed in aggregation algorithms.

Ref.	Mechanism
[50]	Secure vector summing strategy
[51]	Using geometric median estimated using a Weiszfeld-type algorithm
[67]	Refined form of the alternating direction multiplier (ADMM)
[69]	Hardware-based trusted execution environment instead of complex cryptographic tools
[71]	Homomorphisms of the secret exchange
[73]	Masking each user's model update

4.1.5. Communication Cost

Federated machine learning, in its original version, offers a communication reduction approach where, instead of exchanging row data, which can sometimes be huge, it only exchanges model updates, which are typically smaller compared to the initial data. However, the training process in FL can take place in networks of enormous size, probably even around the world, as is the case with FedAvg, which was originally used to train Google keyboard services to improve text prediction. Apart from that, the network state and bandwidth can depend very much on the connection service providers, so one has to worry about the communication costs even if only model updates and no raw data are exchanged.

To this end, in [50], the authors proposed a technique based on the use of quantization, which involved reducing the amount of information exchanged between FL entities. Specifically, they used fixed-point quantization, in which data values are represented as fixed-point numbers with a finite number of bits, and achieved up to a 100-fold reduction in communication costs with their approach compared to the standard approach of secure aggregation without quantization. Similarly, in [60], the authors reduced communication costs by quantizing gradients on client devices before transmitting them to the server, and then aggregating the quantized gradients on the server in a "lazy manner," thereby reduc-

ing the size of the message exchanged and the communication costs. The same approach is taken by the HeterOSAg [73] algorithm.

In addition, SCAFFOLD [52] reduced communication costs by exchanging the control variate term [85] between server and clients instead of sending and receiving the entire model. This term was designed to reduce the variance of the stochastic gradient descent updates [86] and improve the convergence rate of the training process. In addition, the FedMA algorithm [56] has succeeded in reducing the communication cost through the matched averaging aggregation algorithms, where clients with similar distributions are aggregated together, speeding up the convergence, reducing the execution rounds, and reducing the overall communication cost even if the communication cannot be reduced in one round.

In contrast, in [57], the authors proposed to use an analog network coding technique to reduce the communication cost in federated learning over wireless networks. In this approach, the gradients are transmitted with a much lower communication bandwidth by encoding the gradients from multiple wireless devices into a single analog waveform that is transmitted over a wireless network using a technique called a physical layer network (PNC). Then, the received wave is decoded at the central server to recover the gradients from the different devices so that they can be aggregated to update the global model. In [59], the authors managed to reduce the communication cost by applying selective aggregation, where in each round some clients are selected based on their data distribution to perform the aggregation, reducing both the communication cost in a round and in the overall FL cycle.

Moreover, the aggregation algorithm SAFA [61] proposed the introduction of a semi-asynchronous protocol, where clients continue to train their local models while sending updates to the server. The key idea to reduce the communication cost is that instead of waiting for all clients to send their updates before aggregating them, the central server aggregates the clients' updates with a small delay to allow more updates to arrive, thus reducing the communication overhead and latency.

In addition, the authors of [64] proposed FedPSO, in which clients' local models are optimized using particle swarm optimization (PSO) and then only the optimized parameters are transmitted to the central server instead of transmitting the entire model. This lowers the communication cost by significantly reducing the amount of data transmitted between the central server and the clients. However, the algorithm proposed in [65], called LEGATO, reduces the communication cost by performing gradient aggregation on a per-layer basis instead of aggregating the gradient of the entire model. Finally, model compression has been used in several approaches to reduce communication costs, e.g., in FedBoost [54] and in [58]. In summary, the communication costs were increased in the respective aggregation algorithms. The mechanisms for reducing communication costs in the aggregation algorithms of FL are summarized in Table 7 below.

Table 7. Communication cost reduction mechanisms followed in aggregation algorithms.

Ref.	Mechanism
[50,60,73]	Quantization
[52]	Exchanging the control variate term
[56]	Matched averaging
[57]	Analog network coding technique
[59]	Selective aggregation
[61]	Semi-asynchronous protocol
[64]	Particle swarm optimization (PSO)
[65]	Gradient aggregation on a per-layer basis
[54,58]	Model compression

4.1.6. Computation Cost

Federated learning, as a collaborative artificial intelligence technology requires additional computational costs due to the additional communication, aggregation, and management processes performed throughout the FL cycle. However, this problem was addressed with the proposed aggregation algorithms that followed the FedAvg implementation. For example, in [58], the authors proposed the use of the gradient masking [87] technique to reduce computational costs. In this technique, each client encrypts its local gradient updates with a mask generated by the server, which in turn performs secure aggregation over the masked updates to train the global model. Applying aggregation over the masked gradients reduces the computational cost on the server side, yet there is still debate about the additive computational overhead required for masking and mask generation on both the client and server sides.

Moreover, in [61], the authors used a selective technique in developing their aggregation algorithm called SAFA. The server selects a subset of clients to share the model with for training, reducing the size of the data to be retrieved and aggregated. In [65], on the other hand, the reduction in computational overhead comes from the reduction in communication, where the amount of data exchanged between the server and clients is reduced, thus reducing the computational overhead. In addition, the personalization described in [66], in which clients do not receive a uniform model depending on their data distribution and characteristics, reduces the computational cost because each client receives a model that fits its data, so the server only needs to perform minimal executions to train the global model. Moreover, in [69], the authors used a sparse vector technique to compress the updates sent by the clients, which reduces the computational cost in the FL cycle. Furthermore, in [70], the authors reduced computational costs by using a circular multi-group aggregation structure to speed up the model aggregation process. In this approach, customer data are split into multiple groups, with each group assigned a unique aggregation order. Then, the groups are aggregated in a circular fashion so that each group is aggregated with a different subset of groups in each round, resulting in a visible reduction in the computational cost. Finally, the LightSecAgg [74] algorithm reduced the computational cost by reducing the dimensionality of the updates through random projections and hashing, while maintaining privacy as in traditional secure aggregation methods.

4.1.7. Fault Tolerance

The federated learning environment involves servers and clients collaborating in training a global model without sharing client data. However, the participating clients may lose connection to the network for various reasons. In this case, the training process of the global model may be affected, and the server may even wait indefinitely for them to reconnect, and even these stragglers may affect the accuracy of the global model. This case is defined as model staleness [88]. The longer the delay continues, the more outdated the model becomes, since the central server's model is not updated with the latest local models. To deal with this problem, the authors proposed in [61] a semi-asynchronous protocol that preserves the local training results. To this end, the authors used the futility percentage metric to measure the percentage of local progress wasted due to model synchronization forced by the server. Furthermore, the EPPDA aggregation algorithm was described by the authors of [71] as a fault-tolerant algorithm where aggregation continues regardless of how many clients abort the process. Finally, in [74], the authors proposed a new aggregation approach called LightSecAgg to overcome the bottleneck resulting from dropped users. They considered changing the design of their aggregation process from "random-seed reconstruction of the dropped users" to "one-shot aggregate mask reconstruction of active users via mask encoding/decoding". The proposed aggregation algorithms reflect a major advance in the way a FL global server handles the faulty clients and reduces their impact on the accuracy of the resulting global model.

4.1.8. Learning Quality

In a federated learning system, clients may have different data that can affect the quality of the globally trained model. Because clients may have different amounts and qualities of data, it can be a challenge to ensure that the contributions of each client are appropriately evaluated and that the resulting global model is of high quality based on the results aggregated from the local models. To ensure acceptable learning quality, the authors of [63] proposed a quality-aware aggregation scheme that weights each client's contribution based on the quality of its local data and the accuracy of its locally trained model. In other words, they created an index for the contribution of each client's local model, with those with higher indices contributing more to the global model, leading to a higher quality result. A similar approach was also proposed by the authors of [64,68].

4.1.9. Scalability

One of the biggest challenges in federated machine learning is its scalability. Unlike classical machine learning, which requires only one central server for training, FL can involve up to millions of devices in the training process. Therefore, developing a scalable aggregation algorithm that can handle an increasing number of clients is a major requirement in this area. In this context, LEAGTO [65] has been proposed as a scalable aggregation algorithm since it reduces the communication and computation costs. Similarly, TurboAggregate [70] is proposed as a scalable algorithm that can grow with a higher number of clients due to reduced computation and optimized code.

4.1.10. Personalization

Consequently, the aggregation algorithms of federated learning are meant for distribution and collaboration between different clients to train a global model. The diversity and heterogeneity of clients participating in the federated learning process make the development of a personalized aggregation algorithm urgent. However, personalization is a topic that can be considered from different perspectives, such as the following:

- Ability to handle heterogeneous data and hardware;
- Capability to adapt for the network settings such as bandwidth on the client's side;
- Other factors.

In this context, the aggregation algorithms FedYOGI [53], FedMA [56], FAIR [63], LEGATO [65], MHAT [66], and TurboAggregate [70] were proposed as personalized aggregation algorithms that managed to adapt to the particular circumstances on the client side. However, these algorithms were not originally intended for personalization, which is the only reason they were not classified as contributing to the personalization domain in Table 4. In contrast, the FedDist [62] aggregation algorithm was tested for personalization, which made it one of its targets.

4.2. Further Limitations

Federated learning aggregation algorithms are still at a very early stage. It has been only six years since the concept was introduced to FL with its FedAvg aggregation algorithm for averaging. However, tremendous efforts have been made to improve these algorithms. As mentioned earlier, FedAvg struggled with several obstacles such as slow convergence, difficult tuning, high communication and computational costs, dealing with client heterogeneity, scalability, and more [75,76]. These problems have been intensively studied, and the algorithms developed after FedAvg managed to solve several problems, as explained in the previous section. However, there are still some challenges and limitations in the area of aggregation algorithms, which will be discussed in this section.

4.2.1. Global Model Quality

There is agreement that larger amounts of training data can improve the accuracy of a learned model in both traditional machine learning and deep learning. On the other hand, in a distributed environment such as federated learning, the amount of data on

each client is not necessarily the same, and it may be insufficient for local training at a given time, which in turn reduces the accuracy of the local model and the global model accordingly. Traditionally, there are some solutions in ML that can be followed to improve the output of a smart model by improving the quality and quantity of data, such as resampling and standardization, which have successfully improved the accuracy of models in several examples such as [89,90]. However, these techniques are not guaranteed to improve the overall quality of the globally trained model, as preprocessing methods may vary depending on the heterogeneity of the client data, as some are able to handle certain missing data while others cannot. This may create an impetus to find more robust solutions to improve the overall quality of the global model.

4.2.2. Security Limitations

Although federated learning aims to create intelligent models that do not collect user data, it is still vulnerable to data leaks caused by attacks. This is possible due to the transfer of gradients and partial parameters, whether between clients and servers in the centralized architecture or between the clients themselves in the decentralized architecture. These parameters are attackable at three levels: at the inputs, at the learning process, and at the learned model. Typically, the attacks are carried out by attackers originating from malicious clients, and the types of attacks can be grouped as follows [80]:

- Poisoning attacks: these are conducted by injecting noise into the FL system, and are also split into two categories:
  - Data poisoning attacks: these are the most common attacks against ML models and can be either targeted toward a specific class or non-targeted. In a targeted attack, the noisy records of a specific target class are injected into local data so that the learned model will act badly on this class;
  - Model poisoning attacks: these are similar to data poisoning attacks, where the adversary tries to poison the local models instead of the local data.
- Inference attacks: in some scenarios, it is possible to infer, conclude, or restore the party local data from the model updates during the learning process;
- Backdoor attacks: secure averaging allows parties to be anonymous during the model update process. Using the same functionality, a party or group of parties can introduce backdoor functionality in in FL global model. Then, a malicious entity can use the backdoor to mislabel certain tasks such as choosing a specific label for a data instance with specific characteristics. For sure, the proportion of the compromised devices and FL model capacity affects in the intensity of such attacks.

Despite the fact that the developed aggregation algorithms have found robust solutions to poisoning attacks such as the Byzantine attack [91], inference and backdoor attacks are still observed in this area, which requires further development and research. In addition, some techniques and methods in the aggregation algorithm domain are still unknown, such as the polymorphic encryption,

"PE", which has proven to be a viable technology for exchanging encrypted data with high confidence in privacy, as explained in [92].

4.2.3. Evaluation Complexity and Lack of Standards

In classical machine learning and deep learning processes, models are usually evaluated using specific and defined metrics such as accuracy, precision, recall, specificity, negativity, and others. In contrast, evaluating a federated learning system requires parameters that may include privacy level, communication cost, and robustness to attacks. In addition, there are as yet no uniform standards that can be referenced to measure the feasibility of an FL system.

4.2.4. Software and Hardware Heterogeneity

The differences in hardware and software used by individual clients present a significant obstacle to algorithms for aggregating learning across the FL system. Clients can vary widely in terms of their data availability, feature representation, computing power, and network connectivity. For example, poor generalization performance can result from overfitting the local model due to imbalanced data distribution. In addition, the convergence of the model and the performance of the global model may be affected by the different feature representations of the clients, which may lead to inconsistencies or incompatibilities in the feature sets. Differences in the processing power of client devices can also lead to performance inconsistencies during training, with some devices taking more time to complete operations or being unable to run the model at all. Overall, heterogeneity can have a detrimental effect on the efficiency and precision of the overall trained model.

4.2.5. User Adoption

Furthermore, one of the biggest obstacles to integrating federated machine learning into real-world implementations is user acceptance, adoption, and participation. Although FL is known as a privacy-preserving technology, FL is still new and not yet adopted by users due to privacy concerns, discomfort, ethical concerns, and other contextual factors.

So far, Figure 7 below illustrates a summary of the main limitations in the federated machine learning aggregation algorithms.

Figure 7. Federated learning aggregation algorithms limitations.

4.3. Future Perspectives

In addition to what has been achieved with the available aggregation algorithms, further efforts can be directed to improve some features, such as further reducing communication and computational costs, improving scalability, and others. In addition, less studied areas such as scalability, learning quality, and personalization should also be considered in future studies to develop more efficient and accurate aggregation algorithms for federated learning. Moreover, with regard to the limitations in the field of aggregation algorithms for federated learning mentioned in the previous section, there are a number of future prospects that can help improve the field and increase its efficiency.

4.3.1. Boost Learning Quality

Confidence in machine learning and its application in daily life is achieved through various aspects, including high accuracy, explainability, feasibility, and others. However,

accuracy is always a major concern in this field. Therefore, there is a great need to improve the "learning quality" of federated learning aggregation algorithms in order to improve the feasibility and usability, thus increasing the acceptance of the technology in daily life. Accepted learning quality for global models can be achieved by improving the quality of locally trained models by handling heterogeneity, improving generalization, preprocessing customer data, and other steps. Improving the locally trained models can help improve the quality of the global model.

4.3.2. Improving Security and Privacy

Federated learning was originally introduced as a privacy-preserving technology to collaboratively train smart models without sacrificing the privacy and confidentiality of user data. Accordingly, the ability to bypass privacy controls in the FL system undermines the foundation on which this field is built and causes it to lose value. Therefore, it is necessary to strengthen the robustness of aggregation algorithms against attacks, especially inference and backdoor attacks, to prevent malicious entities from reflecting exchanged messages, whether in the form of model updates, gradients, or parameters, in an effort to uncover the data used for local training. In this regard, several technologies can be explored, such as polymorphic encryption [92] and quantum-resistant cryptography [93].

4.3.3. Proposing Standards and Norms

Machine learning norms and standards are very useful for evaluating an intelligent model. In the classical version of ML, accuracy, precision, and recall, among other parameters, are important measures used to evaluate a model. However, when it comes to federated learning and aggregation algorithms, these parameters are not enough because privacy, communication and computational costs, scalability, generalization, and other parameters are also important in evaluating these algorithms. Therefore, it is necessary to propose methods to unify these standards so that future aggregation algorithms can be evaluated based on these specifications.

4.3.4. Enhance Heterogeneity Handling Abilities

The benefits of federated learning techniques extend beyond privacy preservation to several important goals, such as overcoming the data islanding dilemma. However, as resource divergence increases, so does the likelihood of heterogeneity. Therefore, it is necessary to improve the ability to handle heterogeneity in aggregation algorithms. Various techniques can be considered for this purpose, such as the following:

- Resource Allocation [94]: This involves the optimal distribution of computational load and communication bandwidth among clients, taking into account their capabilities and limitations. This can reduce the impact of heterogeneity, minimize training time, and improve the convergence and accuracy of the model;
- Data Clustering: Implemented by grouping clients into clusters based on the similarity of their data distribution or other criteria, this allows the system to leverage the similarities between devices and reduce the impact of heterogeneity;
- Meta-Learning [95]: This involves determining the optimal learning algorithm or hyperparameters for each client based on its past performance or other metadata. This helps to adapt to client heterogeneity and also improves the overall performance and scalability of the federated learning process.

4.3.5. Boost Technology Adoption into Real-Word Scenarios

Federated machine learning is increasingly being studied and is also trending in the scientific research community and among researchers, yet it is not widely used in the real world as it is in research. This may seem normal, especially because it is still in its early stages; however, there are also many opportunities for it to be embedded more and more in real-world scenarios. Smart wearables, for example, have proven to be extremely viable in a number of areas different domains, such as health for example [12,96], and the ability

to embed federated learning into these devices is likely to revolutionize their efficiency by providing access to more data while preserving user privacy. Embedding FL in smart wearables has been extensively studied in [28].

4.3.6. Integrate Different Areas of Contribution

The various aggregation algorithms presented in this study have contributed to the concept of collaboration in training a global model in several areas. Whether it is the aggregation itself, increasing the speed of convergence, reducing computational and communication costs, or even other areas, much has been achieved. However, almost all of the algorithms mentioned have contributed to one or two areas as detailed in Table 4, and only SAFA [61] and LEGATO [65] have contributed to four areas, with the former focusing on aggregation, lowering communication and computation costs, and fault tolerance, and the latter focusing on lowering communication and computation costs, security, and learning quality. Therefore, there is a need to work on aggregation algorithms that integrate more and more domains, such as security, learning quality, scalability, personalization and, of course, aggregation with reduction of communication and computational costs. The ability to integrate these areas into one algorithm will certainly be a big step in this area.

4.3.7. Embedding Latest Technologies into FL: Quantum Computing as an Example

Quantum computing is a new type of computing that takes advantage of the principles of quantum mechanics to perform certain calculations much more efficiently than classical computers. Embedding quantum computing into federated learning will help advance this field from multiple perspectives [97–99]:

- Speeding Up Computation: Quantum computers are capable of solving certain tasks much faster than traditional computers, such as factoring large numbers or scanning unsorted databases. Quantum computers could potentially help speed up the training of machine learning models in the context of federated learning, especially for complicated tasks or large datasets. This could improve the efficiency and feasibility of federated learning for real-world applications;
- Quantum Communication: Quantum communication technologies, such as quantum teleportation and quantum key distribution, could be used to securely transfer model changes between nodes of the federated learning system. This could improve the privacy and security of federated learning, which is one of its main advantages;
- Quantum Encryption: Quantum encryption technology, such as quantum key distribution, could be used to improve the security of communications between nodes of the federated learning system. This could be particularly useful in federated learning environments where privacy and security are critical;
- Improved Optimization: Some optimization problems, such as training machine learning models, can be solved more effectively by using quantum technologies. As a result, federated learning algorithms can become more efficient and effective.

Finally, the limitations known in the field of federated learning aggregation algorithms and the possible future recommendations to solve these problems are presented in Figure 8 below.

Aggregation algorithms for federated learning are a topic that is attracting more and more attention nowadays. Recently proposed algorithms have succeeded in reducing convergence, communication, and computation costs on the one hand, and handling heterogeneous data on the other. Moreover, security and fault tolerance have been strongly emphasized by researchers in this area, while learning quality, scalability, and personalization issues have been less considered. Therefore, federated learning aggregation algorithms are still vulnerable to various challenges such as learning quality of the global model, security limitations and vulnerability to inference and backdoor attacks, evaluation complexity, lack of norms and standards, and other issues as described previously. However, these problems can be addressed with different concepts and notions, such as embedding security techniques, polymorphic encryption as an example, or using emerging technologies such

as quantum computing or other solutions. All challenges and future perspectives of FL aggregation algorithms have been described in detail in this section.

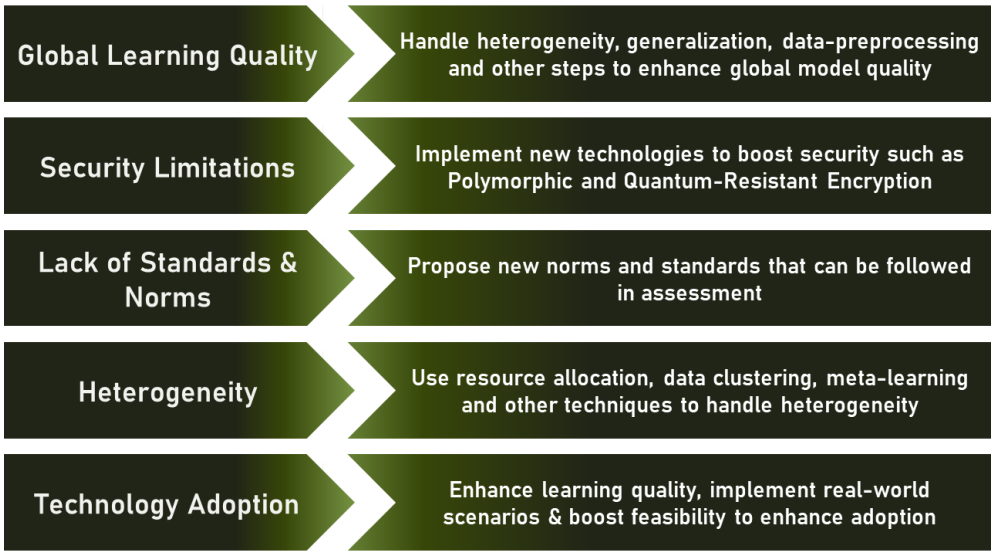

Figure 8. Federated learning aggregation algorithms limitations and solutions.

5. Conclusions

Federated ML and associated aggregation algorithms are emerging as a practical, privacy-preserving ML technology that will improve the effectiveness of smart models and facilitate their integration into people's daily routines. The exchange of models or their parameters between server and clients, rather than the exchange of raw data, makes these technologies feasible. An essential part of the federated learning cycle is the aggregation algorithm, i.e., the method by which the clients' knowledge is integrated and the global model is updated accordingly. Many aggregation methods have been developed and published, each using its own method of data integration. Each aggregation algorithm adds something new to the body of knowledge. The rapid development of aggregation algorithms in their short history is a sign of the great interest in this topic. Nevertheless, such algorithms have several serious drawbacks, including vulnerability to heterogeneity, inference, and backdoor attacks. These problems motivate further studies in this area. This article summarizes the state of the art in aggregation algorithms, analyzes their properties and shortcomings, and suggests numerous perspectives for further investigation.

Author Contributions: Conceptualization: M.M. and M.A.; formal analysis: M.M.; investigation: M.M.; methodology: M.M. and M.A.; supervision: M.A., A.B., H.I. and A.R.; visualization: M.M.; writing—original draft: M.M.; writing—review and editing: M.A., A.B., H.I. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), grant number 06351.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We acknowledge the support of Centre d'Entrepreneuriat et de Valorisation des Innovations (CEVI).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stearns, P.N. *The Industrial Revolution in World History*; Routledge: Oxfordshire, UK, 2020.
2. Campbell-Kelly, M. *Computer, Student Economy Edition: A History of the Information Machine*; Routledge: Oxfordshire, UK, 2018.
3. Moor, J. The Dartmouth College artificial intelligence conference: The next fifty years. *AI Mag.* **2006**, *27*, 87.
4. Frankish, K.; Ramsey, W.M. (Eds.) *The Cambridge Handbook of Artificial Intelligence*; Cambridge University Press: Cambridge, UK, 2014.
5. Aggarwal, K.; Mijwil, M.M.; Al-Mistarehi, A.H.; Alomari, S.; Gök, M.; Alaabdin, A.M.Z.; Abdulrhman, S.H. Has the future started? The current growth of artificial intelligence, Machine Learning, and deep learning. *Iraqi J. Comput. Sci. Math.* **2022**, *3*, 115–123.
6. Bell, J. What Is Machine Learning? In *Machine Learning and the City: Applications in Architecture and Urban Design*; John Wiley & Sons: Hoboken, NJ, USA, 2022; pp. 207–216.
7. Hardt, M.; Benjamin, R. Patterns, predictions, and actions: A story about Machine Learning. *arXiv* **2021**, arXiv:2102.05242.
8. Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef] [PubMed]
9. Sharma, N.; Sharma, R.; Jindal, N. Machine learning and deep learning applications-a vision. *Glob. Transit. Proc.* **2021**, *2*, 24–28. [CrossRef]
10. Pallathadka, H.; Mustafa, M.; Sanchez, D.T.; Sajja, G.S.; Gour, S.; Naved, M. Impact of machine learning on management, healthcare and agriculture. *Mater. Today Proc.* **2023**, *80*, 2803–2806. [CrossRef]
11. Ghazal, T.M.; Hasan, M.K.; Alshurideh, M.T.; Alzoubi, H.M.; Ahmad, M.; Akbar, S.S.; Al Kurdi, B.; Akour, I.A. IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet* **2021**, *13*, 218. [CrossRef]
12. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review. *Sensors* **2023**, *23*, 828. [CrossRef]
13. Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of Machine Learning and IoT in smart transportation. *Future Internet* **2019**, *11*, 94. [CrossRef]
14. Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine learning and deep learning methods for cybersecurity. *IEEE Access* **2018**, *6*, 35365–35381. [CrossRef]
15. Nagarhalli, T.P.; Vaze, V.; Rana, N.K. Impact of machine learning in natural language processing: A review. In Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 4–6 February 2021; pp. 1529–1534
16. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]
17. Larrañaga, P.; Atienza, D.; Diaz-Rozo, J.; Ogbechie, A.; Puerto-Santana, C.; Bielza, C. *Industrial Applications of Machine Learning*; CRC Press: Boca Raton, FL, USA, 2018.
18. Verbraeken, J.; Wolting, M.; Katzy, J.; Kloppenburg, J.; Verbelen, T.; Rellermeyer, J.S. A survey on distributed Machine Learning. *ACM Comput. Surv.* **2020**, *53*, 1–33. [CrossRef]
19. Panayiotou, T.; Savvas, G.; Tomkos, I.; Ellinas, G. Centralized and distributed Machine Learning-based QoT estimation for sliceable optical networks. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–7.
20. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities andchallenges. *Neurocomputing* **2017**, *237*, 350–361. [CrossRef]
21. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45. [CrossRef]
22. Injadat, M.; Moubayed, A.; Nassif, A.B.; Shami, A. Machine learning towards intelligent systems:applications, challenges, and opportunities. *Artif. Intell. Rev.* **2021**, *54*, 3299–3348. [CrossRef]
23. Char, D.S.; Shah, N.H.; Magnus, D. Implementing Machine Learning in health care—Addressingethical challenges. *N. Engl. J. Med.* **2018**, *378*, 981. [CrossRef]
24. Albrecht, J.P. How the GDPR will change the world. *Eur. Data Prot. Law Rev.* **2016**, *2*, 287. [CrossRef]
25. Parasol, M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and onChina's big data and Smart City dreams. *Comput. Law Secur. Rev.* **2018**, *34*, 67–98. [CrossRef]
26. IBM. Security Cost of Data Breach Report. 2021. Available online: https://www.ibm.com/downloads/cas/ojdvqgry (accessed on 1 March 2023).
27. Konečný, J.; McMahan, H.B.; Yu, F.X.; Richtárik, P.; Suresh, A.T.; Bacon, D. Federated learning: Strategies for improving communication efficiency. *arXiv* **2016**, arXiv:1610.05492.
28. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Reviewing Federated Machine Learning and Its Use in Diseases Prediction. *Sensors* **2023**, *23*, 2112. [CrossRef]
29. Malekijoo, A.; Fadaeieslam, M.J.; Malekijou, H.; Homayounfar, M.; Alizadeh-Shabdiz, F.; Rawassizadeh, R. Fedzip: A compression framework for communication-efficient federated learning. *arXiv* **2021**, arXiv:2102.01593.
30. Liu, Z.; Guo, J.; Yang, W.; Fan, J.; Lam, K.-Y.; Zhao, J. Privacy-Preserving Aggregation in Federated Learning: A Survey. *IEEE Trans. Big Data* **2022**. [CrossRef]

31. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated Machine Learning: Concept and applications. *ACM Trans. Intell. Syst. Technol. (TIST)* **2019**, *10*, 1–19. [CrossRef]
32. Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li, W.; Gao, Y. A survey on Federated Learning. *Knowl.-Based Syst.* **2021**, *216*, 106775. [CrossRef]
33. Li, Q.; Wen, Z.; Wu, Z.; Hu, S.; Wang, N.; Li, Y.; Liu, X.; He, B. A survey on Federated Learning systems: Vision, hype and reality for data privacy and protection. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 3347–3366. [CrossRef]
34. Jiang, J.; Hu, L. Decentralised Federated Learning with adaptive partial gradient aggregation. *CAAI Trans. Intell. Technol.* **2020**, *5*, 230–236. [CrossRef]
35. Yao, X.; Huang, T.; Zhang, R.X.; Li, R.; Sun, L. Federated learning with unbiased gradient aggregation and controllable meta updating. *arXiv* **2019**, arXiv:1910.08234.
36. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* **2018**, *4*, e00938. [CrossRef]
37. Cheng, B.; Titterington, D.M. Neural networks: A review from a statistical perspective. *Stat. Sci.* **1994**, *9*, 2–30.
38. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics PMLR, Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.
39. Wang, T.; Zheng, Z.; Lin, F. Federated Learning Framew Ork Based on Trimmed Mean Aggregation Rules. *SSRN Electron. J.* **2022**. [CrossRef]
40. Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H.B.; Patel, S.; Ramage, D.; Segal, A.; Seth, K. Practical secure aggregation for Federated Learning on user-held data. *arXiv* **2016**, arXiv:1611.04482.
41. Wei, K.; Li, J.; Ding, M.; Ma, C.; Yang, H.H.; Farokhi, F.; Jin, S.; Tony, Q.S.Q.; Poor, H.V. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3454–3469. [CrossRef]
42. Xu, J.; Wang, S.; Wang, L.; Yao, A.C.C. Fedcm: Federated learning with client-level momentum. *arXiv* **2021**, arXiv:2106.10874.
43. Reyes, J.; Di Jorio, L.; Low-Kam, C.; Kersten-Oertel, M. Precision-weighted Federated Learning. *arXiv* **2021**, arXiv:2107.09627.
44. West, M. Bayesian aggregation. *J. R. Stat. Soc. Ser. A* **1984**, *147*, 600–607. [CrossRef]
45. Kerkouche, R.; Ács, G.; Castelluccia, C. Federated learning in adversarial settings. *arXiv* **2020**, arXiv:2010.07808.
46. Chen, S.; Shen, C.; Zhang, L.; Tang, Y. Dynamic aggregation for heterogeneous quantization in Federated Learning. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 6804–6819. [CrossRef]
47. Liu, L.; Zhang, J.; Song, S.; Letaief, K.B. Hierarchical quantized Federated Learning: Convergence analysis and system design. *arXiv* **2021**, arXiv:2103.14272.
48. Ma, X.; Zhang, J.; Guo, S.; Xu, W. Layer-wised model aggregation for personalized Federated Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10092–10101.
49. Chen, H.Y.; Chao, W.L. Fedbe: Making bayesian model ensemble applicable to Federated Learning. *arXiv* **2020**, arXiv:2009.01974.
50. Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H.B.; Patel, S.; Ramage, D.; Segal, A.; Seth, K. Practical secure aggregation for privacy-preserving Machine Learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017; pp. 1175–1191.
51. Pillutla, K.; Kakade, S.M.; Harchaoui, Z. Robust aggregation for Federated Learning. *IEEE Trans. Signal Process.* **2022**, *70*, 1142–1154. [CrossRef]
52. Karimireddy, S.P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; Suresh, A.T. Scaffold: Stochastic controlled averaging for Federated Learning. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 5132–5143.
53. Reddi, S.; Charles, Z.; Zaheer, M.; Garrett, Z.; Rush, K.; Konečný, J.; Kumar, S.; McMahan, H.B. Adaptive federated optimization. *arXiv* **2020**, arXiv:2003.00295.
54. Hamer, J.; Mohri, M.; Suresh, A.T. Fedboost: A communication-efficient algorithm for Federated Learning. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 3973–3983.
55. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated optimization in heterogeneous networks. *Proc. Mach. Learn. Syst.* **2020**, *2*, 429–450.
56. Wang, H.; Yurochkin, M.; Sun, Y.; Papailiopoulos, D.; Khazaeni, Y. Federated learning with matched averaging. *arXiv* **2020**, arXiv:2002.06440.
57. Guo, H.; Liu, A.; Lau, V.K. Analog gradient aggregation for Federated Learning over wireless networks: Customized design and convergence analysis. *IEEE Internet Things J.* **2020**, *8*, 197–210. [CrossRef]
58. Choi, B.; Sohn, J.Y.; Han, D.J.; Moon, J. Communication-computation efficient secure aggregation for Federated Learning. *arXiv* **2020**, arXiv:2012.05433.
59. Ye, D.; Yu, R.; Pan, M.; Han, Z. Federated learning in vehicular edge computing: A selective model aggregation approach. *IEEE Access* **2020**, *8*, 23920–23935. [CrossRef]
60. Sun, J.; Chen, T.; Giannakis, G.B.; Yang, Q.; Yang, Z. Lazily aggregated quantized gradient innovation for communication-efficient Federated Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2031–2044. [CrossRef]
61. Wu, W.; He, L.; Lin, W.; Mao, R.; Maple, C.; Jarvis, S. SAFA: A semi-asynchronous protocol for fast Federated Learning with low overhead. *IEEE Trans. Comput.* **2020**, *70*, 655–668. [CrossRef]

62. Sannara, E.K.; Portet, F.; Lalanda, P.; German, V.E.G.A. A Federated Learning aggregation algorithm for pervasive computing: Evaluation and comparison. In Proceedings of the 2021 IEEE International Conference on Pervasive Computing and Communications (PerCom), Kassel, Germany, 22–26 March 2021; pp. 1–10.
63. Deng, Y.; Lyu, F.; Ren, J.; Chen, Y.C.; Yang, P.; Zhou, Y.; Zhang, Y. Fair: Quality-aware Federated Learning with precise user incentive and model aggregation. In Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications, Vancouver, BC, Canada, 10–13 May 2021; pp. 1–10.
64. Park, S.; Suh, Y.; Lee, J. Fedpso: Federated Learning using particle swarm optimization to reduce communication costs. *Sensors* **2021**, *21*, 600. [CrossRef]
65. Varma, K.; Zhou, Y.; Baracaldo, N.; Anwar, A. Legato: A layerwise gradient aggregation algorithm for mitigating byzantine attacks in Federated Learning. In Proceedings of the 2021 IEEE 14th International Conference on Cloud Computing (CLOUD), Chicago, IL, USA, 5–11 September 2021; pp. 272–277.
66. Hu, L.; Yan, H.; Li, L.; Pan, Z.; Liu, X.; Zhang, Z. MHAT: An efficient model-heterogenous aggregation training scheme for Federated Learning. *Inf. Sci.* **2021**, *560*, 493–503. [CrossRef]
67. Jeon, B.; Ferdous, S.M.; Rahman, M.R.; Walid, A. Privacy-preserving decentralized aggregation for Federated Learning. In Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Virtual, 9–12 May 2021; pp. 1–6.
68. Wang, Y.; Kantarci, B. Reputation-enabled Federated Learning model aggregation in mobile platforms. In Proceedings of the ICC 2021-IEEE International Conference on Communications, Xiamen, China, 28–30 July 2021; pp. 1–6.
69. Zhao, L.; Jiang, J.; Feng, B.; Wang, Q.; Shen, C.; Li, Q. Sear: Secure and efficient aggregation for byzantine-robust Federated Learning. *IEEE Trans. Dependable Secur. Comput.* **2021**, *19*, 3329–3342. [CrossRef]
70. So, J.; Güler, B.; Avestimehr, A.S. Turbo-aggregate: Breaking the quadratic aggregation barrier in secure Federated Learning. *IEEE J. Sel. Areas Inf. Theory* **2021**, *2*, 479–489. [CrossRef]
71. Avizienis, A.; Laprie, J.-C.; Randell, B.; Landwehr, C. Basic concepts and taxonomy of dependable and secure computing. *IEEE Trans. Dependable Secur. Comput.* **2014**, *1*, 11–33. [CrossRef]
72. Nguyen, J.; Malik, K.; Zhan, H.; Yousefpour, A.; Rabbat, M.; Malek, M.; Huba, D. Federated learning with buffered asynchronous aggregation. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Virtual, 28–30 March 2022; pp. 3581–3607.
73. Elkordy, A.R.; Avestimehr, A.S. HeteroSAg: Secure aggregation with heterogeneous quantization in Federated Learning. *IEEE Trans. Commun.* **2022**, *70*, 2372–2386. [CrossRef]
74. So, J.; He, C.; Yang, C.S.; Li, S.; Yu, Q.; E Ali, R.; Guler, B.; Avestimehrm S. Lightsecagg: A lightweight and versatile design for secure aggregation in Federated Learning. *Proc. Mach. Learn. Syst.* **2022**, *4*, 694–720.
75. Li, T.; Sahu, A.K.; Talwalkar, A.; Smith, V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **2020**, *37*, 50–60. [CrossRef]
76. Rahman, K.J.; Ahmed, F.; Akhter, N.; Hasan, M.; Amin, R.; Aziz, K.E.; Islam, A.K.M.M.; Mukta, M.S.H.; Islam, A.K.M.N. Challenges, applications and design aspects of Federated Learning: A survey. *IEEE Access* **2021**, *9*, 124682–124700. [CrossRef]
77. Lynch, J.F. Analysis and application of adaptive sampling. In Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Dallas, TX, USA, 15–17 May 2000; pp. 260–267.
78. He, B.; Huang, J.X.; Zhou, X. Modeling term proximity for probabilistic information retrieval models. *Inf. Sci.* **2011**, *181*, 3017–3031. [CrossRef]
79. Zhu, H.; Xu, J.; Liu, S.; Jin, Y. Federated learning on non-IID data: A survey. *Neurocomputing* **2021**, *465*, 371–390. [CrossRef]
80. Lyu, L.; Yu, H.; Yang, Q. Threats to Federated Learning: A survey. *arXiv* **2020**, arXiv:2003.02133.
81. Weiszfeld, E.; Plastria, F. On the point for which the sum of the distances to n given points is minimum. *Ann. Oper. Res.* **2009**, *167*, 7–41. [CrossRef]
82. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends® Mach. Learn.* **2011**, *3*, 1–122.
83. Chakrabarti, S.; Knauth, T.; Kuvaiskii, D.; Steiner, M.; Vij, M. Trusted execution environment with intel sgx. In *Responsible Genomic Data Sharing*; Academic Press: Cambridge, MA, USA, 2020; pp. 161–190.
84. Benaloh, J.C. Secret sharing homomorphisms: Keeping shares of a secret secret. In *Advances in Cryptology—CRYPTO'86: Proceedings*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 251–260.
85. Müller, T.; Rousselle, F.; Keller, A.; Novák, J. Neural control variates. *Acm Trans. Graph. (TOG)* **2020**, *39*, 1–19. [CrossRef]
86. Ketkar, N.; Ketkar, N. Stochastic gradient descent. In *Deep Learning with Python: A Hands-On Introduction*; Apress: Berkeley, CA, USA, 2017; pp. 113–132.
87. Boenisch, F.; Sperl, P.; Böttinger, K. Gradient masking and the underestimated robustness threats of differential privacy in deep learning. *arXiv* **2021**, arXiv:2105.07985.
88. Dai, W.; Zhou, Y.; Dong, N.; Zhang, H.; Xing, E.P. Toward understanding the impact of staleness in distributed Machine Learning. *arXiv* **2018**, arXiv:1810.03264.
89. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability. *Procedia Comput. Sci.* **2022**, *203*, 231–238. [CrossRef]

90. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Predicting Cardiovascular Events with Machine Learning Models and Heart Rate Variability. *Int. J. Ubiquitous Syst. Pervasive Netw. (JUSPN)* **2023**, *18*, 49–59.
91. Lamport, L.; Shostak, R.; Pease, M. The Byzantine generals problem. In *Concurrency: The Works of Leslie Lamport*; Transactions on Programming Languages and Systems; ACM: New York, NY, USA, 2019; pp. 203–226
92. Booher, D.D.; Cambou, B.; Carlson, A.H.; Philabaum, C. Dynamic key generation for polymorphic encryption. In Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 7–9 January 2019; pp. 482–487.
93. Vella, H. The Race for Quantum-Resistant Cryptography [quantum-cyber security]. *Eng. Technol.* **2022**, *17*, 56–59. [CrossRef]
94. Jamil, B.; Ijaz, H.; Shojafar, M.; Munir, K.; Buyya, R. Resource allocation and task scheduling in fog computing and internet of everything environments: A taxonomy, review, and future directions. *ACM Comput. Surv. (CSUR)* **2022**, *54*, 1–38. [CrossRef]
95. Feng, Y.; Chen, J.; Xie, J.; Zhang, T.; Lv, H.; Pan, T. Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects. *Knowl.-Based Syst.* **2022**, *235*, 107646. [CrossRef]
96. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review. *Sensors* **2022**, *22*, 7472. [CrossRef] [PubMed]
97. Yin, H.L.; Fu, Y.; Li, C.L.; Weng, C.X.; Li, B.H.; Gu, J.; Lu, Y.-S.; Huang, S.; Chen, Z.-B. Experimental quantum secure network with digital signatures and encryption. *arXiv* **2021**, arXiv:2107.14089.
98. Xie, Y.M.; Lu, Y.S.; Weng, C.X.; Cao, X.Y.; Jia, Z.Y.; Bao, Y.; Wang, Y.; Fu, Y.; Yin, H.-L.; Chen, Z.-B. Breaking the rate-loss bound of quantum key distribution with asynchronous two-photon interference. *PRX Quantum* **2022**, *3*, 020315. [CrossRef]
99. Liu, Z.P.; Zhou, M.G.; Liu, W.B.; Li, C.L.; Gu, J.; Yin, H.L.; Chen, Z.B. Automated machine learning for secure key rate in discrete-modulated continuous-variable quantum key distribution. *Opt. Express* **2022**, *30*, 15024–15036. [CrossRef] [PubMed]

CHAPTER 4

Securing Federated Learning; Approaches, Mechanisms and Opportunities

This article is undergoing peer review process in the Journal of the ACM.

Résumé: Malgré les progrès réalisés par le FL dans la préservation de la vie privée des utilisateurs, il n'en est encore qu'à ses débuts et est vulnérable à de nombreuses difficultés et obstacles. Dans ce chapitre, l'accent est mis sur les problèmes de sécurité et de confidentialité liés aux algorithmes d'agrégation dans l'apprentissage fédéré, en commençant par une vue d'ensemble des menaces et attaques potentielles. Il procède à une analyse complète des algorithmes d'agrégation soucieux de la sécurité et examine leurs vulnérabilités, en appuyant la principale préoccupation de la recherche sur des preuves empiriques.

Securing Federated Learning; Approaches, Mechanisms and Opportunities

MOHAMMAD MOSHAWRAB*, Université du Québec à Rimouski, Canada
MEHDI ADDA, Université du Québec à Rimouski, Canada
ABDENOUR BOUZOUANE, Université du Québec à Chicoutimi, Canada
HUSSEIN IBRAHIM, Institut Technologique de Maintenance Industrielle, Canada
ALI RAAD, Islamic University of Lebanon, Lebanon

With the ability to analyze data, artificial intelligence technology and its offshoots have made difficult tasks easier. The tools of these technologies are now used in almost every aspect of life. For example, Machine Learning (ML), an offshoot of artificial intelligence, has become the focus of interest for researchers in industry, education, healthcare and other disciplines and has proven to be as efficient as, and in some cases better than, experts in answering various problems. However, the obstacles to ML 's progress are still being explored, and Federated Learning (FL) has been presented as a solution to the problems of privacy and confidentiality. In the FL approach, users do not disclose their data throughout the learning process, which improves privacy and security. In this article, we look at the security and privacy concepts of FL and the threats and attacks it faces. We also address the security measures used in FL aggregation procedures. In addition, we examine and discuss the use of homomorphic encryption to protect FL data exchange, as well as other security strategies. Finally, we discuss security and privacy concepts in FL and what additional improvements could be made in this context to increase the efficiency of FL algorithms.

CCS Concepts: • **General and reference** → **Surveys and overviews**.

Additional Key Words and Phrases: Federated Learning, Security, Privacy, Aggregation Algorithms, Homomorphic Encryption, Securing Mechanisms, Threats, Attacks

ACM Reference Format:
Mohammad Moshawrab, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. 2024. Securing Federated Learning; Approaches, Mechanisms and Opportunities. 1, 1 (July 2024), 32 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Machine Learning (ML), considered an offshoot of Artificial Intelligence (AI), allows computers to "self-learn" from training data. This allows them to gain information over time without having to be explicitly programmed. By recognising patterns in data and learning from them, ML algorithms can develop their own predictions. In short, ML algorithms and models gain knowledge through experience. Later reviews show the various domains in which ML has demonstrated its efficiency and usability, such as healthcare [1–3], smart cities [4, 5], industry[6], Internet of Things (IoT) [7, 8],

Authors' addresses: Mohammad Moshawrab, mohammad.moshawrab@uqar.ca, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC, Canada, G5L 3A1; Mehdi Adda, mehdi_adda@uqar.ca, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC, Canada, G5L 3A1; Abdenour Bouzouane, abdenour_bouzouane@uqac.ca, Université du Québec à Chicoutimi, 555 Boulevard de l'Université, Chicoutimi, QC, Canada, G7H 2B1; Hussein Ibrahim, hussein.ibrahim@itmi.ca, Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, QC, Canada, G4R 5B7; Ali Raad, ali.raad@iul.edu.lb, Islamic University of Lebanon, Wardaniyeh, Lebanon, P.O. Box: 30014.

Manuscript submitted to ACM

e-commerce [7, 8], Natural Language Processing (NLP) [9], and others [10–12]. However, there are still a number of difficulties that must be overcome in order to advance Machine Learning, many of which are described and discussed in detail in the literature. These difficulties can be divided into several groups, including:

- Data Related Challenges [15, 16]: encompassing problems at the level of data collection stage including:
  - Data Availability/Accessibility: data may not be available or accessible;
  - Data Locality (Data Islands): data is scattered into different and non-related entities;
  - Data Readiness: data may be heterogeneous, noisy or imbalanced;
  - Data Volume: data volume causes different challenges such as 'Curse of Dimensionality';
  - Feature representation and selection: selecting the optimal features.
- Models Related Challenges [17–19]: which are defined as challenges faced at the level of ML model building, including:
  - Accuracy and Performance: increasing accuracy of models;
  - Models Evaluation: correctly evaluate the models? performance;
  - Variance and Bias: that affects the results and trust;
  - Explainability: resolving the back-box identity of ML models;
  - Model Selection: choose the best model that fits for the problem being studied.
- Implementation Related Challenges[17–19]: such challenges are faced at the level of implementing ML models in real-life applications, including:
  - Real-Time Processing: adapting models to act on real-time basis;
  - Execution time and complexity: ML models may need high computation powers.
- General Challenges [15, 18, 20]: other issues such as:
  - Users? Data Privacy and Confidentiality: data are protected with many regulations;
  - User Technology Adoption and Engagement;
  - Ethical Constraints.

The challenges of ML have been intensively studied. Since the workflow of ML mainly consists of data management, model training, model review, and model deployment, the data in ML play a central role. Because the performance of ML 's models is highly dependent on the availability of data, the collection of real-world data is the most challenging aspect of ML model development for several reasons, particularly with respect to privacy and confidentiality. Not only individuals, but also society, governments, and organizations are strengthening privacy and data security protections, for which various regulations have been enacted, such as the European Union's General Data Protection Regulation (GDPR) [19], the Chinese People's Republic of China's Cybersecurity Law[20], the General Principles of Civil Law of the People's Republic of China [21], the PDPA in Singapore [22], and others. Although these regulations facilitate the protection of personal data, they pose new difficulties for ML because it is now more difficult to collect data for model training, which makes it more difficult to improve the performance accuracy and personalization of these models. For this reason, maintaining data privacy and confidentiality is not a sole obstacle for ML, but simultaneously raises issues of data availability, performance, personalization, and thus acceptance and trust.

1.1 Federated Learning: A Privacy Preserving Technology

In an attempt to protect user privacy, Google has recently introduced the idea of "Federated Machine Learning" or "Federated Learning (FL)"[23]. The main concept behind FL is to prevent the sharing of user data by peripherals. FL is therefore defined as a type of collaborative distributed/decentralised ML privacy-preserving technology where a model is trained without the need to transfer data from the edges to a central server, but models are sent to peripherals to be trained on local data, and then sent back to a central aggregation server to build the global model without knowledge of the embedded data. Federated Averaging (FedAVG), the first proposed FL model [23], provided a technique for combining locally trained models into a single global model. This process is iteratively repeated until the accuracy of the combined model reaches the target level. Federated Machine Learning enables privacy in Machine Learning, where sensitive data remains under the control of its original owners by ensuring that data is stored locally and data transfer between parties is kept to a minimum. The architecture of FedAvg is shown in Figure 1 below. In FedAvg, a central sever, named as "Manager", send a Machine Learning model for clients, where they train it with their data, and send it back to the Manager that aggregates all the received models and update the global model accordingly.

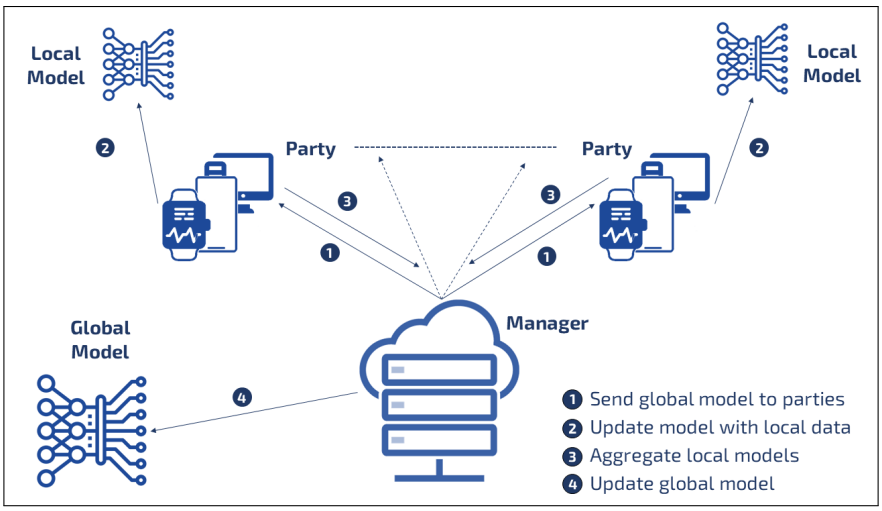

Fig. 1. FedAvg Architecture

1.1.1 Underlying Architecture. Typically, a Federated Machine Learning environment (described in 1) mainly consists of four groups of entities, namely the server, the parties, the communication framework, and the aggregation algorithm [22, 23]. Each of these entities performs a specific role in the Federated Learning process. These entities can be defined as follows:

- Central Server (Manager): the entity responsible for managing connections between entities in the FL environment and for aggregating the knowledge acquired by FL clients;
- Parties (Clients): any computing device with data that can be used to train the global model, including but not limited to: Personal Computers, Servers, Smartphones, Smartwatches, Computerized Sensor Devices, and many more;

- Communication Framework: consists of the tools and devices used to connect servers and parties, and can vary between an internal network, an intranet, or even the Internet;
- Aggregation Algorithm: the entity responsible for aggregating the knowledge obtained by the parties after training with their local data and using the aggregated knowledge to update the global model.

Then, the classical approach of the learning process is achieved in the environment of FL by repeating the following steps:

(1) The central server receives the connection from the clients and sends them the initial global model;
(2) The parties receive an initial copy of the model, train it with their local data, and send the results back to the central server;
(3) The central server receives the locally trained models, which are aggregated with the correct algorithm;
(4) The central server updates the global model based on the aggregation results and sends the updated version to the clients;
(5) Repeat the above steps until the model converges or the server decides to stop.

In Figure 2 below, the underlying architecture, entities, and process steps are illustrated for a better description of the FL environment.

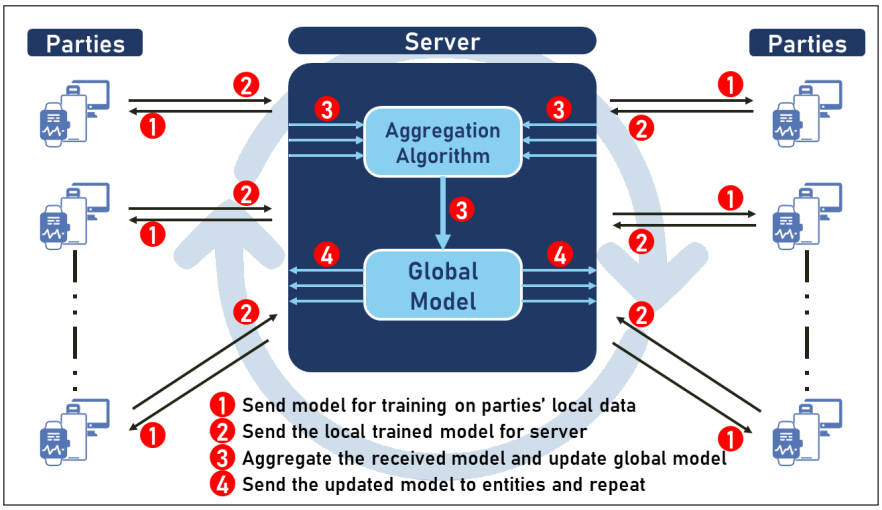

Fig. 2. Federated Learning process and environment

1.1.2 Challenges in FedAvg. Despite the progress FL has made in protecting privacy, FL is still in its infancy and vulnerable to many difficulties and obstacles. However, the performance of FedAVG is poorly understood and encounters a number of problems, including [24–29]:

- Performance Issues:
  - Suffering from 'client-drift' and convergence;
  - Tuning difficulty;
  - High communication and computation cost;
  - Significant variability in terms of the systems characteristics on each device in the network;

  - Non-identically distributed data across the network;
  - Heterogeneity of devices, users and network channels;
  - Sensitivity to local models;
  - Scalability issues.
- Security & Privacy Issues: FL is still under the risk of several breaching attacks such as
  - Poisoning attacks;
  - Inference attacks;
  - Backdoor attacks.

1.2 Privacy & Security Deficiency in FL

First and foremost, security and data protection should be separated, even though security in the broadest sense is part of data protection. Security refers to the ability to transmit and receive data securely without being monitored, altered, or tampered with. If a plan is secure and communication between participants occurs over a secure channel, it is as secure as a face-to-face conversation. On the other hand, the term"Information Privacy " in the context of digital data protection refers to the idea that people should be able to control how their digital information is collected and used. In the case of personal data, this is particularly important. Both the idea of privacy and the field of information technology (IT) have evolved over time. The way information is shared has changed drastically with networking and computers [32, 33]. In this context, security and privacy in Federated Learning technology are discussed below.

1.2.1 Security & Privacy in FL. The analysis of security and privacy in Federated Learning literature leads to the taxonomy proposed below [24–31]:

- **Security:** occurs in the communication process to ensure that two individuals communicate with each other within a network in the same way as they would in a face-to-face environment, and can be divided into:
  - Confidence: ensures that the adversary is not able to obtain information from the transmitted ciphertext;
  - Authentication: guarantees that the recipient of the message is the one intended by the sender of the message;
  - Integrity: verifies that the message is not added, removed, or modified during transmission.
- **Privacy:** refers to the use of the exchanged data, only by the parties authorized to do so, and can be discussed in three categories:
  - **Consent:** to confirm that the shared data is intended only for those users who consent to the sharing of their own data, such as those who sign up to participate in FL;
  - **Precision:** The results of some data activities are shared, but it must be determined which parts of the data are to be shared. For example, in FL, the data is not shared, but the local model trained with local data is;
  - **Preservative:** to ensure the safety of the data against leaks caused by reverse analyses on local models.
- **Robustness:** Resistance to various attacks and breaches, discussed in more detail later in Section 2.

The security and privacy aspects of Federated Learning are summarized in Figure 3 below. The proposed taxonomy was derived from available reviews and implementations in the literature [24–31].

1.2.2 Privacy Leakage in FL. FL offers privacy-preserving model training that requires no data transfer and allows users to join or leave the FL system at any time. However, the transmission of model updates during the training process may expose sensitive information [34–36] or even cause a deep leak [37], either to third parties or to the central server

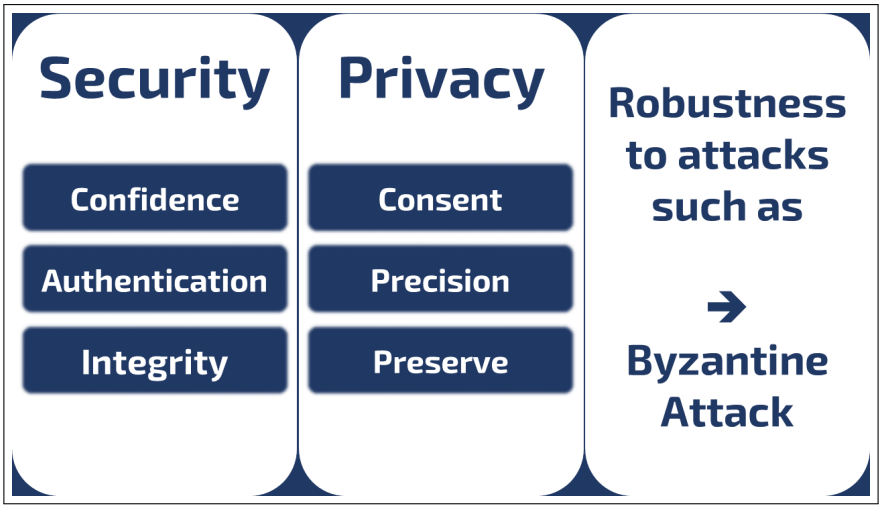

Fig. 3. Security, Privacy and Robustness Taxonomy

[38, 39], as recent research suggests that FL doesn't always provide sufficient privacy protection. For example, [40] shows that even a small portion of the original gradient can reveal information about local data. Moreover, a hostile attacker can quickly and completely obtain the training data for the gradient [37]. Such threats pose a major threat to FL and make it important to understand the concepts underlying these attacks. This is necessary because the FL protocol has vulnerabilities against both:

- **Possibly malicious server:** that can observe individual updates over time, disrupt training, and manage participants' views of global parameters;
- **Any participant:** that can observe the global parameters and manage the parameter uploads. For example, hostile individuals may intentionally change their inputs or introduce covert backdoors into the overall model.

1.3 Article Contributions

This article delves deeply into the critical realm of Federated Machine Learning, emphasizing its paramount importance in addressing the pressing concerns surrounding user data privacy. In a comprehensive exploration, it meticulously scrutinizes the multifaceted dimensions of Federated Learning from both security and privacy standpoints. By meticulously examining threats and vulnerabilities inherent in this domain, the article offers invaluable insights into the protective measures essential for safeguarding Federated Learning systems. The article endeavors to fill critical gaps in the existing literature by offering novel insights and comprehensive analyses across various topics, as outlined in its contributions. While existing literature has extensively covered the security landscape of Federated Learning, this article distinguishes itself by addressing previously unexplored aspects, thereby enriching the scholarly discourse by:

- Proposing a systematic classification framework delineating the varying levels of data security within Federated Learning, categorized into three distinct groups;
- Conducting a thorough review of the myriad threats and attacks prevalent in Federated Learning environments, illuminating potential vulnerabilities:

- Providing an exhaustive analysis of Federated Learning aggregation algorithms, with a specific focus on enhancing privacy, security, and robustness;
- Delving into the intricacies of privacy-enhancing mechanisms integrated within these algorithms, elucidating their efficacy in fortifying Federated Learning against malicious exploits;
- Investigating the utilization of cutting-edge technologies such as Homomorphic Encryption to bolster the security posture of Federated Learning paradigms;
- Exploring alternative methodologies poised to augment the security infrastructure of Federated Learning systems, thereby diversifying the arsenal of defensive measures;
- Concluding with a forward-looking discussion on the future trajectories and emerging trends in the realm of securing Federated Learning, offering invaluable insights for further research and development.

By meticulously dissecting the intricacies of Federated Machine Learning security, this article not only advances the scholarly discourse but also underscores its indispensable role in shaping the future landscape of data privacy and security.

2 FEDERATED LEARNING THREATS & ATTACKS

Federated Learning is vulnerable to several types of attacks that are already known in the Machine Learning domain. A thorough analysis of the literature provides insight into these attacks. However, a deep understanding of attacks in the Federated Learning domain requires a solid understanding of privacy threats in the digital world in general and in the Machine Learning domain in particular. In the Machine Learning context, threats, sometimes referred to as vulnerabilities, refer to potential security flaws or weaknesses that an attacker can exploit. These deficiencies may include inadequate data security, lack of proper authentication systems, and insufficient access restrictions. An attack, on the other hand, is the intentional and aggressive exploitation of these threats that results in damage to the ML system or unauthorized access to sensitive data. An example of a threat to a Machine Learning system is an unsecured database of training data, while an attack is when an unauthorized person attempts to gain access to or even steal that data. Understanding and addressing threats and attacks is critical to maintaining the security and trustworthiness of Machine Learning systems [29, 30, 37]. In this context, threats can be classified into the following three groups [29]. These threats, including but not limited to those discussed in the list below and shown in Figure 4:

- **Insider vs. Outsider Threat:** Since "Insiders" are the parties within the FL system and "Outsiders" are other parties, "Insider Attack" is an attack originating from either the FL server or one of the subscribers, while "Outsider Attack" is defined as an attack initiated by eavesdroppers in the communication channel between subscribers and the FL server or by users of the final FL service. However, due to access restrictions, the insider attack is usually stronger than the outsider attack, so the latter has been less studied in the literature. Therefore, insider attacks can be summarized as follows:
  - **Single Attack [41, 42]:** in which a single, malicious, non-colluding individual aims to make the model misclassify a given set of inputs with certainty;
  - **Sybil Attack [42, 43]:** which aims to launch more effective attacks against FL, attackers can mimic multiple fake subscriber accounts or select already compromised individuals;
  - **Byzantine Attack [44–48]:** a Byzantine Attack, or Byzantine Failure, is a situation in which one or more individuals experience technical glitches or communication problems and, as a result, submit incomplete information to the parameter server, which can affect the accuracy of the overall model. Such a failure

can take the form of an attack, which we refer to as a "Byzantine Attack," in which malicious actors purposefully compromise a FL system by strategically providing dishonest responses. These attacks fall into two categories:

  - **Gaussian Attack [49]:** is performed by a single worker in a FL system, regardless of local training datasets. The individuals performing this attack draw their responses randomly from a Gaussian distribution;
  - **Fall of Empires Attack [50]:** Designed to overcome strong aggregation algorithms, it requires a minimum number of Byzantine workers (i.e., individuals performing the attack), depending on the strength of the resistant algorithms. It also requires that the Byzantine workers know the answers sent by the truthful employees.

- **Semi-Honest vs. Malicious:** Attackers in the semi-honest context are considered passive or honest but curious. They seek to learn the private states of the other parties while adhering to FL protocol. Passive adversaries are assumed to only view the aggregated or averaged gradient, but not the training data or gradient of the other honest players. In the malicious situation, an active or malicious adversary attempts to learn the secret states of the honest players and arbitrarily deviates from the FL protocol by modifying, replaying, or deleting messages. This powerful adversarial model enables the adversary to launch exceptionally deadly attacks;
- **Training Phase vs. Inference Phase [51, 52]:** Attacks in the training phase use data poisoning or model poisoning to learn, influence, or corrupt the FL model itself. In the first case, the integrity of the training data collection is compromised, while in the second case, the learning process is compromised. The attacker can also perform a series of inference attacks on the update of a single participant or on the set of updates of all participants. Inference attacks, on the other hand, often do not interfere with the target model, but either make it produce false results or gather information about the model's properties. The success of such attacks depends primarily on how well the attacker understands the model. Moreover, if the target model is provided as a service, the FL model broadcast phase makes the model accessible to any malicious client.

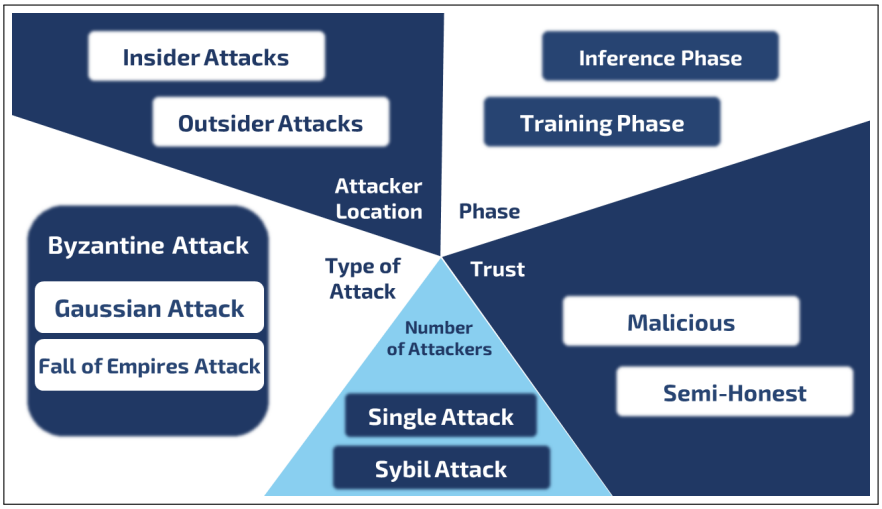

Fig. 4. Known Threats in Federated Learning Field

2.1 Poisoning Attacks in Federated Learning

Poisoning attacks can be either random or targeted [53]. Random attacks aim to reduce the accuracy of the model FL, while targeted attacks aim to make the FL model to output the target label given by the attacker, the latter being more difficult when the attacker has a specific target. In addition, poisoning attacks can be performed on the data or on the model. Both attacks attempt to change the behavior of the target model in an undesirable way. If the attackers can compromise the FL server, they can easily perform both targeted and untargeted poisoning attacks on the trained model. Data and poisoning attacks can therefore be explained as follows:

- **Data Poisoning:** also known as data corruption, occurs in local data collection and is broadly divided into two types: Clean Label [54] and Dirty Label [55]. In clean label attacks, it is assumed that the attacker cannot change the label of the training data because there is a mechanism to confirm that the data belongs to the correct class, and the poisoning of the data must go unnoticed. In contrast, in dirty label poisoning, the attacker can insert a set of data samples into the training set that he wants to misclassify with the intended target label. In addition, any member of FL can perform data poisoning attacks. Thus, the impact on the FL model depends on how many system members participate in the attacks and how much training data is poisoned;
- **Model Poisoning:** occurs during local model training and aims to contaminate local model updates before they are sent to the server or implant secret backdoors in the global model [42]. The attacker's goal in targeted model poisoning is to cause the model FL to misclassify a set of selected inputs with high confidence. Note that these inputs are not modified to cause misclassification at test time, as is the case with attacks by adversarial attacks [56]. Rather, the misclassification is the result of the attacker's manipulation of the training process. In FL, Model Poisoning even trumps Data Poisoning, since attacks by Data Poisoning eventually affect a portion of the updates that are fed to the model at each iteration [43]. This is essentially equivalent to a centralized poisoning attack that poisons a portion of the entire training data. Model poisoning attacks require strong technological capabilities and a large amount of processing power.

2.2 Inference Attacks in Federated Learning

Sharing gradients during FL training can lead to a significant loss of privacy [36, 37, 40, 57]. Since deep learning models appear to internally recognize numerous features of the data that are not clearly related to the core tasks, model updates may provide additional information about undesirable features of the participants' training data to hostile participants. The attacker can also store the snapshot of the FL model parameters and perform property inference based on the difference between subsequent snapshots, which corresponds to the aggregate updates of all participants without the attacker. The fundamental problem is that the gradients are derived from the participants' private data.
The gradients of a given layer in deep learning models are created based on the characteristics of that layer and the error of the layer above it. The gradients of the weights in successive fully connected layers are the inner products of the error of the layer above and the features. Similarly, the gradients of the weights in a convolutional layer are convolutions of the error of the overlying layer and features [36]. As a result, observations can be used to update the model to derive a significant amount of private information, such as class representatives, membership, and attributes associated with a subset of the training data. Worse, an attacker can infer labels from shared gradients and recover the original training samples without knowing anything about the training set [37]. Inference attacks fall into the following categories:

- **Inferring Membership:** The goal of membership inference attacks is to determine whether a particular data element was used to train the model [58];
- **Inferring Class Representatives:** takes place when a malicious participant can specifically compromise any other participant, and the attack takes advantage of the real-time nature of the FL learning process, which allows the attacker to train a network that generates prototype samples of the targeted training data that should be private. The generated samples appear to be from the same distribution as the training data;
- Inferring Properties: In this attack, an attacker can perform both passive and active property inference attacks to infer properties of other participants' training data that are independent of the features describing the classes of the FL model[36]:
  - Property inference attacks require the attacker to have additional training data labeled with the exact property they wish to infer;
  - A passive attacker can only monitor updates and make inferences by training a binary property classifier;
  - An active adversary can use multitask learning to trick the model FL into learning a better separation between data with and without the property, resulting in more information being extracted;
  - An adversarial participant can even infer when a feature appears and disappears in the data during training;
  - It can recover pixel-perfect original images and token-matched original texts.

In Figure 5 below, the attacks that are known in Federated Learning field are illustrated.

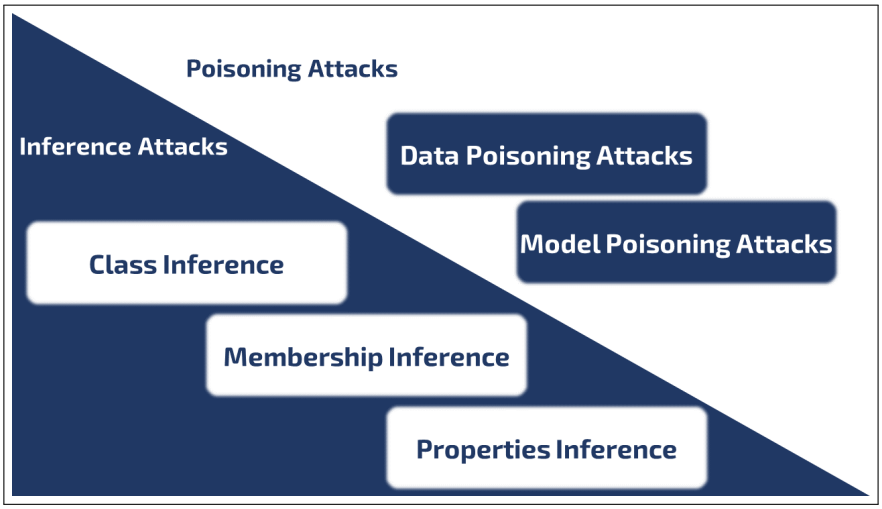

Fig. 5. Known Attacks in Federated Learning Field

3 SECURING FL AGGREGATION ALGORITHMS

Federated Learning technology is known as a privacy-preserving technology that builds Machine Learning models without collecting users' private data. The first federated aggregation algorithm, proposed by Google and called FedAvg [23], was originally dedicated to the idea of aggregating multiple locally trained models. However, later reviews have shown that this algorithm is vulnerable to many challenges, including various attacks and violations [24–29]. Therefore, several aggregation algorithms have been proposed to address these issues. However, not all aggregation algorithms

have been developed to address privacy and security concerns. Some were developed to optimize communication and communication costs, such as FedBoost [59], FedProx [60], FedMA [61], and others. Other approaches have been used to improve personalization, such as FedDist [62]. Thus, in this section, we discuss the aggregation algorithms of FL that aim to improve privacy and security in Federated Learning environments.

3.1 Securing FL Aggregation Against Active Adversaries

The FL algorithm proposed in [23] is vulnerable to poisoning attacks. In [63], the authors proposed a protocol for secure vector summation that has a fixed number of rounds, minimal communication cost, failure robustness, and only one server with limited trust. In this design, the server has two tasks: relaying communication between other parties and computing the final result. In addition, the authors offer two variants of their protocol, the first of which is more efficient and can prove secure against honest but curious adversaries in the simple model. The alternative, on the other hand, ensures anonymity against active adversaries (including an actively hostile server), but requires an additional round and has been shown to be safe in the random oracle model. Using a simulation-based demonstration, it was shown in both cases that the server learns the user input only in aggregate form [23]. To secure the communication between the involved parties, a cryptographic primitive was implemented in the following phases:

- **Secret Sharing:** which relies on Shamir's ***"t-out-of-n"*** Secret Sharing [64], which allows a user to divide a secret ***"s"*** into ***"t"*** shares that can be used to reconstruct the secret without revealing the secret from shares smaller than ***"t"***, does not provide information about the secret ***"s"***;
- **Key Agreement:** consists of three functions, the first of which generates some public parameters, the second of which allows each party to generate a privateand public key pair, and the third of which allows each user to combine his private key with the public key to obtain a private shared key;
- **Authenticated Encryption:** a symmetric encryption that combines confidentiality and integrity guarantees for messages exchanged between two parties;
- **Pseudorandom Generator:** ensures that, given a uniformly random seed, its output is computationally indistinguishable from a uniformly sampled element of the output space as long as the seed is hidden from the discriminator;
- **Signature Scheme:** Now, for the signature to prove the origin of a message, it must be the case that someone without the secret key cannot create a valid signature for a message they have not yet seen signed, which is called UF-CMA security [65];
- **Public Key Infrastructure:** allows clients to register identities and sign messages with their identities so that other clients can verify the signature but not forge it.

Based on the performance analysis performed by the authors, this approach shows several advantages and disadvantages which are listed here:

- **advantages:**
  - **Privacy-Preserving:** By eliminating the need to collect user data, user privacy is preserved;
  - **Security:** by using a cryptographic primitive that prevents communication with unauthorised users;
  - **Dropped Users Management:** The server receives messages from all users who have not dropped this round, and terminates if the number of messages received is less than the desired number.
- **drawbacks:**

- **robustness against active attackers:** The security protocol ensures that when the server learns user input, it always merges it with other users' values, but it does not protect against malicious clients who want to prevent the server from learning any sum at all;
- **forcing well-formed input:** The protocol also does not ensure that user input is well-formed or within certain bounds, allowing malicious users to enter arbitrary values of their choosing, resulting in the server's output also being ill-formed;
- **Communication Overhead:** Users must exchange random vectors, which could require a quadratic communication overhead if naive.

3.2 Robust Federated Aggregation (RFA)

Following the success of the Federated Learning approach as a privacy-friendly technology, the authors of [66] proposed Robust Federated Aggregation (RFA), a new approach for FL, which makes the aggregation process more robust to possible poisoning of local data or model parameters of participating devices. Since corrupted or poisoned devices can only affect the global model through updates, the authors contributed to the aggregation step and proposed an improved aggregation algorithm for Federated Learning. The proposed method is based on the geometric median, which can be easily computed using a Weiszfeld-type algorithm [67] and is independent of the extent of damage, and aggregates model updates without revealing the specific contribution of each device. The authors' experiments have shown that RFA can compete with traditional aggregation when the extent of corruption is low, while it has higher resilience when the extent of corruption is high.
Their model, RFA, is based on the principle of aggregation with the Geometric Median (GM) defined as the minimizer of vectors with an optimal collapse point of 1/2, where at least half of the points must be changed for the geometric median to correspond to any point. Moreover, the RFA algorithm is obtained by replacing the mean aggregation of FedAvg with this GM-based robust aggregation oracle. As a result, RFA is independent of the convexity of the local targets regardless of the actual amount of corruption in the problem, and the aggregate is robust. However, it turns out that robustness is incompatible with the two main goals of Federated Learning communication efficiency and privacy. Therefore, the authors propose two variants of RFA, namely one-step RFA, which aims to reduce communication costs, and personalized RFA, which aims to deal with heterogeneity. In addition, the authors explained the tension between robustness, communication, and privacy, concluding the following:

- Any FL algorithm among existing Secure Multi-Party Computation techniques can only exhibit two of the three aspects of privacy, communication, and robustness
- FedAvg is efficient in terms of communication and privacy, but not robust
- In general, any linear aggregation scheme is not robust, and therefore any robust aggregation must be non-linear;
- Only linear functions of inputs are communication efficient for the secure multiparty computation primitives based on secret sharing on which privacy protection is built.

Therefore, the presented algorithm, called RFA, is based on the geometric median and the smoothed Weiszfeld approach to aggregate the vectors. The approach proved to be robust to corrupted updates and the proposed variants also showed optimization of communication overhead.

3.3 LayerwisE Gradient AggregatTiOn (LEGATO)

Furthermore, in [68], the authors discussed the need for robust aggregation algorithms that can survive Byzantine attacks. In Byzantine attacks, workers are defined as individuals that send malicious gradients to corrupt the global model. The approach proposed by the authors was triggered by the increasing challenges in FL aggregation, where:

- Several known robust aggregation techniques, especially in non-IID environments, are unable to defend against Byzantine attacks;
- The need to develop aggregation algorithms that are:
  - intelligently detect whether a worker sending a "different" response is a malicious worker;
  - can train a global model with reasonable performance given a local data distribution without an IID;
  - uses all information collected from workers to diagnose worker behavior.

Given that existing robust aggregation algorithms are often very computationally intensive, the authors justified the development of their model by the need for robust, yet communication-efficient methods. Therefore, they introduced LayerwisE Gradient AggregatTiOn (LEGATO), a scalable and generalizable FL aggregation algorithm. LEGATO uses a dynamic gradient reweighting approach that is novel in its treatment of gradients based on layer-specific resilience and is beneficial for convergence of gradient descents in the absence of an attack. The authors secured their algorithm via the layer-by-layer approach, which works on each layer of the model. Therefore, it is worth noting that their approach is limited in implementing ML models built from layers, such as Deep Learning and Neural Network models. Consequently, the new steps of LEGATO begin when the server receives gradients from all workers where:

- **First:** The gradient log is updated by the server to include the latest gradients collected from workers;
- **Then:** It assigns to each slice a robustness factor standardized over all slices, which is the inverse of the standard deviation of those norms over all recorded rounds;
- **Finally:** All these reweighted gradients are averaged over all workers, and the resulting aggregated gradient is used as the round gradient.

The proposed approach has thus been shown to be robust to Byzantine attacks while also being considered communication efficient. However, it suffers from several drawbacks, which can be summarized as follows:

- Its limitation to Neural Networks;
- Its weakness against Gaussian variance attacks;
- Its lack of a definition for "Extreme Outliers".

3.4 Privacy-preserving Decentralized Aggregation (SecureD-FL)

So far, it has been discussed that Federated Learning aggregation algorithms are vulnerable to various poisoning attacks. This being said, Secure Multiparty Computation [63, 69, 70] Differential Privacy [71, 72] and combinations of both [73–75] are techniques to address these privacy issues. However, these techniques involve significant computational overhead, require the use of a trusted third party to provide the secret key, or compromise the quality of the trained models due to the noise introduced. Most importantly, these systems require the use of a central aggregation server that acts as a single point of failure and poses a privacy risk in the event of a hacking attack. Therefore, in [76], the authors developed a privacy-preserving decentralized aggregation protocol for Federated Learning called "SecureD- FL". Their proposed aggregation algorithm is based on an improved version of the Alternating Direction Method of Multiplier (ADMM) [77]. The proposed algorithm controls the communication between participants in each aggregation round to

reduce privacy loss and guarantee privacy against honest but curious adversaries, with this communication pattern inspired by combinatorial block design theory.

The algorithm in [76] proposed a novel communication pattern between FL system participants inspired by the theory of combinatorial block design [78]. The basic idea is that the algorithm determines which group of participants (called a group) should interact in each aggregation round to minimize privacy loss. The grouping algorithm is explained with an example:

- *assume having a set of partitions of the nine users 1, . . . , 9 in groups (of size s = 3) with a gap constraint. Each of the partitions corresponds to a communication scheme in an ADMM iteration. The members of a group (triangles) are free to communicate their parameters among themselves in one iteration. These partitions create a communication gap across the ADMM aggregation. Therefore, users do not disclose private information when the aggregation converges in less than twice the number of partitions at least" [76].*

In Figure 6, the communication protocol is shown for a group of 9 users divided into 3 groups. In the figure, parts a, b, and c represent communication with gaps between individuals with unequal distances. In this case, two adjacent individuals can communicate more than once in a full communication cycle (8 iterations), unlike part d where the connection between the same two individuals occurs only after 8 iterations. This reduction in the repetition of communication between individuals contributes to less leakage of private information.

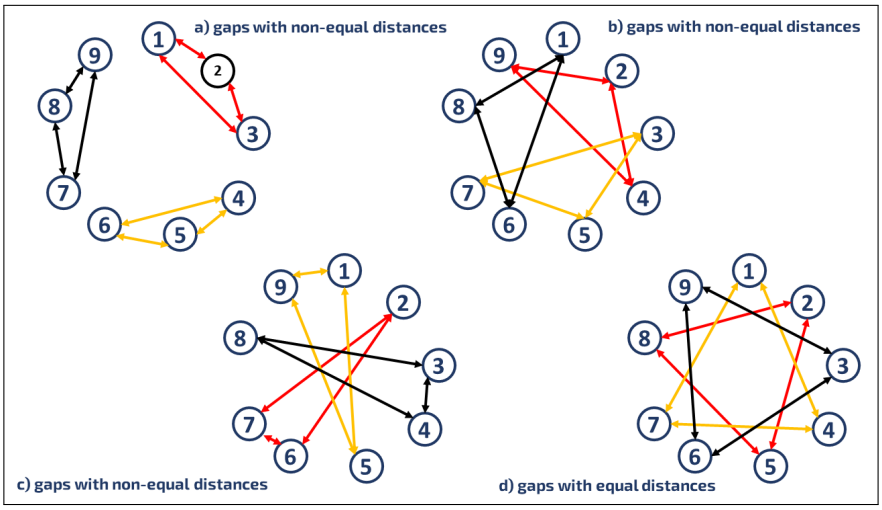

Fig. 6. Communication Control Explained for Set of 9 Users

Following the classical Federated Learning algorithm, the proposed architecture consists of multiple rounds. Since no central server is required, the steps are as follows:

- Each individual trains the global model using its local data and updates the model parameters;
- The individuals synchronize the locally trained models;
- The individuals work together to compute the summed model via the ADMM-based secure aggregation algorithm without having to send it to a central server;
- Repeat the above steps until the model converges.

However, by using the communication control protocol, the communication between individuals proceeds as follows:

- In each iteration, each individual performs its minimization and shares its parameters with other individuals in the same group to calculate the partial sum of the group;
- Different groups exchange their partial sums to calculate the final sum of the model;
- Individuals update their parameters at each iteration with the final sum.

As a result, individuals from FL are able to build an aggregate model with fewer repeated communications, which increases robustness to a loss of privacy.

3.5 Secure and Efficient Aggregation for Byzantine-Robust Federated Learning (SEAR)

Byzantine attackers are not the only problem that Federated Learning technology can suffer from. For example, the server is able to infer private content from the client's data. It can recover this data by Generative Adversarial Network (GAN) [79] or pixel-wise accurate images by using gradients [37]. Moreover, implementing cryptographic primitives such as Homomorphic Encryption (HE) and Secure Multi-Party Computation (MPC) incurs expensive communication and computation costs [80]. Therefore, in [80], the authors proposed a new private and secure aggregation algorithm called SEAR. Their proposed algorithm uses a hardware-based trusted execution environment instead of cryptographic time-consuming tools.

First, SEAR used the Intel SGX [81] Trusted Execution Environment (TEE) to aggregate the locally trained models in a secure and trusted hardware environment. A Trusted Execution Environment is a secure area of the central processor where the confidentiality and integrity of the code and data loaded into it can be well preserved. There are two TEEs based on different processor architectures: ARM TrustZone [82] and Intel Software Guard Extensions (SGX) [81]. In Intel SGX, the trusted part is called the enclave and the protected memory area is called the Processor Reserved Memory range (PRM), which cannot be accessed by code outside the enclave. Since the local models are encrypted and only a trusted execution environment, the enclave, has the key to recover them, sensitive information is never disclosed to the aggregation.

However, the physical memory size of PRM is limited to 128 MB on current Intel CPUs, which limits the number of locally trained models that can be aggregated simultaneously. This poses a major challenge since thousands or even more individuals may be involved in aggregation simultaneously in the FL environment. Therefore, the authors proposed two data storage modes that can be used within the enclave:

- **Row Major Data Storage Mode:** the parameters uploaded by a client are stored in a contiguous storage area suitable for the aggregation algorithm, which is implemented by accessing each client's vector once;
- **Column Major Data Storage Mode:** stores the parameters in the same dimensions in a contiguous array.

Since the row-oriented data storage mode is time-consuming due to the time required for EPC paging, the column-oriented mode is proposed as a solution to this drawback. In this mode, PRM is able to shop more dimensions without changing the total memory consumption. In addition, in [80], the authors considered preventing information leakage through side channels, such as power consumption [83], rollback attacks [84], or other timing attacks [85].

3.6 Efficient Privacy-Preserving Data Aggregation (EPPDA)

FL is vulnerable to data poisoning attacks and also to reverse attacks that can analyze users' model and expose their private data. For example, the aggregation server, as a legal FL participant, can decrypt individuals' locally trained models. Moreover, the instability of the communication network affects the system FL. Therefore, in [86], the authors proposed an

efficient, fault-tolerant privacy preserving data aggregation scheme that requires less communication and computation. Their model, called Efficient Privacy-Preserving Data Aggregation (EPPDA), exploits the homomorphisms of the secret exchange [87] to minimize the iterations of the secret exchange and thus reduce the consumption of communication, computation, and storage resources. In this context, reducing communication, computation, and memory resources leads to improved system efficiency, especially when the number of training times increases. Moreover, secret sharing can secure users' secret data to reduce the influence of some malicious users, which makes EPPDA a private fault-tolerant algorithm. The cryptographic primitives used in EPPDA can be summarized in the following steps:

- **Secret Sharing:** which is based on the Shamir Secret Sharing [64] discussed in Section 3.1;
- **Key Exchange Protocol:** which helps both communication parties generate a session key in a public channel;
- **Authenticated Encryption:** which allows both communication parties to communicate with a shared secret in a public channel;
- **Signature Scheme:** verifies the source of the message.

3.7 Secure Aggregation with Heterogeneous Quantization (HeteroSAg)

FL offers many advantages, but also suffers from key challenges such as communication bottlenecks, system failures, malicious users, and Byzantine faults. Therefore, the authors in [88] Secure Aggregation with Heterogeneous Quantization (HeteroSAg) proposed a privacy-friendly and heterogeneous efficient FL aggregation algorithm. HeteroSAg, as proved by the authors, proves:

- Protect the privacy of each user's local model updates by masking each user's model update so that the mutual information between the masked model and the unique model is zero;
- Enable the use of heterogeneous quantization, which allows edge users to adjust their quantizations proportionally to their available communication resources, resulting in a much better tradeoff between training accuracy and communication time;
- Achieves resilience against Byzantine attacks by adding distance-based defences;
- reduces bandwidth expansion.

HeteroSAg enables secure aggregation with heterogeneous quantization. The efficiency of HeteroSAg and its robustness against Byzantine attacks lie in the FL system cycle that runs the Segment Grouping Strategy. The main concept behind HeteroSAg privacy efficiency lies in this strategy, which is based on dividing edge users into groups and segmenting local model updates for these users. This segmentation helps with the following:

- computation on segments with specified user cooperation, so that segments can be quantized by different quantizers instead of applying the safe aggregation procedure to all local model update vectors;
- enabling safe model aggregation with heterogeneous quantization while preventing the server from unmasking the full average model from a subset of users.

In summary, HeteroSAg has been shown to be an efficient aggregation algorithm that provides a much better tradeoff between training accuracy and communication time. Moreover, the proposed HeteroSAg method can be used to mitigate Byzantine attacks and drastically reduce the bandwidth growth of the secure state-of-the-art aggregation protocol.

3.8 FLDetector: Securing FL via Detecting Malicious Clients

At FL, existing defenses have focused heavily on Byzantine-robust or provably robust methods, even in the presence of malicious clients. However, a major limitation of these defenses is that they can withstand only a small number

of malicious clients. In [89], the authors present FLDetector, which was developed to address this critical challenge of defending against model poisoning attacks with a large number of malicious clients. Their method takes a unique approach by focusing on detecting such malicious clients. The methodology is based on a fundamental observation: in the context of model poisoning attacks, model updates originating from a client in multiple iterations exhibit inconsistencies. In this context, FLDetector identifies potentially malicious clients by checking the consistency of their model updates. Essentially, the server uses historical model updates to predict a client's model update for each iteration. It raises a red flag and classifies a client as malicious if the model update received from that client does not match the predicted model updates across multiple iterations. This innovative approach enables FLDetector to tackle the difficult scenario of combating model poisoning attacks orchestrated by a variety of malicious clients. By assessing the consistency of model updates, FLDetector not only provides a robust defense mechanism, but also provides insight into the limitations and opportunities for mitigating such threats within the Federated Learning landscape. However, it's important to recognize that FLDetector, like any defense strategy, has its own benefits and limitations, which are explored below with the benefits:

- Benefits
  - Improved Defense Against Large-Scale Attacks: it offers a defense mechanism against model poisoning attacks involving a substantial number of malicious clients, addressing a critical challenge in FL domain;
  - Unique Detection Approach: The method's unique approach, focusing on the consistency of model updates across iterations, sets it apart from existing defenses that rely on Byzantine-robust or provably robust methods;
  - Predictive Model Update Analysis: By utilizing historical model updates to predict a client's model update for each iteration, FLDetector demonstrates the ability to proactively identify potentially malicious clients;
  - Early Detection: The system raises a red flag and designates a client as malicious if inconsistencies persist across multiple iterations, enabling early detection and mitigation of threats.
- Limitations
  - Detection Sensitivity: FLDetector can identify fraudulent clients based on inconsistent model updates, but it may also raise false alarms in cases of benign inconsistencies like communication problems or system noise;
  - Complexity and Resource Requirements: Predictive model update analysis and consistency testing may increase server-side computational complexity and resource overhead, impacting FL system performance;
  - Effectiveness Under Evolving Attacks: The method's effectiveness could be limited if adversaries adapt their strategies to introduce more subtle and harder-to-detect inconsistencies in their model updates;
  - Trade-Off Between False Positives and False Negatives: Fine-tuning the balance between false positives (flagging innocent clients as malicious) and false negatives (failing to detect true dangerous clients) may be difficult.

3.9 FLCert: Security by Clients Grouping Strategy

Similarly, the authors in [90] Proposed FLCert, whose main concept is to categorize customers into groups and facilitate the learning of a global model for each client group using established FL methods. Then, the system uses a majority voting mechanism between these global models to classify test inputs. The approach considers two different methods for grouping clients, resulting in two variants: FLCert-P, in which clients are randomly grouped, and FLCert-D, in which

clients are deterministically divided into disjoint groups. Through extensive experimentation on multiple datasets, the results show that the labels predicted by FLCert for test inputs are demonstrably not affected by the influence of a limited number of malicious clients, regardless of the specific poisoning attack strategies. This breakthrough in providing provable security guarantees positions FLCert as a robust and promising defense-in-depth strategy in the FL landscape that addresses critical vulnerabilities and advances security standards in the field. However, as with any approach, it is important to recognize the potential benefits and limitations, which are listed below:

- Benefits:
  - Provable Security: it offers a solid security guarantee against poisoning attacks;
  - Ensemble Approach: It leverages client grouping and majority voting among global models;
  - Two Variants: it presents flexibility with FLCert-P and FLCert-D catering to various use cases.
- Limitations:
  - Assumed Malicious Clients: FLCert assumes a known upper limit on malicious clients, limiting its applicability in scenarios with uncertain adversarial activity;
  - Complexity: Implementing client grouping and ensemble learning introduces computational complexity and resource demands, impacting efficiency;
  - Grouping Methods: The effectiveness of client grouping methods may vary depending on the data distribution, necessitating careful selection;
  - False Positives: Like any defense mechanism, finding the right balance between false positives and false negatives in identifying malicious clients may pose a challenge.

3.10 ELSA: Security by Distribution of Trust

In the context of FL, they cannot consider malicious actors within the system, which is a major obstacle to making FL an ideal solution for privacy-preserving Machine Learning applications. As a solution, the authors in [91] proposed ELSA, a breakthrough secure aggregation protocol to overcome these challenges. ELSA not only ensures efficiency, but also combats the presence of malicious actors at its core. ELSA introduces a novel secure aggregation protocol based on distributed trust between two servers that keeps individual client updates secret as long as a server remains honest. This design not only protects against malicious clients, but also ensures end-to-end efficiency. What distinguishes ELSA from previous protocols is its innovative approach. Instead of servers interactively generating cryptographic correlations, clients act as untrusted traders of these correlations without compromising the security of the protocol. This innovation results in a much faster protocol that provides even higher security compared to previous work. Moreover, ELSA introduces novel techniques that maintain privacy even when a server is malicious, with only a small increase in runtime and negligible communication overhead compared to the case of reasonably honest servers. This groundbreaking work significantly improves end-to-end runtime over previous approaches with similar security guarantees. A number of benefits and limitations of ELSA are proposed in the following list:

- Benefits:
  - Efficiency: it offers a much faster secure aggregation protocol compared to previous approaches, making it suitable for real-world FL scenarios;
  - Malicious Actor Resilience: The protocol addresses the presence of malicious actors at its core, ensuring robust security even in the face of adversarial clients or servers;

  - Distributed Trust: ELSA leverages distributed trust across two servers, keeping client updates private as long as one server remains honest, enhancing privacy and security;
  - Improved Security: It achieves stronger security at high efficiency compared to prior work, making it a noteworthy advancement in secure aggregation techniques;
  - Negligible communication overhead compared to semi-honest server scenarios.
- Limitations:
  - Assumed Trust: The security of ELSA relies on the assumption of at least one honest server. In cases where both servers are compromised, the protocol's security guarantees may be compromised;
  - Additional Runtime Cost: While ELSA maintains privacy when a server is malicious, it does come with a runtime cost of 7-25%, which may impact the speed of Federated Learning processes;
  - Specific Model Consideration: The performance improvements mentioned in the text, may be contingent on the specific ML models used, and results could vary with different model architectures or dataset sizes.
  -

3.11 Multi-RoundSecAgg: Securing by Random User Selection Strategy

In [92], the authors showed that the conventional practice of random user selection in FL can lead to the leakage of users' individual models within a number of rounds proportional to the total number of users. To address this critical challenge, they introduced a new secure aggregation framework known as Multi-RoundSecAgg that provides privacy guarantees over multiple rounds. This framework goes beyond the single-round privacy paradigm and introduces a new metric to quantify the privacy guarantees of FL over successive training rounds. They also develop a systematic user selection strategy that ensures the long-term privacy of each user, regardless of the number of training rounds. Importantly, their framework incorporates fairness considerations and maintains an average number of participating users in each round. Their experiments, conducted on datasets such as MNIST [101] & CIFAR [105], illustrate the effectiveness and practicality of our multi-roundSecAgg framework. This work represents a significant advance in the field, addressing the privacy challenges posed by multi-round scenarios from FL and providing improved privacy guarantees for Federated Learning systems.

- Benefits:
  - Enhanced Privacy over Multiple Rounds: it recognizes and mitigates privacy vulnerabilities that may arise from partial user participation over time;
  - Novel Privacy Metric: it introduces a new metric to quantify privacy guarantees across multiple training rounds enhances the assessment and understanding of long-term privacy preservation in Federated Learning;
  - Structured User Selection: it ensures the sustained privacy of each user, irrespective of the number of training rounds, addressing a critical concern in Federated Learning;
  - Fairness and Participation: it takes into account considerations of fairness and maintains an average number of participating users at each round, promoting equitable involvement and balanced contributions in the learning process.
- Limitations:
  - Complexity: The introduction of multi-round privacy guarantees and structured user selection strategies may increase the computational and operational complexity of FL systems;

- Resource Demands: Implementation could require additional computational and storage resources, which may pose challenges in resource-constrained environments;
- Generalization: The framework's effectiveness and privacy guarantees may depend on specific use cases and data distributions, and its generalization to all scenarios may require further exploration;
- Fairness Considerations: While the framework accounts for fairness in user participation, achieving perfect fairness in all practical scenarios may still be a challenge;
- Scalability: The scalability of Multi-RoundSecAgg to large-scale Federated Learning scenarios with numerous participants and data sources remains an area of consideration and potential limitation.

4 SECURING FEDERATED LEARNING WITH HOMOMORPHIC ENCRYPTION

The challenges of Federated Learning require the implementation of various solutions, especially to improve security and privacy. In this context, encryption algorithms have been used to provide additional security for transactions between different FL individuals. Encryption techniques are divided into two types: Secret key algorithms and public key algorithms. All of today's encryption algorithms fall into one of the two categories: Secret key encryption techniques, where the same key, called the secret key, is used to encrypt and decrypt a message, and public key encryption techniques, where one key, called the public key, is used to encrypt a message and another key, called the private key, is used to decrypt it[93–98].

Homomorphic encryption (HE) in this context is a type of encryption that makes it possible to perform certain types of calculations with the ciphertext and obtain an encrypted result that, when decrypted, matches the result of operations with the plaintext. In the development of current communication systems, this is a desired property [93–97]. RSA [95] is known as the first public-key encryption algorithm with a homomorphic scheme. Moreover, the debate on homomorphic encryption schemes can be summarized in the list below [93–98]:

- **Benefits:**
  - Elimination of the need for trusted third parties, keeping data secure and confidential in untrusted contexts, such as public clouds or third parties. Data is encrypted at all times, reducing the possibility that sensitive information could ever be hacked;
  - Elimination of the tradeoff between data usability and data privacy, where there will be no need to obfuscate or remove elements to ensure data protection;
  - resistance to quantum attacks.
- **Limitations:**
  - Poor performance: due to issues such as slow computational speed and accuracy, fully Homomorphic encryption remains commercially impractical for computationally intensive applications. The research community generally agrees that research in fully homomorphic encryption still has a long way to go, although it is useful today in conjunction with other privacy-enhancing technologies such as secure multiparty computing.

In addition, the authors confirm in [80] what has already been discussed about the poor performance of HE for heavy computation. In particular, they discussed the use of HE in Federated Learning aggregation algorithms, noting that HE supports simple operations with encrypted data and typically incurs expensive computation and communication costs for complex problems. Since defending against existing attacks requires repeated comparison operations and distance calculations, securing FL with HE is time consuming, making it impractical for known attacks such as the Byzantine

attack. However, several attempts have been made to secure the FL algorithms with homomorphic encryption. These implementations are discussed in this section. It should be noted, however, that these implementations are not FL aggregation algorithms, but are used only to secure communications in the FL system, which is the main reason these implementations were not mentioned in the previous section.

4.1 Securing FL Communications With HE; State-of-The Art

Homomorphic Encryption has been a hot research topic in recent years. Research institutions, whether individuals, laboratories, or companies, have used this technique in various digital domains. Federated Learning, for example, has played a part in this interest, as HE is used to secure communication across FL individuals.

4.1.1 Using HE as a Standalone Securing Solution. For example, the authors of [99] were the first to use HE to secure the FL system. They originally described their solution as a three-party end-to-end solution secure against an honest but curious adversary. Their solution consisted of two phases: privacy-preserving entity resolution and federated logistic regression using messages encrypted with an additive Homomorphic method. Since HE allows operations over integers, the authors developed an encoding technique that translates floating-point numbers into modular integers while preserving addition and multiplication operations to implement algorithms over floating-point numbers. They used an encoding scheme similar to the floating-point representation, where a number is encoded as a pair of an encoded significant and an unencoded exponent. Their solution allows the FL model to be trained without sharing user's data, and is as accurate as a naive, non-private solution that collects all the data in one place, making it scalable to millions of entities, each with hundreds of features. However, the authors did not provide any analysis regarding the time and computational complexity of their proposed system. In the same way, the authors in [100] secured the messages exchanged between the aggregation server and the clients using HE. They secured the system FL against inference attacks and demonstrated the efficiency of their solution by testing experiments with two private financial datasets.

Likewise, authors also proposed POSEIDON in [101], which is an extension for SPINDLE [102]. POSEIDON, as defined by the authors, is a new system that enables neural network training and evaluation in a Federated Learning environment. The proposed solution secures the exchanged messages with Multiparty Lattice-Based Homomorphic Encryption [103]. They evaluated their model with several datasets: Breast Cancer Wisconsin Dataset (BCW) [104], EMNIST Dataset [105], the Epileptic seizure recognition (ESR) Dataset [106], the default of credit card clients (CREDIT) Dataset [107], the street view house numbers (SVHN) Dataset [108], and the CIFAR-10 and CIFAR-100 [109].

Therefore, in [110], the authors proposed a privacy-friendly FL (PEFL) architecture that uses HE as the underlying technology and provides a way to penalize poisoners through effective gradient data extraction of the logarithmic function. PEFL, as proposed by the authors, is the first attempt to detect poisoning behavior in FL using ciphertext. PEFL was evaluated using the EMNIST [105] and CIFAR [109] datasets. Similarly, in [111], the authors proposed their Federated Learning security mechanism based on additive homomorphic encryption (DTAHE) techniques. The proposed model allows the aggregation server to multiply the individual inputs by arbitrary coefficients and aggregate them to build a complete contiguous layer or on the individual inputs. Similarly, in [112], the authors proposed a FL security framework that uses fully homomorphic encryption. In particular, they used an approximate floating-point compatible scheme that benefits from packing and scaling the ciphertext. The authors evaluated the solution on the UK Biobank (UKBB) neuroimaging dataset [113] and the results proved the improved learning performance while maintaining the security of the FL transactions. Moreover, in [114], the authors proposed a security scheme for Federated Learning

Table 1. HE, as Standalone Solution, Implementations to Secure FL Algorithms

Ref	Securing Scheme	Dataset Used
[99]	Homomorphic Encryption to secure data exchange in FL System	-
[100]		-
[101]		Breast Cancer Wisconsin dataset (BCW) [104] EMNIST dataset [105] Epileptic seizure recognition (ESR) dataset [106] The default of credit card clients (CREDIT) dataset [107] Street View House Numbers (SVHN) dataset [108] CIFAR-10 and CIFAR-100 [109]
[110]		EMNIST [105], CIFAR [109]
[111]		-
[112]		UK Biobank (UKBB) neuroimaging dataset [113]
[114]		Human Against Machine with 10,000 training images (HAM10000) dataset [115]

based on homomorphic encryption. The proposed model was introduced to secure collaborative Deep Learning models in an Internet of Things-based healthcare system. The proposed model was evaluated on the Human vs. Machine with 10,000 Training Images (HAM10000) dataset [115] and obtained promising privacy preserving results. Moreover, in [116], [117], and [118], the authors also proposed the use of homomorphic encryption to build a Federated Learning security system. Their models built on homomorphic encryption have succeeded in creating a secure and trustworthy data exchange environment for Federated Learning systems. The discussed implementations are summarized in Table 1 below.

4.1.2 Combining HE with Other Security Technologies. Moreover, in [73], the authors proposed to secure the aggregation algorithms of FL by homomorphic encryption. Since they did not propose a new aggregation algorithm, their proposed solution is an alternative method using differential privacy, Homomorphic Encryption and Secure Multiparty Computation (SMC) to balance the tradeoff between accuracy and privacy. This combination allows for a reduction in the increase in noise injection as the number of participants increases, while maintaining a predefined trust rate. The concept of transaction assurance in the proposed FL solution lies in the role played by Differential Privacy and HE. The former is used by the participants to add a certain amount of noise, which is calculated based on various metrics. Then, the cryptosystem HE is used to encrypt the noisy message, which is then sent to the aggregator, which uses it to sum the global model. By combining the different security approaches, the authors presented their model as a scalable approach that defends against inference attacks while generating highly accurate models. These results were verified by several experiments that proved their model to be superior to the state-of- the art. However, the complexity of time computation was not considered in this study.

Similarly, the authors combined in [119] Homomorphic Encryption and Verifiable Computing (VC), a cryptographic method used to ensure the integrity of computations on authenticated data, to secure Federated Learning communications. The proposed model was tested with the Federated Extended EMNIST dataset [105]. However, the proposed solution was developed only for Neural Networks and compatibility with other types of Machine Learning models was not discussed.

On the other hand, in [120], the authors proposed an algorithm based on Blockchain Federated Learning. They secured

Table 2. HE, Combined with Other Solutions, to Secure FL Algorithms

Ref	Combined With	Dataset Used
[73]	Secure Multiparty Computation (SMC)	-
[119]	Verifiable Computing (VC)	EMNIST dataset[105]
[120]	Blockchain	-

the data exchange with differential privacy and Homomorphic Encryption. They applied different models, securing distributed Random Forest with differential privacy and distributed AdaBoost with HE, which provided multiple privacy in data and model sharing. Finally, they integrated the methods with Blockchain and Federated Learning and applied extensive experimental results that proved that their working mechanism had the better performance on the selected indicators. These implementations are summarized in Table 2.

4.1.3 HE with Reduced Communication and Computation Cost. In contrast, the increase in computing and communication costs due to HE was studied by the authors in [121]. In their study, they proposed BatchCrypt, a system solution for cross-silo Federated Learning. Their model was designed to secure communications in a FL system while reducing the overhead caused by HE. To achieve this, rather than encoding individual quantized gradients with full precision, they encoded a batch of quantized gradients into a long integer and encoded them all at once. They also developed new quantization and encryption strategies, as well as a unique gradient truncation mechanism to enable gradient-by-gradient aggregation of ciphertexts of encrypted batches. They then integrated BatchCrypt as a plugin module in FATE [122], a cross-silo industrial FL framework. Evaluations in geographically distributed data centers show that BatchCrypt achieves significant training acceleration, ranging from 23 to 93 times, while reducing communication costs from 66 to 101 times. Moreover, the loss of accuracy of the model due to quantization errors was less than 1%.
In addition, the authors in [123] Dubhe, a customizable, adaptable, and resilient FL fuse mechanism with low encryption and communication overhead. Dubhe improves training performance while posing no security risks by using homomorphic encryption. The authors evaluated their method on the EMNIST [105] and CIFAR [109] datasets, and the results showed that it outperformed other approaches in terms of unbiasedness. Similarly, the authors presented in [124] FLASHE, a HE scheme suitable for cross-silo FL that is able to capture the bare minimum of security and functionality by eliminating asymmetric key design and using only modular addition operations with random integers. They also evaluated their model against the EMNIST [105] and CIFAR [109] Datasets, and the results showed a 63-fold and 48-fold reduction in computational and communication costs, respectively. Similarly, the authors in [125] PFMLP, a security mechanism for Federated Learning that ensures that all FL individuals transmit their encrypted gradients through homomorphic encryption. They evaluated their model against the EMNIST dataset [105] and demonstrated a computational cost reduction of up to 25-28%.
In addition, a similar solution was also proposed in [126], where they proposed PCFL, a privacy-preserving and communication-efficient method for Federated Learning in the Internet of Things. PCFL consists of three key components: spatial sparsification with gradients, bidirectional compression, and a privacy-preserving protocol based on Homomorphic Encryption to protect data privacy and be resilient to various collusion scenarios. They evaluated their model with the EMNIST [105] and the results show that PCFL outperforms state-of-the-art methods by more than doubling the communication reduction while maintaining high model accuracy and slightly reducing the convergence rate. The above implementations are summarized in Table 3 below.

Table 3. HE with Communication and Computation Cost Reduction

Ref	Enhancement Ratio	Dataset Used
[121]	Training: 23x-93x - Communication: 66x-101x	EMNIST [105] & CIFAR [109]
[123]	negligible encryption and communication overhead	EMNIST [105] & CIFAR [109]
[120]	Computation:63x - Communication:48x	-
[125]	Computation: 25?28%	EMNIST [105]
[126]	Communication: 2x	EMNIST [105]

4.2 Other Approaches to Secure FL

Despite the feasibility that Homomorphic Encryption has shown in securing Federated Learning, other technologies have been used in this context. For example, in [127], the authors proposed to secure FL systems using the Covert Communication-based Federated Learning (CCFL) approach. Their method relies on the emerging communication security technique of covert communication, which disguises the existence of wireless communication activities. CCFL can reduce the ability of attackers to extract useful information from the Federated Learning Network training (FLN) training protocol, which is a crucial process in most existing attacks, and thus holistically improve FLN privacy. The authors extensively tested CCFL under real-world conditions, optimizing the latency of FL under certain security criteria.

In contrast, the authors of [128] advocated protecting FL frameworks from attackers by detecting and minimizing their influence on the model, especially in the context of bidirectional label flipping threats with cooperation. Exploiting correlations between local models, they presented "two graph-theoretic algorithms" based on the Minimum Spanning Tree and the k-Densest graph. Their method can minimize the impact of attackers even when they account for up to 70% of all FL individuals, whereas previous efforts could only allow 50% of these individuals to be attackers. Using experiments with the EMNIST dataset, the efficiency of the approach is [105].

Finally, several implementations were performed to secure FL algorithms using blockchain technology. Thus, in [129], [130], and [131], the authors proposed several blockchain solutions aimed at securing Federated Learning algorithms. FL Based on the results obtained in these studies, blockchain as a decentralized technology has demonstrated its ability to improve the performance of FL without the need for a centralized server and solve several problems and challenges, such as communication cost, disclosure of private information, the irregularity of uploading model parameters to the aggregator, and others.

5 DISCUSSING SECURITY IN FL AGGREGATION ALGORITHMS

Federated Learning technology is emerging as an efficient, robust, and viable Machine Learning technology while maintaining privacy. The interest in improving the aggregation algorithms and securing the data sharing mechanisms of Federated Learning has increasingly attracted the attention of researchers worldwide. Attempts to improve the security, privacy, and robustness of these aggregation algorithms may lead to greater confidence in this technology, which in turn will encourage the adoption of FL in various areas of life.

5.1 Securing Aggregation Algorithms

Federated Learning Aggregation algorithms are of great interest in terms of security and privacy, especially because they are vulnerable to poisoning attacks, inference attacks, and other breaches. Therefore, researchers are looking for

various ways to secure these algorithms, as described in detail in Section 3. However, it was found that most of the proposed algorithms focused either on robustness against attacks, especially the Byzantine attack, or on the security branches of trust, authentication, and integrity. Of the seven algorithms discussed, only SecureD- FL [76] addressed securing the FL aggregation algorithms against the inference attacks within the proposed communication model, and none of the algorithms proposed solutions for both class and membership inference. This conclusion sparks interest in working on new algorithms that can help solve these challenges.

The Table 4 below summarizes the various aggregation algorithms intended to solve security challenges. The roles played by these algorithms are presented in three categories: Security, Privacy, and Robustness, based on the taxonomy mentioned in Section 1.2.1.

Reasons for ignoring inference attacks were not discussed for these algorithms. As shown in the table below, four of the implementations discussed poisoning attacks, particularly the Byzantine attack, while neglecting the severity of inference attacks where at least some information about the individual's data can be extracted from the exchanged local model.

Moreover, none of these aggregation algorithms have considered the use of Homomorphic Encryption as a solution to security threats. HE has been widely considered as a solution in Federated Learning, but it has never been embedded in an aggregation algorithm. As mentioned earlier, the authors in [80], who proposed the aggregation algorithm SEAR, have confirmed that HE has a high communication and computation overhead on the one hand, and is limited for complex problems on the other hand. This problem is worth to be considered as a challenge and therefore solutions in this point can be feasible and efficient. Therefore, we can conclude with the following summary regarding security approaches in FL aggregation algorithms (in the lists below, the acronym RF is used as an abbreviation for a research finding):

- **RF1:** *There is a need to improve the **privacy** of aggregation algorithms by leveraging resistance to inference attacks, which can occur at the communication channel level or even through the central aggregation server itself.*

5.2 Securing Communication Among FL Individuals

The Federated Learning system usually consists of individuals and an aggregation server. In some studies, the aggregation algorithm has been moved to the individuals themselves, eliminating the need for a central server. Securing this communication is discussed in detail in Section 4. Homomorphic Encryption, Secure Multiparty Computation (SMC), Verifiable Computing, and Blockchain are examples of tools that have been considered for developing systems to secure communications in FL environments. Given the high communication and computational costs associated with using HE, several implementations have considered reducing these costs to improve performance and increase the usability of FL algorithms.

However, these proposed solutions presented their work as a security layer or as an add-on component to the FL system. In this context, the ability to generalize and adapt to different aggregation algorithms is not guaranteed. This is also evidenced by the fact that some, or rather most, of the proposed mechanisms were developed only for Neural Networks models. Therefore, they have not been tested with linear models such as Support Vector Machines (SVMs), which are of great interest in the Machine Learning world.

In addition, the reduction in communication and computational costs has been studied with certain datasets, and it is not known if the performance improvement occurs when other datasets are used, especially when the analyzed databases are heterogeneous.

Table 4. FL Aggregation Algorithms Oriented for Security Issues

Ref	Name	Summary	Security			Privacy			Robustness
			Confidence	Authentication	Integrity	Consent	Precision	Preserve	
[63]	-	Used Shamir's Secret Sharing [64] to split the secret into multiple shares	✓	✓	✓				
[66]	RFA	built on the principle of aggregation with the Geometric Median (GM) which is computed using Weiszfeld-type algorithm [67]							✓
[68]	LEGATO	uses a dynamic gradient reweighing approach that treats gradients based on layer-specific resilience							✓
[76]	SecureD-FL	the algorithm determines which set of participants (named group) should interact in each round of aggregation in order to minimize privacy leakage						✓	
[80]	SEAR	the proposed algorithm used hardware-based trusted execution environment instead of cryptographic time-computation consuming tools							✓
[86]	EPPDA	benefit from the homomorphisms of secret sharing [87] to minimize the secret sharing iterations and therefore reduce communication, calculation, and storage resources usage	✓	✓	✓				
[88]	HeteroSAg	the algorithms uses Segment Grouping Strategy that is based on dividing individuals into groups and segmenting local model updates for these users							✓
[89]	FLDetector	innovatively detects malicious clients, using model update consistency, bolstering security in Federated Learning							✓
[90]	FLCert	provides provable security against poisoning attacks, featuring ensemble learning with resource efficiency and privacy							✓
[91]	ELSA	revolutionizes secure aggregation, offering efficiency and resilience, surpassing other approaches in runtime							✓
[92]	Multi-RoundSecAgg	boosts FL privacy over multiple rounds			✓				

Furthermore, outsourcing encryption algorithms, communication control, and other security techniques may require adding a new individual to the FL system to control or manage these mechanisms, such as an encryption server or communication controller. This point in turn raises other debates about the integrity of these added individuals and their vulnerability to attacks or threats, which also require additional security.

Consequently, we can conclude the summary below regarding security mechanisms in FL systems:

- **RF2:** *Most security schemes are focused on **Neural Networks** and other types of ML models have rarely been considered, if at all*;
- **RF3:** *Compatibility of the schemas with different Machine Learning models is **not guaranteed***;
- **RF4:** *Communication/Computational cost has been evaluated with a limited number of datasets and this improvement is **not guaranteed** with other datasets, especially heterogeneous datasets which are a major concern in FL systems*;
- **RF5:** *Adding individuals to control security managers such as Encryption Managers or Communication Controllers will raises debate about the **security of those individuals** and their vulnerability to breaches and attacks.*

5.3 Unrevealed Secrets; Techniques Unseen in FL Security Yet

Federated Learning is on the rise. This is a fact that can be observed when analyzing the implementations carried out in the last few years. Although it is still in its infancy, the number of studies published in this field gives hope for a great development in the near future, which will promote and facilitate its application in different areas of life. However, in securing the FL system, various technologies have not been considered. They were not considered feasible or impractical in this field and, as far as we know, have not been used in any of the implementations of FL. Of these technologies, we mention Polymorphic Encryption "PE", which has been shown to be a viable technology for exchanging encrypted data with high confidence in privacy, as discussed in [132] and [133]. The PE technique can be used to protect the FL system against inference attacks that may occur at the level of communication between the individuals themselves or between the individuals and the aggregation server. However, the communication and computational costs as well as the resistance to poisoning attacks when using PE are interesting areas worth investigating further. Therefore, we can summarize this idea with the following research results:

- **RF6:** *There are other security technologies that have not been considered to secure the FL system,* ***Polymorphic Encryption*** *as an example.*

In Figure 7 below, the research findings, which are the results obtained after analyzing the privacy and security approaches in Federated Learning are presented.

Fig. 7. Research Findings for Analyzing Privacy and Security in Federated Learning

5.4 Future Perspectives

Given what has been said in this article and the discussion in this section, it can be concluded that Federated Learning has been extensively studied in terms of resilience to malicious individuals and robustness to poisoning attacks such as the Byzantine attack. However, there are a number of aspects that can be considered to improve the privacy of Federated Learning and increase its feasibility and usability. These aspects are summarized in the following list (the acronym FP used in the list below is short for future perspective):

- **FP1:** Consider applying privacy mechanisms to improve the robustness of existing or new FL aggregation algorithms against inference attacks;
- **FP2:** Generalize the security mechanisms to cover ML models other than Neural Networks such as SVMs and others;
- **FP3:** Ensure that the improvements in communication and computation time reduction achieved by the available studies are maintained for datasets other than those commonly used in these studies (mainly EMNIST and CIFAR);
- **FP4:** Demonstrate that third-party security managers (i.e., communication control unit, Homomorphic Encryption controller, etc.) are secure enough to be embedded in a Federated Learning system;
- **FP5:** Investigate embedding efficient security mechanisms, such as polymorphic encryption, into FL systems and compare their performance with available implementations.

CONCLUSION

Federated Learning is rapidly emerging as a potential technique to increase the confidence and adoption of Machine Learning in various aspects of life. The privacy and security techniques used in Federated Learning algorithms have been discussed in detail in this article. It has been shown that the available FL aggregation algorithms perform well in terms of security and robustness, but poorly in terms of privacy and resistance to inference attacks. In this context, security approaches such as homomorphic encryption, polymorphic encryption, and other tools can be considered as secure ways to improve the performance, privacy, and security of Federated Learning, increasing its use in real-world applications.

REFERENCES

[1] Ramkumar, P. N., Haeberle, H. S., Bloomfield, M. R., Schaffer, J. L., Kamath, A. F., Patterson, B. M., &Krebs, V. E. (2019). Artificial intelligence and arthroplasty at a single institution: real-world applications of Machine Learning to big data, value-based care, mobile health, and remote patient monitoring. The Journal ofarthroplasty, 34(10), 2204-2209.

[2] Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine Learning for medical imaging. Radiographics,37(2), 505.

[3] Bhardwaj, R., Nambiar, A. R., & Dutta, D. (2017, July). A study of Machine Learning in healthcare. In 2017IEEE 41st Annual Computer Software and Applications Conference (COMPSAC) (Vol. 2, pp. 236-241). IEEE.

[4] Ghazal, T. M., Hasan, M. K., Alshurideh, M. T., Alzoubi, H. M., Ahmad, M., Akbar, S. S., ... & Akour, I.A. (2021). IoT for smart cities: Machine Learning approaches in smart healthcare?A review. Future Internet,13(8), 218.

[5] Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A review of Machine Learning and IoT in smart transportation. Future Internet, 11(4), 94.

[6] Larra?naga, P., Atienza, D., Diaz-Rozo, J., Ogbechie, A., Puerto-Santana, C., & Bielza, C. (2018). Industrial applications of Machine Learning. CRC press.

[7] Sarker, I. H. (2021). Machine Learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 1-21.

[8] Sharma, N., Sharma, R., & Jindal, N. (2021). Machine Learning and deep learning applications-a vision. Global Transitions Proceedings, 2(1), 24-28.

[9] Nagarhalli, T. P., Vaze, V., & Rana, N. K. (2021, February). Impact of Machine Learning in natural language processing: A review. In 2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV) (pp. 1529-1534). IEEE.

[10] Pallathadka, H., Mustafa, M., Sanchez, D. T., Sajja, G. S., Gour, S., & Naved, M. (2021). Impact of Machine Learning on management, healthcare and agriculture. Materials Today: Proceedings.

[11] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in agriculture: Areview. Sensors, 18(8), 2674.

[12] Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., ... & Wang, C. (2018). Machine Learning and deep learning methods for cybersecurity. Ieee access, 6, 35365-35381.

[13] L?heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine Learning with big data:Challenges and approaches. Ieee Access, 5, 7776-7797.

[14] Paleyes, A., Urma, R. G., & Lawrence, N. D. (2020). Challenges in deploying Machine Learning: a survey ofcase studies. ACM Computing Surveys (CSUR).

[15] Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine Learning on big data: Opportunities andchallenges. Neurocomputing, 237, 350-361.

[16] Wuest, T., Weimer, D., Irgens, C., & Thoben, K. D. (2016). Machine Learning in manufacturing: advantages,challenges, and applications. Production & Manufacturing Research, 4(1), 23-45.

[17] Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. (2021). Machine Learning towards intelligent systems:applications, challenges, and opportunities. Artificial Intelligence Review, 54(5), 3299-3348.

[18] Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing Machine Learning in health care?addressingethical challenges. The New England journal of medicine, 378(11), 981.

[19] Albrecht, J. P. (2016). HowtheGDPRwill change the world. Eur. Data Prot. L. Rev., 2, 287.

[20] Parasol, M. (2018). The impact of China?s 2016 Cyber Security Law on foreign technology firms, and onChina?s big data and Smart City dreams. Computer law & security review, 34(1), 67-98.

[21] Gray,W., & Zheng, H. R. (1986). General Principles of Civil Law of the People?s Republic of China. TheAmerican Journal of Comparative Law, 34(4), 715-743.

[22] Zhang, C., Xie, Y., Bai, H., Yu, B., Li,W., & Gao, Y. (2021). A survey on Federated Learning. Knowledge-BasedSystems, 216, 106775

[23] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficientlearning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282).PMLR.

[24] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated Learning: Challenges, methods, and futuredirections. IEEE Signal Processing Magazine, 37(3), 50-60.

[25] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advancesand open problems in Federated Learning. Foundations and Trends in Machine Learning, 14(1?2), 1-210.

[26] Ding, J., Tramel, E., Sahu, A. K., Wu, S., Avestimehr, S., & Zhang, T. (2022, May). Federated Learningchallenges and opportunities: An outlook. In ICASSP 2022-2022 IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP) (pp. 8752-8756). IEEE.

[27] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and applications. ACMTransactions on Intelligent Systems and Technology (TIST), 10(2), 1-19.

[28] Rahman, K. J., Ahmed, F., Akhter, N., Hasan, M., Amin, R., Aziz, K. E., ... & Islam, A. N. (2021). Challenges,applications and design aspects of Federated Learning: A survey. IEEE Access, 9, 124682-124700.

[29] Lyu, L., Yu, H., & Yang, Q. (2020). Threats to Federated Learning: A survey. arXiv preprint arXiv:2003.02133.

[30] Bambauer, D. E. (2013). Privacy versus security. J. Crim. L. & Criminology, 103, 667.

[31] Acquisti, A. (2004). Privacy and security of personal information. In Economics of Information Security (pp.179-186). Springer, Boston, MA.

[32] Regan, P. M. (2002). Privacy as a common good in the digital world. Information, Communication & Society,5(3), 382-405.

[33] Kernighan, B. W. (2021). Understanding the digital world: What you need to know about computers, theinternet, privacy, and security. Princeton University Press.

[34] Bhowmick, A., Duchi, J., Freudiger, J., Kapoor, G., & Rogers, R. (2018). Protection against reconstructionand its applications in private Federated Learning. arXiv preprint arXiv:1812.00984.

[35] Fredrikson, M., Jha, S., & Ristenpart, T. (2015, October). Model inversion attacks that exploit confidenceinformation and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC conference on computerand communications security (pp. 1322-1333).

[36] Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019, May). Exploiting unintended feature leakage incollaborative learning. In 2019 IEEE symposium on security and privacy (SP) (pp. 691-706). IEEE.

[37] Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. Advances in neural information processingsystems, 32.

[38] McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2017). Learning differentially private recurrentlanguage models. arXiv preprint arXiv:1710.06963.

[39] Agarwal, N., Suresh, A. T., Yu, F. X. X., Kumar, S., & McMahan, B. (2018). cpSGD: Communication-efficientand differentially-private distributed SGD. Advances in Neural Information Processing Systems, 31.

[40] Aono, Y., Hayashi, T., Wang, L., & Moriai, S. (2017). Privacy-preserving deep learning via additively homomorphicencryption. IEEE Transactions on Information Forensics and Security, 13(5), 1333-1345.

[41] Bhagoji, A. N., Chakraborty, S., Mittal, P., & Calo, S. (2019, May). Analyzing Federated Learning through anadversarial lens. In International Conference on Machine Learning (pp. 634-643). PMLR.

[42] Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020, June). How to backdoor federatedlearning. In International Conference on Artificial Intelligence and Statistics (pp. 2938-2948). PMLR.

[43] Fung, C., Yoon, C. J., & Beschastnikh, I. (2018). Mitigating sybils in Federated Learning poisoning. arXivpreprint arXiv:1808.04866.

[44] Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). Machine Learning with adversaries:Byzantine tolerant gradient descent. Advances in Neural Information Processing Systems, 30.

[45] Chen, Y., Su, L., & Xu, J. (2017). Distributed statistical Machine Learning in adversarial settings: Byzantinegradient descent. Proceedings of the ACM on Measurement and Analysis of Computing Systems, 1(2), 1-25.

[46] Chen, L., Wang, H., Charles, Z., & Papailiopoulos, D. (2018, July). Draco: Byzantine-resilient distributedtraining via redundant gradients. In International Conference on Machine Learning (pp. 903-912). PMLR.

[47] Yin, D., Chen, Y., Kannan, R., & Bartlett, P. (2018, July). Byzantine-robust distributed learning: Towardsoptimal statistical rates. In International Conference on Machine Learning (pp. 5650-5659). PMLR.

[48] Lamport, L., Shostak, R., & Pease, M. (2019). The Byzantine generals problem. In Concurrency: the worksof leslie lamport (pp. 203-226).

[49] Xie, C., Koyejo, O., & Gupta, I. (2018). Generalized byzantine-tolerant sgd. arXiv preprint arXiv:1802.10116.
[50] Xie, C., Koyejo, O., & Gupta, I. (2020, August). Fall of empires: Breaking byzantine-tolerant sgd by innerproduct manipulation. In Uncertainty in Artificial Intelligence (pp. 261-270). PMLR.
[51] Biggio, B., Nelson, B., & Laskov, P. (2011, November). Support vector machines under adversarial label noise.In Asian conference on Machine Learning (pp. 97-112). PMLR.
[52] Barreno, M., Nelson, B., Sears, R., Joseph, A. D., & Tygar, J. D. (2006, March). Can Machine Learning besecure?. In Proceedings of the 2006 ACM Symposium on Information, computer and communications security(pp. 16-25).
[53] Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I., & Tygar, J. D. (2011, October). Adversarial machinelearning. In Proceedings of the 4th ACM workshop on Security and artificial intelligence (pp. 43-58).
[54] Shafahi, A., Huang, W. R., Najibi, M., Suciu, O., Studer, C., Dumitras, T., & Goldstein, T. (2018). Poisonfrogs! targeted clean-label poisoning attacks on neural networks. Advances in neural information processingsystems, 31.
[55] Gu, T., Dolan-Gavitt, B., & Garg, S. (2017). Badnets: Identifying vulnerabilities in the Machine Learningmodel supply chain. arXiv preprint arXiv:1708.06733.
[56] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguingproperties of neural networks. arXiv preprint arXiv:1312.6199.
[57] Su, L., & Xu, J. (2019). Securing distributed gradient descent in high dimensional statistical learning. Proceedingsof the ACM on Measurement and Analysis of Computing Systems, 3(1), 1-41.
[58] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks againstMachine Learning models. In 2017 IEEE symposium on security and privacy (SP) (pp. 3-18). IEEE.
[59] Hamer, J., Mohri, M., & Suresh, A. T. (2020, November). Fedboost: A communication-efficient algorithm forFederated Learning. In International Conference on Machine Learning (pp. 3973-3983). PMLR.
[60] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization inheterogeneous networks. Proceedings of Machine Learning and Systems, 2, 429-450.
[61] Wang, H., Yurochkin, M., Sun, Y., Papailiopoulos, D., & Khazaeni, Y. (2020). Federated Learning with matchedaveraging. arXiv preprint arXiv:2002.06440.
[62] Sannara, E. K., Portet, F., Lalanda, P., & German, V. E. G. A. (2021, March). A Federated Learning aggregationalgorithm for pervasive computing: Evaluation and comparison. In 2021 IEEE International Conference onPervasive Computing and Communications (PerCom) (pp. 1-10). IEEE.M., & Harchaoui, Z. (2019). Robust aggregation for Federated Learning. arXiv preprintarXiv:1912.13445.
[63] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017, October). Practical secure aggregation for privacy-preserving Machine Learning. In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175-1191).
[64] Shamir, A. (1979). How to share a secret. Communications of the ACM, 22(11), 612-613.
[65] Yu, H., Wang, Z., Li, J., & Gao, X. (2018). Identity-based proxy signcryption protocol with universal composability. Security and Communication Networks, 2018.
[66] Pillutla, K., Kakade, S. M., & Harchaoui, Z. (2019). Robust aggregation for Federated Learning. arXiv preprint arXiv:1912.13445.
[67] Weiszfeld, E., & Plastria, F. (2009). On the point for which the sum of the distances to n given points isminimum. Annals of Operations Research, 167(1), 7-41.
[68] Varma, K., Zhou, Y., Baracaldo, N., & Anwar, A. (2021, September). LEGATO: A LayerwisE Gradient AggregaTiOnAlgorithm for Mitigating Byzantine Attacks in Federated Learning. In 2021 IEEE 14th InternationalConference on Cloud Computing (CLOUD) (pp. 272-277). IEEE.
[69] Chen, V., Pastro, V., & Raykova, M. (2019). Secure computation for Machine Learning with SPDZ. arXivpreprint arXiv:1901.00329.
[70] Agrawal, N., Shahin Shamsabadi, A., Kusner, M. J., & Gascon, A. (2019, November). QUOTIENT: twopartysecure neural network training and prediction. In Proceedings of the 2019 ACM SIGSAC Conference onComputer and Communications Security (pp. 1231-1247).
[71] Shokri, R., & Shmatikov, V. (2015, October). Privacy-preserving deep learning. In Proceedings of the 22ndACM SIGSAC conference on computer and communications security (pp. 1310-1321).
[72] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October).Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer andcommunications security (pp. 308-318).
[73] Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., & Zhou, Y. (2019, November).A hybrid approach to privacy-preserving Federated Learning. In Proceedings of the 12th ACM workshop onartificial intelligence and security (pp. 1-11).
[74] Xu, R., Baracaldo, N., Zhou, Y., Anwar, A., & Ludwig, H. (2019, November). Hybridalpha: An efficientapproach for privacy-preserving Federated Learning. In Proceedings of the 12th ACM Workshop on ArtificialIntelligence and Security (pp. 13-23).
[75] Ryffel, T., Trask, A., Dahl, M., Wagner, B., Mancuso, J., Rueckert, D., & Passerat-Palmbach, J. (2018). Ageneric framework for privacy preserving deep learning. arXiv preprint arXiv:1811.04017.
[76] Jeon, B., Ferdous, S. M., Rahman, M. R., & Walid, A. (2021, May). Privacy-preserving decentralized aggregationfor Federated Learning. In IEEE INFOCOM 2021-IEEE Conference on Computer CommunicationsWorkshops (INFOCOM WKSHPS) (pp. 1-6). IEEE.
[77] Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statisticallearning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning,3(1), 1-122.

[78] Stinson, D. R. (2008). Combinatorial designs: constructions and analysis. ACM SIGACT News, 39(4), 17-21.
[79] Hitaj, B., Ateniese, G., & Perez-Cruz, F. (2017, October). Deep models under the GAN: information leakagefrom collaborative deep learning. In Proceedings of the 2017 ACM SIGSAC conference on computer andcommunications security (pp. 603-618).
[80] Zhao, L., Jiang, J., Feng, B., Wang, Q., Shen, C., & Li, Q. (2021). Sear: Secure and efficient aggregationfor byzantine-robust Federated Learning. IEEE Transactions on Dependable and Secure Computing, 19(5),3329-3342.
[81] McKeen, F., Alexandrovich, I., Berenzon, A., Rozas, C. V., Shafi, H., Shanbhogue, V., & Savagaonkar, U. R.(2013). Innovative instructions and software model for isolated execution. Hasp@ isca, 10(1).
[82] Li, W., Xia, Y., & Chen, H. (2019). Research on arm trustzone. GetMobile: Mobile Computing and Communications,22(3), 17-22.
[83] Brasser, F., Muller, U., Dmitrienko, A., Kostiainen, K., Capkun, S., & Sadeghi, A. R. (2017). Software grandexposure:SGX cache attacks are practical. In 11th USENIX Workshop on Offensive Technologies (WOOT 17).
[84] Moghimi, A., Irazoqui, G., & Eisenbarth, T. (2017, September). Cachezoom: How SGX amplifies the power ofcache attacks. In International Conference on Cryptographic Hardware and Embedded Systems (pp. 69-90).Springer, Cham.
[85] Schwarz, M., Weiser, S., Gruss, D., Maurice, C., & Mangard, S. (2017, July). Malware guard extension:Using SGX to conceal cache attacks. In International Conference on Detection of Intrusions and Malware, andVulnerability Assessment (pp. 3-24). Springer, Cham.
[86] Song, J., Wang, W., Gadekallu, T. R., Cao, J., & Liu, Y. (2022). Eppda: An efficient privacy-preserving dataaggregation Federated Learning scheme. IEEE Transactions on Network Science and Engineering.
[87] Benaloh, J. C. (1986, August). Secret sharing homomorphisms: Keeping shares of a secret secret. In Conferenceon the theory and application of cryptographic techniques (pp. 251-260). Springer, Berlin, Heidelberg.
[88] Elkordy, A. R., & Avestimehr, A. S. (2022). Heterosag: Secure aggregation with heterogeneous quantizationin Federated Learning. IEEE Transactions on Communications, 70(4), 2372-2386.
[89] Zhang, Zaixi, Xiaoyu Cao, Jinyuan Jia, and Neil Zhenqiang Gong. "FLDetector: Defending Federated Learning against model poisoning attacks via detecting malicious clients." In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2545-2555. 2022.
[90] Cao, Xiaoyu, Zaixi Zhang, Jinyuan Jia, and Neil Zhenqiang Gong. "Flcert: Provably secure Federated Learning against poisoning attacks." IEEE Transactions on Information Forensics and Security 17 (2022): 3691-3705.
[91] Rathee, Mayank, Conghao Shen, Sameer Wagh, and Raluca Ada Popa. "Elsa: Secure aggregation for Federated Learning with malicious actors." In 2023 IEEE Symposium on Security and Privacy (SP), pp. 1961-1979. IEEE, 2023.
[92] So, Jinhyun, Ramy E. Ali, Ba?ak Guler, Jiantao Jiao, and A. Salman Avestimehr. "Securing secure aggregation: Mitigating multi-round privacy leakage in Federated Learning." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 8, pp. 9864-9873. 2023.
[93] Paillier, P. (1999, May). Public-key cryptosystems based on composite degree residuosity classes. In Internationalconference on the theory and applications of cryptographic techniques (pp. 223-238). Springer, Berlin,Heidelberg.
[94] Yi, X., Paulet, R., & Bertino, E. (2014). Homomorphic encryption. In Homomorphic encryption and applications(pp. 27-46). Springer, Cham.
[95] Rivest, R. L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-keycryptosystems. Communications of the ACM, 21(2), 120-126.
[96] Rothblum, R. (2011, March). Homomorphic encryption: From private-key to public-key. In Theory of cryptographyconference (pp. 219-234). Springer, Berlin, Heidelberg.
[97] Li, B., & Micciancio, D. (2021, October). On the security of homomorphic encryption on approximate numbers.In Annual International Conference on the Theory and Applications of Cryptographic Techniques (pp. 648-677). Springer, Cham.
[98] Fontaine, C., & Galand, F. (2007). A survey of homomorphic encryption for nonspecialists. EURASIP Journalon Information Security, 2007, 1-10.
[99] Hardy, S., Henecka, W., Ivey-Law, H., Nock, R., Patrini, G., Smith, G., & Thorne, B. (2017). Private federatedlearning on vertically partitioned data via entity resolution and additively homomorphic encryption. arXivpreprint arXiv:1711.10677.
[100] Ou, W., Zeng, J., Guo, Z., Yan, W., Liu, D., & Fuentes, S. (2020). A homomorphic-encryption-based verticalFederated Learning scheme for rick management. Computer Science and Information Systems, 17(3), 819-834.
[101] Sav, S., Pyrgelis, A., Troncoso-Pastoriza, J. R., Froelicher, D., Bossuat, J. P., Sousa, J. S., & Hubaux, J. P.(2020). POSEIDON: Privacy-preserving federated neural network learning. arXiv preprint arXiv:2009.00349.
[102] Froelicher, D., Troncoso-Pastoriza, J. R., Pyrgelis, A., Sav, S., Sousa, J. S., Bossuat, J. P., & Hubaux, J.P. (2021). Scalable privacy-preserving distributed learning. Proceedings on Privacy Enhancing Technologies,2021(2), 323-347.
[103] Mouchet, C., Troncoso-Pastoriza, J., Bossuat, J. P., & Hubaux, J. P. (2020). Multiparty homomorphic encryptionfrom ring-learning-with-errors. Cryptology ePrint Archive.
[104] UCI Machine Learning Repository: Data Set. Retrieved October 10, 2022, fromhttps://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original).
[105] Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). EMNIST: an extension of MNIST to handwrittenletters
[106] UCI Machine Learning Repository: Epileptic Seizure Recognition Data Set. Retrieved October 1, 2022, fromhttps://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition
[107] Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy ofprobability of default of credit card clients. Expert systems with applications, 36(2), 2473-2480.
[108] Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, A. Y. (2011). Reading digits in natural imageswith unsupervised feature learning.
[109] Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images.

[110] Liu, X., Li, H., Xu, G., Chen, Z., Huang, X., & Lu, R. (2021). Privacy-enhanced Federated Learning againstpoisoning adversaries. IEEE Transactions on Information Forensics and Security, 16, 4574-4588.

[111] Tian, H., Zhang, F., Shao, Y., & Li, B. (2021). Secure linear aggregation using decentralized thresholdadditive homomorphic encryption for Federated Learning. arXiv preprint arXiv:2111.10753.

[112] Stripelis, D., Saleem, H., Ghai, T., Dhinagar, N., Gupta, U., Anastasiou, C., ... & Ambite, J. L. (2021,December). Secure neuroimaging analysis using Federated Learning with homomorphic encryption. In 17thInternational Symposium on Medical Information Processing and Analysis (Vol. 12088, pp. 351-359). SPIE.

[113] Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., ... & Smith, S. M.(2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Natureneuroscience, 19(11), 1523-1536.

[114] Zhang, L., Xu, J., Vijayakumar, P., Sharma, P. K., & Ghosh, U. (2022). Homomorphic Encryption-basedPrivacy-preserving Federated Learning in IoT-enabled Healthcare System. IEEE Transactions on NetworkScience and Engineering.

[115] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-sourcedermatoscopic images of common pigmented skin lesions. Scientific data, 5(1), 1-9.

[116] Fan, C. I., Hsu, Y. W., Shie, C. H., & Tseng, Y. F. (2022). ID-Based Multi-Receiver Homomorphic ProxyRe-Encryption in Federated Learning. ACM Transactions on Sensor Networks (TOSN).

[117] Ku, H., Susilo, W., Zhang, Y., Liu, W., & Zhang, M. (2022). Privacy-Preserving Federated Learning in medicaldiagnosis with homomorphic re-Encryption. Computer Standards & Interfaces, 80, 103583.

[118] Park, J., & Lim, H. (2022). Privacy-Preserving Federated Learning Using Homomorphic Encryption. AppliedSciences, 12(2), 734.

[119] Madi, A., Stan, O., Mayoue, A., Grivet-Sebert, A., Gouy-Pailler, C., & Sirdey, R. (2021, May). A secureFederated Learning framework using homomorphic encryption and verifiable computing. In 2021 Reconciling Data Analytics, Automation, Privacy, and Security: A Big Data Challenge (RDAAPS) (pp. 1-8). IEEE.

[120] Jia, B., Zhang, X., Liu, J., Zhang, Y., Huang, K., & Liang, Y. (2021). Blockchain-enabled Federated Learningdata protection aggregation scheme with differential privacy and homomorphic encryption in IIoT. IEEETransactions on Industrial Informatics, 18(6), 4049-4058.

[121] Zhang, C., Li, S., Xia, J., Wang, W., Yan, F., & Liu, Y. (2020). BatchCrypt: Efficient homomorphicencryption for Cross-Silo Federated Learning. In 2020 USENIX annual technical conference (USENIX ATC 20)(pp. 493-506).

[122] FATE. (2019, October 31). Retrieved October 15, 2022, from https://fate.fedai.org/

[123] Zhang, S., Li, Z., Chen, Q., Zheng, W., Leng, J., & Guo, M. (2021, August). Dubhe: Towards data unbiasednesswith homomorphic encryption in Federated Learning client selection. In 50th International Conference onParallel Processing (pp. 1-10).

[124] Jiang, Z., Wang, W., & Liu, Y. (2021). Flashe: Additively symmetric homomorphic encryption for cross-siloFederated Learning. arXiv preprint arXiv:2109.00675.

[125] Fang, H., & Qian, Q. (2021). Privacy preserving Machine Learning with homomorphic encryption and federatedlearning. Future Internet, 13(4), 94.

[126] Fang, C., Guo, Y., Hu, Y., Ma, B., Feng, L., & Yin, A. (2021). Privacy-preserving and communication-efficientFederated Learning in Internet of Things. Computers & Security, 103, 102199.

[127] Xie, Y. A., Kang, J., Niyato, D., Van, N. T. T., Luong, N. C., Liu, Z., & Yu, H. (2022). Securing federatedlearning: A covert communication-based approach. IEEE Network.

[128] Ranjan, P., Gupta, A., Cor'o, F., & Das, S. K. (2022). Securing Federated Learning against OverwhelmingCollusive Attackers. arXiv preprint arXiv:2209.14093.

[129] Li, Z., Yu, H., Zhou, T., Luo, L., Fan, M., Xu, Z., & Sun, G. (2021). Byzantine resistant secure blockchainedFederated Learning at the edge. Ieee Network, 35(4), 295-301.

[130] Yuan, S., Cao, B., Peng, M., & Sun, Y. (2021, March). ChainsFL: Blockchain-driven Federated Learningfrom Design to Realization. In 2021 IEEE Wireless Communications and Networking Conference (WCNC)(pp. 1-6). IEEE.

[131] Li, Y., Chen, C., Liu, N., Huang, H., Zheng, Z., & Yan, Q. (2020). A blockchain-based decentralized federatedlearning framework with committee consensus. IEEE Network, 35(1), 234-241.

[132] Rajput, A. S., & Raman, B. (2021). Privacy-Preserving Distribution and Access Control of PersonalizedHealthcare Data. IEEE Transactions on Industrial Informatics, 18(8), 5584-5591.

[133] Booher, D. D., Cambou, B., Carlson, A. H., & Philabaum, C. (2019, January). Dynamic key generation forpolymorphic encryption. In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference(CCWC) (pp. 0482-0487). IEEE.26

Received 10 March 2024; revised 10 March 2024; accepted 10 March 2024

CHAPTER 5

EMBEDDING HOMOMORPHIC & POLYMORPHIC ENCRYPTION in FL

Résumé: Cette recherche propose de nouveaux frameworks qui utilisent des techniques de chiffrement polymorphe et homomorphe pour sécuriser les environnements FL contre différentes attaques malveillantes, en particulier les attaques par inférence. Ces frameworks permettent d'entraîner, en toute sécurité, cinq modèles d'apprentissage machine intelligents. PolyFLAG_SMV, PolyFLAM, PolyFLAP, HP_FLAP ont été créés en raison du besoin urgent d'améliorer les protocoles de sécurité dans les frameworks d'apprentissage fédéré contre les attaques par inférence, ainsi que de la possibilité pratique de combiner le chiffrement homomorphe et polymorphe comme réponse à ce problème. Cette nouvelle idée combine astucieusement les idées fondamentales des techniques de chiffrement récentes dans sa conception. Cela crée une barrière de défense impénétrable autour du FL, garantissant une sécurité totale. Cette section détaille à la fois la base conceptuelle et la conception détaillée du frameworks proposé. Cela donne une image complète des nombreuses parties de la stratégie, qui sont toutes destinées à renforcer l'intégrité de l'environnement FL. Ce chapitre explore en profondeur les frameworks décrits dans cette thèse, en couvrant la motivation du problème, les concepts préliminaires, les explications détaillées de quatre nouveaux frameworks, les évaluations complètes par rapport aux approches actuelles, et les applications pratiques validées par des données du monde réel et des études de cas d'utilisation.

PolyFLAG_SVM: A Polymorphic Federated Learning Aggregation of Gradients Support Vector Machines Framework

Published in Procedia Computer Science Journal 2023
Volume 224; Pages 139-146; doi: 10.1016/j.procs.2023.09.021

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 224 (2023) 139–146

Procedia Computer Science

www.elsevier.com/locate/procedia

The 20th International Conference on Mobile Systems and Pervasive Computing (MobiSPC)
July 24-26, 2023, Halifax, Nova Scotia, Canada

PolyFLAG_SVM: a Polymorphic Federated Learning Aggregation of Gradients Support Vector Machines Framework

Mohammad Moshawrab[a,c,*], Mehdi Adda[a], Abdenour Bouzouane[b], Hussein Ibrahim[c], Ali Raad[d]

[a]*Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, G5L 3A1, Québec, Canada*

[b]*Département D'informatique et de Mathématique, Université du Québec à Chicoutimi, Chicoutimi, 555 boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada*

[c]*Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, G4R 5B7, Québec, Canada*

[d]*Dean of the Faculty of Science and Arts, Islamic University of Lebanon, Wardaniyeh, Lebanon*

Abstract

The critical importance of user privacy in the context of Machine Learning is a hot research topic because it hinders data collection. Consequently, Federated Learning (FL) has emerged as a solution to the privacy problem. Instead of collecting users' data to train smart models, FL exchanges models with clients, which are trained and sent back to the server for aggregation and global model updating. However, FL still faces some hurdles, such as vulnerability to inference and poisoning attacks. For this reason, this paper proposes **PolyFLAG_SVM**: Polymorphic Federated Learning Aggregation of Gradients Support Vector Machines Framework. **PolyFLAG_SVM** is a novel, secure, communication-efficient framework that provides several variants of Support Vector Machines models that follow the Gradient Descent Update technique. The confidence in the security of the proposed model is the result of the polymorphism of the encryption keys used, which guarantees that a cracked or leaked key is useless, since it is not used twice within the FL cycle. Moreover, the proposed framework is communication efficient due to the small size of messages exchanged between servers and clients. The proposed model is explained in detail and evaluated appropriately in this paper.

Keywords: Polymorphic Encryption; Federated Machine Learning; Federated Learning; Privacy-Preserving; Security-Enhanced; Polymorphism

1. Introduction

Machine learning (ML) methods have been widely adopted due to their efficacy in data analysis and pattern extraction. The ability of ML algorithms for processing massive amounts of data in different formats has also improved, contributing to this achievement [1]. As a result, ML has been adopted into different domains of humans daily lives, not limited to healthcare, finance, industry, education, agriculture and plenty other domains [2].

* Corresponding author; Tel.: +1-581-624-9394
E-mail address: mohammad.moshawrab@uqar.ca

10.1016/j.procs.2023.09.021

1.1. Challenges in Machine Learning Domain

Despite the success of ML tools and techniques in various fields, the technology itself is still vulnerable to various challenges [3]. These may be data-related, model-related, or general challenges. However, these challenges are not weighted, and none has been introduced as the most critical. However, ML usually follows a known workflow that can be described as: data collection and preprocessing, model selection, model training, evaluation, fine-tuning, and finally production and deployment. In this context, various regulations are introduced to manage user data collection, such as the European Union General Data Protection Regulation (GDPR) [4], China's Cyber Security Law of the People's Republic of China [5], and others. Consequently, the restriction of data collection, whether due to regulations and laws or user lack of engagement, definitely affects the entire workflow. Moreover, it is common among ML researchers that the ability to collect more data helps to increase the accuracy of the trained smart models. On the other hand, the data needed to train a ML model in the real world may not be in the same physical or even virtual locations at the same time. For example, to predict Cardiovascular Diseases, it would be more useful to simultaneously analyze medical images, Electronic Health Records (EHRs) and continuous vital signs. However, these data sets may not be available in the same health institute, or not accessible simultaneously due to privacy restrictions, making training ML model with this collection an impossible task. This phenomenon is known in the field ML as "data islands."

1.2. Federated Learning: A Privacy Issue Solution

To address the privacy issue, Google proposed Federated Learning, a privacy-preserving technology that trains smart models without compromising user privacy [6]. FL is a collaborative, decentralized ML technology that enables training of ML models without transmitting user data to a central server. Instead, models are sent from the server to clients who train the models locally and send them back to the server for aggregation to update the global model. This process is repeated until the global model converges [3, 7, 8]. FL allows models to be trained, preserves data privacy, and even allows smart models to access more data while maintaining privacy.

2. Problem Statement: Security Threats in FL Domain

Federated Learning achieves privacy by decentralizing ML and reducing data transfer between clients and server. This method, additionally, lowers transmission costs as raw data is typically larger than transmitted models. Google confirmed that data sent to FL servers was reduced by 99.6% compared to a centralized ML environment [9] .

2.1. FL under the Scope: Challenges and Issues

FL has been successful, but prone to several problems, which have been studied extensively in the literature. For example, studies [3, 10, 11] found issues with first FL aggregation algorithm, FedAvg [6], such as data and hardware heterogeneity, sensitivity to local model, scalability, slow convergence, complexity, communication costs, and vulnerability to malicious attacks. These challenges have driven researchers to find solutions to improve FL's feasibility and usability.

2.2. Security in FL Domain

FL is vulnerable to malicious attacks, despite being a privacy-preserving ML technology [3, 10, 11]. Messages exchanged in FL are vulnerable to attacks at three levels: input, learning process, and the learned model. These attacks include poisoning, inference, and backdoor attacks. Poisoning attacks can compromise learning quality, inference attacks expose users' private data, and backdoor attacks grant unauthorized access to the FL system [12].

2.3. Securing FL Frameworks: State of the Art

Researchers are interested in improving the security of FL to promote their usability and feasibility. Several attempts have been made in this regard. For example, in [13], the authors proposed a secure vector summation strategy using a protocol with a fixed number of rounds that reduces computational cost and is robust to faulty clients. In their approach, only a single server can be trusted to hold the exchanged data. Their proposed framework showed high security against honest but curious adversaries, and it also guarantees anonymity even when faced with active adversaries, like a hostile server. Moreover, in [14], the authors proposed Robust Federated Aggregation (RFA) that aimed to protect the FL aggregation process against poisoning attacks. To achieve their goal, they aggregated the exchanged models based on the geometric median, which can be computed using a Weiszfeld-type algorithm [15]. RFA was able to compete with the traditional FedAvg algorithm and was more resistant to data poisoning attacks.
On the other hand, the authors in [16] developed SecureD-FL, which is a FL framework that is based on a refined

form of the Alternating Direction Multiplier (ADMM) [17]. Their proposed framework uses a communication mode where the algorithm decides in each round of execution which subset of users should exchange data to minimize the disclosure of private data during the aggregation process. In addition, the authors of [18] proposed SEAR, a Secure and Efficient Aggregation for byzantine-robust federated learning that aggregates the local models in a secure and trusted hardware environment, specifically Intel SGX Trustworthy Execution Environment (TEE) [19], which is a secure CPU area, where the data and programs being executed are kept secret and cannot be modified. Moreover, in [20], the authors proposed the Efficient Privacy-Preserving Data Aggregation (EPPDA), which uses homomorphisms of secret exchange[21] in the FL environment. Their algorithm is secured, and reduce the influence of some malicious clients. The cryptographic primitives used in their approach can be summarized as follows: secret sharing, key exchange protocol, authenticated encryption and signature methods.
Finally, in [22], the authors proposed the aggregation algorithm HeteroSAg, which uses masking to secure the exchanged messages such that the mutual information between the masked model and the unique model is zero. The resilience of HeteroSAg against Byzantine attacks depends on the FL cycle, which implements a segment grouping strategy based on dividing edge users into groups and segmenting local model updates for those users. The security approaches followed in the state of the art of secured FL frameworks are summarized in Table 1 below.

Table 1: State-of-the-art of secured FL aggregation algorithms

Ref#	Mechanism
[13]	Secure Vector Summing Strategy
[14]	Using geometric median estimated using a Weiszfeld-type algorithm
[16]	Refined form of the Alternating Direction Multiplier (ADMM)
[18]	Hardware-based trusted execution environment instead of complex cryptographic tools
[20]	Homomorphisms of the secret exchange
[22]	Masking each user's model update

2.4. Problem and Motivation

Poisoning attacks have been extensively studied in secure FL aggregation algorithms, while inference attacks have not received as much attention. Techniques such as Polymorphic Encryption (PE) [23] have been shown to be effective in reducing the impact of inference attacks by securing data exchanges, but have not been used in previous studies. The criticality of inference attacks and the feasibility of PE inspired the proposed framework, **PolyFLAG_SVM**. This framework is the first to use PE for securing FL aggregation frameworks, which makes it novel and unique.

2.5. Preliminaries: Polymorphism and Gradient Descent Update

Polymorphic Encryption can serve as a viable protection technology against inference attacks. In addition, FL has several alternatives for exchanging models, such as exchanging model parameters or gradients. In this section, both polymorphism and gradients are explained to show their subsequent use in the proposed model.

2.5.1. Polymorphic Encryption

Polymorphism is the ability of an object or function to take on multiple forms or behaviors. Encryption, on the other hand, is the process of converting normal data into unreadable form to prevent unauthorized access or use. Commonly known encryption algorithms include AES (Advanced Encryption Standard) [24], RSA (Rivest-Shamir-Adleman) and others [25]. Consequently, PE can be defined as an encryption scheme that changes the algorithm or encryption keys to increase security. In PE, it is difficult for attackers to break the encryption even if they get hold of the ciphertext, unlike traditional encryption methods with fixed algorithms and keys.

2.5.2. Gradient Descent Update

Gradient Descent is a ML optimization algorithm that iteratively changes the parameters of a trained model to determine the minimum of a cost function. The algorithm computes the gradient of the cost function with respect to the parameters and updates it in the direction of steepest descent, controlled by a learning rate. Variants of gradient descent include the batch, stochastic, and mini-batch methods, each of which has its own strengths and weaknesses [26]. In FL, the amount of data exchanged between server and clients is significantly reduced when exchanging gradients instead of models, resulting in greater scalability and efficiency.

3. Our Approach: PolyFLAG_SVM

The need to improve the security of FL frameworks against inference attacks and along with the feasibility of Polymorphic Encryption in this regard motivated the proposal of **PolyFLAG_SVM**, which embeds PE in its structure to secure the FL environment. The concept and design of the framework are explained in this section.

3.1. PolyFLAG_SVM: Main Concept

A typical FL framework includes a central server and clients, where the server sends a global model to the clients, who train it using their local data and send the updated model back to the server. The server combines the received models to create an updated global model, and repeats the process until convergence. In **PolyFLAG_SVM**, the exchanged messages are polymorphically encrypted using the AES-256 algorithm [24] to ensure that the exchanged data is protected and secured. The polymorphism in the proposed framework is created by using different encryption keys to encrypt and decrypt each message exchanged between the server and the clients, even using different keys for each message exchanged between the server and a single client. In this context, the two main concepts that generate polymorphism are the table of encryption keys (ToKs) and the initial encryption key, which are explained below.

3.1.1. Table of Encryption Keys (ToKs)

In **PolyFLAG_SVM**, when the client sends the first connection request, the server replies with the table of encryption keys, each indexed with a unique ID. These keys are later used to encrypt the exchanged messages. Each message is marked with the index that refers to the key used in the encryption on the sender's side and to be used in the decryption on the receiver's side. AES-256 keys consist of 32 characters, and cracking them is somehow impossible. Nevertheless, the mechanism used in **PolyFLAG_SVM** guarantees that a cracked or leaked key is not a threat, since it will not be used again in the FL process, neither with the same client nor with other clients. This theory is explained in more detail in the following sections. Thus, if a malicious client manages to crack one of the keys, it will not gain any benefit from it because it will most likely no longer be used. Moreover, it is worth noting that each client receives a different ToKs when it connects to the server, and even the same client receives a different ToKs for each connection session. On the other hand, this key table is transmitted to the client after the connection is established and before the ToKs are received, which requires another encryption mechanism to ensure that malicious entities cannot access it. Otherwise, the entire security approach would be useless. To achieve this, the initial encryption key, referred to as ***initial_key***, which is used to encrypt the ToKs, is also polymorphically generated, as described in the next section.

3.1.2. Initial Encryption Key

The initial encryption key or ***initial_key*** is used to encrypt the ToKs that are later used to encrypt the messages exchanged. Given the criticality of the ToKs data, it is critical that different initial keys are generated for each client. To achieve this, **PolyFLAG_SVM** follows defined steps to generate the ***initial_key*** before it is used to encrypt the ToKs. It is worth noting that each connection session, even for the same client, has its own key. Moreover, the keys are not transmitted over the network, but are generated on both the server and client sides based on the same mechanism to increase the security level. The steps to create ***initial_key*** are the same on the server and client side and are:

1. After the client connects to the server, it creates a random string 32 characters long; called ***random_secret***
2. The client mixes its connection data (IP & address) with the generated ***random_secret*** to create a 32 character string, where:
   (a) The first 8 characters are the reverse of the last 8 characters of ***random_secret***
   (b) Second 4 characters are last 4 characters of socket data
   (c) Third 8 characters are the middle 8 characters of ***random_secret***
   (d) Fourth 4 characters are the first 4 characters of the socket data
   (e) Fifth 8 characters are the reverse of the first 8 characters of the ***random_secret***

Concatenating the above substrings results in a 32-character string representing the ***initial_key***. After this key is obtained, it is hashed using SHA-256 algorithm [25], and the first 32 characters of the hashing result are then used to encrypt the ToKs. The addition of the hashing step increases security against the possibility of being cracked. Since the socket data is known to both the client and the server, the steps can be followed by both to generate the same key if they have the same ***random_secret***. However, this secret is randomly generated on the client side, so it is almost impossible to generate the same string on the server side, making it mandatory to send this secret to the server. To secure the transmission, the ***shuffled_secret*** is formed by following the below steps and then passed to the server:

1. First 8 characters are the reverse of third 8 characters of ***random_secret***
2. Second 8 characters are the first 8 characters of ***random_secret***
3. Third 8 characters are reverse of last 8 characters of ***random_secret***
4. Last 4 characters are the second 8 characters of the ***random_secret***

Following these steps ensures that the ***shuffled_secret*** is useless to malicious entities unless they can know the steps followed to restore the original order, which is not possible if those entities do not have access to the code. In this case,

security is of no concern at all. After the server receives the ***shuffled_secret***, its original order is restored by reversing the shuffling steps, and then the server follows the same steps that the client followed to create the ***initial_key***. Since both the server and the client have the same key, it is now possible to encrypt the ToKs and send them to the client, which decrypts them to use them to secure the exchanged messages. It is worth noting that even if the client connected via same IP and address for two times, the ***initial_key*** will not be the same because of the ***random_secret*** involved in its creation, along with the shuffling, mixing and hashing steps. Creating ***initial_key*** is shown in Figure 1 below.

3.2. Framework Design

In light of the above, the workflow followed by **PolyFLAG_SVM** is illustrated in the following steps, which can be seen in Figure 1 below:

1. server starts FL process on its side;
2. client connects to the server;
3. client generates the ***random_secret*** and ***initial_key*** and sends the first to the server in a ***"Connect"*** message;
4. server receives the message and creates the table of random encryption keys; simply the ToKs;
5. server regenerates the ***initial_key*** based on the received ***random_secret*** in the ***"Connect"*** message
6. server encrypts the ToKs using the first 32 characters of the hashed ***initial_key*** and sends them to the client;
7. client receives the encrypted ToKs and decrypts them using ***initial_key*** *(After this step, the client selects an unused key from the ToKs to encrypt its message, and encapsulates the sent message with the ID of the used key)*;
8. client replies to server with an encrypted ***"Ready"*** message;
9. server receives the message and responds with an initial ***"Model"*** message;
10. client receives the first ***"Model"*** message and trains the model on the local data;
11. client replies to the server with its encrypted gradients;
12. server checks if all clients have sent their gradients; and
    (a) If so, it starts the aggregation process, updates the global gradients, and sends them back to the clients;
    (b) If not, it sends an encrypted ***"Hibernate"*** message to the clients to wait until the above condition is met.
13. The clients receive the updated gradients and re-train their models based on them;
14. Steps 11, 12, and 13 are repeated until the model converges or until the server decides to stop.

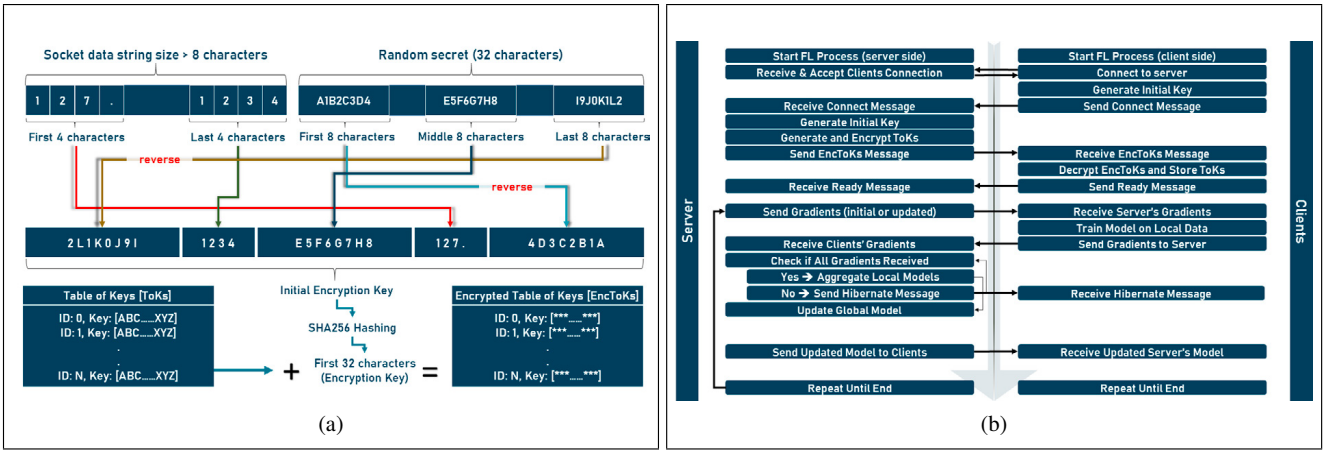

Fig. 1: (**a**) Polymorphic Initial Encryption Key; (**b**) PolFLAG_SVM Workflow

4. Experimental Evaluation and Discussion

The proposed framework, **PolyFLAG_SVM** gains its robustness against inference attacks by embedding Polymorphic Encryption to secure messages exchanged in the FL cycle. The theoretical guarantees of PE, evaluation of the proposed model, challenges and future prospects are discussed in this section. Despite the fact that the proposed framework provides a secure Federated Learning environment, it is necessary to consider adding authentication services to the proposed framework in future, to ensure that none of the connected clients is a malicious entity. Authentication services, which are not a point of focus in this study, may include, but are not limited to: password-based authentication, two-factor authentication (2FA), Public Key Infrastructure (PKI), Single Sign-On (SSO), biometric authentication, and others.

4.1. Theoretical Guarantees of PolyFLAG_SVM

In the proposed model, messages are encrypted using the AES-256 algorithm, which is considered one of the most secure encryption methods available today. The key used in this algorithm has a size of 32 characters and is considered almost impossible to crack, since there are more than 10^{77} possibilities for each key, which according to [24] takes billions of years to crack with a supercomputer. However, attacks as "quantum" attacks [27] can threaten the security of AES even if it is not possible to crack the key in a short time. To counter this threat, each message exchanged in **PolyFLAG_SVM** is encrypted with a unique key from the ToKs. In addition, this table is encrypted with the ***initial_key***, which is created using the methodology explained earlier. Moreover, the keys, whether the ToKs or the ***initial_key***, are created differently for each client and for each connection session. Even if a client connects twice, or two different clients connect to the same socket at two different times, the probability that the keys used will be reused is almost zero due to the randomization explained earlier. In summary, the theoretical guarantee provided by **PolyFLAG_SVM** is as follows: ***"AES-256 keys are known to be unbreakable. However, if a key is leaked or has been cracked in some way, it causes no threat because it will nearly not be used again in the FL cycle"***.

4.2. Evaluation

Several metrics were used to evaluate the proposed framework. The framework is evaluated based on guaranteed polymorphism, communication cost, and learning quality, as explained in this section.

4.2.1. Datasets Used

To assess the framework, three different datasets, that fits for binary classification, were used which are:

- a simulated dataset, generated using the SKLearn dataset library [28] (9000 records & 20 data features);
- SHAREEDB Cardiovascular Diseases prediction dataset [29], (139 records & 26 features);
- Dataset Surgical binary classification [30](14636 records and 24 features).

4.2.2. Polymorphism

First, the encryption keys used to encrypt each message were tracked and compared to ensure that a single key is never used twice. For example, for the first data set, two clients were selected, and the ***initial_keys*** and 5 keys used to encrypt the messages were recorded and compared. The results obtained are shown in Table 2 below.

Table 2: Encryption keys polymorphism in **PolyFLAG_SVM**

-	**server**	**client 1**	**client 2**
Initial Key	depending on the client	6b7e9d53e14bbac3787a90d0aaa1cf22	699858a7ba15eeb61fa5441c945a9fae
Randomly selected 5 keys	Esjk2Gk84zdVK9tlXPCnhlKuf8hsoVSQ	ZgGN4tnYlRZX4xGVWEo9CjtZe6pMPsui	ZOGW0S7d8M5DkgzJ8yZuQxEwyRBC3p2S
	NEu7K5IdD1AKHPW0fLhtjGVWdSOjXDXj	xImgrk0gZuRl8kVqlg7eTUVVXPsrWnLR	IOhhIAg6sYmybB7BH6F9FDJT1nK3pndU
	v67HKldenUarfh9eKf8tYHr5vQlp4IhY	Esjk2Gk84zdVK9tlXPCnhlKuf8hsoVSQ	JEGpEMSRaaRfpkrQHYgvDlMLZf9vjC3E
	ZOGW0S7d8M5DkgzJ8yZuQxEwyRBC3p2S	a5SAqLZE4snkIhVEBXoWH3jGHS9d5NAf	EKiim6TWfG9YvvMM4hejy2406o29RMAq
	y04aAD42M3J8SyJYTwKcya5PysixvR59	cg9ijnmsU9rKCFPfoBLWsmfOWQPmL9ay	OECIZeuDny8AjnADqksx8PShmRUakoR7

4.2.3. Communication Cost

Reducing communication costs is critical to improve system efficiency. Following this, **PolyFLAG_SVM** exchanges gradients rather than the entire models. Since the former are much smaller compared to the latter, **PolyFLAG_SVM** provides a communication efficient framework. To analyze the communication cost of the framework, data exchanged between the server and 3 (randomly selected) clients from the process of training the global model on the datasets used were examined for 10 training rounds, where the messages exchanged didn't exceed 5400 Bytes in total as shown in Table 3 below. The obtained results show that **PolyFLAG_SVM** is very efficient in terms of reducing communication costs. In this context, the following should be mentioned:

- server will receive only four types of messages: *"Connect"*, *"Ready"*, *"Gradients"* and *"Disconnect"*;
- the size of *"Connect"*, *"Ready"* & *"Disconnect"* messages are fixed since they always contain the same content;
- the size of *"Gradients"* message varies with the dataset and features selected for local models training;
- the size of a *"Gradients"* message is relatively small;
- the communication cost in the system exhibits linear growth with respect to the number of clients and number of training rounds followed, which boosts scalability in terms of communication cost.

4.2.4. Learning Quality

The quality of learning in **PolyFLAG_SVM** was not set as a goal, but rather security. However, the accuracy of a smart model is one of the important points to be considered in the evaluation. Next, the commonly used performance metrics were collected for the different datasets used, and the results are shown in Table 4 below. The results show that the numbers, while not outstanding, are encouraging and can be considered for future improvement.

Table 3: Communication cost in **PolyFLAG_SVM**

Dataset	Simulated Dataset			SHAREEDB			Surgical-Binary		
Client	client 1	client 2	client 3	client 1	client 2	client 3	client 1	client 2	client 3
Connect mesage (Bytes)	82	82	82	82	82	82	82	82	82
Ready message (Bytes)	91	91	91	91	91	91	91	91	91
Gradients message /1 round (Bytes)	462	462	462	510	510	510	494	494	494
Disconnect message (Bytes)	107	107	107	107	107	107	107	107	107
Total size /10 rounds (Bytes)	4900	4900	4900	5380	5380	5380	5220	5220	5220

Table 4: Performance metrics for different datasets (LR: Learning Rate, LP: Lambda Parameter, TR: Training Rounds)

Dataset	Parameters	Accuracy	Precision	Recall	F1 Score	Specificity	NPV
Simulated data	LR: 0.1; LP: 0.1; TR: 50	86.67%	82.86%	93.55%	87.88%	79.31%	92.00%
SHAREEDB dataset	LR: 0.00001; LP: 0.001; TR: 50	69.39%	67.94%	72.95%	70.36%	65.85%	71.05%
Surgical-Binary	LR: 0.00001; LP: 0.001; TR: 50	66.48%	47.92%	80.7%	60.13%	60.0%	87.21%

4.3. Challenges

Despite polymorphism, **PolyFLAG_SVM** can suffer from different challenges such as:

- Restriction to Support Vector Machines (SVM): SVM has proven its efficiency in solving ML problems, outperforming other models such as in [31, 32]. However, the limitation of **PolyFLAG_SVM** to the SVM model may limit its use;
- Heterogeneity: the proposed framework supports "Horizontal FL data", which describes the case where different clients process data with the same features. However, other approaches were not considered in our study;
- Complexity and Computation Cost: encryption algorithms are computationally intensive algorithms;
- Scalability: due to the additional computational cost, the scalability of this system may suffer when the number of clients is high, specifically on the server side;
- Learning Quality: the main purpose of the proposed framework FL is security and not learning quality;
- Resources Constraints: the clients in FL usually have limited computational resources, which may be a problem when implementing this framework in practice.

4.4. Future Perspectives

The challenges discussed above have been addressed in previous frameworks for aggregating FL, so they are not considered shortcomings. Future improvements can be achieved by combining our proposed framework with existing techniques to create even more powerful FL frameworks. Improvements to **PolyFLAG_SVM** could include::

- Framework Generalization: by embedding more models, e.g. neural networks, linear regression, or even more;
- Handling Heterogeneity: to deal with heterogeneity of devices or data, different approaches can be used, such as Resources Allocation [33] and Meta-Learning [34];
- Reducing Computation Cost: this is possible by using different techniques, such as parallel programming;
- Enhancing Scalability: this is the result of dealing with heterogeneity and reducing computation cost, where solving these two problems guarantees the scalability of the framework for a large number of clients;
- Boosting Learning Quality: by implementing different techniques for data pre-processing on the client side.

Conclusion

The **PolyFLAG_SVM** framework benefits from Polymorphic Encryption to secure messages exchanged between servers and clients in a Federated Learning environment. The security guarantees in the proposed framework derive from the polymorphism of encryption keys, where each message exchanged between the server and a client is encrypted with a different key. If a key is cracked or leaked for any reason, it is useless because it is almost never used twice in the FL cycle. Also, analysis of the communication cost of **PolyFLAG_SVM** has shown that it is efficient due to the small size of the messages exchanged. Even though **PolyFLAG_SVM** focuses only on security, it can benefit from other existing approaches to increase learning quality, handle heterogeneity, and improve scalability.

References

[1] Bell, Jason. "What is Machine Learning?." Machine Learning and the City: Applications in Architecture and Urban Design (2022): 207-216.
[2] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. "Reviewing Multimodal Machine Learning and Its Use in Cardiovascular Diseases Detection." Electronics 12, no. 7 (2023): 1558.
[3] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. "Reviewing Federated Machine Learning and Its Use in Diseases Prediction." Sensors 23, no. 4 (2023): 2112.
[4] Albrecht, J.P. How the GDPR will change the world. Eur. Data Prot. Law Rev. 2016, 2, 287.
[5] Parasol, M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams. Comput. Law Secur. Rev. 2018, 34, 67–98.

[6] McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics PMLR, Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.

[7] Li, Q.;Wen, Z.;Wu, Z.; Hu, S.;Wang, N.; Li, Y.; Liu, X.; He, B. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. IEEE Trans. Knowl. Data Eng. 2021.

[8] Mammen, P.M. Federated learning: Opportunities and challenges. arXiv 2021, arXiv:2101.05428.

[9] Federated Learning: Collaborative Machine Learning without Centralized Training Data. (2017, April 6). Federated Learning: Collaborative Machine Learning Without Centralized Training Data – Google AI Blog. https://ai.googleblog.com/2017/04/federated-learning-collaborative.html. Accessed on (February 5, 2023).

[10] Li, Tian, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. "Federated learning: Challenges, methods, and future 1295 directions." IEEE signal processing magazine 37, no. 3 (2020): 50-60. 1296

[11] Rahman, KM Jawadur, Faisal Ahmed, Nazma Akhter, Mohammad Hasan, Ruhul Amin, Kazi Ehsan Aziz, AKM Muzahidul 1297 Islam, Md Saddam Hossain Mukta, and AKM Najmul Islam. "Challenges, applications and design aspects of Federated Learning: 1298 A survey." IEEE Access 9 (2021): 124682-124700.

[12] Lyu, Lingjuan, Han Yu, and Qiang Yang. "Threats to federated learning: A survey." arXiv preprint arXiv:2003.02133 (2020).

[13] Bonawitz, Keith, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. "Practical secure aggregation for privacy-preserving Machine Learning." In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1175-1191. 2017.

[14] Pillutla, Krishna, Sham M. Kakade, and Zaid Harchaoui. "Robust aggregation for Federated Learning." IEEE Transactions on Signal Processing 70 (2022): 1142-1154.

[15] Weiszfeld, Endre, and Frank Plastria. "On the point for which the sum of the distances to n given points is minimum." Annals of Operations Research 167, no. 1 (2009).

[16] Jeon, Beomyeol, S. M. Ferdous, Muntasir Raihan Rahman, and Anwar Walid. "Privacy-preserving decentralized aggregation for Federated Learning." In IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 1-6. IEEE, 2021.

[17] Boyd, Stephen, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. "Distributed optimization and statistical learning via the alternating direction method of multipliers." Foundations and Trends® in Machine learning 3, no. 1 (2011): 1-122.

[18] Zhao, Lingchen, Jianlin Jiang, Bo Feng, Qian Wang, Chao Shen, and Qi Li. "Sear: Secure and efficient aggregation for byzantine-robust Federated Learning." IEEE Transactions on Dependable and Secure Computing 19, no. 5 (2021): 3329-3342.

[19] McKeen, Frank, Ilya Alexandrovich, Alex Berenzon, Carlos V. Rozas, Hisham Shafi, Vedvyas Shanbhogue, and Uday R. Savagaonkar. "Innovative instructions and software model for isolated execution." Hasp@ isca 10, no. 1 (2013).

[20] Song, Jingcheng, Weizheng Wang, Thippa Reddy Gadekallu, Jianyu Cao, and Yining Liu. "Eppda: An efficient privacy-preserving data aggregation Federated Learning scheme." IEEE Transactions on Network Science and Engineering (2022).

[21] Benaloh, Josh Cohen. "Secret sharing homomorphisms: Keeping shares of a secret secret." In Advances in Cryptology—CRYPTO'86: Proceedings, pp. 251-260. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.

[22] Elkordy, Ahmed Roushdy, and A. Salman Avestimehr. "HeteroSAg: Secure aggregation with heterogeneous quantization in Federated Learning." IEEE Transactions on Communications 70, no. 4 (2022): 2372-2386.

[23] Booher, D. Duane, Bertrand Cambou, Albert H. Carlson, and Christopher Philabaum. "Dynamic key generation for polymorphic encryption." In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0482-0487. IEEE, 2019.

[24] Daemen, Joan, and Vincent Rijmen. "Reijndael: The advanced encryption standard." Dr. Dobb's Journal: Software Tools for the Professional Programmer 26, no. 3 (2001): 137-139.

[25] Alenezi, Mohammed N., Haneen Alabdulrazzaq, and Nada Q. Mohammad. "Symmetric encryption algorithms: Review and evaluation study." International Journal of Communication Networks and Information Security 12, no. 2 (2020): 256-272.

[26] Ruder, Sebastian. "An overview of gradient descent optimization algorithms." arXiv preprint arXiv:1609.04747 (2016).

[27] Bonnetain, Xavier, María Naya-Plasencia, and André Schrottenloher. "Quantum security analysis of AES." IACR Transactions on Symmetric Cryptology 2019, no. 2 (2019): 55-93.

[28] sklearn.datasets.make_classification. Scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.datasets.make_classification.html. (Accessed on 15 Feb. 2023)

[29] Smart Health for Assessing the Risk of Events via ECG Database v1.0.0. Smart Health for Assessing the Risk of Events via ECG Database v1.0.0. https://physionet.org/content/shareedb/1.0.0/. (Accessed on 1 March 2023).

[30] Dataset Surgical binary classification. Dataset Surgical Binary Classification — Kaggle. https:///datasets/omnamahshivai/surgical-dataset-binary-classification. (Accessed on 15 March 2023)

[31] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. "Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability." Procedia Computer Science 203 (2022): 231-238.

[32] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad." Predicting Cardiovascular Events with Machine Learning Models and Heart Rate Variability". International Journal of Ubiquitous Systems and Pervasive Networks (JUSPN), 18 no. 2 (2023): 49-59.

[33] Jamil, Bushra, Humaira Ijaz, Mohammad Shojafar, Kashif Munir, and Rajkumar Buyya. "Resource allocation and task scheduling in fog computing and internet of everything environments: A taxonomy, review, and future directions." ACM Computing Surveys (CSUR) 54, no. 11s (2022): 1-38.

[34] Feng, Yong, Jinglong Chen, Jingsong Xie, Tianci Zhang, Haixin Lv, and Tongyang Pan. "Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects." Knowledge-Based Systems 235 (2022): 107646.

PolyFLAM & PolyFLAP: Federated Learning Aggregation Frameworks Secured with Polymorphic Encryption

---

This article is undergoing peer review process in the Elsevier Journal of Information Security and Applications

PolyFLAM & PolyFLAP: Federated Learning Aggregation Frameworks Secured with Polymorphic Encryption

Mohammad Moshawrab[a,*], Mehdi Adda[a], Abdenour Bouzouane[b], Hussein Ibrahim[c], Ali Raad[d]

[a]*Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, G5L 3A1, Québec, Canada*
[b]*Département D'informatique et de Mathématique, Université du Québec à Chicoutimi, Chicoutimi, 555 boulevard de l'Université, Chicoutimi, G7H 2B1, Québec, Canada*
[c]*Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, G4R 5B7, Québec, Canada*
[d]*Dean of Faculty of Science and Arts, Islamic University of Lebanon, Wardaniyeh, Lebanon*

Abstract

Maintaining user privacy in machine learning is a primary research concern given the implications for data collection. In this context, Federated Learning (FL) has emerged as a solution, sharing trained models instead of user data. However, FL faces several challenges, including security and privacy hurdles, especially inference attacks. To address this, new frameworks called PolyFLAM & PolyFLAP are proposed. PolyFLAM: "Polymorphic Federated Learning Aggregation of Models" and PolyFLAP: "Polymorphic Federated Learning Aggregation of Parameters" are supported by polymorphic encryption. These frameworks include several models: support vector machines, logistic regression, Gaussian Naïve Bayes, Stochastic Gradient Descent, and Multi Layer Perceptron. Security depends on a unique key polymorphism that doesn't compromise privacy in case of leakage or compromise. In this paper, the proposed models are described and evaluated in detail.

*Corresponding author: Tel.: +1(581)624-9394; E-mail address: mohammad.moshawrab@uqar.ca;

Email addresses: mehdi_adda@uqar.ca (Mehdi Adda), abdenour_bouzouane@uqac.ca (Abdenour Bouzouane), Hussein.Ibrahim@itmi.ca (Hussein Ibrahim), ali.raad@iul.edu.lb (Ali Raad)

Keywords: Federated Machine Learning, Federated Learning, Aggregation Algorithms, Polymorphic Encryption, Encryption, Security, Privacy

1. Introduction

Artificial Intelligence (AI) is a rapidly advancing technology that is becoming increasingly integrated into various industries and aspects of daily life, resulting in significant changes and advancements in lifestyles and professional activities. This reality is obvious and observable and requires no proof. Despite the long duration of AI research, dating back to the 1950s when Alan Turing famously asked, "Can computers think?!" [1], there is no single definition for this field. For example, a simple definition for AI is provided by the authors in [2], where they describe it as programs that are no less competent than a human in any given setting. In contrast, the authors in [3] describe it as a set of tools and methods that use principles and mechanisms from various fields such as computation, mathematics, logic, and biology to address the challenges associated with realizing, modeling, and mimicking human intelligence and cognitive processes. Since then, AI has been a broad field of research, leading to various derivatives such as machine learning (ML), deep learning (DL), federated learning (FL), and others. Machine learning, for example, allows computers to "learn" from training information and incrementally improve their understanding without the need for explicit programming or with the least amount of supervision.
ML algorithms strive to identify patterns in data and derive knowledge from them to formulate independent predictions. ML Algorithms and models acquire knowledge through encounters with the real world. In traditional contexts, a computer program is developed by engineers and provided with a set of instructions that facilitate the transformation of incoming data into the desired outcome. In contrast, ML is designed to learn on its own with minimal or no human intervention, gradually expanding its knowledge. The impressive performance of ML, combined with its enormous potential in classification and regression problems and its ability to use both supervised and unsupervised learning methods, have made it attractive to researchers [4, 5]. Subsequent research has shown that ML has a wide range of applications in areas such as: E-commerce and product recommendation, image, speech and pattern recognition, user behavior analysis and context-aware smartphone applications [4, 5], health services [6, 7, 8], traffic prediction and transportation [4, 9], Internet of Things (IoT) and smart cities [9], cybersecurity [10],

Natural Language Processing and sentiment analysis [11], sustainable agriculture [12], industrial applications [13], and many others.

1.1. Challenges in Machine Learning Domain

The precise results obtained in classification or regression gradually promote the integration of these methods into aspects of daily life. The practicality of using AI tools, especially ML, has been underpinned by their exceptional efficiency and the potential of their application in various domains. Nevertheless, ML continues to struggle with a number of challenges that are described in detail in the existing scientific literature. However, these challenges cannot be categorized uniformly, but are classified according to different viewpoints. This section presents the prevailing challenges and places them in a proposed framework that classifies them based on factors related to data, models, implementation, and other general dimensions.

- General Challenges [14, 15]
  - User Data Privacy and Confidentiality
  - User Technology Adoption and Engagement
  - Ethical Constraints
- Models Related Challenges [14, 15]
  - Accuracy and Performance
  - Model Evaluation
  - Variance and Bias
  - Explainability
- Data-Related Challenges [16, 17]
  - Data Availability and Accessibility [18]
  - Data Locality [19]
  - Data Readiness [18]
  - Data Heterogeneity
  - Noise and Signal Artifacts
  - Missing Data
  - Classes Imbalance
  - Data Volume Course of Dimensionality
  - Bonferroni principle [20]
  - Feature Representation and Selection
- Implementation-Related Challenges [18, 21]
  - Real-Time Processing
  - Model Selection
  - Execution Time and Complexity

The challenges within ML and related fields are the subject of extensive study, with researchers seeking to address these challenges collectively rather than focusing on any one. It's difficult to definitively state that any one of the above challenges is the most significant or has the most detrimental impact on the machine learning field. Nonetheless, the machine learning workflow primarily includes phases such as data collection and preprocessing, feature engineering, model training, model evaluation, and model deployment. The structure of this workflow highlights the central role of data in machine learning, as it's the first step in the process; without its completion, subsequent phases cannot proceed. Moreover, the performance of ML 's models is directly linked to the availability of data. While achieving highly accurate intelligent models depends on the technical architecture of the models themselves, the quality and availability of the data, preprocessing, and several other factors, it's generally accepted that data availability contributes to increased and improved accuracy [16, 17].

1.2. Federated Learning: A Privacy Issue Solution

In the real world, due to several factors, the process of data collection is a major challenge, if not the biggest challenge, in developing machine learning models, and privacy and confidentiality are of paramount importance. This concern goes beyond individual privacy to include societal, governmental, and organizational dimensions, all of which reinforce efforts to protect privacy and security of data. These efforts have led to the introduction of numerous regulations and laws around the world, such as the European Union's General Data Protection Regulation (GDPR) [22], China's Cyber Security Law of the People's Republic of China [23], the General Principles of the Civil Law of the People's Republic of China [24], Singapore's PDPA [25], and countless other laws implemented around the world.

While these regulations undeniably help protect private information, they also introduce some complexities into the landscape of ML. Collecting data for model training becomes much more difficult, which in turn hinders advances in model performance accuracy and personalized model results. Consequently, privacy and confidentiality issues not only present an isolated challenge, but also set in motion a number of additional hurdles for ML. These include challenges related to data availability, model performance, personalization, and ultimately building trust and acceptance. The critical importance of privacy in information sharing has led to extensive research resulting in several proposed privacy algorithms such as differential privacy [26],

anonymity k-order [27], and homomorphic encryption [28]. However, these methods do not provide optimal solutions, as demonstrated by observed machine learning attacks, e.g., model inversion [29] and membership inference attacks [30], where raw data is extracted by accessing the model.

1.2.1. Federated Learning: An Overview

To address privacy issues without restricting data collection, Google recently introduced a novel concept in machine learning called federated machine learning or federated learning (FL) [31]. The basic premise of federated learning is that it does not require the sharing of user data between different devices. This concept can be defined as collaborative, distributed, and decentralized machine learning with privacy preservation. In federated learning, an intelligent model is trained without the need to transfer data from edge devices to a central server. Instead, models are sent to these devices, where they are trained on local data. Then, these refined models are sent back to a central server for aggregation, which assembles the global model without having visibility into the specific embedded data. The technical infrastructure of federated learning is shown in Figure 1 below.

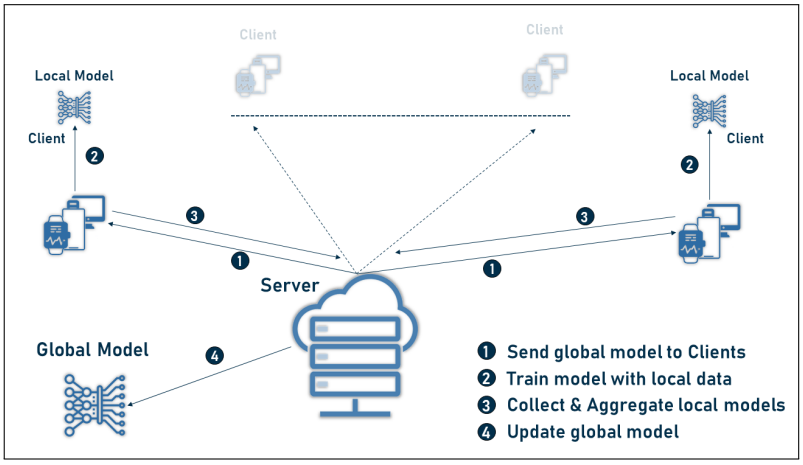

Figure 1: Federated Learning technical architecture

The concept of federated learning provides an effective solution to user privacy concerns. It not only addresses these concerns, but also unlocks the potential to collect more data for training machine learning models, which helps improve accuracy and efficiency. In addition, Federated Learning facilitates training models with data from disparate and unrelated sources,

referred to as "data islands." In addition, Federated Learning enables the management of disparate data spread across different data spaces, each characterized by its unique attributes. This approach also facilitates what is known as "learning transfer," which allows models to share knowledge without transmitting users' private data. However, it is important to note that FL is still in its infancy and faces a number of challenges. This necessitates targeted research efforts to improve its capabilities.
In response to this need, this article presents two innovative frameworks for federated learning, both of which involve the use of Polymorphic Encryption[32] to strengthen the security of FL. Section 2 addresses the problem, specifically the existing privacy inadequacies in FL. It also explains the motivation for developing these frameworks. A key aspect is the introduction of polymorphic encryption, a new addition in federated learning. Section 3 presents the proposed frameworks in detail, explaining the mechanisms incorporated in them and providing comprehensive explanations of their inner processes. Section 4 discusses and evaluates the proposed frameworks from different perspectives. In this context, tests are performed under real conditions to prove their efficiency. Finally, section 5 addresses the challenges that hinder the development of the proposed frameworks, while providing perspectives for their future development.

2. Problem Statement: Security Threats in FL Domain

Federated learning is a robust solution for ensuring data privacy by taking a decentralized approach to machine learning and minimizing extensive data sharing between clients and servers. Equally advantageous, FL succeeds in reducing transmission costs, as the raw data usually exceeds the size of the transmitted models or their parameters.

2.1. FL under the Scope: Challenges and Issues

Federated learning has proven to be highly successful in a variety of applications, but it is not immune to challenges, a topic that has been extensively studied in the academic literature. This thorough investigation has brought to light a number of issues that have been discussed in detail in studies [33, 34, 35]. In particular, the original aggregation algorithm FL named FedAvg [31] has been studied with respect to several limitations. These include issues such as data and hardware heterogeneity, sensitivity to local models, scalability limitations, incremental convergence rates, computational

and communication overheads, and vulnerability to malicious attacks. The diversity of these challenges has led to increased research efforts aimed at improving the practicality of FL, and has prompted researchers to develop solutions that deftly address these challenges.

2.2. Security in FL Domain

Although Federated Learning functions as a privacy-preserving ML technology, it remains vulnerable to malicious attacks [33, 34, 35]. The security of messages exchanged within the FL cycle can be divided into the input phase, the learning process itself, and the resulting learned model. This vulnerability leads to a spectrum of attacks, including but not limited to poisoning, inference, and backdoor attacks. Poisoning attacks can adversely affect the quality of learning outcomes, inference attacks expose users' private data, and backdoor attacks allow unauthorized intrusion into the FL system [36].

2.2.1. Poisoning Attacks

Poisoning attacks, whether random or targeted [37], aim to either reduce model accuracy (random) or manipulate the model to output a label specified by the attacker (targeted). These attacks can target data or the model, both of which negatively impact the overall behavior of FL. Compromised FL environments allow attackers to perform targeted and untargeted poisoning attacks that include both data and model poisoning attacks.

- **Data Poisoning:** also known as data corruption, has two main forms: Clean Label [38] and Dirty Label [39]. Clean label attacks assume that the labels cannot change, requiring stealthy poisoning, while dirty label attacks can insert misclassified data with the intended target labels. Data poisoning can be performed by any Federated Learning participant, and the impact on the FL model depends on the number of attackers and the amount of data poisoned.

- **Model Poisoning:** Local model training leads to model poisoning by contaminating updates before committing them to the server or embedding secret global model backdoors [40]. Targeted model poisoning aims to securely misclassify selected inputs without modifying them as in adversarial attacks [41], which is achieved by manipulating the training process. Model poisoning in federated learning surpasses the effects of data poisoning by affecting model updates during each iteration [42].

It mimics centralized poisoning on a subset of the entire training data. Performing model poisoning requires significant technical resources.

2.2.2. Inference Attacks

In federated learning, sharing parameters during training raises privacy concerns [43, 44]. Deep learning models unintentionally internalize various data features beyond the core tasks, potentially revealing sensitive data aspects of the participants. Attackers can infer features by comparing model parameter snapshots, revealing aggregate updates of all participants except the attacker. The problem lies in gradients computed from participants' private data. Gradients in deep learning models arise from layer attributes and errors abovethe layer, providing opportunities for inference attacks [43]. These observations can reveal private data attributes, including class representatives and membership, or even allow recovery of labels without knowledge of the training set [44]. Inference attacks categorically include:

- **Inferring Membership:** The goal of membership inference attacks is to determine whether a particular data element was used to train the model [45]
- **Inferring Class Representatives:** Occurs when a malicious participant intentionally compromises another participant and exploits the real-time learning of FL to train a network that generates private prototype samples of targeted training data. These generated samples mimic the distribution of the training data.
- **Inferring Properties:** in this attack, an attacker can perform both passive and active property inference attacks to infer properties of other participants' training data that are independent of the features describing the classes of the FL model [45]:
  - Property inference attacks require the attacker to have additional training data labeled with the exact property they wish to infer
  - A passive attacker can only monitor updates and make inferences by training a binary property classifier
  - An active adversary can use multitask learning to trick the model FL into learning a better separation between data with and without the property, resulting in more information being extracted

– An adversarial participant can even infer when a feature appears and disappears in the data during training
– It can recover pixel-perfect original images and token-matched original texts

2.3. Securing FL Frameworks: State of the Art

Researchers are interested in improving the safety of FL to promote its usability and feasibility. Several attempts have already been made in this regard. For example, in [46], the authors propose a secure vector summation strategy using a protocol with a fixed number of rounds that reduces computational costs and is robust to faulty clients. In their approach, only a single server can be trusted to hold the exchanged data. Their proposed framework provides high security against honest but curious adversaries and guarantees anonymity even when faced with active adversaries, such as an enemy server. Moreover, in [47], the authors proposed Robust Federated Aggregation (RFA) that aims to protect the aggregation process FL against poisoning attacks. To achieve their goal, they aggregated the exchanged models based on the geometric median, which can be calculated using a Weiszfeld-type algorithm [48]. RFA was able to compete with the traditional FedAvg algorithm and was more resistant to data poisoning attacks.
On the other hand, the authors in [49] developed SecureD- FL, a FL framework based on a refined form of the Alternating Direction Multiplier (ADMM) [50]. Their proposed framework uses a communication mode in which the algorithm decides in each round of execution which subset of users should exchange data in order to minimize the disclosure of private data during the aggregation process. In addition, the authors of [51] proposed SEAR,a secure and efficient aggregation for byzantine-stable federated learning that aggregates the local models in a secure and trusted hardware environment, specifically, the Intel SGX Trustworthy Execution Environment (TEE) [52], a secure CPU domain area, where the executed data and programs are kept secret and cannot be modified. Moreover, in [53], the authors proposed the Efficient Privacy-Preserving Data Aggregation (EPPDA), which uses homomorphisms of the secret exchange [54] in the FL environment. Their algorithm is secure and reduces the impact of some malicious clients. The cryptographic primitives used in their approach can be summarized as follows: Secret exchange, key ex-change protocol, authenticated encryption, and signature methods.

Finally, in [55], the authors proposed the aggregation algorithm HeteroSAg, which uses masking to secure the exchanged messages such that the mutual information between the masked model and the unique model is zero. HeteroSAg's resilience to Byzantine attacks depends on the FL cycle, which implements a segment grouping strategy based on dividing edge users into groups and segmenting local model updates for those users. The security approaches followed in the state of the art of secured FL frameworks are summarized in Table 1 below.

Table 1: State-of-the-art of secured FL aggregation algorithms

Ref#	**Mechanism**
[46]	Secure Vector Summing Strategy
[47]	Using geometric median estimated using a Weiszfeld-type algorithm
[49]	Refined form of the Alternating Direction Multiplier (ADMM)
[51]	Hardware-based trusted execution environment instead of complex cryptography
[53]	Homomorphisms of the secret exchange
[55]	Masking each user's model update

2.4. Problem and Motivation

Extensive research has focused primarily on understanding poisoning attacks in federated learning secure aggregation algorithms, whereas attention to inference attacks has been relatively limited. Although techniques such as polymorphic encryption (PE) [32] promise in reducing the impact of inference attacks by making data exchanges more secure, they have been little explored in previous FL frameworks. Considering the critical importance of inference attacks and inspired by the effectiveness of PE, this paper introduces two frameworks called **PolyFLAM** and **PolyFLAP**. What makes these frameworks special is that they're the first to combine PE with FL aggregation, which makes them innovative and novel solutions in this area. This integration not only sets them apart from others, but also opens new possibilities for improving the security of FL. On the other hand, these frameworks provide secure FL in different models, as will be explained later.

2.5. Polymorphic Encryption

Polymorphism can be understood as the remarkable ability of objects or functions to take on different forms or behaviors and adapt to different con-

texts. In contrast, encryption is a complicated process of converting regular data into an unintelligible format to protect it from malicious use or access. The most popular encryption algorithms are AES (Advanced Encryption Standard) [56], RSA (Rivest-Shamir-Adleman), and others [57, 58], all of which contribute to data security. This leads to the concept of polymorphic encryption, a sophisticated encryption paradigm that introduces a dynamic dimension by changing the encryption algorithm or keys to enhance overall security. Unlike traditional encryption methods, which are characterized by fixed algorithms and keys, PE poses a major challenge to attackers because even possession of the ciphertext provides minimal advantage in decryption, underscoring the robustness of the technique to attackers.

3. PolyFLAM & PolyFLAP: FL Frameworks Secured with Polymorphic Encryption

The compelling need to improve security protocols within federated learning frameworks against inference attacks, and the feasibility of polymorphic encryption in response, were the primary motivations for PolyFLAM and PolyFLAP. These frameworks integrate the core principles of polymorphic encryption into their architecture to improve the security and privacy of FL, ensuring their unassailable security. In this section both the conceptual foundations and the design of the proposed frameworks are explained in detail.

3.1. Main Concept

A typical federated learning system consists of a central server and multiple clients. The server sends a global model to the clients, which train their own model using their local data. After training, the clients send their updated models back to the server, which combines them into a single improved global model. The process is repeated until the global model reaches a point of stability. In the cases of **PolyFLAM** & **PolyFLAP**, the exchanged messages are subjected to a special type of encryption, called polymorphic encryption, using the algorithm AES -256 [56]. In this way, the security and protection of the exchanged data is ensured. The uniqueness of the proposed framework lies in the use of different encryption keys for each message exchanged between the server and the clients. This approach generates polymorphism that adds an additional layer of security. Moreover, even for a single client, different keys are used for each message exchanged with the server. The main sources of this polymorphism are the encryption key table

Table of Encryption Keys (ToKs) and the initial encryption key, which are described in the following section.

3.1.1. Table of Encryption Keys (ToKs)

In both proposed frameworks, when a client makes a connection request, the server responds by providing a Table of Encryption Keys (ToKs). Each key in this table is assigned a unique ID for indexing. These keys play a critical role in encrypting the messages that are later exchanged. In this process, each message is assigned an index corresponding to the key used for encryption on the sender's side and decryption on the receiver's side. The AES -256 keys consist of 32 characters (bytes) and are extremely resistant to cracking attempts, which ensures a high level of security. In the context of **PolyFLAM** & **PolyFLAP**, even the case of a key being cracked or leaked does not pose a significant threat. This is because the implemented mechanism ensures that the compromised key, if present, is not reused in the federated learning process, either with the same client or with other clients. This concept is explained in more detail in the following sections. In practise, a malicious client that successfully cracks a key would gain minimal benefit from it, since that key is unlikely to have any further use.
It's important to emphasise that each client receives its own set of ToKs when connecting to the server. Even the same client receives a new set of ToKs each time it connects. Moreover, the transmission of ToKs to the client after the connection is established requires an additional encryption mechanism to protect against malicious entities. This precaution is critical because the effectiveness of the entire security scheme would be compromised if the ToKs were cracked or leaked. To counteract this, the initial encryption key, called the "initial_key," used to encrypt the ToKs is also generated polymorphically, a concept that is explained in more detail in the following section.

3.1.2. Initial Encryption Key

The initial encryption key, referred to as the "initial_key," plays an important role in encrypting the Table of Encryption Keys (ToKs) that is subsequently used to encrypt messages. Given the sensitive nature of ToKs data, it is imperative that separate initial_keys generated for each client. To achieve this, both **PolyFLAM** & **PolyFLAP** have well-defined procedures to generate the initial_key prior to its use in ToKs encryption. It is important to note that each connection session, even for the same client, uses a unique key, since random characters are used to generate the initial_key creation. It

is worth noting that these keys are not transmitted over the network. Instead, they are generated independently on both the server and the client, using a unified mechanism. This approach significantly increases the level of security. The procedure for generating the initial_key remains identical on both the server and client sides and includes the following steps (the process described below on the side of an entity, where the entity can be the server or a client):

1. Once the connection is established, the client generates a 32-character string called the "random_secret"
2. This string is then combined with the client's connection data (socket data), which includes the IP address and address details. This merging process results in a new 32-character string that conforms to the following structure:
   (a) The first 8 characters result from the inversion of the last 8 characters of random_secret;
   (b) The following 4 characters are extracted from the last 4 characters of the socket data;
   (c) The following 8 characters are made of the middle 8 characters of the random_secret;
   (d) The next 4 characters correspond to the first 4 characters of the socket data;
   (e) Finally, the last 8 characters are obtained by reversing the first 8 characters of the random_secret.

By concatenating the above substrings, a 32-character string is formed. This string is used as input to the SHA -256 algorithm [57], which produces a hashing result. The initial_key, is then derived from the first 32 characters of this hashing result. The inclusion of the hashing process increases security by reducing vulnerability to potential cracking attempts. Since both the client and the server know the socket data, they can independently repeat the steps to create an identical key when they receive the same random_secret. However, since this random_secret is randomly generated on the client side, reproducing an identical string on the server side is highly impossible. Consequently, it is necessary to share this secret with the server. To ensure secure transmission, the "shuffled_secret" is constructed according to the following steps and then forwarded to the server so that it can use it to regenerate the random_secret:

1. The first 8 characters are opposite to the third 8 characters of random_secret

2. The second 8 characters are the first 8 characters of random_secret
3. The third 8 characters are the inverse of the last 8 characters of random_secret
4. The last 4 characters are the second 8 characters of random_secret

By following these steps, the shuffled_secret becomes virtually useless to malicious entities unless they know the process required to restore the original sequence. This is not possible unless these entities can access the underlying code. Once the server receives the shuffled_secret, it reverses the steps of the shuffle to restore the original sequence, thus building the random_secret as it was on the client's side. Then the server mimics the client's actions and repeats the same sequence of steps to generate the initial_key. Now that both the server and the client have the identical initial_key, the encryption of the Table of Encryption Keys (ToKs) can be performed. Then these encrypted ToKs are sent to the clients, which decrypt them and use them to secure the exchanged messages.
It is important to note that even if a client connects through the same IP address in different sessions at different times, the initial_key would not be consistent. This is because randomness was included in the key creation process, in addition to complex shuffling, mixing, and hashing. The visual representation of the initial_key creation process can be seen in Figure 2.

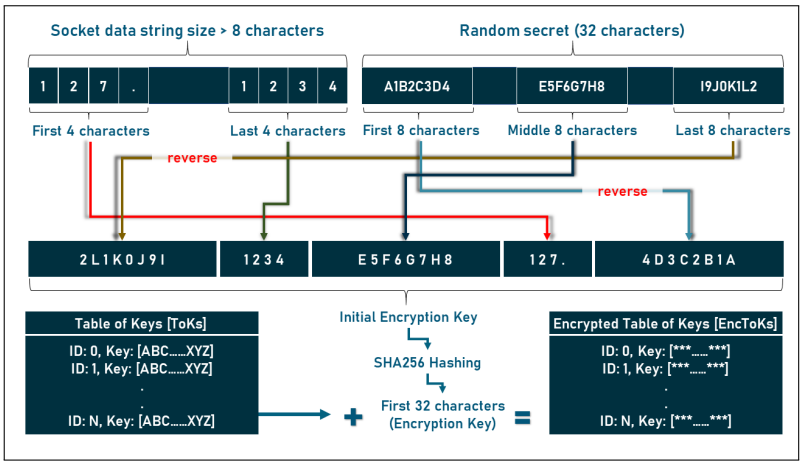

Figure 2: Initial Encryption Key generation mechanism

3.2. Supported ML Models

PolyFLAM and PolyFLAP are innovative frameworks for federated learning that expand the horizons of model training. These frameworks provide a diverse set of five different machine learning models that give users the flexibility to effectively tackle a variety of data analysis problems. The models offered are:

- Support Vector Machines (SVMs) [59]: A powerful classification algorithm that determines the optimal hyperplane to divide data into different classes
- Logistic Regression [60]: A widely used binary classification algorithm that estimates the probability that a given input belongs to a particular class
- Gaussian Naive Bayes [61]: This algorithm relies on the Naive Bayes theorem and the Gaussian distribution to classify data points based on their feature values
- Stochastic Gradient Descent (SGD Classifier) [62]: An iterative optimization algorithm used for training linear classifiers, often applied to large data sets
- Neural Network (Multi-Layer Perceptron) [63]: A versatile Deep Learning architecture that simulates the interconnected structure of the human brain and is capable of processing complex patterns and relationships in data

These models are suitable for a variety of machine learning tasks and provide users with the flexibility to choose the model that best fits their specific needs.

3.3. PolyFLAM vs. PolyFLAP

PolyFLAM and PolyFLAP differ significantly in one key respect, namely the nature of the messages exchanged between clients and servers. PolyFLAM follows the strategy of transmitting entire models to clients, while PolyFLAP takes a more efficient route by transmitting only model parameters, which has several advantages and improvements. In the case of PolyFLAM, the complete model is transmitted to the clients, while PolyFLAP optimizes communication by transmitting only short sets of parameters. This deliberate shift from full model transmission to parameter exchange reduces both the complexity of the system and the time required to encode messages. It also significantly reduces communication overhead, resulting in a more streamlined federated learning process. Each model type generates a set

of parameters during the local training process, which are explained in the Table2 below.

Table 2: Parameters generated by each model on local training

Model	**Parameter**	**Description**
Support Vector Machines	Support vectors	data points that significantly influence the determination of the separating hyperplane
	Coefficients	weights assigned to features, contributing to the hyperplane's orientation
	Intercept	also known as the bias term, it shifts the hyperplane's position, aiding in better classification
Logistic Regression	Coefficients	weights determine the influence of individual features on the log-odds of the predicted outcome
	Intercept	bias term that adjusts the threshold for classifying instances
Gaussian Naive Bayes	Class priors	represent the prior probabilities of different classes in the training data
	Theta	mean values of features for each class, used in the Gaussian probability density function
	Sigma	variance of features for each class, also utilized in Gaussian probability calculations
SGD Classifier	Coefficients	similar to other models, these weights influence the classification decision
	Intercept	a bias term that adjusts the decision threshold
Multi Layer Perceptron	Coefficients	regulate the connections between neurons in the neural network layers
	Intercept	similar to bias terms in other models, it offsets the overall computation

These parameters, which collectively represent the core attributes of their respective models, are shared between clients and servers to jointly refine the global model during the federated learning process. The parameters exchanged between clients and servers contain the essence of the model complexity. On the server side, these received parameters are skillfully integrated and aggregated, enabling iterative refinement of the global model. This col-

laborative process ensures that the collective knowledge of the various clients contributes to the creation of a better informed and trained global model.

3.4. Frameworks Design

With this in mind, the PolyFLAM & PolyFLAP workflow is described in the following steps, which are also shown in Figure 3 below. Recall that both frameworks have the same workflow, except for the type of messages exchanged between server and clients, which are models in the case of PolyFLAM and parameters in the case of PolyFLAP.

1. server starts FL process on its side;
2. client connects to the server;
3. client generates the random_secret and initial_key and sends the first to the server in a "Connect" message;
4. server receives the message and creates the table of random encryption keys (ToKs);
5. server regenerates the initial_key based on the received random_secret in the "Connect" message;
6. server encrypts the ToKs using the first 32 characters of the hashed initial_key and sends them to the client;
7. client receives the encrypted ToKs and decrypts them using initial key (After this step, the client selects an unused key from the ToKs to encrypt its message, and encapsulates the sent message with the ID of the used key);
8. client replies to server with an encrypted "Ready" message;
9. server receives the message and responds with an initial "Model/Parameters" message;
10. client receives the first "Model/Parameters" message and trains the model on the local data;
11. client replies to the server with its encrypted model (in case of PolyFLAM) or encrypted model parameters (in case of PolyFLAP);
12. server checks if all clients have sent their models/parameters; and
    (a) If so, it starts the aggregation process, updates the global model/parameters, and sends them back to the clients;
    (b) If not, it sends an encrypted "Hibernate" message to the clients to wait until the above condition is met.
13. The clients receive the updated gradients and re-train their models based on them;

14. Repeat Steps 11, 12, and 13 until the model converges or until the server decides to stop.

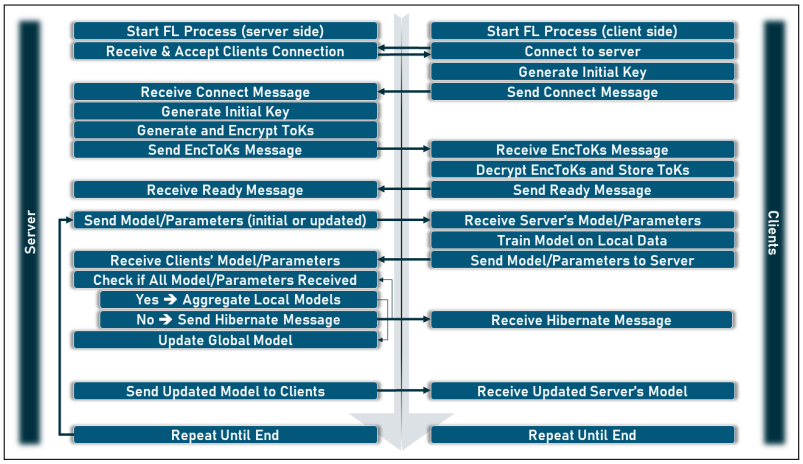

Figure 3: PolyFLAM & PolyFLAP followed workflow

4. Experimental Evaluation and Discussion

By using polymorphic encryption, this study provides two FL frameworks with increased resistance to inference attacks, strengthening the secrecy of messages exchanged within a federated learning cycle. This section focuses on an in-depth evaluation of the proposed innovative frameworks: PolyFLAM & PolyFLAP. It is worth noting that while the proposed frameworks undoubtedly create a secure environment for FL efforts, it is better to consider including authentication services in future revisions. This proactive step ensures that the FL system has a robust defence mechanism against potentially malicious entities. Although beyond the scope of this study, the scope of authentication services includes, but is not limited to: the traditional foundation of password-based authentication, the additional security layer of two-factor authentication (2FA), the robust security of public key infrastructure (PKI), simplified access through single sign-on (SSO), the innovative area of biometric authentication, and a variety of other options [64].

4.1. Security Analysis

The messages are encrypted in the proposed framework using the AES -256 algorithm, which is widely considered to be one of the most unassailable

cryptographic systems known today. The cryptographic key of this algorithm has a length of 32 characters and is nearly impenetrable, as there are an incredible 10^{77} possible variants for each individual key. According to [56], trying to crack such a key with the computing power of a supercomputer would take billions of years. Attacks using quantum computers, as described in "Quantum Attacks" [65], however, are already on the horizon and may break through the protective framework of AES, even if rapid key cracking is still a long way off.

To counter this emerging threat, each message exchanged within the domains of PolyFLAM & PolyFLAP is encrypted with a unique key taken from the Table of Keys (ToKs). At the same time, the basic initial_key, which is polymorphically generated by the process described earlier, strengthens the security of the ToKs by encryption. It is important to emphasize that the key management, which includes both the ToKs and the initial_key, uses a unique instantiation for each client and each subsequent connection session. This cautious approach also applies to situations where clients reconnect or different clients use the same connection at different times. Thanks to the randomness already explained, the probability of a key being reused is extremely low and approaches zero, so there is no risk of any of the keys used being leaked or cracked. In summary, the security of PolyFLAM & PolyFLAP is paramount: ***"Although AES -256 keys are very difficult to crack, the risk caused by a compromised or leaked key is almost zero, since this key is almost never used again during the FL cycle."***

4.2. Frameworks Complexity

Complexity analysis of PolyFLAM & PolyFLAM framework is a critical examination of the efficiency and computational requirements of these solutions. By evaluating the time complexity of essential processes such as communication, encryption, and aggregation, a comprehensive understanding emerges that provides insights into the scalability and performance characteristics of the proposed federated learning frameworks. To get a better overview of the complexity analysis of PolyFLAM & PolyFLAP, it is important to be aware of the different functions and processes involved in these frameworks. Figure 4 shows the different threads and functions involved in the execution of both frameworks, which are similar in both frameworks, except for the differences in the messages exchanged between server and

clients, where were models are exchanged in PolyFLAM and parameters are exchanged in PolyFLAP.

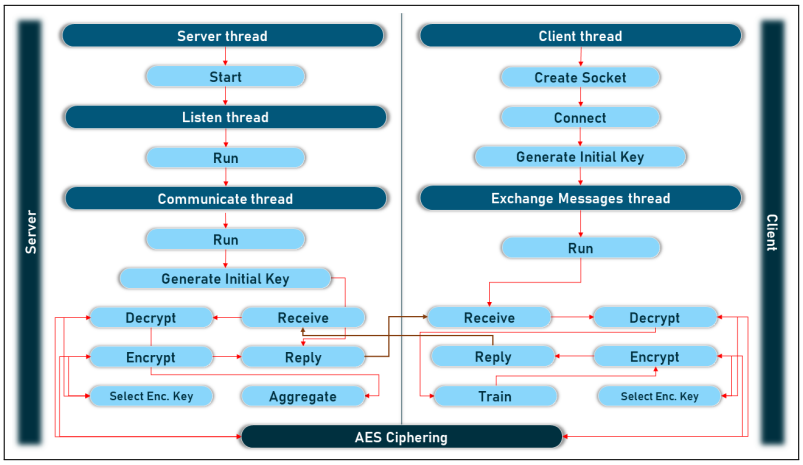

Figure 4: PolyFLAM & PolyFLAP threads and functions

In this context, it is crucial to clarify that the functions executed at both server and clients can be summarized as below:

- Server: The functions executed are:
  1. Run server thread and start the FL Cycle (function S1)
  2. Run listen thread and await clients connection (function S2)
  3. Run the communication thread and exchange messages with clients (function S3)
  4. Generate Table of Keys (function S4)
  5. Receive and accept client's connection (function S5)
  6. Generate initial_key based on clients shuffled secret (function S6)
  7. Encrypt Table of Keys (function S7)
  8. Send Encrypted Table of Keys to client (function S8)
  9. Receive client's ready message (function S9)
  10. Decrypt Ready message (function S10)
  11. Encrypt Model/Parameters (function S11)
  12. Send Encrypted Model/Parameters to client (function S12)
  13. Receive trained Model/Parameters from clients (function S13)
  14. Check if Model/Parameters are received from all clients
      - if yes, aggregate all Models/Parameters (function S14)
      - if no, Encrypt and send Hibernate message to clients and await receiving all Models/Parameters (function S15)

15. Encrypt aggregated Models/Parameters (function S16)
16. Send Encrypted aggregated Models/Parameters to clients (function S17)
17. Repeat all steps from S13 to S17 until the global model converge

- Client: The functions executed are:
  1. Run client thread and create socket (function C1)
  2. Connect to server (function C2)
  3. Generate Initial Encryption Key (function C3)
  4. Run Exchange Messages thread (function C4)
  5. Encrypt "Connect" message (function C5)
  6. Send encrypted "Connect" message (function C6)
  7. Receive encrypted "Table of Keys" from server (function C7)
  8. Decrypt "Table of Keys" (function C8)
  9. Encrypt "Ready" message (function C9)
  10. Send encrypted "Ready" message (function C10)
  11. Receive encrypted "Model/Parameters" from server (function C11)
  12. Decrypt "Models/Parameters" from server (function C12)
  13. Train model using the local data (function C13)
  14. Encrypt "Model/Parameters" from local training (function C14)
  15. Send encrypted "Model/Parameters" to server (function C15)
  16. Receive and Decrypt the new message (function C16) and if the message is:
      - "Hibernate" await until receiving another "Model/Parameter" message (function C17)
      - "Model/Parameter" then repeat steps C13 to C17 as per the number of training rounds

4.2.1. Time Complexity

In the field of Federated Learning, PolyFLAM and PolyFLAP follow defined steps with complexities defined by the following parameters:

- N (number of participating clients)
- IK (generation of initial key)
- ToKs (Table of Keys size
- E (encryption/decryption factors)
- R (number of training iterations rounds)
- A (aggregation complexity)
- MP (model/parameters complexity)
- T (training on local data)

To describe the time complexity of the framework on the server side, the O() parameter is used to form the necessary formulas. This parameter, commonly known as Big-O notation, is a mathematical notation used to describe the upper bound of the growth rate of the time complexity of an algorithm as the size of the input data increases. For example, O(1) represents a simple operation such as the initiation of the FL cycle, which occurs once on the server. Other messages have a different complexity, as described below:

- messages of fixed sizes such as connect, ready and done are impacted by the number of participating clients: O(N)
- messages depending on number of rounds and participating clients which are:
  - hibernate message that are sent to all participating clients except for the last to send its parameters: R * O(N-1)
  - models or parameters messages exchanged between server and clients are sent to all clients during all training rounds: R * O(N)

Following the steps performed on the server side and using the notations described above, the complexity function on the server side can be described as follows:

$$\begin{aligned} Server_{Complexity} = O(1) + (O(IK) * O(N)) + \\ (O(E) * (O(ToKs) * O(N))) + \\ (R * O(E) * O(MP) * O(N)) + \\ (R * O(E) * O(A)) + (O(E) * O(N)) \end{aligned} \tag{1}$$

The time complexity analysis of the federated learning cycle executed on the client side involves a comprehensive evaluation of various operations, each of which is affected by different time complexities. Notable operations with constant time complexity, denoted as O(1), include client thread initiation. However, unlike the server, the operations are not multiplied by the number of clients, but by the number of training rounds. Consequently, the complexity on the client side can be represented as follows:

$$Server_{Complexity} = O(1) + (R * O(IK)) + (R * O(E) * O(MP) * O(N)) + R * O(T) + O(N) \quad (2)$$

The complexity profile shows that the efficiency of the federated learning cycle scales linearly with the number of clients and communication rounds. The linear nature of the complexity indicates that as the size of the inputs (number of clients, rounds of communication) increases, the time required for the process also increases proportionally. This is generally preferable to a quadratic or higher complexity, which would lead to a much higher time requirement as the size of the inputs increases.

4.3. Communication Overhead

Running the PolyFLAM and PolyFLAP frameworks introduces an inherent communication overhead as part of the orchestration between the server and the clients. This additional overhead results primarily stems from the additional messages that are exchanged, serving as the management keys for collaboration. These messages include important components such as the Table of Keys (ToKs), as well as other messages that signal readiness, hibernation, connection establishment, and disconnection. Particularly, a notable fraction of the overall overhead arises from the encrypted Table of Keys, since the other messages are of fixed and small sizes. Considering that the size of the Table of Keys (ToKs) messages is multiplied by 32 bytes due to the use of a 256-bit key (AES), and taking into account the number of clients participating in the federated learning process, the total communication overhead can be represented approximately as follows, where K is the number of encryption keys and C is the number of participating clients:

$$Communication_Overhead = C * K * 32Bytes \quad (3)$$

This equation summarises the major factors contributing to the communication overhead caused by PolyFLAM and PolyFLAP encryption mechanisms and message exchanges.

4.4. Model Accuracy and Convergence

It is important to emphasize that the PolyFLAM and PolyFLAP frameworks were developed with the goal of strengthening the security and integrity

of the federated learning environment, not improving learning quality. Although improving machine learning models is important, these frameworks were developed with the goal of providing effective protection against potential vulnerabilities and privacy breaches in FL decentralized collaborative learning scenarios. The frameworks protect sensitive data and confidential model information from potential attacks by relying on polymorphic encryption strategies that guarantee that an encryption key is never used twice within a FL cycle, even for the same client. This technique demonstrates a proactive strategy for establishing trust in FL contexts and ensures that the collaborative process takes place within a fortified, robust, and trust-driven framework.

4.5. Space & Storage Utilization

The PolyFLAM and PolyFLAP frameworks introduce an overhead in the exploration of spatial complexity that must be considered. Central to this overhead is the inclusion of Table of Keys (ToKs), a cryptographic cornerstone. While the basic memory components of PolyFLAM and PolyFLAP are consistent with the foundation of federated learning frameworks, ToKs have a noticeable impact. ToKs require additional storage, but also play an important role in securing messages exchanged between servers and clients.

4.6. Evaluation using Real-World Data

Real-world test data sets are used to evaluate the PolyFLAM and PolyFLAP frameworks. These frameworks, strengthened by cryptographic capabilities and precise orchestration, serve as the vanguard of federated learning in a world where theoretical ideas converge with practical implementations. The upcoming research aims to move beyond the conceptual level and into a realm where actual data is used to validate the usability and efficiency of the proposed frameworks.

4.6.1. Testing Environment

To evaluate the effectiveness of the proposed framework, a simulated federated learning network was built, delineated by its hardware and software components:

- Hardware Configuration: the simulated network configuration included a server equipped with an Intel Core i7 processor and 16 GB of memory.

This server, running Microsoft Windows 10 Home, managed the orchestration of the network. At the same time, the client role was performed by different computers, each with different hardware specifications to simulate the heterogeneity of the real world;

- Software Framework: PolyFLAM and PolyFLAP were developed using Python (ver. 3.9) as the basic programming language.

During data partitioning, the records described in the next section were divided among the different customers during each FL cycle. If a dataset contains 1000 records and four clients participate in the training cycle, a fair partitioning would mean that each client performs local training on 250 different datasets. This careful division of data ensures equality and a consistent basis for comparative evaluation throughout the implementation of the system.

4.6.2. Datasets Used

A wide range of datasets specifically selected for binary classification tasks were carefully used to thoroughly evaluate the effectiveness and robustness of the frameworks. The three datasets include a simulated dataset generated using the SKLearn dataset library [66]. Using this generated dataset, which includes 9,000 records and 20 data properties, the core capabilities of the frameworks are thoroughly tested. In addition, the SHAREEDB Cardiovascular Diseases prediction dataset [67] proves to be a critical component of the evaluation as it goes deeper into the real-world complexity domain. This dataset highlights the ability of the framework to adapt to real-world medical data, as it contains 139 records and 26 variables that capture the details of cardiovascular health. The Surgical binary classification dataset [68] expands on its contributions with 14,636 records and 24 features that complement the analysis. This diverse dataset highlights the framework's ability to handle the complexity of a more complex, real-world scenario. Together, these carefully selected datasets provide the framework for a thorough evaluation and allow for better examination of framework performance across multiple dimensions of complexity and scale.

4.6.3. Security Analysis: Proof of Polymorphism

A rigorous process was used to evaluate the resilience of the framework. The encryption keys for communications had to be constantly monitored and

verified. By prohibiting the reuse of encryption keys, the resilience of the framework to key reuse was tested at this critical stage. This was demonstrated with two clients using the initial data set. The original and five additional ToKs keys used for message encryption were carefully recorded and compared. Table 3 summarizes the conclusions of this PolyFLAM cryptographic investigation and gives the collected findings from this comprehensive experiment. This technique is critical to determine the cryptographic robustness of the framework and its ability to secure complicated data transmissions.

Table 3: Encryption keys polymorphism in **PolyFLAM**

-	**server**	**client 1**	**client 2**
Initial Key	depending on the client	btkzo1PLJsQjVVRxr0u7mytmup9proQ3	PmBlfl3k4k7rgoDm1etNWJ6IsWyKLezS
Randomly selected 5 keys	wpmDQnYv8ncZhvNKaeXUvtFlZ9pcuM2p	3PwImZIqZT8o2DQVBfKnpile6B7nwGcp	YDRf1hXOq5POw311LflEBl3zcCFak41t
	uHAgHdMLG9cqPlvqMkHItBwkWJTFzMZL	Bi7a1kZHKJbJdVgA1WcTEoJILoEPCj4j	6sqKWRZk9coIYXhH6uogaBCI8C8TRGNd
	Wlhty1PtLy86wxH4lTxhjFTbt7dhHdT6	YZDBazmbbBCGXox8KXQzepJH3N2sUOkG	QdjTzsKiH4EwOp6R3CZ8UC2U2l7r2tKG
	qkgC6eIwiyyAq0Hfv4ajOqpszeuYxUbu	srmBy6rUmfH23Qn8GEPhTM9egvMBSd8S	Q318ez6o1n2YZ1AKM1uTDaPPDDvjqQuj
	C0KFcCE18VeFkho8qwPzKu6DA6hoUZBY	qxr9Qz3sFIAYGWL69CwJtBsnTdcetaqn	ZEwkF5PSvAxCP6LVMHWEqegV1tWHgO4r

4.6.4. Communication Cost

As part of the study and evaluation of PolyFLAM & PolyFLAP, communication costs were tracked and recorded. The communication stream includes different message types, both on the server and on the client, as shown in the list below:

- Server will be sending the below messages to each client
  - "Encrypted ToKs" (S1)
  - "Model/Parameters" (S2)
  - "Hibernate" (S3)
  - "Disconnect" (S4)
- Client will be sending the below messages to the server
  - "Connect" (C1)
  - "Ready" (C2)
  - "Model/Parameters" (C3)
  - "Disconnected" (C4)

The subtleties of message size are closely related to parameters such as the number of training rounds (R) and the number of participating clients (C). Consequently, the quantification of communication costs can be succinctly formulated as follows:

$$CommunucationCost_{Server} = C * (S1 + R * (S2 + S3 + S4)) \tag{4}$$

$$CommunicationCost_{Client} = C1 + C2 + R * C3 + C4 \tag{5}$$

It is worth noting that messages S3, S4, C1, and C2 have a fixed size due to their characteristic properties. The variability of S1 depends on the dimensions of the number of keys in the ToKs, with each factor contributing 32 bytes. It is important to emphasise that most of the communication volume comes from S2 and C3, which encapsulate a complicated model/parameter size. The recorded communication costs when running PolyFLAM & PolyFLAP are shown in Table 4 and Table 5 below. The communication costs are tracked between the server and a randomly selected client in a randomly selected training round for messages sent and received in the three databases selected for testing. Table 4 shows the size of each message sent by the server and the client, as described below:

Table 4: PolyFLAM & PolyFLAP communication cost per message, model, and dataset.

Entity	Dataset			Simulated Dataset	SHAREEDB	Surgical-Binary Dataset
Sent (Server)	"Encrypted ToKs"			480	480	480
	"Model/Parameters"	PolyFLAM	SVM	88060	10481	75604
			LR	864	912	896
			NB	1498	1786	1690
			SGD	954	1002	986
			MLP	34697	38748	39270
		PolyFLAP	SVM	89294	8853	72022
			LR	470	510	502
			NB	1101	1293	1229
			SGD	560	608	592
			MLP	10173	11709	11197
	"Hibernate"			139	139	139
	"Disconnect"			136	136	136
Sent (Client)	"Connect"			83	83	83
	"Ready"			91	91	91
	"Model/Parameters"	PolyFLAM	SVM	94279	11006	80119
			LR	926	974	990
			NB	1678	1966	1902
			SGD	1166	1214	1230
			MLP	37758	42350	39582
		PolyFLAP	SVM	89293	8782	71975
			LR	462	510	494
			NB	1054	1246	1214
			SGD	510	558	574
			MLP	10174	11710	11198
	"Disconnected"			107	107	107

On the other hand, Table 5 shows the total messages exchanged between server and client and also the reduction in communication cost between PolyFLAM and PolyFLAP one by one:

Table 5: PolyFLAM & PolyFLAP communication cost aggregation and reduction ratio

Framework	Direction	Model	Simulated Dataset	SHAREEDB	Surgical-Binary Dataset
PolyFLAM	Sent (Server)	SVM	88815	11236	76359
		LR	1619	1667	1651
		NB	2253	2541	2445
		SGD	1709	1757	1741
		MLP	35452	39503	40025
	Sent (Client)	SVM	94560	11287	80400
		LR	1207	1255	1271
		NB	1959	2247	2183
		SGD	1447	1495	1511
		MLP	38039	42631	39863
PolyFLAP	Sent (Server)	SVM	90049	9608	72777
		LR	1225	1265	1257
		NB	1856	2048	1984
		SGD	1315	1363	1347
		MLP	10928	12464	11952
	Sent (Client)	SVM	89574	9063	72256
		LR	743	791	775
		NB	1335	1527	1495
		SGD	791	839	855
		MLP	10455	11991	11479
Cost Reduction	Sent (Server)	SVM	-1%	14%	5%
		LR	24%	24%	24%
		NB	18%	19%	19%
		SGD	23%	22%	23%
		MLP	69%	68%	70%
	Sent (Client)	SVM	5%	20%	10%
		LR	38%	37%	39%
		NB	32%	32%	32%
		SGD	45%	44%	43%
		MLP	73%	72%	71%

4.6.5. Learning Quality

The two frameworks PolyFLAM & PolyFLAP comprise a versatile ensemble of five different models tailored for machine learning training: Support Vector Machine (SVM), Logistic Regression (LR), Gaussian Naive Bayes (Gaussian NB), Stochastic Gradient Descent (SGD), and Multi-Layer Perceptron (MLP). To comprehensively evaluate the effectiveness and robustness

of these methods, a series of experiments were conducted on three different datasets described previously. The results were tracked and plotted as shown in Table 6. The acronyms used in the table can be described as follows:

- AC: Accuracy
- PR: Precision
- RE: Recall
- F1: F1 Score
- SP: Specificity
- NPV: Negative Predictive Value

Table 6: PolyFLAM & PolyFLAP Learning Quality Results

Framework	Model	Dataset	AC	PR	RE	F1	SP	NPV
PolyFLAM	SVM	Simulated	91.76%	98.75%	91.30%	89.36%	91.89%	94.44%
		SHAREEDB	48.57%	40.91%	7.38%	12.50%	89.43%	49.33%
		Surgical-Binary	66.48%	40.40%	95.24%	56.74%	57.86%	97.59%
	LR	Simulated	86.67%	85.29%	90.62%	87.88%	82.14%	88.46%
		SHAREEDB	45.71%	47.56%	87.70%	61.67%	4.07%	25.00%
		Surgical-Binary	72.53%	46.15%	66.67%	54.55%	74.45%	87.18%
	NB	Simulated	98.33%	100.00%	96.43%	98.18%	100.00%	96.97%
		SHAREEDB	50.20%	50.00%	95.90%	65.73%	4.88%	54.55%
		Surgical-Binary	68.13%	39.47%	71.43%	50.85%	67.14%	88.68%
	SGD	Simulated	90.00%	88.89%	88.89%	88.89%	90.91%	90.91%
		SHAREEDB	49.39%	42.86%	4.92%	8.82%	93.50%	49.78%
		Surgical-Binary	63.19%	35.79%	85.00%	50.37%	57.04%	93.10%
	MLP	Simulated	76.67%	70.00%	93.33%	80.00%	60.00%	90.00%
		SHAREEDB	49.39%	43.75%	5.74%	10.14%	92.68%	49.78%
		Surgical-Binary	64.84%	20.83%	10.00%	13.51%	85.61%	71.52%
PolyFLAP	SVM	Simulated	91.67%	96.00%	85.71%	90.57%	96.88%	88.57%
		SHAREEDB	48.16%	40.74%	9.02%	14.77%	86.99%	49.08%
		Surgical-Binary	42.31%	31.21%	84.62%	45.60%	25.38%	80.49%
	LR	Simulated	95.00%	97.22%	94.59%	95.89%	95.65%	91.67%
		SHAREEDB	62.04%	60.90%	66.39%	63.53%	57.72%	63.39%
		Surgical-Binary	42.86%	25.00%	86.84%	38.82%	31.25%	90.00%
	NB	Simulated	91.67%	92.11%	94.59%	93.33%	86.96%	90.91%
		SHAREEDB	48.57%	40.91%	7.38%	12.50%	89.43%	49.33%
		Surgical-Binary	74.73%	0.00%	0.00%	0.00%	100.00%	74.73%
	SGD	Simulated	91.67%	93.10%	90.00%	91.53%	93.33%	90.32%
		SHAREEDB	48.16%	33.33%	4.10%	7.30%	91.87%	49.13%
		Surgical-Binary	76.37%	68.42%	26.00%	37.68%	95.45%	77.30%
	MLP	Simulated	71.67%	90.48%	55.88%	69.09%	92.31%	61.54%
		SHAREEDB	53.06%	51.58%	93.44%	66.47%	13.01%	66.67%
		Surgical-Binary	28.57%	27.27%	96.00%	42.48%	3.03%	66.67%

The PolyFLAM and PolyFLAP frameworks were developed primarily not with the sole goal of improving learning quality, but rather with the goal of strengthening federated learning against potential attacks, especially inference attacks. Nonetheless, it is critical to recognise that learning quality retains its importance as a key metric, particularly in the context of machine

learning models intended for predictive applications. In particular, accuracy plays a prominent role as it is a critical criterion for the utility and effectiveness of a model. The results presented above justify an analysis from different points of view.
Moreover, there is a clear trend where datasets with a larger number of records consistently show higher accuracy across the five different models in both PolyFLAM and PolyFLAP. This result is consistent with expectations, as a larger volume of records in a dataset leads to additional data availability for local training at the client node. This subsequently leads to an improvement in the local training quality and also in the global model quality. In particular, the results observed with the simulated dataset are remarkable, showing accuracies that exceed the threshold of 90% across various quality parameters. In contrast, the surgical deepnet dataset achieves comparatively lower accuracy, at 76% during the optimal iteration. In turn, the SHAREEDB dataset exhibits the least pronounced performance, as its highest accuracy across models does not exceed 62%. This clearly shows the influence of dataset size on model performance and learning quality. This observed phenomenon can be discussed from two strategic perspectives:

- Potential to improve learning quality: observed results encourage improving PolyFLAM and PolyFLAP so that they perform well when dealing with relatively small data sets;

- encouragement of client contributions: Since the two proposed frameworks preserve user privacy, they can be considered as a solution to attract more participants to a FL training cycle, thus providing an opportunity to increase data availability and improve model performance.

In summary, the results highlight the potential of these frameworks to go beyond their primary focus on security and also contribute to improving the quality of learning. Moreover, the scalability and privacy-friendly characteristics of these frameworks suggest that they can provide even more robust results as the participating entities expand. Therefore, due to their different concepts, PolyFLAM and PolyFLAP cannot be compared to the state of the art of classical ML models applied to these datasets such as[69, 70, 71] due to the difference of concepts,but they can certainly be compared to the current approaches being taken to secure FL environments against attacks and threats.

4.7. Comparison to The State-of-The-Art

In this section, a thorough comparison will be presented between the novel federated machine learning frameworks (FL) presented in this study and existing state-of-the-art approaches. This comparison highlights the particular focus on data security and privacy achieved by incorporating polymorphic and homomorphic encryption techniques. This comparison is presented in Table 7 below:

Table 7: Comparison of FL Security Approaches

Criteria	**Proposed Frameworks**	**Homomorphic Encryption Only**	**SecureD-FL**	**SEAR**	**HeteroSAg**
Encryption Techniques	Polymorphic & Homomorphic Encryption	Homomorphic Encryption	Homomorphic Encryption	Trusted Execution Environment (TEE)	Homomorphic Encryption
Unique Encryption Keys for Parameters	Yes (Polymorphic Encryption)	No (Single Key)	Yes (Homomorphic Encryption)	Yes (TEE-Based Encryption)	No (Single Key)
Data Privacy & Access Control	Strong Data Privacy & Access Control	Limited Access Control	Strong Data Privacy & Access Control	Strong Data Privacy & Access Control	Limited Access Control
Security Against Key Compromises	Highly Resilient (Granular Key Usage)	Vulnerable to Key Compromise	Highly Resilient (Granular Key Usage)	Highly Resilient (TEE-Based)	Vulnerable to Key Compromise
Robustness Against Attacks	Multi-Layered Security Approach	Limited Security Layers	Multi-Layered Security Approach	Multi-Layered Security Approach	Enhanced Security Layers
Communication Efficiency	Efficient with Enhanced Security	Efficient but Less Granular	Efficient with Enhanced Security	Efficient with Hardware-Based TEE	Efficient with Enhanced Security
Byzantine Attack Resilience	Strong Resilience	Limited Resilience	Strong Resilience	Strong Resilience	Strong Resilience
Inference Attack Resilience	High Resilience	Limited Resilience	High Resilience	Limited	Moderate Resilience
Bandwidth Efficiency	Enhanced Efficiency	Standard Efficiency	Enhanced Efficiency	Enhanced Efficiency	Enhanced Efficiency

5. Challenges and Future Perspectives

As the federal learning environment evolves, a number of obstacles and interesting options for future development emerge. This chapter addresses the many challenges associated with the development, deployment, and evaluation of the proposed federated learning frameworks PolyFLAM and PolyFLAP. These issues range from heterogeneity to the requirement for effective communication methods. In addition, the chapter highlights the future prospects

that will face the proposed frameworks and, by extension, the entire field FL. Federated learning will change machine learning paradigms and data-driven innovation by addressing current problems and highlighting future opportunities.

5.1. Challenges

In the context of PolyFLAM and PolyFLAP, a number of challenges arise that are closely related to the implementation and refinement of these federated learning frameworks.

5.1.1. Heterogeneity

Heterogeneity poses a particular challenge for the proposed framework, especially due to the fact that it only supports "horizontal FL data" This notion covers scenarios where different clients process data with identical characteristics. Although the framework covers this particular data type well, it is important to recognise that alternative approaches were not considered or tested in our study. This highlights the need for further research to include the broader landscape of data heterogeneity.

5.1.2. Complexity and Computation Cost

The issue of complexity and processing cost is critical, especially given the costly nature of encryption techniques. Since these algorithms are intended to ensure the integrity of the transmitted data, they necessarily require significant computing resources. While ensuring the security of the data, the rigorous computations required for encryption increase the complexity of the framework. As a result, striking a balance between robust security measures and efficient data processing becomes a critical problem that requires new ways to reduce computational costs while maintaining the integrity of the system.

5.1.3. Scalability

The issue of scalability is a major problem due to the increased processing requirements of encryption. As the framework deals with an increasing number of clients, especially on the server side, the additional computational overhead can limit the scalability of the system. The influx of clients places additional demands on the server's processing capacity, which can lead to bottlenecks and performance degradation. To achieve smooth scalability,

it is critical to explore effective optimization approaches that reduce computational load while maintaining system responsiveness and supporting an increasing number of clients.

5.1.4. Learning Quality

One critical issue that emerges is the quality of learning in the proposed frameworks. While the main goal of PolyFLAM and PolyFLAP in this context is to improve security and robustness against inference attacks, the inherent tradeoff between security and learning quality should not be neglected. Emphasising security techniques such as encryption and privacy may divert attention from improving model learning quality. Striking a delicate balance between strong security and optimal learning outcomes is an ongoing problem that requires careful evaluation of the impact of security measures on the efficiency of the learning process and the future performance of the overall model.

5.1.5. Resources Limitations

Resource constraints pose a significant challenge, especially in the federated learning paradigm where clients typically operate with limited computational resources, which may be the case if the clients are smartphones or smart wearables instead of powerful computers. This challenge becomes even more significant when considering the implementation of the proposed framework. The demands imposed by encryption and other security measures could overwhelm the already limited resources of the clients. This scenario raises concerns about the feasibility and practicality of deploying the framework in real-world scenarios, given the potential burden on client devices. To overcome this challenge, strategies must be developed to optimize the use of available resources and ensure that the system remains operational while addressing the constraints of client environments.

5.2. Future Perspectives

In moving to future perspectives, it's important to acknowledge that the challenges discussed above are by no means new territory in academic discourse. Various researchers have addressed these obstacles and offer innovative solutions to be explored. With careful consideration, a promising path emerges in which the proposed PolyFLAM and PolyFLAP frameworks converge with established techniques. This convergence has the potential not only to overcome existing challenges, but also to usher in an era of increased efficiency and versatility for federated learning systems.

5.2.1. Handling Heterogeneity

Addressing the challenge of heterogeneity arising from different devices and data requires innovative approaches. A number of strategies can be explored to effectively manage this variability. One option is to use resource allocation techniques [72], which intelligently allocate computing resources based on the capabilities of individual devices. This approach optimizes the use of resources and enables a more balanced and efficient federated learning process. In addition, the integration of Meta-Learning methods [73] represents a promising avenue. Meta-learning allows models to learn and adapt quickly to new data distributions, and thus has the potential to improve the adaptability of the system to the heterogeneity of client devices and data sources. The synergistic fusion of these approaches with the proposed frameworks could lead to a more agile and effective framework for federated learning, capable of addressing the difficulties posed by heterogeneity.

5.2.2. Computation Cost & Time Reduction

The challenge of computational costs can be mitigated through the strategic use of various techniques. One notable approach is the use of parallel programming methods. By breaking down complex computations into smaller tasks that can be executed simultaneously, parallel programming makes more efficient use of the processing power of modern devices. This leads to an acceleration of model training and a reduction in computation time, effectively reducing the burden on computing resources. Incorporating parallel programming techniques into the proposed PolyFLAM and PolyFLAP frameworks has the potential to significantly reduce computational costs while improving system scalability and responsiveness.

5.2.3. Enhancing Scalability

The prospect of improving scalability is linked to effectively solving the problems of heterogeneity and computational cost. As these challenges are addressed through approaches such as resource allocation and parallel programming, a symbiotic relationship emerges. By addressing device and data heterogeneity, the system is enabled to serve a variety of clients. At the same time, reducing computational costs through techniques such as parallel programming ensures that the system remains responsive as the number of participants grows. This convergence of solutions paves the way for a more scalable federated learning system capable of accommodating significant numbers of clients while maintaining performance and efficiency. The

interplay of these strategies has the potential to create a robust and adaptable ecosystem that meets real-world needs. In addition, the concept of third-party vendors can be incorporated into the framework to move some tasks outside the server, such as key generation or encryption, and keep network management and aggregation under the control of the central server.

5.2.4. Boosting Learning Quality

The quality of learning results can be improved by using different techniques for data preprocessing on the client side. As data is prepared prior to training, strategic preprocessing steps can be incorporated to improve the quality of the input data. Techniques such as feature scaling, outlier removal, and data augmentation can be applied to improve the quality and relevance of the data. By ensuring that the data fed into the training process is well prepared and free of noise or irregularities, the overall learning quality can be greatly enhanced. The integration of these preprocessing techniques into the proposed PolyFLAM and PolyFLAP frameworks may have the potential to fine-tune the learning process in addition to their improved safety function, leading to improved model convergence and overall performance.

Conclusion

The PolyFLAM and PolyFLAP frameworks use polymorphic encryption to enhance the security of message exchanges between servers and clients in a federated learning context. The security guarantees arise from the diversity of encryption keys, with each server-client message encrypted with a different key. Therefore, in the event of key compromise, there is minimal risk as key reuse within the FL cycle is virtually eliminated. While PolyFLAM and PolyFLAP prioritize security, they incur additional computational and communication costs due to the computationally intensive encryption operations. However, they can synergize with established methods to parallelize computation, increase learning efficiency, account for heterogeneity, and improve scalability to overcome such challenges and improve their usability and feasibility.

References

[1] Turing, A.M. Computing machinery and intelligence. In Parsing the Turing Test; Springer: Dordrecht, The Netherlands, 2009; pp. 23–65.

[2] Hernández-Orallo, J.; Minaya-Collado, N. A formal definition of intelligence based on an intensional variant of algorithmic complexity. In Proceedings of International Symposium of Engineering of Intelligent Systems (EIS98), Tenerife, Spain, 11–13 February 1998; pp. 146–163.

[3] Frankish, K.; Ramsey, W.M. (Eds.). The Cambridge Handbook of Artificial Intelligence; Cambridge University Press: Cambridge, UK, 2014.

[4] Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. SN Comput. Sci. 2021, 2, 1–21.

[5] Sharma, N.; Sharma, R.; Jindal, N. Machine learning and deep learning applications-a vision. Glob. Transitions Proc. 2021, 2, 24–28.

[6] Pallathadka, H.; Mustafa, M.; Sanchez, D.T.; Sajja, G.S.; Gour, S.; Naved, M. Impact of machine learning on management, healthcare and agriculture. Mater. Today Proc. 2021, in press.

[7] Ghazal, T.M.; Hasan, M.K.; Alshurideh, M.T.; Alzoubi, H.M.; Ahmad, M.; Akbar, S.S.; Al Kurdi, B.; Akour, I.A. IoT for smart cities: Machine learning approaches in smart healthcare—A review. Future Internet 2021, 13, 218.

[8] Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.L. Machine learning for medical imaging. Radiographics 2017, 37, 505.

[9] Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. Future Internet 2019, 11, 94.

[10] Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine learning and deep learning methods for cybersecurity. IEEE Access 2018, 6, 35365–35381.

[11] Nagarhalli, T.P.; Vaze, V.; Rana, N.K. Impact of machine learning in natural language processing: A review. In Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, Tirunelveli, India, 4–6 February 2021; pp. 1529–1534.

[12] Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. Sensors 2018, 18, 2674.

[13] Larrañaga, P.; Atienza, D.; Diaz-Rozo, J.; Ogbechie, A.; Puerto-Santana, C.; Bielza, C. Industrial Applications of Machine Learning; CRC Press: Boca Raton, FL, USA, 2018.

[14] Paleyes, A.; Urma, R.G.; Lawrence, N.D. Challenges in deploying machine learning: A survey of case studies. ACM Comput. Surv. 2020, 55, 1–29.

[15] Char, D.S.; Shah, N.H.; Magnus, D. Implementing machine learning in health care—Addressing ethical challenges. N. Engl. J. Med. 2018, 378, 981.

[16] L'heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A. Machine learning with big data: Challenges and approaches. IEEE Access 2017, 5, 7776–7797.

[17] Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. Neurocomputing 2017, 237, 350–361.

[18] Injadat, M.; Moubayed, A.; Nassif, A.B.; Shami, A. Machine learning towards intelligent systems: Applications, challenges, and opportunities. Artif. Intell. Rev. 2021, 54, 3299–3348.

[19] Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. Future Internet 2019, 11, 94.

[20] Leskovec, J.; Rajaraman, A.; Ullman, J.D. Mining of Massive Data Sets; Cambridge University Press: Cambridge, UK, 2020.

[21] Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. Prod. Manuf. Res. 2016, 4, 23–45.

[22] Albrecht, J.P. How the GDPR will change the world. Eur. Data Prot. L. Rev. 2016, 2, 287.

[23] Parasol, M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams. Comput. Law Secur. Rev. 2018, 34, 67–98.

[24] Gray, W.; Zheng, H.R. General Principles of Civil Law of the People's Republic of China. Am. J. Comp. Law 1986, 34, 715–743.

[25] Chik, W.B. The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform. Comput. Law Secur. Rev. 2013, 29, 554–575

[26] Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, 4–7 March 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 265–284.

[27] El Emam, K.; Dankar, F.K. Protecting privacy using k-anonymity. J. Am. Med. Inform. Assoc. 2008, 15, 627–637.

[28] Li, P.; Li, J.; Huang, Z.; Li, T.; Gao, C.Z.; Yiu, S.M.; Chen, K. Multi-key privacy-preserving deep learning in cloud computing. Future Gener. Comput. Syst. 2017, 74, 76–85.

[29] Fredrikson, M.; Jha, S.; Ristenpart, T. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; (pp. 1322–1333).

[30] Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership inference attacks against machine learning models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), IEEE, San Jose, CA, USA, 22–26 May 2017; pp. 3–18.

[31] McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics PMLR, Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.

[32] Booher, D. Duane, Bertrand Cambou, Albert H. Carlson, and Christopher Philabaum. "Dynamic key generation for polymorphic encryption."

In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0482-0487. IEEE, 2019.

[33] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. "Reviewing Federated Machine Learning and Its Use in Diseases Prediction." Sensors 23, no. 4 (2023): 2112.

[34] Li, Tian, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. "Federated learning: Challenges, methods, and future directions." IEEE signal processing magazine 37, no. 3 (2020): 50-60. 1296

[35] Rahman, KM Jawadur, Faisal Ahmed, Nazma Akhter, Mohammad Hasan, Ruhul Amin, Kazi Ehsan Aziz, AKM Muzahidul Islam, Md Saddam Hossain Mukta, and AKM Najmul Islam. "Challenges, applications and design aspects of Federated Learning: 1298 A survey." IEEE Access 9 (2021): 124682-124700.

[36] Lyu, Lingjuan, Han Yu, and Qiang Yang. "Threats to federated learning: A survey." arXiv preprint arXiv:2003.02133 (2020).

[37] Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I., & Tygar, J. D. (2011, October). Adversarial machinelearning. In Proceedings of the 4th ACM workshop on Security and artificial intelligence (pp. 43-58).

[38] Shafahi, A., Huang, W. R., Najibi, M., Suciu, O., Studer, C., Dumitras, T., & Goldstein, T. (2018). Poisonfrogs! targeted clean-label poisoning attacks on neural networks. Advances in neural information processingsystems, 31.

[39] Gu, T., Dolan-Gavitt, B., & Garg, S. (2017). Badnets: Identifying vulnerabilities in the machine learningmodel supply chain. arXiv preprint arXiv:1708.06733.

[40] Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020, June). How to backdoor federatedlearning. In International Conference on Artificial Intelligence and Statistics (pp. 2938-2948). PMLR.

[41] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguingproperties of neural networks. arXiv preprint arXiv:1312.6199.

[42] Fung, C., Yoon, C. J., & Beschastnikh, I. (2018). Mitigating sybils in federated learning poisoning. arXivpreprint arXiv:1808.04866.

[43] Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019, May). Exploiting unintended feature leakage incollaborative learning. In 2019 IEEE symposium on security and privacy (SP) (pp. 691-706). IEEE.

[44] Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. Advances in neural information processingsystems, 32.

[45] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks againstmachine learning models. In 2017 IEEE symposium on security and privacy (SP) (pp. 3-18). IEEE.

[46] Bonawitz, Keith, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. "Practical secure aggregation for privacy-preserving Machine Learning." In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1175-1191. 2017.

[47] Pillutla, Krishna, Sham M. Kakade, and Zaid Harchaoui. "Robust aggregation for Federated Learning." IEEE Transactions on Signal Processing 70 (2022): 1142-1154.

[48] Weiszfeld, Endre, and Frank Plastria. "On the point for which the sum of the distances to n given points is minimum." Annals of Operations Research 167, no. 1 (2009).

[49] Jeon, Beomyeol, S. M. Ferdous, Muntasir Raihan Rahman, and Anwar Walid. "Privacy-preserving decentralized aggregation for Federated Learning." In IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 1-6. IEEE, 2021.

[50] Boyd, Stephen, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. "Distributed optimization and statistical learning via the alternat- ing direction method of multipliers." Foundations and Trends® in Machine learning 3, no. 1 (2011): 1-122.

[51] Zhao, Lingchen, Jianlin Jiang, Bo Feng, Qian Wang, Chao Shen, and Qi Li. "Sear: Secure and efficient aggregation for byzantine-robust Federated Learning." IEEE Transactions on Dependable and Secure Computing 19, no. 5 (2021): 3329-3342.

[52] McKeen, Frank, Ilya Alexandrovich, Alex Berenzon, Carlos V. Rozas, Hisham Shafi, Vedvyas Shanbhogue, and Uday R. Savagaonkar. "Innovative instructions and software model for isolated execution." Hasp@ isca 10, no. 1 (2013).

[53] Song, Jingcheng, Weizheng Wang, Thippa Reddy Gadekallu, Jianyu Cao, and Yining Liu. "Eppda: An efficient privacy-preserving data aggre- gation Federated Learning scheme." IEEE Transactions on Network Science and Engineering (2022).

[54] Benaloh, Josh Cohen. "Secret sharing homomorphisms: Keeping shares of a secret secret." In Advances in Cryptology—CRYPTO'86: Pro- ceedings, pp. 251-260. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.

[55] Elkordy, Ahmed Roushdy, and A. Salman Avestimehr. "HeteroSAg: Secure aggregation with heterogeneous quantization in Federated Learning." IEEE Transactions on Communications 70, no. 4 (2022): 2372-2386.

[56] Daemen, Joan, and Vincent Rijmen. "Reijndael: The advanced encryption standard." Dr. Dobb's Journal: Software Tools for the Professional Programmer 26, no. 3 (2001): 137-139.

[57] Alenezi, Mohammed N., Haneen Alabdulrazzaq, and Nada Q. Mohammad. "Symmetric encryption algorithms: Review and evaluation study." International Journal of Communication Networks and Information Security 12, no. 2 (2020): 256-272.

[58] Daemen, Joan, and Vincent Rijmen. "Reijndael: The advanced encryption standard." Dr. Dobb's Journal: Software Tools for the Professional Programmer 26, no. 3 (2001): 137-139.

[59] Hearst, Marti A., Susan T. Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. "Support vector machines." IEEE Intelligent Systems and their applications 13, no. 4 (1998): 18-28.

[60] LaValley, Michael P. "Logistic regression." Circulation 117, no. 18 (2008): 2395-2399.

[61] Hand, David J., and Keming Yu. "Idiot's Bayes—not so stupid after all?." International statistical review 69, no. 3 (2001): 385-398.

[62] Ketkar, Nikhil, and Nikhil Ketkar. "Stochastic gradient descent." Deep learning with Python: A hands-on introduction (2017): 113-132.

[63] Murtagh, Fionn. "Multilayer perceptrons for classification and regression." Neurocomputing 2, no. 5-6 (1991): 183-197.

[64] Barkadehi, Mohammadreza Hazhirpasand, Mehrbaksh Nilashi, Othman Ibrahim, Ali Zakeri Fardi, and Sarminah Samad. "Authentication systems: A literature review and classification." Telematics and Informatics 35, no. 5 (2018): 1491-1511.

[65] Bonnetain, Xavier, Mar´ıa Naya-Plasencia, and Andre´ Schrottenloher. "Quantum security analysis of AES." IACR Transactions on Symmetric Cryptology 2019, no. 2 (2019): 55-93.

[66] sklearn.datasets.make classification. Scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.datasets.make_ classification.html. (Accessed on 15 Feb. 2023)

[67] Smart Health for Assessing the Risk of Events via ECG Database v1.0.0. Smart Health for Assessing the Risk of Events via ECG Database v1.0.0. https://physionet.org/content/shareedb/1.0.0/. (Accessed on 1 March 2023).

[68] Dataset Surgical binary classification. Dataset Surgical Binary Classification — Kaggle. https:///datasets/omnamahshivai/ surgical-dataset-binary-classification. (Accessed on 15 March 2023)

[69] Lynch, Damian, and M. Suriya. "PE-DeepNet: A deep neural network model for pulmonary embolism detection." International Journal of Intelligent Networks 3 (2022): 176-180.

[70] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. "Cardiovascular events prediction using artificial intelligence models and heart rate variability." Procedia Computer Science 203 (2022): 231-238.

[71] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad." Predicting Cardiovascular Events with Machine Learning Models and Heart Rate Variability". International Journal of Ubiquitous Systems and Pervasive Networks (JUSPN), 18 no. 2 (2023): 49-59

[72] Jamil, Bushra, Humaira Ijaz, Mohammad Shojafar, Kashif Munir, and Rajkumar Buyya. "Resource allocation and task scheduling in fog computing and internet of everything environments: A taxonomy, review, and future directions." ACM Computing Surveys (CSUR) 54, no. 11s (2022): 1-38.

[73] Feng, Yong, Jinglong Chen, Jingsong Xie, Tianci Zhang, Haixin Lv, and Tongyang Pan. "Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects." Knowledge-Based Systems 235 (2022): 107646.

HP_FLAP: Homomorphic & Polymorphic Federated Learning Aggregation of Parameters Framework

---

This article is undergoing peer review process in the Cybersecurity Journal

HP_FLAP: Homomorphic & Polymorphic Federated Learning Aggregation of Parameters Framework

Abstract

Given the implications of data collection in machine learning research, protecting user privacy is paramount. Federated learning (FL), which involves sharing trained models instead of user data, has emerged as a solution in this context. However, FL faces security and privacy challenges, particularly in terms of vulnerability to inference attacks. Therefore, in this paper, a novel aggregation framework for federated learning called HP_FLAP is proposed as a countermeasure. HP_FLAP: "Homomorphic & Polymorphic Federated Learning Aggregation of Parameters" is supported by both homomorphic and polymorphic encryption to provide a secure environment for federated learning. This framework incorporates a variety of models, including logistic regression, Gaussian Naive Bayes, Stochastic Gradient Descent, and Multi-Layer Perceptron. In the proposed framework, security is enhanced by embedding homomorphic and polymorphic encryption. On the one hand, homomorphic encryption allows the server to summarize the encrypted collected parameters without decrypting them, which improves the security and privacy of the system. In addition, polymorphic encryption relies on a unique key polymorphism, where each set of parameters or messages is encrypted with a different key. Therefore, if a key has been compromised or leaked, it does not pose a threat to the overall security of the system. This paper provides a detailed description and evaluation of the proposed models.

Keywords: Federated Machine Learning, Aggregation Algorithms, Polymorphic Encryption, Homomorphic Encryption, Security, Privacy

1 Introduction

Artificial intelligence (AI) is a rapidly advancing technology that is increasingly finding its way into many fields, leading to significant advances in both personal and professional life. This fact is indisputable and needs no further review. Despite the long time devoted to the study of artificial intelligence, which can be traced back to the 1950s when Alan Turing asked his famous question "Can computers think?", there is no single accepted definition of [1]. Thus, in [2], the authors offer a concise definition of AI that characterizes it as software programs that exhibit competence comparable

to humans in a given context. On the other hand, the authors define in [3] AI as a set of tools and methods that draw on principles and mechanisms from various fields such as computation, mathematics, logic, and biology. These are used to solve the difficulties associated with achieving, representing, and simulating human intelligence and cognitive processes. Artificial intelligence has evolved into a broad field of study that has spawned several branches such as machine learning (ML), deep learning (DL), federated learning (FL), and others. ML using machine learning, for example, computers can gain knowledge from training data and incrementally expand their knowledge through implicit programming.
Machine learning algorithms aim to identify patterns in data and use this knowledge to make autonomous predictions later. In conventional environments, engineers develop computer programs and insert a set of instructions that enable the transformation of input data into the intended output. In contrast, ML is designed to acquire knowledge autonomously and improve its understanding over time with little or no human intervention. The remarkable performance of ML, as well as its extensive capabilities in solving classification and regression problems and its effectiveness in applying supervised and unsupervised learning techniques, have made it very attractive to researchers in a variety of fields [4, 5]. Subsequent studies have revealed a wide range of ML applications in various fields. These include e-commerce and product recommendation, image, speech and pattern recognition, user behavior analysis and context-aware smartphone applications [4, 5], health services [6–8], transportation [4, 9], Internet of Things (IoT) and smart cities [9], cybersecurity and threat intelligence [10], natural language processing [11], sustainable agriculture [12], industrial applications [13], and more.

1.1 Machine Learning Challenges

The increasing integration of classification and regression approaches in various fields is driven by the consistent achievement of accurate results. The effectiveness and versatility of using AI techniques, especially those based on machine learning, are well known. However, ML still encounters several obstacles that have been extensively discussed and analyzed in the academic literature. However, these problems cannot be classified into a single category, but are grouped based on certain aspects. This section presents the most prevalent problems, which are classified into a proposed framework that classifies them according to criteria related to data, models, implementation, and other general dimensions. This classification is reflected in Table 1 below:

As a result, researchers have studied the problems at ML and in related areas and are working together to solve them. It is difficult to say which of the above difficulties is more important or harmful to machine learning models. However, the workflow of ML often follows a certain order, namely data acquisition and preprocessing, feature engineering, model training, evaluation, and deployment. This workflow recognizes the importance of data to machine learning as the first step, without which the rest of the process cannot move forward. The availability of data is therefore also critical to the performance of machine learning models. Although model accuracy depends on technical architecture, data quality, feature processing, and other factors, higher

Table 1 Machine Learning challenges

Category	Challenge
General Challenges [14, 15]	User Data Privacy and Confidentiality User Technology Adoption and Engagement Ethical Constraints
Models Related Challenges [14, 15]	Accuracy and Performance Model Evaluation Variance and Bias Explainability Data Availability and Accessibility [18]
Data-Related Challenges [16, 17]	Data Locality [19] Data Readiness [18] Data Heterogeneity Noise and Signal Artifacts Missing Data Classes Imbalance Data Volume Course of Dimensionality Bonferroni principle [20] Feature Representation and Selection
Implementation Related Challenges [18, 21]	Real-Time Processing Model Selection Execution Time and Complexity

data availability for training has been shown to result in models with better accuracy [16, 17].

1.2 Privacy Preserving Federated Learning

Data collection can be said to be an important, if not the biggest, problem in developing real-world machine learning models for several reasons, including confidentiality and privacy. This problem extends not only to the privacy of individuals, but also to the privacy and security measures of society, government, and business. In this context, efforts have resulted in global regulations and laws, such as the European Union's General Data Protection Regulation (GDPR) [22], China's Cyber Security Law [23], Chinese Civil Law's General Principles [24], Singapore's PDPA [25], and many others that restrict collection of information.

These laws protect private data but complicate the ML process. Obtaining data for model training is increasingly difficult, which hinders model performance and tailored outcomes. Thus, privacy and confidentiality are not only problems in their own right, but also set in motion a number of other ML challenges. These include issues of data accessibility, model efficiency, personalization, and ultimately gaining people's trust and consent. The importance of protecting privacy in information sharing has led to the exploration of algorithms such as differential privacy, k-order anonymity, and homomorphic encryption (HE) [26–28]. However, these approaches have not proven invincible, as evidenced by machine learning-related attacks such as model inversion [29] and attacks on membership inference [30].

1.2.1 An overview of Federated Learning

Google has recently developed a revolutionary approach to machine learning called "Federated Machine Learning" or "Federated Learning" (FL) [31]that can help solve privacy problems. One of the main ideas behind Federated Learning is to eliminate the need to share user data across different platforms. This method uses decentralized, distributed, and collaborative strategies for training ML models while protecting individual privacy. In the FL environment, a model is trained without the need to send data from edge devices to a central server. Instead, models are sent to devices where they can be trained with their own local data. A central aggregation server then receives these improved models and merges them into a global model without having access to the granular embedded data. Figure1 below depicts the technical infrastructure of federated learning.

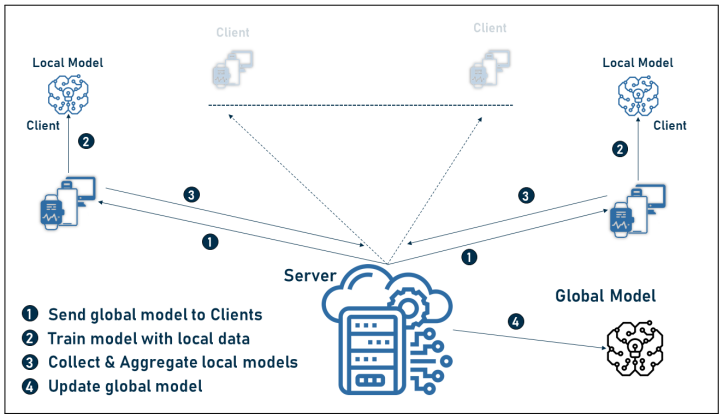

Fig. 1 Federated Learning Technical Architecture

Federated Learning solves user privacy problems. It addresses such limitations by allowing more data for training machine learning models, improving accuracy and efficiency. In addition, Federated Learning facilitates training of models with data from multiple sources called "data islands" These islands can send their datasets to a central model, increasing the overall efficiency of the models. In addition, federated learning allows models to handle different data from different data spaces, each with its own characteristics. This method also enables "learning transfer," where models can share information without sharing users' sensitive data. Federated learning, however, is still new and faces several challenges that require in-depth study to improve its capabilities.

Considering the above requirements, this paper presents a novel framework for federated learning that employs both Homomorphic Encryption[28] and Polymorphic Encryption[32] to enhance the security of the FL environment. Section 2 of this paper explores the main goal behind creating HP_FLAP by outlining the problem, especially the current shortcomings related to privacy in federated learning systems. It also explains the motivation behind the creation of HP_FLAP is elaborated upon. A crucial

element is the inclusion of both homomorphic and polymorphic encryption, which is seen as an innovative contribution in the field of federated learning. Section 3 aims to provide a thorough explanation of the proposed framework with a complete description of its internal operations. Section 4 presents a comprehensive analysis, critically evaluating and assessing the proposed framework from many perspectives. Practical investigations are conducted in real-world environments to validate its effectiveness. Finally, section 5 discusses the obstacles that hinder the proposed framework, but also identifies potential opportunities for its future development.

2 Problem Statement

Federated Learning presents itself as a solid privacy solution by taking a decentralized approach to machine learning, reducing the need for extensive data transfer between clients and servers. In this context, FL effectively reduces transmission costs by minimizing data transfer, as raw data often exceeds the size of the communicated models or their associated parameters.

2.1 Challenges and Issues in FL Domain

While federated learning has shown promise in a variety of contexts, it is also fraught with difficulties and challenges that have been extensively explored in the academic literature, e.g., in studies such as [33–35]. According to these and other studies, the early aggregation technique FL, FedAvg [31], has been investigated for a number of shortcomings. These challenges include issues such as data and hardware heterogeneity, sensitivity to local models, limits to scalability, slow convergence rates, complicated architectures, high communication costs, and the possibility of malicious attacks. Due to the complexity of these issues, research has proliferated to improve the usability of FL and make it accessible to a wide audience.

2.2 Security in FL Domain

Federated learning is vulnerable to malicious attacks [33–35]. Currently, there are three points of attack for the messaging of FL: the input phase, the learning phase, and the learned model itself. Therefore, many types of attacks are known in this context, including poisoning, inference, and backdoor attacks. While the quality of the learning results can be compromised by poisoning attacks, the user's privacy can be compromised by inference attacks, and the whole FL system can be attacked by backdoor attackers accordingly [36].

2.2.1 Poisoning Attacks

Poisoning attacks, whether random or targeted, attempt to reduce model accuracy (random) or alter the model to output a label specified by the attacker (targeted) [37]. These attacks can target either data or the model, in either case adversely affecting the overall performance and quality of the model. In addition, attackers can use compromised FL servers to launch both random and targeted poisoning attacks against trained models, including both data and poisoning attacks.

- Data Poisoning: or data corruption and has two main forms, 'Clean Label' [38] and 'Dirty Label' [39]. 'Clean Label' attacks require covert poisoning because the labels cannot change, while 'Dirty Label' attacks allow misclassification of data with target labels. Data poisoning by any federated learning entity can affect the final model depending on the number of attackers and data contamination;
- Model Poisoning: local model training can cause model poisoning by contaminating updates before server transmission or creating hidden global backdoors [40]. Targeted model poisoning manipulates the training process to misclassify selected inputs without modifying them, similar to adversarial attacks [41]. Model poisoning in federated learning affects model updates during each iteration more than data poisoning [42]. Centralized poisoning is simulated with a fraction of the training data. In addition, performing model poisoning requires significant technical resources.

2.2.2 Inference Attacks

Sharing models or parameters during training in federated learning also raises privacy issues [43, 44] . For example, Deep Learning models sometimes include data features in the exchanged models or parameters that aren't directly related to their main tasks. This could reveal personal data about users. In this context, attackers can find out features by comparing snapshots of model parameters, which shows the total changes made by everyone except the attacker. The problem is that slopes are computed from users' private data. Technically, the slopes in Deep Learning models come from the characteristics of the layers and the errors in the layers above them, which allows attackers to make inferences and invert some private user data. These observations can reveal private data characteristics such as class representatives and membership, or even enable label recovery without knowledge of the training set [44]. Inference attacks always involve the following:

- Inferring Membership: the goal of membership inference attacks is to determine whether a particular data element was used to train the model [45];
- Inferring Class Representatives: this occurs when a malicious participant intentionally compromises another participant and exploits the real-time learning of FL to train a network that generates private prototype samples of targeted training data. These generated samples mimic the distribution of the training data;
- Inferring Properties: in this attack, an attacker can perform both passive and active property inference attacks to infer properties of other participants' training data that are independent of the features describing the classes of the FL model. [45]:
  - Property inference attacks require the attacker to have additional training data labeled with the exact property they wish to infer;
  - A passive attacker can only monitor updates and make inferences by training a binary property classifier;
  - An active attacker can use multitasking to make the model FL learn to better separate data with and without the property, resulting in more information being extracted;
  - An adversarial participant can even infer when a feature appears and disappears in the data during training;

– It can recover pixel-precise original images and token-matched original texts.

2.3 Secured FL Frameworks: State of the Art

To increase the practicality and applicability of FL, researchers are looking for ways to improve the security of the system. There have been several attempts in this direction. To minimize computational costs and resist defective clients, the inventors of [46] have developed a technique with a specified number of rounds for secure vector summation. Their method relies on a single trusted server that stores all data from the communicating parties. High security was demonstrated against honest but curious adversaries, and anonymity was guaranteed even in the presence of active adversaries, such as an enemy server. In addition, Robust Federated Aggregation (RFA) was developed to protect the FL aggregation process from poisoning attempts [47]. To this end, a Weiszfeld-like technique was used to compute the geometric median of the traded models [48]. With its improved resistance to data poisoning attacks, RFA was able to compete with the standard FedAvg algorithm.
SecureD- FL, on the other hand, was developed by the authors of [49] and is a FL framework based on a modified version of the Alternating Direction Multiplier (ADMM) [50]. Their proposed framework uses a type of communication where the algorithm determines at the beginning of each execution cycle which subset of users should share data in order to limit the disclosure of sensitive information while still allowing for the efficient collection of data. Furthermore, [51] proposed SEAR,a Secure and Efficient Aggregation for byzantine-robust federated learning, which aggregates the local models in a secure and trusted hardware environment, specifically Intel SGX Trustworthy Execution Environment (TEE) [52], a secure CPU area, where the executed data and programs are kept secret and cannot be modified. Efficient Privacy-Preserving Data Aggregation (EPPDA) was also introduced in [53], which exploits homomorphisms of secret sharing in the FL environment, as shown in [54]. Secret sharing, a key exchange protocol, authenticated encryption, and signature mechanisms are the four pillars of their cryptography methodology.
Finally, the authors in [55] presented the aggregation method HeteroSAg, which employs masking to secure the exchanged messages in a way that ensures the mutual information between the masked model and the unique model is zero. The FL cycle, which implements a segment grouping method by partitioning edge users into groups and segmenting local model updates for those users, is crucial to HeteroSAg's robustness against Byzantine assaults. Table 2 below summarizes the security methods used by leading secured FL frameworks.

2.4 Problem and Motivation

While much work has been done to better handle poisoning threats in secure aggregation methods of federated learning, researchers have paid far less attention to inference attacks. Although methods such as polymorphic encryption and homomorphic encryption show promise in making data transmissions more secure and thus reducing the impact of inference attacks, they have not been extensively studied in previous research. This paper presents HP_FLAP, in response to growing concerns

Table 2 State-of-the-art of secured FL aggregation algorithms

Ref#	Mechanism
[46]	Secure Vector Summing Strategy
[47]	Using geometric median estimated using a Weiszfeld-type algorithm
[49]	Refined form of the Alternating Direction Multiplier (ADMM)
[51]	Hardware-based trusted execution environment instead of complex cryptography
[53]	Homomorphisms of the secret exchange
[55]	Masking each user's model update

about inference attacks and partly as a result of the success of PE and HE. Its originality and groundbreaking nature stems from the fact that it is the first framework to integrate both PE and HE with a secure FL aggregation. This partnership not only sets it apart from the competition, but also creates new opportunities to improve the security of FL. As will be shown later, the proposed framework also provides secure FL in a variety of paradigms.

2.5 Polymorphic Encryption

Polymorphism is the ability of objects or functions to change their form or behaviour to adapt to different circumstances. On the other hand, encryption involves the complex process of converting conventional data into an unintelligible format that prevents unauthorised access or use. Numerous encryption methods such as AES (Advanced Encryption Standard) and RSA (Rivest-Shamir-Adleman) contribute to complex data security [56, 57]. Polymorphic encryption is therefore defined as a complex encryption paradigm that adds a dynamic component by changing the encryption algorithm or keys, improving overall security. PE is resistant to intrusion, especially if the embedded encryption keys are too long. It is known that it is not easy to crack an encryption algorithm like AES by decrypting its key. Therefore, iterative use of an encryption key or even a cypher algorithm would certainly increase the level of security.

2.6 Homomorphic Encryption

Homomorphic encryption [28]is a new type of encryption that allows computations to be performed on encrypted data without having to decrypt it first. With traditional encryption methods, the data must be decrypted before it can be used. However, with homomorphic encryption, the data remains secret while mathematical processes can be performed directly on the encrypted data. There are different types of homomorphic encryption, such as partial homomorphic encryption and full homomorphic encryption. This discovery has huge implications for privacy and security, especially when it comes to cloud computing and data sharing. By using homomorphic encryption, private data can remain protected during processing. This can be used for secure data analytics, private machine learning and private outsourcing of data processing.

Although HE is not as fast as other methods because it requires difficult mathematical processes, researchers are still trying to make it faster and more useful for the real world. Homomorphic encryption can be divided into two main categories:

- Partial Homomorphic Encryption: cryptographic technique that allows certain mathematical operations, such as addition or multiplication, to be performed on encrypted data;
- Fully Homomorphic Encryption: is a cryptographic technique that enables the execution of various mathematical operations on encrypted data, which include both addition and multiplication operations. This method ensures that the secrecy of the data is maintained throughout the computational process.

In the proposed framework, HP_FLAP embeds Fully Homomorphic Encryption to allow a wide range of operations on parameters obtained from local training on each client.

3 HP_FLAP: FL Framework Secured with Homomorphic and Polymorphic Encryption

HP_FLAP framework was developed due to the urgent need to improve security protocols in federated learning frameworks against inference attacks, and the practicality of combining homomorphic and polymorphic encryption in response to this problem. This new idea cleverly combines the core ideas of homomorphic and polymorphic encryption in its design. This creates an impenetrable defense barrier around the FL landscape and ensures that it is completely secure. In this section, both the conceptual basis and the detailed design details of the proposed framework are explained in detail. This provides a complete picture of the many parts of the strategy, all of which serve to strengthen the integrity of the FL environment.

3.1 Main Concept

A traditional federated learning configuration has a central server and a set of clients. The server transmits a global model to the clients, which then train using their own local datasets. After the training process is complete, the clients transmit their trained models or associated parameters back to the server. The server, in turn, integrates the received models to generate an improved global model and iterates this process until an equilibrium state is reached. In the case of HP_FLAP, the exchanged messages are subjected to a particular type of encryption that combines both homomorphic and polymorphic encryption. In the proposed framework, the AES -256 algorithm [58] is embedded to encrypt the messages exchanged between the server and the participating clients. On the one hand, homomorphic encryption allows the server to aggregate the encrypted parameters without having to decrypt them, thus securing the parameters generated from the clients? local training data. On the other hand, polymorphic encryption ensures that different keys are used to encrypt different messages in the FL cycle.
The uniqueness of the proposed framework lies in its ability to aggregate multiple models or their associated parameters without requiring decryption and, by changing

the encryption key, secure messages exchanged between servers and clients. Furthermore, this is achieved by using different encryption keys for each message transmitted between the server and the clients. This method results in both homomorphism and polymorphism, which increases protection at two levels. Furthermore, in the context of a single client, unique keys are used for each message transmitted between that client and the server. The primary origin of polymorphism is the Table of Encryption Keys (ToKs) and the original encryption key. The following section elaborates on these principles, which serve as fundamental factors for the development of homomorphism and polymorphism within the proposed framework.

3.1.1 Encryption Keys Tables: ToKs and HEToKs

When a client makes a connection request, the server responds with a Table of Encryption Keys (ToKs). Each key in this table is assigned a unique ID for indexing. These keys play a crucial role in encrypting the exchanged messages. In this process, each message is assigned an index corresponding to the key used for encryption on the sender's side and decryption on the receiver's side. The AES -256 keys consist of 32 characters (bytes) and are extremely resistant to cracking attempts, which ensures a high level of security. In the context of HP_FLAP, even the case of a key being cracked or leaked does not pose a significant threat. This is because the implemented mechanism ensures that the compromised key, if any, is not reused in the FL process, either with the same customer or with other customers. This concept is explained in more detail in the following sections. In practise, a malicious client that successfully cracks a key would gain minimal benefit from it, since that key is unlikely to have any further use.
It's important to emphasise that each client is provided with its own ToKs when connecting to the server. Even the same client receives a new ToKs with each new connection session. Moreover, the transmission of ToKs to the client after the connection is established requires an additional encryption mechanism to protect against malicious entities. This precaution is critical because the effectiveness of the entire security scheme would be compromised if the ToKs were cracked or leaked. To counteract this, the initial encryption key, referred to as the "initial_key" and used to encrypt the ToKs, is also generated polymorphically, a mechanism that is explained in more detail in the following section.
In addition, the server is responsible for creating the Homomorphic Encryption Table of Keys (HEToKs) and distributing them to the participating clients. Similar to the ToKs, this table consists of a collection of 32-byte strings used to encrypt models or their associated parameters created after local training performed by the clients. The encrypted parameters are encapsulated in a message and transmitted to a server, where they are aggregated without decryption. The server then sends the aggregated parameters back to the clients so that they can continue with the training. In this current context, note the following about HEToKs:

- All clients receive identical HEToKs, and the keys in this table are each associated with a unique identifier;
- During each iteration of the global model aggregation process, the server uses a random selection mechanism to choose an identifier associated with one of the HEToKs

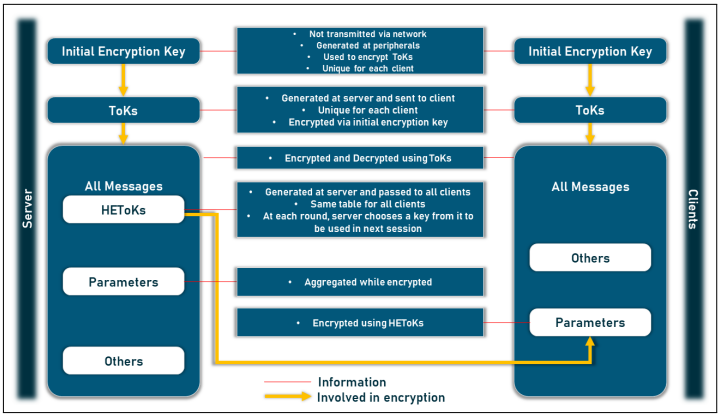

Fig. 2 Initial encryption key, ToKs and HEToKs explained graphically

keys. This ID is then sent to the clients to inform them of the specific key that will be used to generate polymorphism in the following round;

- HEToKs are securely encapsulated in the communication channel between servers and clients. These HEToKs are subjected to an additional layer of encryption using a randomly selected key from the ToKs instead of using the original encryption key.

In order to provide a more comprehensive explanation of the ideas of ToKs, initial encryption key, and HEToKs, Figure 2 below illustrates a graphical summary of the functionality associated with these concepts.

3.1.2 Initial Encryption Key

The encryption process uses the initial encryption key known as the "initial_key" to encrypt the Table of Encryption Keys. These ToKs are then used to encrypt messages exchanged between the server and clients. Since the ToKs are sensitive data, it is critical to create unique initial keys for each client to create more polymorphism and achieve better security. To achieve this, HP_FLAP utilize carefully defined technique to construct the initial key before it is used in the encryption process of the ToKs. It is significant that each connection session, even for the same client, generates a different key because random characters are used. It is worth noting that the transmission of these keys does not take place over the network. In contrast, the generation of these entities is done autonomously on both the server and client sides, using a uniform procedure. This strategy significantly increases the level of security. The procedure for generating the initial_key is uniform on both the server and client sides and is explained in more detail below for an entity, be it a server or a client:

1. After the connection is established, the client creates a 32-character string called "random_secret".
2. Then, the aforementioned string is combined with the client's connection information, which includes the IP address and associated address information. The

combination of these strings results in the creation of a new 32-character string that conforms to the following structure:

(a) The first 8 characters are obtained by inverting the last 8 characters of the random_secret.
(b) The following 4 characters are taken from the last 4 characters of the socket data.
(c) The following eight characters are composed of the middle eight characters of the randomly generated secret. The next four characters match the first four characters of the data received through the socket.
(d) Finally, the last eight characters are obtained by reversing the first eight characters of the random_secret.

By combining the previously described substrings, a string of 32 characters is created. The provided string serves as input to the SHA -256 algorithm [57], which generates a hash value. The initial_key, which is critical to the encryption process in ToKs, is obtained by extracting the first 32 characters from the resulting hash value. The use of the hash method increases security by reducing vulnerability to future cracking attempts. In addition, it should be noted that in the scenario where both the client and the server have access to the socket data, they are able to independently reproduce the necessary procedures to generate an identical key, provided they are given the same random secret (random_secret). However, since the random secret is randomly generated by the client, the probability of replicating a similar string on the server is quite low. Therefore, it is mandatory to communicate this confidential information to the server. To ensure secure transmission, the "shuffled_secret" is created using a series of sequential procedures and then sent to the server.

1. The first 8 characters are the opposite of the third 8 characters of random_secret
2. The second 8 characters are the first 8 characters of random_secret
3. The third 8 characters are the inverse of the last 8 characters of random_secret
4. The last 4 characters are the second 8 characters of random_secret

Following these steps renders the shuffled_secret unusable to malicious entities unless they know how to restore the original sequence. After receiving the shuffled_secret, the server reverses the shuffling operations to recover the original sequence and constructs the random_secret on the client's side. The server then matches the client's activities and repeats the same procedures to generate the first key. Since the server and client now have the same key, the Table of Encryption Keys may be encrypted by the server and sent to the clients. The client then decrypts these encrypted ToKs to protect the communication.
Note that even if a client connects numerous times with the same IP address, the initial_key will not be constant. This results from the use of random_secret during generation, as well as complex shuffling, mixing, and hashing steps. The initial_key generating procedure is shown in Figure 3 below.

3.2 Supported ML Models

HP_FLAP is a cutting-edge Federated Learning framework that opens up new possibilities for training smart models. To help end users effectively tackle a variety

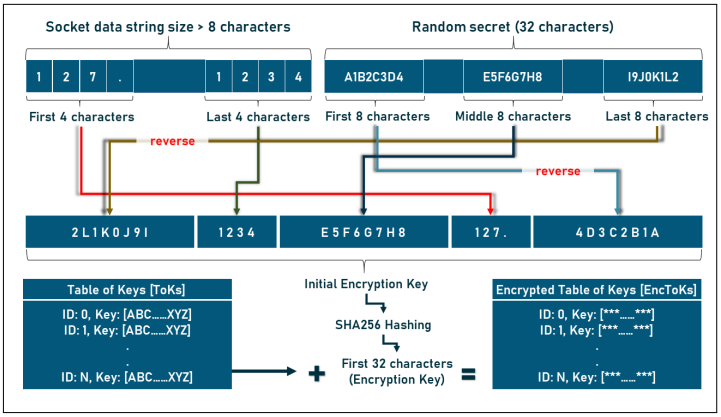

Fig. 3 Initial Encryption Key generation mechanism

of data-driven problems, this framework provides a choice of four separate machine learning models, which are:

- Logistic Regression [59]
- Gaussian Naive Bayes [60]
- Stochastic Gradient Descent (SGD Classifier) [61]
- Neural Network (Multi-Layer Perceptron) [62]

Taken together, these models support a wide range of ML tasks, leaving end users free to choose the model that best suits their needs. Since both models or their associated parameters can be exchanged or aggregated with a FL cycle, each of the above models generates a set of parameters during the local training process, which are explained in the Table 3 below.

During the process of federated learning, clients and servers share these parameters, which collectively reflect the key properties of their own models, to iteratively improve the global model. The essential details of the model are stored in the parameters that clients and servers share. These parameters are expertly merged and aggregated on the server side, even if they are encrypted, enabling iterative refinement of the global model. The collaborative nature of this process ensures that the aggregated knowledge of multiple clients plays a critical role in developing a global model that is more accurate and personalized.

3.3 Frameworks Design

In light of the information presented above, the workflow that is followed by HP_FLAP is explained in the following stages, which are also depicted in Figure 4 that can be found further below.

1. server starts FL process on its side;
2. server generate the HEToKs
3. client connects to the server;

Table 3 Parameters generated by each model on local training

Model	Parameter	Description
Support Vector Machines	Support vectors	data points that significantly influence the determination of the separating hyperplane
	Coefficients	weights assigned to features, contributing to the hyperplane's orientation
	Intercept	also known as the bias term, it shifts the hyperplane's position, aiding in better classification
Logistic Regression	Coefficients	weights determine the influence of individual features on the log-odds of the predicted outcome
	Intercept	bias term that adjusts the threshold for classifying instances
Gaussian Naive Bayes	Class priors	represent the prior probabilities of different classes in the training data
	Theta	mean values of features for each class, used in the Gaussian probability density function
	Sigma	variance of features for each class, also utilized in Gaussian probability calculations
SGD Classifier	Coefficients	similar to other models, these weights influence the classification decision
	Intercept	a bias term that adjusts the decision threshold
Multi Layer Perceptron	Coefficients	regulate the connections between neurons in the neural network layers
	Intercept	similar to bias terms in other models, it offsets the overall computation

4. client generates the random_secret and initial_key and sends the first to the server in a "Connect" message;
5. server receives the message and creates the table of random encryption keys (ToKs);
6. server regenerates the initial_key based on the received random_secret in the "Connect" message;
7. server encrypts the ToKs using the first 32 characters of the hashed initial_key
8. server wrap the HEToKs and encrypted ToKs in a message and sends them to the client;

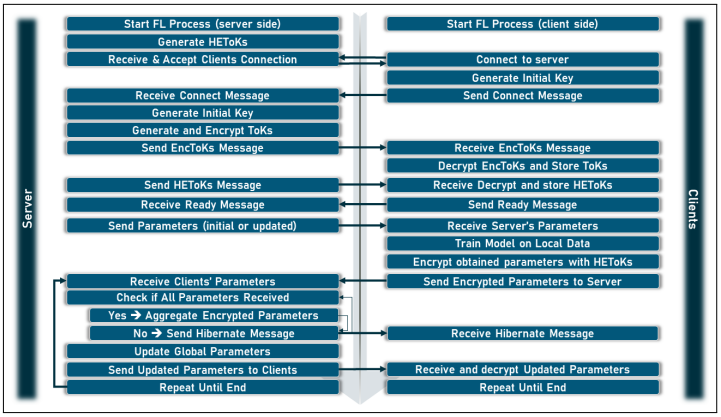

Fig. 4 HP_FLAP workflow

9. client receives the encrypted ToKs and decrypts them using initial key (After this step, the client selects an unused key from the ToKs to encrypt its message, and encapsulates the sent message with the ID of the used key);
10. client receive the HEToKs to be used later in encrypting the parameters
11. client replies to server with an encrypted "Ready" message;
12. server receives the message and responds with an initial "Model/Parameters" message;
13. client receives the first "Parameters" message and trains the model on the local data;
14. client encrypt the obtained parameters via a key obtained from HEToKs;
15. client replies to the server with its encrypted model parameters;
16. server checks if all clients have sent their parameters; and
    (a) If so, it starts the aggregation process without decrypting the parameters, updates the global parameters, elect a new HEToKs key to be used in next session, and sends all info back to the clients;
    (b) If not, it sends an encrypted ?Hibernate? message to the clients to wait until the above condition is met.
17. The clients receive the updated parameters, decrypt them and re-train their models based on them;
18. Repeat Steps 15, 16, and 17 until the model converges or until the server decides to stop.

4 Experimental Evaluation and Discussion

Message privacy within the federated learning system is strengthened by this research, introducing a novel FL framework that is resistant to inference attacks by combining homomorphic and polymorphic encryption. This section provides a detailed analysis and evaluation of the proposed framework. Despite the fact that HP_FLAP establishes a secure framework for FL activities, it is recommended that future versions

consider the incorporation of authentication services. This preventive measure ensures that the FL system has a robust defense mechanism against malicious entities. Traditional password-based authentication, the added security of two-factor authentication (2FA), the robust assurance of Public Key Infrastructure (PKI), the simplified access facilitated by Single Sign-On (SSO), the cutting-edge realm of biometric authentication, and a variety of others are beyond the scope of this study, but included in the authentication services scope[63]that may be considered in future versions of HP_FLAP.

4.1 Security Analysis

HP_FLAP encrypts messages using the AES -256 algorithm, one of the most secure cryptographic systems. The 32-digit cryptographic key of this algorithm is essentially unbreakable with 10^{77} unique permutations per key. According to [56], cracking this key with a supercomputer would take billions of years. Quantum threats, such as "quantum attacks" [64], however, could threaten the security of AES even if they have not yet succeeded in cracking the AES key in a fast way.
To counter this threat, the polymorphism of the encryption keys can serve as a solid solution. In the HP_FLAP context, messages are encrypted with a unique key from the Table of Keys (ToK). At the same time, the initial_key encrypts the ToKs for further security. It is important that each client uses a unique set of encryption keys, including the ToKs and the initial_key. In addition, the clients' parameters are encrypted, and the server aggregates them without decryption.
In addition, the keys, whether ToKs or the initial_key, are generated uniquely for each client and each connection session. Even if a client connects to the same socket twice, or two clients connect to the same socket at different times, the probability that the keys will be reused is close to zero due to the randomization described earlier. In summary, the theoretical guarantee of HP_FLAP is as follows: "AES -256 keys are known to be unbreakable. However, if a key is compromised or released, it does not pose a threat because it is almost never used again in the FL cycle. Furthermore, the parameters are aggregated while encrypted, without the need to decrypt them."

4.2 Framework Complexity

The complexity study of the 'HP_FLAP' framework entails a rigorous examination of the efficiency and computational requirements of the processes embedded within it. By evaluating the time complexity associated with basic operations such as communication, encryption, and aggregation, a comprehensive understanding is obtained that provides insight into the scalability and performance attributes underlying the proposed federated learning framework. To gain a comprehensive understanding of the complexity analysis of HP_FLAP, it is important to understand the various functions and processes of the framework. The diagram shown in Figure 5 illustrates the different threads and functions involved in the execution of HP_FLAP.

In this context, it is crucial to clarify that the functions executed at both server and clients can be summarized as below:

- Server: The functions executed are:

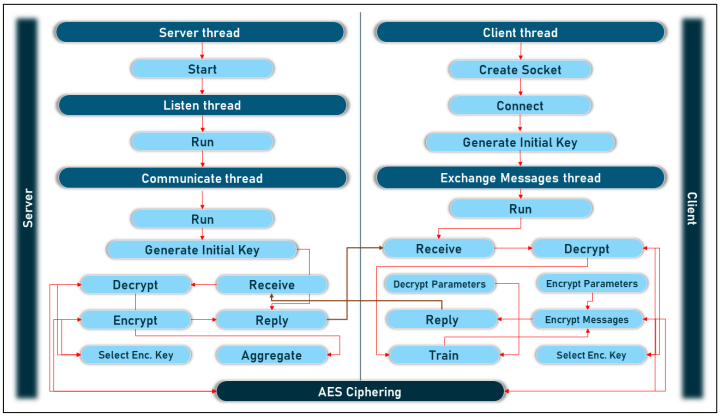

Fig. 5 HP_FLAP threads and functions

1. Run server thread and start the FL Cycle (function S1)
2. Generate Homomorphic Encryption Table of Keys HEToKs (function S2)
3. Run listen thread and await clients connection (function S3)
4. Run the communication thread and exchange messages with clients (function S4)
5. Generate Table of Keys (function S5)
6. Receive and accept client's connection (function S6)
7. Generate Initial Encryption Key based on clients shuffled secret(function S7)
8. Encrypt Table of Keys (function S8)
9. Send Encrypted Table of Keys and HEToKs to clients (function S9)
10. Receive client's ready message (function S10)
11. Decrypt Ready message (function S11)
12. Encrypt parameters (function S12)
13. Send encrypted parameters message to client (function S13)
14. Receive trained parameters message from clients (function S14)
15. Check if parameters are received from all clients
    - if yes, aggregate all encrypted parameters (function S15)
    - if no, encrypt and send Hibernate message to clients and await receiving all parameters (function S16)
16. Send Encrypted aggregated Parameters to clients (function S17)
17. Repeat all steps from S14 to S17 until the global model converge

- Client: The functions executed are:

1. Run client thread and create socket (function C1)
2. Connect to server (function C2)
3. Generate Initial Encryption Key (function C3)
4. Run Exchange Messages thread (function C4)
5. Encrypt "Connect" message (function C5)
6. Send encrypted "Connect" message (function C6)
7. Receive encrypted "Table of Keys" from server (function C7)

8. Decrypt "Table of Keys" (function C8)
9. Encrypt "Ready" message (function C9)
10. Send encrypted "Ready" message (function C10)
11. Receive encrypted "Parameters" message from server (function C11)
12. Decrypt "Parameters" message from server (function C12)
13. Train model using the local data (function C13)
14. Encrypt "Parameters" obtained from local training with HEToKs (function C14)
15. Embed encrypted parameters in a message and Encrypt it (function C15)
16. Send encrypted "Parameters" message to server (function C16)
17. Receive and Decrypt the new message (function C16) and if the message is:
    - "Hibernate" await until receiving another "Model/Parameter" message (function C17)
    - "Parameter" then Decrypt parameters (function C18)
18. repeat steps C13 to C18 as per the number of training rounds

4.2.1 Time Complexity

In the field of Federated Learning, HP_FLAP follow defined steps with complexities defined by the following parameters:

- N (number of participating clients)
- IK (generation of initial key)
- ToKs (Table of Keys size)
- HEToKs (Homomorphic Encryption Table of Keys)
- HED (Homomorphic Encryption/Decryption)
- E (encryption/decryption factors)
- R (number of training iterations rounds)
- HA (Homomorphic aggregation complexity)
- P (parameters complexity)
- T (training on local data)

To describe the time complexity of the framework on the server side, the O() parameter is used to form the necessary formulas. This parameter, commonly known as Big-O notation, is a mathematical notation used to describe the upper bound of the growth rate of the time complexity of an algorithm as the size of the input data increases. For example, O(1) represents a simple operation such as the initiation of the FL cycle, which occurs once on the server. Other messages have a different complexity, as described below:

- messages of fixed size such as connect, ready, and done depend on the number of participating clients O(N)
- messages that depend on the number of rounds and participating clients are:
  - hibernate message sent to all participating clients except the last one, which sends its parameters: R * O(N-1)
  - parameter messages exchanged between server and clients are sent to all clients during all training rounds: R * O(N)

Following the steps performed on the server side and using the notations described above, the complexity function on the server side can be described as follows

$$\begin{aligned} Server_{Complexity} = O(1) + (O(IK) * O(N)) + \\ (O(E) * (O(ToKs) * O(N))) + \\ (O(E) * O(HEToKs)) + \\ (R * O(E) * O(P) * O(N)) + \\ (R * O(E) * O(HA)) + (O(E) * O(N)) \end{aligned} \tag{1}$$

The time complexity analysis of the federated learning cycle executed on the client side involves a comprehensive evaluation of various operations, each of which is affected by different time complexities. Notable operations with constant time complexity, denoted as O(1), include client thread initiation. However, unlike the server, the operations are not multiplied by the number of clients, but by the number of training rounds. Consequently, the complexity on the client side can be represented as follows:

$$\begin{aligned} Server_{Complexity} = O(1) + (R * O(IK)) + \\ (R * O(E) * O(MP) * O(N)) + \\ R * O(T) + O(N) \end{aligned} \tag{2}$$

The complexity profile shows that the efficiency of the federated learning cycle scales linearly with the number of clients and communication rounds. The linear complexity profile indicates that as the size of the inputs (number of clients, rounds of communication) increases, the time required for the process also increases proportionally. This is generally preferable to a quadratic or higher complexity, which would lead to a much higher time requirement as the input size increases.

4.3 Communication Overhead

The communication overhead created by the operations of the HP_FLAP framework as part of the orchestration between the server and the clients is unavoidable. The additional communication operations that are exchanged to facilitate collaboration are largely responsible for this additional cost. Both the Table of Keys (ToKs) and the Homomorphic Encryption Table of Keys (HEToKs) are included in these messages, as are brief messages about readiness, hibernation, and connection establishment. The Tables of Keys ToKs and HEToKs account for a significant portion of the total communication overhead. Using a 256-bit key AES increases the size of the ToKs and HEToKs messages by a factor of 32. Multiplied by the number of clients participating in the federated learning process, the total communication cost can therefore be roughly expressed using the equation below, where K is the number of encryption keys and C is the number of participating clients:

$$Communication_Overhead = C * K * 32Bytes \tag{3}$$

This equation summarises the main factors contributing to the communication overhead caused by the encryption mechanisms and message exchanges within HP_FLAP.

4.4 Model Accuracy and Convergence

It is important to remember that HP_FLAP was not originally developed to improve the quality of learning, but to strengthen security and reliability in the federated learning environment, especially against inference attacks. This framework was developed with the intention of providing reliable protection against vulnerabilities and privacy violations in decentralized collaborative learning environments, regardless of the importance of improving machine learning models. The framework ensures that an encryption key is never reused with the same client or other clients during a FL cycle by employing homomorphic and polymorphic encryption techniques that make it difficult for intruders to gain access to sensitive data and private model information. Parameters learned from locally trained intelligent models are aggregated on the server without decryption, which provides additional security to the system. This method is an example of a preventive approach to building trust in FL environments and helps to ensure that all work is performed in a secure and trusted environment.

4.5 Space & Resource Utilization

In exploring the concept of space complexity, the HP_FLAP framework introduces a significant additional cost that is worthy of consideration and attention. The core of this framework is the inclusion of the Table of Keys (ToKs) and the homomorphic encryption table (HEToKs), which serves as the fundamental element of cryptography in this framework. The storage components of HP_FLAP are consistent with the basic principles of federated learning frameworks. However, the inclusion of ToKs and HEToKs comes at an additional storage cost. Even though they require this storage capacity, they perform a crucial function in securing the communication of messages between servers and clients, which is the main goal of this study.

4.6 Evaluation using Real-World Data

HP_FLAP framework is evaluated using test datasets derived from real-world scenarios. This framework, strengthened by advanced cryptographic techniques and careful coordination, acts as a precursor to FL in a context where theoretical concepts intersect with real-world applications. To test HP_FLAP, several tests were conducted, as described below.

4.6.1 Testing Environment

To evaluate the effectiveness of the proposed framework, a precisely designed simulated federated learning network was created, characterized by its individual hardware and software elements.

- Hardware configuration: the hardware arrangement consisted of a server running in a computing environment with an Intel Core i7 processor and 16 GB of random

access memory (RAM). The server, which ran the Microsoft Windows 10 Home operating system, played an important role in orchestrating the activities of the network. At the same time, clients were running on other computers, each with different hardware specifications to simulate real-world heterogeneity;
- software configuration: the development of HP_FLAP depended on the use of Python version 3.9 as the basic programming language. The complex development process was supported by the use of PyCharm, a powerful integrated development environment (IDE) designed for efficient code creation and maintenance.

In each round of testing, the data sets described in the following sections were divided equitably among the different clients. To give an example: If a dataset contains 1000 records and 4 clients are involved in the training cycle, a fair distribution would require each client to train locally on 250 different records. This careful distribution of data ensures uniformity and a stable metric against which to measure the performance of the system.

4.6.2 Datasets Used

Three data sets selected for binary classification tasks were used to evaluate the effectiveness and durability of the system. The three datasets contain a simulated dataset that was professionally produced using the SKLearn dataset library [65]. This produced dataset with its 9000 records and 20 data features is used to critically test the key capabilities of the framework. The SHAREEDB Cardiovascular Diseases prediction dataset [66] is another important component of the evaluation because it takes into account the complexity of the real world. With 139 records and 26 variables capturing the nuances of cardiovascular health, this dataset highlights the adaptability of the framework to real-world medical data. The Surgical binary classification dataset [67] adds by significantly increasing the amount of data available for analysis with 14,636 records and 24 features. The adaptability of the framework to a more complex, real-world scenario is underscored by this diverse dataset. Together, these carefully selected datasets form the basis for a comprehensive evaluation that provides an in-depth look at framework performance across multiple dimensions of complexity and scale.

4.6.3 Security Analysis: Proof of Polymorphism

A rigorous procedure was established for determining the cryptographic resilience of the frameworks under study. This involved close monitoring and evaluation of the encryption keys used for communication security. This step aimed to avoid vulnerabilities caused by the repetition of keys by prohibiting the reuse of encryption keys. In an exemplary test with two clients trained with the first data set, both the original encryption keys and five additional randomly selected message encryption keys were documented and compared. The results of this evaluation for HP_FLAP are shown in Table 4 and illustrate the valuable insights gained from this thorough procedure. This process is critical to confirming the cryptographic resilience of the framework and its ability to provide security and confidentiality for complex data transactions.

Table 4 Encryption keys polymorphism in **HP_FLAP**

-	server	client 1	client 2
Initial Key	depending on the client	7ZhSavGs4UumODkQJUrkcqX4xHsMXG8e	5o6NTq98PtXXclItETR0pFpp7q4zsv8I
Randomly selected 5 keys	S9thR5ZhU4FA2npqwQNj3dF1WivNuyJu	Crie0nwT0OM4QljHASR7wNHMPgbSSpUo	xvqAEzyDimNUW1SHCXbLepgwCvYs6Qzi
	QDVw410ohdBDRcLyDs9ooTunEuF0kNw7	Imz50JCPuQPa8244WkQFy3EZWeZlHytB	DZco40E0MdUJHJ6Illhv8lOHD3336hSQF
	WPweWn4Bv9FNat9trpdWK7GU14GJ8waU	SbnETWC4z9Nl26aoW23aLT6Gq8ZpWFUh	jrknL4MjK2O1zghzaMXqJ9SKs8Oox8CW
	a8pYNEdzELgN9IpthV1IM9iKX95fmXnY	2r8UJySrDduvF5KFdSUFGPLsg5gVUuGj	Ua1L6W0HaQSYPHfsfG0WW4Gw2x4zqy2j
	JV2ccd8dwlFCcS0fLn9bHnzDHVOdYsaX	VV8g2yvbOm8YSlwoGCSi8zbhc0uqi8bK	pzzqXBVACQqVtvsoCkwqP7BxIYqrfTMT

4.6.4 Communication Cost

As part of the study and evaluation of HP_FLAP, communication costs were tracked and recorded. The communication stream includes different message types, both on the server and on the client, as shown in the list below:

- Server will be sending the below messages to each client
  - "Encrypted ToKs" (S1)
  - Homomorphic Encryption Table of Keys HEToKs (S2)
  - "Parameters" (S3)
  - "Hibernate" (S4)
  - "Disconnect" (S5)
- Client will be sending the below messages to the server
  - "Connect" (C1)
  - "Ready" (C2)
  - "Parameters" (C3)
  - "Disconnected" (C4)

The complexity of message scaling is closely related to characteristics such as the number of training rounds (R) and the number of participating customers (C). Consequently, the quantification of the communication cost can be clearly stated as follows:

$$CommunucationCost_{Server} = C * (S1 + R * (S2 + S3 + S4 + S5)) \tag{4}$$

$$CommunicationCost_{Client} = C1 + C2 + R * C3 + C4 \tag{5}$$

It is worth noting that messages S4, S5, C1, and C2 have fixed sizes due to their unique characteristics. The variability of S1 and S2 is determined by the number of keys in the ToKs and HEToKs tables, with each factor contributing 32 bytes. It is important to emphasize that most of the communication overhead comes from S3 and C3, which contain complex parameter quantities. The recorded communication cost (in bytes) for running HP_FLAP is presented in Table 5 below. The communication costs are computed between the server and a randomly selected client over a randomly selected

Table 5 PHP_FLAP communication cost per message, model, and dataset (in Bytes).

Entity	Dataset			Simulated Dataset	SHAREEDB	Surgical-Binary Dataset
Sent (Server)	"Encrypted ToKs & HEToKs"			38098	38098	38098
	Parameters	HP_FLAP	LR	19022	23846	22252
			NB	67829	86914	80635
			SGD	19084	23849	22244
			MLP	975048	1128510	1077357
	"Hibernate"			139	139	139
	"Disconnect"			136	136	136
Sent (Client)	"Connect"			83	83	83
	"Ready"			91	91	91
	Parameters	HP_FLAP	LR	19084	23893	22222
			NB	67815	86983	80599
			SGD	19038	23822	22222
			MLP	975031	1128487	1077335
	"Disconnected"			107	107	107

training round, for both sent and received messages, and for the three databases used for testing.

4.6.5 Learning Quality

HP_FLAP is a versatile ensemble of four different training models for machine learning: logistic regression (LR), Gaussian Naive Bayes (Gaussian NB), Stochastic Gradient Descent (SGD), and Multi-Layer Perceptron (MLP). To thoroughly evaluate the effectiveness and robustness of this framework, a series of experiments were conducted on the three previously mentioned datasets. Table 6 shows how the results were tracked and recorded. The acronym of the table can be defined as follows:

- AC: Accuracy
- PR: Precision
- RE: Recall
- F1: F1 Score
- SP: Specificity
- NPV: Negative Predictive Value

The main goal of the HP_FLAP framework is not to improve the quality of learning, but rather to protect the Federated Learning system from possible attacks, especially those that take advantage of inference shortcomings. However, it is important to keep in mind that learning quality is still an important metric, especially when it comes to

Table 6 HP_FLAP Learning Quality Results

Model	Dataset	AC	PR	RE	F1SCORE	SP	NPV
LR	Simulated	92.40%	92.31%	93.47%	92.88%	91.19%	92.51%
	SHAREEDB	75.00%	70.00%	77.78%	73.68%	72.73%	80.00%
	Surgical Deepnet	72.62%	71.32%	74.62%	72.93%	70.68%	74.02%
NB	Simulated	90.67%	89.20%	92.53%	90.84%	88.80%	92.24%
	SHAREEDB	70.00%	85.71%	54.55%	66.67%	88.89%	61.54%
	Surgical Deepnet	65.40%	62.12%	66.67%	64.31%	64.29%	68.70%
SGD	Simulated	82.13%	80.79%	83.42%	82.09%	80.89%	83.51%
	SHAREEDB	75.00%	71.43%	62.50%	66.67%	83.33%	76.92%
	Surgical Deepnet	64.64%	72.00%	52.55%	60.76%	77.78%	60.12%
MLP	Simulated	90.27%	88.34%	92.41%	90.33%	88.19%	92.31%
	SHAREEDB	80.00%	88.89%	72.73%	80.00%	88.89%	72.73%
	Surgical Deepnet	65.40%	68.35%	45.00%	54.27%	82.52%	64.13%

machine learning models used for predictions. Accuracy is very important as it is one of the main factors that determines how useful and successful a model is. The above results deserve to be considered from many different points of view.

Primarily, there is a clear trend showing that datasets with more records provide more accurate results for all four models in the HP_FLAP framework. This result is consistent with expectations, as increasing the number of records in a dataset generally means increasing the amount of data that can be used for localized training at each client node. This change leads to an improvement in the quality of the local training, which in turn increases the quality of the global model as a whole. In particular, the results with the simulated dataset are interesting, as they show more than 90% accuracy across several quality variables. The SHAREEDB dataset, on the other hand, achieves only about 80% accuracy in its best version. The surgical deepnet dataset, on the other hand, has the weakest results, with the highest accuracy of all models not exceeding 72%. This contrast shows that the details of the dataset have a significant impact on how well a model performs.

This observed phenomenon can be dissected from two strategic perspectives:

- Potential to improve learning quality: The current study suggests that HP_FLAP could be improved in the future in a way that improves learning quality even in situations with relatively small data sets. This idea states that the built-in mechanisms of frameworks could be used to improve learning outcomes regardless of how large the data sets are.
- encouragement of client participation: The results also indicate that the effectiveness of the framework could be stronger in situations where more clients are involved. HP_FLAP is based on respecting users' privacy. This could encourage their participation in FL cycles, leading to a wider range of data sources, which could improve overall results.

In short, the results show that HP_FLAP may do more than just improve safety. Some improvements can also be made to improve the quality of learning. Also, the fact that this framework can grow and still maintain user privacy suggests that it could lead to even better results as the number of users using it grows. In summary, HP_FLAP can't be compared to the state of the art of classical ML models applied to these datasets, such as. [68–70], as they are based on different ideas. However, they

can be compared to how FL environments are currently protected from attacks and threats related to privacy and security.

4.7 Comparison to State-of-the Art Approaches

As discussed in the previous chapter, researchers around the world have used a variety of approaches to try to secure the FL environment. HP_FLAP, however, offers new approaches that can help improve the FL domain as a whole. Embedding both polymorphic and homomorphic encryption in FL helps secure messages exchanged between the main server and participating clients.

4.7.1 Proposed Frameworks vs. Baseline FL

FedAvg [31], the original FL algorithm, was proposed to train the popular Google keyboard. It enabled collaborative training without sharing raw data, preserving anonymity and supporting the development of decentralized machine learning applications. "Despite the practical privacy benefits, providing stronger guarantees through differential privacy, secure computation with multiple participants, or their combination is an interesting direction for future work" [46]. Those words came from the authors themselves, who clearly pointed out that FedAvg doesn't embed any security mechanisms in its architecture, apart from the concept of preventing data sharing between servers and clients. Later, several approaches to securing FL algorithms were implemented, but none of them integrated both polymorphic and homomorphic encryption to secure the FL environment.

4.7.2 Comparison with Securing Against Active Adversaries Approach

In [46], the authors presented a novel protocol for ensuring the security of federated learning aggregation in the presence of active attackers. This protocol is specifically designed to provide secure vector summation and includes key aspects such as fixed rounds, minimal communication cost, and the ability to withstand failures. However, it also has some limitations in terms of its ability to withstand active attacks, ensure the use of well-formed input, and manage communication overhead. These limitations need to be further studied and considered in the actual implementation of federated learning systems. However, HP_FLAP combines polymorphic and homomorphic encryption techniques to ensure the protection of data transmission between the server and clients. The combination of polymorphic and homomorphic encryption algorithms is an effective solution to some limitations identified in the previous [46] protocol, particularly with respect to the ability to fend off active attacks and maintain the integrity of data security in the federated learning environment.

4.7.3 Comparison with RFA

Compared to Robust Federated Aggregation (RFA) [47], HP_FLAP focuses on a new approach that seamlessly combines polymorphic and homomorphic encryption. The main goal of RFA is to improve the security of the FL system against poisoning attacks

by using geometric median-based aggregation. HP however However, HP_FLAP guarantees that communications between the server and clients are is secured by employing polymorphic encryption, which ensures that each message is encrypted with distinct keys. This enhances access control and maintains the confidentiality of the data. Furthermore, the utilization of homomorphic encryption for the aggregation of model parameters guarantees the preservation of data confidentiality with utmost rigour during the whole aggregation procedure. This technique demonstrates a high level of effectiveness in protecting against unauthorized access and potential privacy breaches, reinforcing RFA's focus on strengthening resilience against data poisoning attacks. Nonetheless, it is critical to recognize the complex tradeoffs that arise, as explained by the authors of the research article (RFA), which navigate the subtle interaction between the needs of ensuring resilience, optimizing communication efficiency, and protecting privacy in federated learning algorithms (FL).

4.7.4 Comparison to SecueD-FL

SecureD- FL [49] emphasizes decentralized aggregation with a focus on privacy preservation in federated learning (FL), while HP_FLAP focuses on increasing the security of FL through advanced encryption techniques. While SecureD- FL uses Alternating Direction Method of Multiplier (ADMM) and applies combinatorial block design theory to control participants' communication patterns and minimize privacy loss, HP_FLAP focuses on protecting data through encryption. SecureD- FL also seeks to reduce privacy risks and improve privacy against honest but curious adversaries by dynamically grouping participants for communication in each aggregation round. This combinatorial approach is designed to minimize privacy loss while efficiently aggregating model updates. In contrast, HP_FLAP focuses on encrypting all messages exchanged between the server and clients with unique keys thanks to polymorphic encryption to ensure privacy and access control. In addition, homomorphic encryption is used for secure aggregation of model parameters to ensure data confidentiality during aggregation. It is worth noting that SecureD- FL addresses the privacy concerns of FL by optimizing communication patterns and decentralizing aggregation, while HP_FLAP addresses security concerns mainly through encryption techniques. The choice between these approaches should depend on the specific security and privacy requirements of the particular FL application.

4.7.5 Comparison with SEAR

SEAR [51] highlights the server's ability to infer sensitive content from customer data, including the use of Generative Adversarial Networks (GANs) and gradient-based techniques. However, HP_FLAP prioritize data security by using polymorphic and homomorphic encryption. The algorithm SEAR uses the trusted execution environment (TEE) provided by Intel SGX to perform secure aggregation of locally trained models in a trusted hardware environment. This strategy protects the data by encrypting the local models and ensures that only the trusted enclave has the key required for access and recovery. This effectively prevents the disclosure of sensitive information during the aggregation process. However, the limited memory capacity of the reserved processor memory (PRM) in Intel SGX poses a major challenge when aggregating a

large number of models simultaneously.
Conversely, HP_FLAP employs polymorphic encryption to encrypt messages exchanged between clients and the server, employing distinctive keys for each message to amplify access control and data privacy. In addition, homomorphic encryption is used for aggregation of model parameters, maintaining their encrypted state and preserving data confidentiality throughout the aggregation process. These encryption techniques provide a robust and efficient approach to securing FL data without the limitations associated with hardware-based TEEs. Although both strategies address the security concerns of FL, they differ in their fundamental mechanisms. SEAR relies on hardware-based TEEs, while the proposed frameworks focus on encryption to enhance privacy and data security. The choice between these methods should be based on the exact security and privacy requirements of the particular FL application.

4.7.6 Proposed Frameworks vs. EPPDA

In [53], the authors present the EPPDA (Efficient Privacy-Preserving Data Aggregation) model, which exploits the homomorphisms of homomorphic encryption for secret sharing to streamline the iterations of secret sharing and reduce the consumption of communication, computation, and storage resources. This resource optimization is especially beneficial in scenarios with multiple training iterations and ultimately improves system efficiency. In addition, EPPDA incorporates secret sharing to protect user data, reduce the impact of malicious users, and increase fault tolerance.
On the other hand, HP_FLAP place a significant emphasis on privacy and security by encrypting messages exchanged between clients and the server with unique keys via polymorphic encryption and aggregating model parameters while keeping them encrypted via homomorphic encryption. These encryption techniques provide a comprehensive and robust security framework that protects FL data from various threats. Although both approaches address FL security concerns, they differ in their core mechanisms. The choice between these methods should be based on the specific security and privacy requirements of the FL application.

4.7.7 Proposed Frameworks vs. HeteroSAg

In contrast to HP_FLAP, the Heterogeneous Quantization approach presented in [55] is primarily concerned with communication efficiency and resistance to Byzantine attacks in the ecosystem FL. HP_FLAP relies primarily on advanced encryption techniques, including polymorphic and homomorphic encryption, to secure messages exchanged between the server and clients while ensuring privacy and access control. In contrast, HeteroSAg emphasizes privacy preservation, communication efficiency, and Byzantine fault tolerance through innovative techniques. Although both approaches aim to improve FL, their main goals differ significantly, with the proposed frameworks emphasizing data security and privacy, while HeteroSAg focuses on communication efficiency and Byzantine resilience. The choice between these approaches should depend on the specific security and privacy requirements of the particular FL application.

The comparison between the proposed frameworks and the state-of-the art of secured FL algorithms can be summarized in Table 7 below

Table 7 Comparison of FL Security Approaches

Criteria	Proposed Frameworks	SecureD-FL	SEAR	HeteroSAg
Encryption Techniques	Polymorphic & Homomorphic Encryption	Homomorphic Encryption	Trusted Execution Environment (TEE)	Homomorphic Encryption
Unique Encryption Keys for Parameters	Yes (Polymorphic Encryption)	Yes (Homomorphic Encryption)	Yes (TEE-Based Encryption)	No (Single Key)
Data Privacy & Access Control	Strong Data Privacy & Access Control	Strong Data Privacy & Access Control	Strong Data Privacy & Access Control	Limited Access Control
Security Against Key Compromises	Highly Resilient (Granular Key Usage)	Highly Resilient (Granular Key Usage)	Highly Resilient (TEE-Based)	Vulnerable to Key Compromise
Robustness Against Attacks	Multi-Layered Security Approach	Multi-Layered Security Approach	Multi-Layered Security Approach	Enhanced Security Layers
Communication Efficiency	Efficient with Enhanced Security	Efficient with Enhanced Security	Efficient with Hardware-Based TEE	Efficient with Enhanced Security
Byzantine Attack Resilience	Strong Resilience	Strong Resilience	Strong Resilience	Strong Resilience
Inference Attack Resilience	High Resilience	High Resilience	Limited	Moderate Resilience
Bandwidth Efficiency	Enhanced Efficiency	Enhanced Efficiency	Enhanced Efficiency	Enhanced Efficiency

5 Challenges and Future Perspectives

There are currently a variety of subtle challenges and exciting opportunities for future development in the federal learning environment. The difficulties encountered in creating the HP_FLAP Federated Learning Framework are discussed in this chapter. In addition, the chapter sheds light on the future by addressing potential developments that could help improve the proposed framework and, by extension, the entire FL field. By solving pressing problems and envisioning a promising future, Federated Learning is poised to revolutionize machine learning paradigms and data-driven innovation.

5.1 Challenges

In the setting of HP_FLAP, a unique collection of difficulties arises, all of which are intertwined with the development and deployment of this FL framework.

5.1.1 Heterogeneity

The aspect of heterogeneity is a major obstacle in the proposed framework, especially with respect to facilitating "Horizontal Federated Learning (FL) data" This term encompasses situations where different clients handle data that have identical attributes. Although the framework effectively handles this particular category of data, it is important to recognize that this framework does not encompass other potential methods. This underscores the complexity associated with accommodating different

data formats and highlights the need for additional research in the broader area of data heterogeneity.

5.1.2 Complexity and Computation Cost

The issue of complexity and computational cost is of paramount importance, especially given the resource-intensive nature of encryption methods. Because these algorithms are critical to maintaining the integrity of the transmitted data, they inherently require significant computational resources. Incorporating robust encryption algorithms into the operations of the framework adds an additional layer of complexity and thus increases data security measures. Consequently, the task of achieving a harmonious balance between strong security protocols and optimal computational efficiency is a key challenge that requires innovative approaches to minimize computational overhead while maintaining the integrity of the system.

5.1.3 Scalability

The scalability concern arises from the increased computational requirements associated with encryption. The ability of the framework to handle an increasing number of clients, especially on the server side, may be constrained by the additional processing costs, affecting the scalability of the system. The increasing number of clients accessing the server puts more strain on its processing capacity, which can lead to bottlenecks and a drop in performance. To achieve seamless scalability, it is imperative to explore efficient optimization strategies that reduce computational overhead while maintaining system responsiveness and accommodating an increasing number of clients.

5.1.4 Learning Quality

The quality of learning within the proposed framework is a crucial question that arises. The main goal of HP_FLAP is to improve security and resilience against inference attacks. However, it is important to recognize the inherent tradeoff between security and quality of learning, which should not be neglected. Prioritizing security measures, such as encryption and privacy, can detract from improving the quality of model learning. Maintaining a balance between robust security measures and achieving optimal learning outcomes is an ongoing challenge that requires careful evaluation of the impact of security measures on the effectiveness of the learning process and the future performance of the global model.

5.1.5 Resources Limitations

The presence of resource constraints is a major hurdle, especially in the context of federated learning. In this context, clients often work with limited computing resources, typically found in smartphones or smart wearables rather than more powerful computers. The importance of this difficulty becomes even more apparent when considering the practical application of the proposed framework. The use of encryption and other security measures may place additional strain on clients' limited resources. The current situation raises concerns about the feasibility and practicality of implementing

the framework in real-world situations, considering the potential burden on client devices. To overcome this difficulty, solutions need to be developed that maximize the efficient use of existing resources and ensure that the framework remains feasible for implementation, while taking into account the constraints of client settings.

5.2 Future Perspectives

When considering future improvements, it is critical to recognize that the above difficulties have been extensively addressed in academic debate. Several researchers have actively addressed these challenges and provided new answers that need further study. Careful consideration of the future prospects reveals a very interesting development in which the proposed framework fits harmoniously with established methods. The current convergence has the potential not only to remove current barriers, but also to increase the effectiveness and adaptability of FL systems, definitely HP_FLAP.

5.2.1 Handling Heterogeneity

Innovative solutions are needed to address the heterogeneity of different devices and data. Fortunately, many methods can be used to control this variability. For example, leveraging resource allocation methods [71]can intelligently distribute computing resources among devices based on their capabilities. This method optimizes resource usage for balanced and efficient federated learning. In addition, integrating meta-learning approaches [72] is also a promising approach. Meta-learning can improve the adaptability of the system to the heterogeneity of client devices and data sources by allowing the models to learn and adapt quickly to new data distributions. These techniques, together with the proposed framework, could provide a more flexible and effective framework for federated learning that can handle heterogeneity.

5.2.2 Computation Cost & Time Reduction

The problem of high computational cost can be reduced by careful implementation of a number of methods. The use of parallel programming techniques is one such strategy. Parallel programming makes better use of the computational capacity of modern devices by breaking large computations into smaller tasks that can be executed simultaneously. Faster model training and less time spent on computations means less strain on already overloaded computer systems. The proposed HP_FLAP framework could benefit from the use of parallel programming techniques to dramatically reduce computational costs while increasing scalability and responsiveness.

5.2.3 Enhancing Scalability

Improving scalability depends on finding workable solutions to the challenges of heterogeneity and high computational costs. A mutually beneficial relationship emerges when these difficulties are addressed with solutions such as resource allocation and parallel programming. The system's ability to serve a variety of users is enhanced by dealing with the heterogeneity of their devices and data. At the same time, reducing processing costs through methods such as parallel programming ensures that the

system remains responsive as the number of users increases. Together, these solutions pave the way for a federated learning system that can serve a large number of users without sacrificing performance. The combination of these tactics has the potential to create an ecosystem that can withstand the stresses of real life. In addition, outsourcing key creation and management to a third party can be a successful solution for improving scalability.

5.2.4 Boosting Learning Quality

Implementing a variety of client-side data preprocessing strategies can significantly improve the standard of learning performance. Strategic preprocessing methods can be used in data preparation prior to training to improve the quality of input data. Data quality and utility can be improved by using methods such as feature scaling, outlier removal, and data synthesis. The overall quality of learning can be greatly improved by ensuring that the data input to the training process is well prepared and free of noise or anomalies. Integrating data preprocessing with HP_FLAP can improve the learning process and lead to better model convergence and performance.

Conclusion

To ensure the confidentiality of communications between servers and clients in a federated learning environment, the HP_FLAP architecture uses homomorphic and polymorphic encryption. The diversity of encryption keys provides security guarantees by encrypting each server-client communication with a different key. The parameters of locally trained models are summarised on the central server without being decrypted, providing an additional layer of security. Since key reuse within the FL cycle is extremely rare, the consequences of a compromised key are small. HP_FLAP places a high value on security, but the complexity of encryption makes it an expensive computation and transmission method. However, they can complement established approaches to parallelize computation, improve learning efficiency, manage heterogeneity, and increase scalability.

Availability of data and material

Not Applicable

Funding

Not Applicable

Acknowledgements

We acknowledge the support of Natural Sciences and Engineering Research Council of Canada (NSERC), grant number 06351, Fonds Quebecois de la Recherche sur la Nature.

References

[1] Turing, A.M. Computing machinery and intelligence. In Parsing the Turing Test; Springer: Dordrecht, The Netherlands, 2009; pp. 23–65.

[2] Hernandez-Orallo, J.; Minaya-Collado, N. A formal definition of intelligence based on an intensional variant of algorithmic complexity. In Proceedings of International Symposium of Engineering of Intelligent Systems (EIS98), Tenerife, Spain, 11–13 February 1998; pp. 146–163.

[3] Frankish, K.; Ramsey, W.M. (Eds.). The Cambridge Handbook of Artificial Intelligence; Cambridge University Press: Cambridge, UK, 2014.

[4] Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. SN Comput. Sci. 2021, 2, 1–21.

[5] Sharma, N.; Sharma, R.; Jindal, N. Machine learning and deep learning applications-a vision. Glob. Transitions Proc. 2021, 2, 24–28.

[6] Pallathadka, H.; Mustafa, M.; Sanchez, D.T.; Sajja, G.S.; Gour, S.; Naved, M. Impact of machine learning on management, healthcare and agriculture. Mater. Today Proc. 2021, in press.

[7] Ghazal, T.M.; Hasan, M.K.; Alshurideh, M.T.; Alzoubi, H.M.; Ahmad, M.; Akbar, S.S.; Al Kurdi, B.; Akour, I.A. IoT for smart cities: Machine learning approaches in smart healthcare—A review. Future Internet 2021, 13, 218.

[8] Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.L. Machine learning for medical imaging. Radiographics 2017, 37, 505.

[9] Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. Future Internet 2019, 11, 94.

[10] Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine learning and deep learning methods for cybersecurity. IEEE Access 2018, 6, 35365–35381.

[11] Nagarhalli, T.P.; Vaze, V.; Rana, N.K. Impact of machine learning in natural language processing: A review. In Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, Tirunelveli, India, 4–6 February 2021; pp. 1529–1534.

[12] Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. Sensors 2018, 18, 2674.

[13] Larranaga, P.; Atienza, D.; Diaz-Rozo, J.; Ogbechie, A.; Puerto-Santana, C.; Bielza, C. Industrial Applications of Machine Learning; CRC Press: Boca Raton, FL, USA, 2018.

[14] Paleyes, A.; Urma, R.G.; Lawrence, N.D. Challenges in deploying machine learning: A survey of case studies. ACM Comput. Surv. 2020, 55, 1–29.

[15] Char, D.S.; Shah, N.H.; Magnus, D. Implementing machine learning in health care—Addressing ethical challenges. N. Engl. J. Med. 2018, 378, 981.

[16] L?heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A. Machine learning with big data: Challenges and approaches. IEEE Access 2017, 5, 7776–7797.

[17] Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. Neurocomputing 2017, 237, 350–361.

[18] Injadat, M.; Moubayed, A.; Nassif, A.B.; Shami, A. Machine learning towards intelligent systems: Applications, challenges, and opportunities. Artif. Intell. Rev. 2021, 54, 3299–3348.

[19] Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. Future Internet 2019, 11, 94.

[20] Leskovec, J.; Rajaraman, A.; Ullman, J.D. Mining of Massive Data Sets; Cambridge University Press: Cambridge, UK, 2020.

[21] Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. Prod. Manuf. Res. 2016, 4, 23–45.

[22] Albrecht, J.P. How the GDPR will change the world. Eur. Data Prot. L. Rev. 2016, 2, 287.

[23] Parasol, M. The impact of China?s 2016 Cyber Security Law on foreign technology firms, and on China?s big data and Smart City dreams. Comput. Law Secur. Rev. 2018, 34, 67–98.

[24] Gray, W.; Zheng, H.R. General Principles of Civil Law of the People?s Republic of China. Am. J. Comp. Law 1986, 34, 715–743.

[25] Chik, W.B. The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform. Comput. Law Secur. Rev. 2013, 29, 554–575

[26] Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, 4–7 March 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 265–284.

[27] El Emam, K.; Dankar, F.K. Protecting privacy using k-anonymity. J. Am. Med. Inform. Assoc. 2008, 15, 627–637.

[28] Li, P.; Li, J.; Huang, Z.; Li, T.; Gao, C.Z.; Yiu, S.M.; Chen, K. Multi-key privacy-preserving deep learning in cloud computing. Future Gener. Comput. Syst. 2017, 74, 76–85.

[29] Fredrikson, M.; Jha, S.; Ristenpart, T. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; (pp. 1322–1333).

[30] Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership inference attacks against machine learning models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), IEEE, San Jose, CA, USA, 22–26 May 2017; pp. 3–18.

[31] McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B. A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the Artificial Intelligence and Statistics PMLR, Lauderdale, FL, USA, 20–22 April 2017; pp. 1273–1282.

[32] Booher, D. Duane, Bertrand Cambou, Albert H. Carlson, and Christopher Philabaum. ?Dynamic key generation for polymorphic encryption.? In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0482-0487. IEEE, 2019.

[33] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. ?Reviewing Federated Machine Learning and Its Use in Diseases Prediction.? Sensors 23, no. 4 (2023): 2112.

[34] Li, Tian, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. "Federated learning: Challenges, methods, and future 1295 directions." IEEE signal processing magazine 37, no. 3 (2020): 50-60. 1296

[35] Rahman, KM Jawadur, Faisal Ahmed, Nazma Akhter, Mohammad Hasan, Ruhul Amin, Kazi Ehsan Aziz, AKM Muzahidul 1297 Islam, Md Saddam Hossain Mukta, and AKM Najmul Islam. "Challenges, applications and design aspects of Federated Learning: 1298 A survey." IEEE Access 9 (2021): 124682-124700.

[36] Lyu, Lingjuan, Han Yu, and Qiang Yang. "Threats to federated learning: A survey." arXiv preprint arXiv:2003.02133 (2020).

[37] Huang, L., Joseph, A. D., Nelson, B., Rubinstein, B. I., & Tygar, J. D. (2011, October). Adversarial machinelearning. In Proceedings of the 4th ACM workshop on Security and artificial intelligence (pp. 43-58).

[38] Shafahi, A., Huang, W. R., Najibi, M., Suciu, O., Studer, C., Dumitras, T., & Goldstein, T. (2018). Poisonfrogs! targeted clean-label poisoning attacks on neural networks. Advances in neural information processingsystems, 31.

[39] Gu, T., Dolan-Gavitt, B., & Garg, S. (2017). Badnets: Identifying vulnerabilities in the machine learningmodel supply chain. arXiv preprint arXiv:1708.06733.

[40] Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020, June). How to backdoor federatedlearning. In International Conference on Artificial Intelligence and Statistics (pp. 2938-2948). PMLR.

[41] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguingproperties of neural networks. arXiv preprint arXiv:1312.6199.

[42] Fung, C., Yoon, C. J., & Beschastnikh, I. (2018). Mitigating sybils in federated learning poisoning. arXivpreprint arXiv:1808.04866.

[43] Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019, May). Exploiting unintended feature leakage incollaborative learning. In 2019 IEEE symposium on security and privacy (SP) (pp. 691-706). IEEE.

[44] Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. Advances in neural information processingsystems, 32.

[45] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks againstmachine learning models. In 2017 IEEE symposium on security and privacy (SP) (pp. 3-18). IEEE.

[46] Bonawitz, Keith, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. ?Practical secure aggregation for privacy-preserving Machine Learning.? In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1175-1191. 2017.

[47] Pillutla, Krishna, Sham M. Kakade, and Zaid Harchaoui. ?Robust aggregation for Federated Learning.? IEEE Transactions on Signal Processing 70 (2022): 1142-1154.

[48] Weiszfeld, Endre, and Frank Plastria. ?On the point for which the sum of the distances to n given points is minimum.? Annals of Operations Research 167, no. 1 (2009).

[49] Jeon, Beomyeol, S. M. Ferdous, Muntasir Raihan Rahman, and Anwar Walid. ?Privacy-preserving decentralized aggregation for Federated Learning.? In IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 1-6. IEEE, 2021.

[50] Boyd, Stephen, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. ?Distributed optimization and statistical learning via the alternating direction method of multipliers.? Foundations and Trends in Machine learning 3, no. 1

(2011): 1-122.

[51] Zhao, Lingchen, Jianlin Jiang, Bo Feng, Qian Wang, Chao Shen, and Qi Li. ?Sear: Secure and efficient aggregation for byzantine-robust Federated Learning.? IEEE Transactions on Dependable and Secure Computing 19, no. 5 (2021): 3329-3342.

[52] McKeen, Frank, Ilya Alexandrovich, Alex Berenzon, Carlos V. Rozas, Hisham Shafi, Vedvyas Shanbhogue, and Uday R. Savagaonkar. ?Inno- vative instructions and software model for isolated execution.? Hasp@ isca 10, no. 1 (2013).

[53] Song, Jingcheng, Weizheng Wang, Thippa Reddy Gadekallu, Jianyu Cao, and Yining Liu. ?Eppda: An efficient privacy-preserving data aggre- gation Federated Learning scheme.? IEEE Transactions on Network Science and Engineering (2022).

[54] Benaloh, Josh Cohen. ?Secret sharing homomorphisms: Keeping shares of a secret secret.? In Advances in Cryptology?CRYPTO?86: Pro- ceedings, pp. 251-260. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.

[55] Elkordy, Ahmed Roushdy, and A. Salman Avestimehr. ?HeteroSAg: Secure aggregation with heterogeneous quantization in Federated Learn- ing.? IEEE Transactions on Communications 70, no. 4 (2022): 2372-2386.

[56] Daemen, Joan, and Vincent Rijmen. ?Reijndael: The advanced encryption standard.? Dr. Dobb?s Journal: Software Tools for the Professional Programmer 26, no. 3 (2001): 137-139.

[57] Alenezi, Mohammed N., Haneen Alabdulrazzaq, and Nada Q. Mohammad. ?Symmetric encryption algorithms: Review and evaluation study.? International Journal of Communication Networks and Information Security 12, no. 2 (2020): 256-272.

[58] Daemen, Joan, and Vincent Rijmen. "Reijndael: The advanced encryption standard." Dr. Dobb?s Journal: Software Tools for the Professional Programmer 26, no. 3 (2001): 137-139.

[59] LaValley, Michael P. "Logistic regression." Circulation 117, no. 18 (2008): 2395-2399.

[60] Hand, David J., and Keming Yu. "Idiot's Bayes?not so stupid after all?." International statistical review 69, no. 3 (2001): 385-398.

[61] Ketkar, Nikhil, and Nikhil Ketkar. "Stochastic gradient descent." Deep learning with Python: A hands-on introduction (2017): 113-132.

[62] Murtagh, Fionn. "Multilayer perceptrons for classification and regression." Neurocomputing 2, no. 5-6 (1991): 183-197.

[63] Barkadehi, Mohammadreza Hazhirpasand, Mehrbaksh Nilashi, Othman Ibrahim, Ali Zakeri Fardi, and Sarminah Samad. "Authentication systems: A literature review and classification." Telematics and Informatics 35, no. 5 (2018): 1491-1511.

[64] Bonnetain, Xavier, Maria Naya-Plasencia, and Andre Schrottenloher. ?Quantum security analysis of AES.? IACR Transactions on Symmetric Cryptology 2019, no. 2 (2019): 55-93.

[65] sklearn.datasets.make classification. Scikit-learn. https://scikit-learn/stable/modules/generated/sklearn.datasets.make_ classification.html. (Accessed on 15 Feb. 2023)

[66] Smart Health for Assessing the Risk of Events via ECG Database v1.0.0. Smart Health for Assessing the Risk of Events via ECG Database v1.0.0. https://physionet.org/content/shareedb/1.0.0/. (Accessed on 1 March 2023).

[67] Dataset Surgical binary classification. Dataset Surgical Binary Classification ? Kaggle. https:///datasets/omnamahshivai/ surgical-dataset-binary-classification. (Accessed on 15 March 2023)

[68] Lynch, Damian, and M. Suriya. "PE-DeepNet: A deep neural network model for pulmonary embolism detection." International Journal of Intelligent Networks 3 (2022): 176-180.

[69] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. "Cardiovascular events prediction using artificial intelligence models and heart rate variability." Procedia Computer Science 203 (2022): 231-238.

[70] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad.? Predicting Cardiovascular Events with Machine Learning Models and Heart Rate Variability?. International Journal of Ubiquitous Systems and Pervasive Networks (JUSPN), 18 no. 2 (2023): 49-59

[71] Jamil, Bushra, Humaira Ijaz, Mohammad Shojafar, Kashif Munir, and Rajkumar Buyya. ?Resource allocation and task scheduling in fog computing and internet of everything environments: A taxonomy, review, and future directions.? ACM Computing Surveys (CSUR) 54, no. 11s (2022): 1-38.

[72] Feng, Yong, Jinglong Chen, Jingsong Xie, Tianci Zhang, Haixin Lv, and Tongyang Pan. ?Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects.? Knowledge-Based Systems 235 (2022): 107646.

CHAPTER 6

GENERAL CONCLUSION

In conclusion, this thesis has explored the dynamic landscape of Artificial Intelligence (AI) and its derivative, Machine Learning, with a particular focus on the challenges of security and privacy that have emerged alongside the proliferation of data-driven technologies. The overarching theme of this research revolves around the pressing need to enhance the security of Federated Learning, a promising paradigm designed to preserve user privacy in a decentralized manner while maintaining the utility of smart models.

The growth of ML has undoubtedly been remarkable, fueled by advances in computing power and the versatility of the existing ML algorithms in solving complex data analysis tasks across various domains. However, as ML models become increasingly integrated into our daily lives and critical infrastructure, the issues of security and privacy have taken center stage. On one hand, the vulnerability of ML models to various attacks threatens the integrity of the learning process and the privacy of user data. On the other hand, privacy concerns have prompted regulations that restrict access to data, thereby limiting the potential of ML models.

In response to these challenges, FL emerged as a privacy-preserving approach by distributing models to clients for local training, eliminating the need to centralize sensitive user data. While FL is a significant step forward, it remains susceptible to security threats including poisoning, inference and backdoor attacks. While poisoning attacks can affect the quality of the learning in FL system, inference attacks can enable malicious entities to collect models exchanged between the FL main server and clients, thus cracking private data, which may ruin the privacy-preserving identity of FL. This thesis has highlighted the need to enhance the security measures within the FL domain, recognizing the potential of Polymorphic and Homomorphic Encryption as powerful tools to bolster FL's resilience against various attacks, especially inference attacks.

The introduction of four novel FL aggregation frameworks, namely PolyFLAG_SVM, PolyFLAM, PolyFLAP, and HP_FLAP, represents a significant contribution to the field. The first three frameworks embed Polymorphic encryption in their architecture. However, HP_FLAP embeds both Polymorphic and Homomorphic Encryption to secure FL against inference attacks. Integrating those encryption mechanisms ensure that messages exchanged between the server and clients remain safeguarded against malicious entities. The use of Homomorphic Encryption allows secure aggregation of parameters without decryption, while Polymorphic Encryption ensures that each message is encrypted with a distinct key, minimizing the risk of key compromise. This dual-layered security approach has been demonstrated to effectively counteract threats, including inference attacks. In addition, the proposed frameworks offer training different ML models to support solving different problems.

Furthermore, the comprehensive evaluation of these frameworks has provided strong empirical evidence of their efficacy. The assessment encompassed theoretical guarantees, time and space complexity analysis, resource utilization assessments, and learning quality evaluations across diverse datasets. The results have unequivocally demonstrated the substantial enhancement in security, even in the face of compromised or leaked encryption keys.

In summary, the proposed frameworks offer secure, communication-efficient FL aggregation approaches, serving as a foundation upon which further advancements and integrations with existing approaches can be built. These frameworks not only advance the security of Federated Learning but also contribute to the reliability and trustworthiness of the entire FL environment. As the field of AI and ML continues to evolve, the importance of addressing security and privacy concerns in tandem with technological advancements cannot be overstated. This thesis has made significant strides in this direction and lays the groundwork for future research in the quest to harness the full potential of smart models while safeguarding user data and privacy.

6.1 Comparison to State-of-the Art Approaches

Threats and attacks have hindered the progress of FL as much as ML. For this reason, researchers around the world have attempted to secure the FL environment using a variety of approaches, as described previously in chapter 4. However, the frameworks proposed in this research contain new approaches that can contribute to the overall improvement of the FL domain. The results of this research show that embedding both polymorphic and homomorphic

encryption in FL helps secure messages exchanged between the main server and participating clients.

6.1.1 Compared with Baseline FL

FedAvg was the first research paper to define and introduce the concept of Federated Learning [63]. "Despite its practical privacy benefits, providing stronger guarantees via differential privacy, secure multi-party computation, or their combination is an interesting direction for future work" [63]. These words come from the creators of FedAvg itself, where they clearly indicated that it didn't embed any security mechanisms in its architecture, apart from the concept of precluding data sharing between servers and clients. Later, several approaches to securing FL algorithms were implemented, but none of them integrated both polymorphic and homomorphic encryption to secure the FL environment. The framework proposed in this research complements FedAvg by using security mechanisms to secure messages exchanged between servers and clients.

6.1.2 Comparison with Securing Against Active Adversaries Approach

The protocol defined in [38] proposed securing FL against active adversaries, by enabling safe vector summing. Notably, their system is tailored to function well within an environment where there is only one server with limited trust, utilizing cryptographic primitives in multiple stages. Their methodology has several benefits, including the protection of privacy and enhanced security. However, it also exhibits some limits in terms of its ability to withstand active attacks, ensure the use of well-formed input, and manage communication overhead. These limitations necessitate more examination and consideration for the actual implementation of Federated Learning systems. However, the methodology described in this research, which combines polymorphic and homomorphic encryption techniques, represents a significant progress in enhancing the security of Federated Learning (FL) by ensuring the protection of data transmission between the server and clients. The combination of encryption algorithms employed in this study effectively addresses some constraints identified in the previous protocol described in [38], particularly its ability to withstand against inference attacks. This advantage is achieved by minimizing the risk of a successful attack, that if succeeded to crack one of the used encryption keys, which is not an easy task to be done, will not threat the whole FL system, where a key is never used again within the FL cycle, even for the same FL client.

6.1.3 Comparison with RFA

Compared to robust federated aggregation (RFA) described in [39]'s study, our proposed research focuses on a different approach based on encryption mechanisms. While the frameworks proposed in this research provide a unique and synergistic security framework specifically designed for Federated Learning, the primary goal of RFA is to improve the security of the FL system against poisoning attacks, solely through the use of geometric median-based aggregation. The frameworks proposed in this research exhibit a higher degree of resistance to unauthorized access and potential privacy violations, thus extending the RFA's capability to stand against inference and backdoor attacks. The risk emerging from those types of attacks is diminished by minimizing the risk of a leaked or cracked encryption key. As it has been previously explained, an AES key will require a very long time to be cracked, if cracked. In such a case, this key will not cause any threat to the system since it will not be used in the FL cycle again. In addition, the time required to crack such an encryption key is relatively great when compared to the FL cycle.Nevertheless, it is critical to recognize the complex trade-offs between RFA and the proposed frameworks as they navigate the subtle interaction between the needs of ensuring resilience, optimizing communication efficiency, and other considerations.

6.1.4 Comparison with LEGATO

In addition, the FL method named LEGATO [40] was developed to address the pressing issue of developing robust aggregation strategies in the context of Byzantine attacks in the Federated Learning paradigm. Their mechanism aims to improve the convergence of gradient descent algorithms, even in the presence of malicious attacks. However, it is important to note that the scope of LEGATO is mainly limited to Machine Learning models organized in layers, such as Deep Learning and neural networks. The frameworks discussed in this study, on the other hand, are not only resilient against Byzantine attacks, but also against inference and backdoor attacks, as well as against unauthorized access and accidental disclosure of data. In addition, the variety of models supported by the frameworks provides users with the flexibility to use the models for different types of data. However, it is important to recognize that the choice between these methods should depend on the particular security and privacy requirements of the Federated Learning application.

6.1.5 Comparison with SecureD-FL

In addition, the approach presented in [41], named SecureD-FL, leverages the Alternating Direction Method of Multiplier (ADMM) and employs combinatorial block design theory to control participant communication patterns and minimize privacy loss, the proposed frameworks prioritize data privacy through polymorphic and homomorphic encryption. Therefore, the associated mechanism in SecureD-FL is based on communication-control, rather than encryption or security mechanisms. In contrast, the proposed frameworks focus on encrypting all messages exchanged between the server and clients with unique keys, courtesy of polymorphic encryption, to ensure data privacy and access control. Moreover, homomorphic encryption is employed for secure aggregation of model parameters, maintaining data confidentiality during aggregation. The choice between these approaches should hinge on the specific security and privacy requirements of the FL application in question.

6.1.6 Comparison with SEAR

In [42], the algorithm SEAR was presented, which uses the trusted execution environment (TEE) offered by Intel SGX to perform secure aggregation of locally trained models in a trusted hardware environment. This effectively prevents the disclosure of sensitive information during the aggregation process. However, the limited memory capacity of the reserved processor memory (PRM) in Intel SGX poses a major challenge when aggregating a large number of models simultaneously. Conversely, by using encryption techniques, the proposed frameworks provide a robust and efficient approach to securing FL data without the limitations associated with hardware-based TEEs. While both strategies address the security issues of FL, they differ in their fundamental mechanisms. SEAR relies on hardware-based algorithms, while the proposed frameworks focus on encryption to enhance privacy and security. The choice between these methods should be based on the exact security and privacy requirements of the particular FL application, as well as the hardware capacity and structure of the server. In addition, SEAR does not show resistance to inference attacks compared to the proposed frameworks.

6.1.7 Comparison with EPPDA

Moreover, in [43], the authors present the Efficient Privacy-Preserving Data Aggregation (EPPDA) model, which exploits the homomorphisms of homomorphic encryption for secret

sharing to streamline the iterations of secret sharing and reduce the consumption of communication, computation, and storage resources. This resource optimization is especially beneficial in scenarios with multiple training iterations and ultimately improves system efficiency. In addition, EPPDA incorporates secret sharing to protect user data, reduce the impact of malicious users, and increase fault tolerance. Although EPPDA incorporates homomorphic encryption into its mechanism, the frameworks proposed by this research is more efficient due to polymorphic encryption, which requires the use of different keys for each encryption operation. This results in a higher level of security and improved resistance to various types of attacks.

6.1.8 Comparison with HeteroSAg

In addition, Heterogeneous Quantization (HeteroSAg) [44] addresses communication efficiency and resilience to Byzantine attacks in the FL ecosystem. Their approach differs from the frameworks proposed in this research, where the former focuses on communication efficiency and resilience against Byzantine attacks, while the latter prioritizes data security and privacy and HeteroSAg. Moreover, HeteroSAg's security is limited to Byzantine resilience only, while the proposed frameworks extend its security to different types of attacks, including inference and backdoor attacks.

6.1.9 Comparison with FLDetector

However, in stark contrast to existing defense mechanisms, exemplified by the previous implementation FLDetector [45], the frameworks proposed in this research take a distinctive and more technologically oriented approach. FLDetector primarily directs its focus towards the identification of potentially malicious clients, employing a strategy centered on assessing the consistency of model updates. While such methods offer invaluable insights into the realm of threat detection, the frameworks proposed in this research fundamentally diverge in their primary objective. They have devoted their efforts to fortify the underpinning security and privacy aspects of Federated Learning. To achieve this, they have meticulously integrated advanced encryption techniques, specifically homomorphic and polymorphic encryption, into the fabric of Federated Learning frameworks. This strategic integration has empowered these models to provide a comprehensive and robust defense mechanism for securing Federated Learning environments. This approach, thus, addresses not only the crucial issue of malicious client detection but extends its protective mantle to encompass the broader

spectrum of data privacy and integrity, safeguarding Federated Learning in a more holistic and technologically profound manner.

6.1.10 Comparison with FLCert

Compared to the FLCert framework [46], which centers its strategy on classifying customers into groups and utilizing majority voting among global models to resist poisoning attacks by malicious clients, the frameworks developed in this research takes a divergent path. These frameworks prioritize the secure exchange of messages between the server and clients, ensuring that each message is encrypted with a unique key, thereby reducing the risk posed by potential key breaches. Our approach is predominantly anchored in encryption, delivering a comprehensive strategy for safeguarding the privacy and integrity of Federated Learning environments. While FLCert excels in robustness against malicious clients, providing provable security guarantees even in the presence of a limited number of adversarial actors, the frameworks proposed in this research complements the security landscape by focusing on data protection and transmission security. In essence, the choice between these approaches hinges on the specific security and operational requirements governing Federated Learning systems, as they each offer distinct advantages and considerations."

6.1.11 Comparison with ELSA

While the proposed frameworks prioritize data encryption and integrity, ELSA [47] revolutionizes secure aggregation protocols to efficiently combat the presence of malicious actors. ELSA's core innovation lies in its utilization of distributed trust between two servers, which enables the preservation of individual client updates' secrecy as long as one server remains honest. This ensures robust protection against malicious clients, guarantees end-to-end efficiency, results in a significantly faster and more secure protocol compared to previous work. ELSA also introduces techniques that maintain confidentiality, even when a server turns malicious, with only a slight increase in execution time and minimal communication overhead, especially when compared to scenarios with reasonably honest servers. However, ELSA's security is predicated on the presence of at least one honest server, leaving room for concerns in cases where both servers are compromised. However, the frameworks proposed in this research are not affected by the existence of a malicious entity, where none of the FL entities will have access to other's secured data. Therefore, ELSA proved to be more flexible and

performing with less communication cost, but the frameworks proposed here will be more safe even in the context of malicious servers.

6.1.12 Comparison with Multi-RoundSecAgg

In contrast to the focus of our previously proposed Federated Learning frameworks on securing data transmission and privacy through advanced encryption techniques, the Multi-RoundSecAgg framework introduced in [48] addresses a distinct yet equally pressing challenge long-term privacy preservation in Federated Learning. While our frameworks concentrate on encryption and data privacy, Multi-RoundSecAgg excels in long-term privacy preservation, structured user selection, and fairness considerations. However, it introduces complexity in terms of multi-round confidentiality guarantees and structured user selection strategies, potentially increasing computational and operational complexity. Implementing the framework may also require additional computing and storage resources, which could pose challenges in resource-constrained environments. Its effectiveness and privacy guarantees may vary depending on specific use cases and data distributions and achieving perfect fairness in all practical scenarios remains a challenge. Additionally, the modularity of Multi-RoundSecAgg for large-scale Federated Learning scenarios with many participants and data sources may require further investigation.

6.1.13 Comparison with Stand-Alone HE Solutions

In contrast to implementations that embed only homomorphic encryption in Federated Learning (FL) frameworks such as [49–55], the frameworks proposed in this research introduce an additional layer of security by combining both polymorphic and homomorphic encryption. This combination significantly enhances data privacy, access control, and the overall robustness of FL against various threats, including data poisoning and model exposure.
The key differentiation in the proposed frameworks lies in their utilization of polymorphic encryption, wherein each set of parameters exchanged between the server and clients is encrypted with a distinct encryption key. This dynamic key assignment offers several distinct advantages:

- Enhanced Security: By encrypting each set of parameters with a different encryption key, the proposed frameworks ensure that even if one key is leaked or cracked, it poses no substantial risk. Since each key is used only for a specific set of parameters, a

breach of one key does not compromise the security of the entire system. This granular approach significantly enhances security against potential attacks

- Data Privacy: The individual encryption of parameters with unique keys ensures that sensitive data remains confidential. Unauthorized access to one set of parameters does not automatically grant access to others, bolstering data privacy in FL
- Access Control: The dynamic key assignment allows for precise access control. Only entities with the corresponding decryption keys can access and decrypt specific sets of parameters, minimizing the risk of unauthorized data access
- Unbreakable Encryption: The use of different encryption keys for each parameter set contributes to the unbreakable nature of the encryption. Even if an attacker were to gain access to one key, it would not provide a universal decryption capability, rendering the encryption highly secure
- Robustness Against Attacks: The combination of polymorphic and homomorphic encryption in the proposed frameworks creates multiple layers of defense against potential attacks. This multi-tiered security approach significantly enhances the FL system's robustness, making it resilient against a wide range of threats

In contrast, implementations that rely solely on homomorphic encryption typically employ a single encryption key for the entire FL process. While homomorphic encryption offers data security, it lacks the fine-grained security control provided by polymorphic encryption. In the event of a key compromise, the entire system's security is jeopardized. The proposed frameworks address this vulnerability by ensuring that the compromise of one key does not undermine the overall security of the FL system, making them an attractive choice for privacy-conscious and security-focused FL applications.

The comparison between the proposed frameworks and the state-of-the art of secured FL algorithms can be summarized in Table 3 below:

Table 3: Comparison of FL Security Approaches.

Criteria	Proposed Frameworks	Homomorphic Encryption Only	SecureD-FL	SEAR	HeteroSAg
Encryption Techniques	Polymorphic & Homomorphic Encryption	Homomorphic Encryption	Homomorphic Encryption	Trusted Execution Environment (TEE)	Homomorphic Encryption
Unique Encryption Keys for Parameters	Yes (Polymorphic Encryption)	No (Single Key)	Yes (Homomorphic Encryption)	Yes (TEE-Based Encryption)	No (Single Key)
Data Privacy & Access Control	Strong Data Privacy & Access Control	Limited Access Control	Strong Data Privacy & Access Control	Strong Data Privacy & Access Control	Limited Access Control
Security Against Key Compromises	Highly Resilient (Granular Key Usage)	Vulnerable to Key Compromise	Highly Resilient (Granular Key Usage)	Highly Resilient (TEE-Based)	Vulnerable to Key Compromise
Robustness Against Attacks	Multi-Layered Security Approach	Limited Security Layers	Multi-Layered Security Approach	Multi-Layered Security Approach	Enhanced Security Layers
Communication Efficiency	Efficient with Enhanced Security	Efficient but Less Granular	Efficient with Enhanced Security	Efficient with Hardware-Based TEE	Efficient with Enhanced Security
Byzantine Attack Resilience	Strong Resilience	Limited Resilience	Strong Resilience	Strong Resilience	Strong Resilience
Inference Attack Resilience	High Resilience	Limited Resilience	High Resilience	Limited	Moderate Resilience
Bandwidth Efficiency	Enhanced Efficiency	Standard Efficiency	Enhanced Efficiency	Enhanced Efficiency	Enhanced Efficiency

6.2 CHALLENGES & FUTURE PERSPECTIVES

To ensure the privacy of communications between servers and clients in a Federated Learning setting, PolyFLAG_SVM, PolyFLAM, PolyFLAP and HP_FLAP makes use of Homomorphic and Polymorphic Encryption. The polymorphism of encryption keys used to encrypt each message exchanged between the server and the participating clients provides security guarantees to an extent that a leaked or cracked key will not be risky to the whole system, since it will never be used again in the FL cycle. Additionally, in HP_FLAP, parameters from locally trained models are aggregated on the central server without being decrypted, adding an extra level of security. The proposed frameworks places an emphasis on security, but the intricacy of its encryption makes it an expensive method of computation and transmission, thus imposing some challenges and issues. However, they can complement well-established approaches in order to parallelize computing, improve learning efficacy, manage heterogeneity, and scale up to overcome such issues and to enhance their feasibility. This chapter delves into the challenges and future prospects tied to the proposed frameworks, shedding light on potential obstacles and avenues for future research and development. The thesis wraps up

with a comprehensive executive summary that encapsulates the entire study, offering crucial insights, findings, and reflections on its contributions and implications.

6.2.1 Challenges

In the proposed frameworks, a set of difficulties arises, all of which are intertwined with the development and deployment of this FL framework.

6.2.1.1 Limitation to SVM Model

SVM has demonstrated its effectiveness in addressing Machine Learning challenges, surpassing alternative models in some problems. Nevertheless, the restriction of PolyFLAG_SVM exclusively to the SVM model may hinder its versatility. This limitation served as the primary impetus for the development of both PolyFLAM and PolyFLAP which offers five different ML models as explained earlier.

6.2.1.2 Heterogeneity

The aspect of heterogeneity poses a significant obstacle in the envisioned framework, particularly concerning the facilitation of "Horizontal Federated Learning (FL) data." This phrase encompasses situations in which varied clients handle data exhibiting identical attributes. While the framework effectively handles this particular data category, it's important to recognize that this framework did not encompass other potential methodologies. This underscores the intricacy linked with adapting to diverse data formats and underscores the necessity for additional investigation into the wider realm of data heterogeneity.

6.2.1.3 Complexity and Computation Cost

The issue of complexity and computational cost is of utmost importance, particularly in light of the resource-intensive characteristics inherent in encryption methodologies. Due to the critical role of these algorithms in preserving the integrity of communicated data, they inherently require substantial computational resources. The inclusion of robust encryption algorithms in the framework's operations introduces an additional layer of complexity, hence enhancing data security measures. Consequently, the task of achieving a harmonious equilibrium between strong security protocols and optimal computational efficiency emerges as

a pivotal challenge, thereby requiring innovative approaches to minimize computational expenses while upholding the integrity of the system.

6.2.1.4 Scalability

The concern about scalability arises due to the heightened computational requirements associated with encryption. The framework's capacity to handle an increasing number of clients, especially on the server side, may be constrained by the additional processing costs, thereby impeding the system's scalability. The increased number of clients accessing the server places greater strain on its processing capability, potentially leading to the occurrence of bottlenecks and a subsequent decline in performance. In order to attain seamless scalability, it is imperative to explore efficient optimization strategies that decrease computational burden while maintaining system responsiveness and accommodating an increasing client base.

6.2.1.5 Learning Quality

The quality of learning within the proposed framework is a crucial issue that arises. The primary objective of the proposed frameworks is to enhance security and resilience against inference attacks. However, it is important to acknowledge the inherent compromise between security and the quality of learning that should not be neglected. The prioritization of security measures, like as encryption and data protection, may divert focus from the improvement of model learning quality. Maintaining a nuanced equilibrium between robust security measures and achieving optimal learning outcomes is an ongoing challenge that requires meticulous assessment of how security measures impact the effectiveness of the learning process and the future performance of the global model.

6.2.1.6 Resources Limitations

The presence of resource limitations is a major hurdle, particularly in the context of Federated Learning. In this context, clients often work with limited computational resources, which is commonly found in smartphones or smart wearables, rather than more powerful computers. The relevance of this difficulty is heightened when considering the practical application of the suggested framework. The use of encryption and other security measures has the potential to place a further burden on the limited resources of clients. The present situation gives rise to apprehensions over the feasibility and practicality of implementing the framework in real-life situations, considering the possible strain it may place on client devices. To

tackle this difficulty, it is imperative to develop solutions that maximize the efficient use of existing resources, ensuring that the framework stays feasible for implementation while also considering the constraints of client settings.

6.2.2 Future Perspectives

When considering future enhancements, it is crucial to recognize that the aforementioned difficulties have been extensively addressed within the academic debate. Multiple researchers have actively addressed these challenges, providing novel answers that require some further investigation. Upon careful consideration of future prospects, a very captivating trajectory emerges, wherein the suggested framework aligns harmoniously with well-established methodologies. The current convergence has the potential to not only address current obstacles but also provide a period of increased effectiveness and adaptability for Federated Learning systems, and definitely for the proposed frameworks.

6.2.2.1 Handling Heterogeneity

Innovative solutions are needed to address heterogeneity from varied devices and data. Luckily, many methods may be used to control this variability. For instance, Leveraging Resource Allocation methods [64] can smartly allocates computational resources among devices based on their capabilities. This method optimizes resource use for balanced and efficient Federated Learning. In addition, integrating Meta-Learning approaches [65] is a promising approach as well. Meta-Learning can improve system adaptability to client device and data source heterogeneity by allowing models to learn and adapt fast to new data distributions. These techniques together with the proposed framework might provide a more flexible and effective Federated Learning framework that can handle heterogeneity.

6.2.2.2 Computation Cost & Time Reduction

The issue of high computational costs can be reduced by careful implementation of a number of methods. The use of parallel programming techniques is one such strategy. Parallel programming makes better use of the computing capacity of contemporary devices by decomposing large computations into smaller tasks that may be completed simultaneously. Faster model training and less time spent computing mean less stress on already-strapped computer systems. The suggested frameworks might benefit from the use of parallel programming

techniques to drastically reduce computing costs while simultaneously increasing scalability and responsiveness.

6.2.2.3 Enhancing Scalability

Improving scalability depends on finding viable solutions to the challenges of heterogeneity and high computational costs. A mutually beneficial relationship develops when these difficulties are met with solutions like resource allocation and parallel programming. The system's ability to serve a wide variety of users is improved by its approach to the heterogeneity of their devices and data. Simultaneously, lowering processing costs using methods like parallel programming keeps the framework responsive even as the number of users grows. Together, these solutions pave the path for a Federated Learning framework that can scale to serve a large number of users without sacrificing performance. The combination of these tactics has the potential to launch an ecosystem that can handle the stresses of real life. In addition, offloading keys creation and management to a third-party, may be a successful solution to boost scalability.

6.2.2.4 Boosting Learning Quality

Implementing a wide variety of client-side data pre-processing strategies can significantly improve the standard of learning outputs. Strategic pre-processing methods can be added during data preparation before training to improve the quality of input data. Data quality and benefit may be improved through the use of methods including feature scaling, outlier removal, and data synthesizing. The overall quality of learning may be greatly improved by ensuring that the data input into the training process is well-prepared and free of noise or anomalies. Integrating data-preprocessing with the proposed frameworks, may enhance the learning process, leading to better model convergence and performance.

APPENDIX A

PUBLICATIONS

In this dedicated appendix, I take great pride in presenting a comprehensive compilation of my scholarly contributions throughout the course of my rigorous doctoral journey. As I embark on this endeavor to enumerate the culmination of countless hours of research, analysis, and scholarly dedication, I am pleased to include a total of 13 meticulously crafted articles. Each publication within this list reflects not only the culmination of my own academic growth but also the invaluable support and mentorship of my advisors and collaborators. These publications represent the synthesis of knowledge, the pursuit of excellence, and the unwavering commitment to the pursuit of innovative and meaningful research, all of which have been integral to my doctoral experience. It is with great enthusiasm and a sense of achievement that I present this record of my academic contributions, each article a testament to the dedication and hard work that have been at the core of my doctoral journey.

In the Table 4 below, the list of publications is presented with its relevant details. Within this appendix, you will find an organized compilation of my publications, meticulously arranged to provide all pertinent details, and thoughtfully categorized according to their direct relevance to the overarching research pursuits of my doctoral journey. Furthermore, as you progress through this section, each individual article or sets of articles will be thoughtfully introduced, offering concise insights into their primary objectives and contributions to the scholarly discourse. This comprehensive structure has been designed to offer readers a holistic understanding of the scope and impact of my research endeavors, underscoring their significance within the broader academic landscape.

-	Article Title	Status	Type	Publisher	Journal	Note	Publish Date	URL
1	Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review	Published	Journal	MDPI	Sensors	Under Special Issue: Selected Papers from 2022 IEEE International Conference on e-Health and Bioengineering	11-Jan-23	https://www.mdpi.com/1424-8220/23/2/828
2	Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review	Published	Journal	MDPI	Sensors	Under Special Issue: Embedded Sensor Systems for Health	2-Oct-22	https://www.mdpi.com/1424-8220/22/19/7472
3	Detection of Occupational Fatigue in Digital Era; Parameters In Use	Published	Conference	IEEE	Conferences	Conference: IHSH'2022	10-April-23	https://ieeexplore.ieee.org/abstract/document/10092165
4	Machine Learning Models to Predict Cardiovascular Events from Heart Rate Variability Data	Published	Conference	IEEE	Conferences	Conference: IHSH'2022	10-April-23	https://ieeexplore.ieee.org/abstract/document/10092060
5	Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability	Published	Conference	Elsevier	Procedia Computer Science	Conference: MobiSPC2022	12-Aug-22	https://www.sciencedirect.com/science/article/pii/S1877050922006354?via%3Dihub
6	Predicting Cardiovascular Events with Machine Learning Models and Heart Rate Variability	Published	Journal	IASKS	JUSPN	Published as an extension for MobiSPC2022 article	2-Mar-23	https://iasks.org/articles/juspn-v18-i2-pp-49-59.pdf
7	Reviewing Multimodal Machine Learning and its Use in Cardiovascular Diseases Detection	Published	Journal	MDPI	Electronics	Under Special Issue: IoT for Healthcare and Wellbeing: Trends, Challenges, and Applications	26-Mar-23	https://www.mdpi.com/2079-9292/12/7/1558
8	Reviewing Federated Machine Learning and its Use in Diseases Prediction	Published	Journal	MDPI	Sensors	Under Special Issue: Smart Environments for Health and Well-Being	13-Feb-23	https://www.mdpi.com/1424-8220/23/4/2112
9	Reviewing Federated Learning Aggregation Algorithms; Strategies, Contributions, Limitations and Future Perspectives	Published	Journal	MDPI	Electronics	Under Special Issue: Collaborative Artificial Systems	18-May-23	https://www.mdpi.com/2079-9292/12/10/2287
10	Securing Federated Learning; Approaches, Mechanisms and Oppurtunities	Under Review	Journal	ACM	Journal of the ACM	-	-	-
11	PolyFLAG_SVM: A Polymorphic Federated Learning Aggregation of Gradients Support Vector Machines Framework	Published	Conference	Elsevier	Procedia Computer Science	Conference: MobiSPC2023	18-Sep-23	https://www.sciencedirect.com/science/article/pii/S1877050923010682
12	PolyFLAM & PolyFLAP: Federated Learning Aggregation Frameworks Secured with Polymorphic Encryption	Under Review	Journal	Elsevier	Journal of Information Security and Applications	-	-	-
13	HP_FLAP: Homomorphic & Polymorphic Federated Learning Aggregation of Parameters Framework	Under Review	Journal	Springer	Cybersecurity	-	-	-

A.1 Reviewing Smart Wearables in Diseases Management

Embarking on my research journey, particularly during the second academic term, I immersed myself in the dynamic realm of smart health. My objectives were twofold: staying current with the latest research trends in this domain and maximizing the value of my reading time. Aware of the time-consuming nature of academic reading, I ensured that each reading session was not just passive consumption but a strategic step towards knowledge acquisition.
To achieve this, I diligently documented my readings, recorded essential insights, and captured novel ideas. This effort resulted in two well-crafted review articles, both contributing significantly to the field of wearables in smart health.
The first review, conducted systematically following PRISMA guidelines, examined the use of smart wearables in Cardiovascular Diseases. This article, highlighting wearables' role in cardiovascular health, was published in MDPI-Sensors, as discussed later.
The second review, also published in MDPI-Sensors, explored the application of smart wearables in detecting Occupational Physical Fatigue. This work contributes to the growing knowledge in the field and emphasizes the practicality of wearable technology in addressing occupational health concerns.
These publications showcase the synergy between rigorous research and effective time management, illustrating my dedication to advancing smart health and wearables' discourse. This introduction sets the stage for the comprehensive listing and explanation of my publications, highlighting the precision and commitment that defined my doctoral research journey.
In this section, two article are fully presented which are respectively:

- **Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review** *(MDPI-Sensors / Impact Factor 3.9)*
- **Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review** *(MDPI-Sensors / Impact Factor 3.9)*

sensors

MDPI

Systematic Review

Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review

Mohammad Moshawrab [1,*], Mehdi Adda [1], Abdenour Bouzouane [2], Hussein Ibrahim [3] and Ali Raad [4]

[1] Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC G5L 3A1, Canada
[2] Département d'Informatique et de Mathématique, Université du Québec à Chicoutimi, 555 Boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada
[3] Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, QC G4R 5B7, Canada
[4] Faculty of Arts & Sciences, Islamic University of Lebanon, Wardaniyeh P.O. Box 30014, Lebanon
* Correspondence: mohammad.moshawrab@uqar.ca; Tel.: +1-(581)624-9394

Abstract: Background: The advancement of information and communication technologies and the growing power of artificial intelligence are successfully transforming a number of concepts that are important to our daily lives. Many sectors, including education, healthcare, industry, and others, are benefiting greatly from the use of such resources. The healthcare sector, for example, was an early adopter of smart wearables, which primarily serve as diagnostic tools. In this context, smart wearables have demonstrated their effectiveness in detecting and predicting cardiovascular diseases (CVDs), the leading cause of death worldwide. **Objective:** In this study, a systematic literature review of smart wearable applications for cardiovascular disease detection and prediction is presented. After conducting the required search, the documents that met the criteria were analyzed to extract key criteria such as the publication year, vital signs recorded, diseases studied, hardware used, smart models used, datasets used, and performance metrics. **Methods:** This study followed the PRISMA guidelines by searching IEEE, PubMed, and Scopus for publications published between 2010 and 2022. Once records were located, they were reviewed to determine which ones should be included in the analysis. Finally, the analysis was completed, and the relevant data were included in the review along with the relevant articles. **Results:** As a result of the comprehensive search procedures, 87 papers were deemed relevant for further review. In addition, the results are discussed to evaluate the development and use of smart wearable devices for cardiovascular disease management, and the results demonstrate the high efficiency of such wearable devices. **Conclusions:** The results clearly show that interest in this topic has increased. Although the results show that smart wearables are quite accurate in detecting, predicting, and even treating cardiovascular disease, further research is needed to improve their use.

Keywords: cardiovascular diseases; smart wearables; sensors; body sensor networks; machine learning; smart health; wide body area networks

Citation: Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review. *Sensors* **2023**, *23*, 828. https://doi.org/10.3390/s23020828

Academic Editors: Hariton-Nicolae Costin, Cristian Rotariu and Monica Fira

Received: 20 November 2022
Revised: 27 December 2022
Accepted: 9 January 2023
Published: 11 January 2023

1. Introduction

Healthcare has always been one of the most important issues that people have cared about. Given the prevalence of diseases and their impact on people's lives, researchers are always looking for methods to improve medical services and promote public health. In addition, the aging population, shortage of medically trained personnel, lack of equity in services, epidemic planning, and a host of other problems hinder the growth of public health worldwide [1]. However, advances in information and communication technology (ICT) offer effective answers to these challenges. In this context, artificial intelligence (AI) is considered the most promising tool for improving healthcare, as it has the potential to be used in virtually all areas of medicine [2] and will transform healthcare for patients

and communities [3]. This enormous contribution is not due to magic, but to AI's data-processing capabilities, which surpass those of humans, especially when large computations are performed in a short period of time. Even though the majority of AI applications in healthcare were developed after 2008 [4], their importance is obvious. First, AI has improved the learning capabilities of computers and humans, leading to improved diagnostic and healthcare procedures [5]. In addition, AI technologies are able to accept common sense, extract information from raw data, use human-like thought processes, deal with inaccuracies, adapt to a rapidly changing environment, and even act on their knowledge [2]. These characteristics enable AI tools to think and behave similar to humans at a virtually unparalleled level, allowing them to articulate clinical patterns and visions beyond human capabilities [3]. Combining AI capabilities with human intelligence, sometimes referred to as augmented intelligence, is probably the most effective way to improve healthcare services [3].

1.1. Cardiovascular Diseases Latest Figures

Cardiovascular diseases (CVDs) are the leading cause of death and are hence recognized as the most dangerous disease in the world. According to the most recent World Health Organization (WHO) statistics on heart disease, the number of CVD patients worldwide has increased from 271 million to 523 million between 1990 and 2019, and the number of deaths caused by this disease has increased from 12.1 million to 18.6 million during the same period, accounting for 32% of global mortality in 2019 [6]. For example, in the United States, a person dies from heart disease at least every 34 s [7], and in Canada, a person dies at least every 5 min [8]. Moreover, cardiovascular disease is a major cause of both health conflict and economic suffering. According to the Medical Expenditure Panel Survey, the total cost of CVDs in the United States between 2017 and 2018 was estimated at USD 378.0 billion, including USD 226.0 billion in expenditures and USD 151.8 billion in lost future productivity [9]. Figure 1 illustrates the increase in the number of patients and deaths due to cardiovascular disease worldwide between 1990 and 2019.

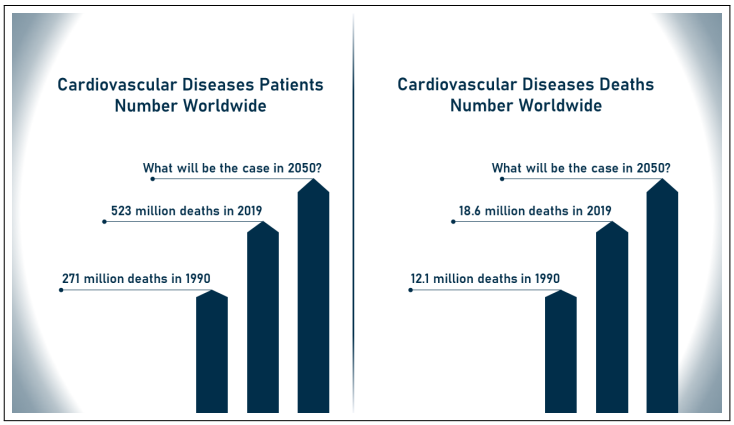

Figure 1. Increase in number of patients and deaths due to CVDs.

1.2. CVDs Detection: From Classic to Technology-Assisted

Due to their potentially fatal nature, cardiovascular diseases need the development of efficient solutions that allow early diagnosis and, ideally, prediction of their onset. The predictive power of modern technologies could help reduce the global prevalence of CVDs. Traditional methods for diagnosing these diseases include electrocardiogram, echocardiography, coronary angiography, stress testing, magnetic resonance imaging, or intracoronary ultrasonography. However, new technologies are improving health services and facilitating the detection of cardiovascular disease, particularly information and communication technologies (ICTs) and the development of artificial intelligence (AI) and its derivatives. The novel approaches of AI in cardiology have proven to be successful in providing fast,

accurate, and less erroneous patient care, which has significant medical and financial implications. It is more effective and widely used, as the tools and applications offered are at the level of an expert using real-world data. In general, AI has fantastic potential to transform cardiology in the near future and is often seen as the next revolutionary step in the field due to its potential to accelerate and improve patient care. Moreover, AI will soon revolutionize cardiovascular health, as its tools have the potential to outperform experts in detecting and predicting cardiovascular disease [10–12]. Therefore, smart wearables that combine AI and ICT are expected to be very useful in cardiovascular disease detection and prediction.

1.3. Smart Wearables: Definitions and Overview

Smart wearables, also known as smart wearable technology or wearable gadgets, are a new breed of compact, rugged, and efficient computing devices made possible by the rapid growth of information and communication technologies and the advancement of electronics, particularly microprocessors. These devices are being hailed as the next generation of ubiquitous technology after smartphones, as they allow access to data at any time and from any location. The topic of smart wearables has evolved rapidly in recent years, and their technologies are now applicable in many other fields [13–16]. This section provides a definition of "smart wearables" and a brief overview of the history of wearable technology. In addition, various categories of smart wearables are discussed in the upcoming sections.

1.3.1. Smart Wearables: Brief History

In 1950, Alan Turing asked the now famous question "Can machines think?" which marked the beginning of the era of "Smart Machines" [17]. Since then, researchers around the world have attempted to answer this question by turning computers into intelligent devices. Despite its widespread use, the term "Smart" is not uniformly defined and is presented in different ways by different scholars [18]. In [19], "Smart" devices are defined as embedded sensors, processors, and network devices that give smart things the ability to behave based on their own knowledge. In addition, Ref. [20] defines them as objects that can learn from their environment and interact with humans. Different definitions focus on the capabilities of the devices. For example, smart wearables are defined by the authors in [21,22] as devices that can be worn by the user at all times to monitor factors such as personal data, vital signs, locations, environment, movements, and more. In this context, a shoe-sized computer developed by Edward Thorp and Claude Shannon in 1961 is widely considered to be the first ever wearable computing device [23,24]. In the 1980s, Steve Mann developed EyeTap glasses that displayed computer-generated images in one eye and added textual information to the user's visual experience [25]. Subsequently, in 1996, the U.S. Department of Defense Navy funded a study to monitor the vital signs of its troops [26,27], which is widely considered a defining moment in the history of smart wearables. Since then, smart wearables have gradually evolved from invasive, heavy, and huge technologies to more adaptable, compact, and lightweight devices. This is because researchers have expanded their projects in this field to different areas of life such as health, fitness, sports, fashion, and even other sectors.

1.3.2. Classification of Smart Wearables

Over the past few decades, there have been more than a thousand studies on smart wearables. However, smart wearables cannot be classified into a specific category. Accordingly, smart wearables are divided into six groups, as described by the authors of Ref. [28]:

- Medical;
- Industrial;
- Lifestyle;
- Fitness;
- Entertainment;

- Gaming.

On the other hand, the smart wearables were categorized by the authors of Ref. [29] according to their personal features rather than their function. They provided examples of the three categories into which they fall:

- Watch-type;
- Necklace or wristband-type;
- Headmount display-type.

However, other technologies, such as electronic patches and health apparel, could not fit within this classification. Therefore, a set of commonly known wearables are displayed in Figure 2 below.

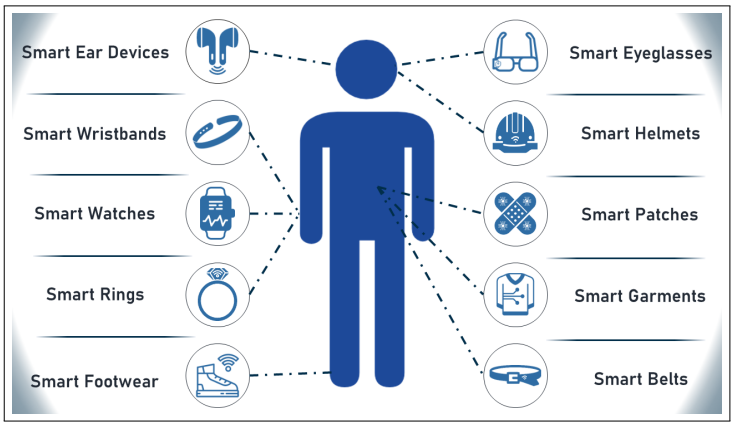

Figure 2. Set of commonly known wearables.

1.4. Role of Smart Wearables in CVDs

Over the past decade, smart wearables have been increasingly used as health solutions. Their effectiveness and proliferation has been fueled by advances in performance, size, style, and durability, among other factors. Examples of smart wearables used to diagnose, track, and treat cardiovascular disease include wristbands, patches, headbands, eyeglasses, and necklaces. The implications of CVD wearables are many. For example, they enable continuous and long-term recording of functional or physiological data, leading to more accurate diagnosis and better health outcomes for patients. In addition, they enable the collection of necessary data in locations other than physicians' offices or hospitals, expanding the capacity of healthcare facilities to serve larger numbers of patients over longer periods of time. More importantly, the continuous monitoring capabilities of smart wearables enable more sophisticated knowledge of an individual patient's physiological state and ongoing activity, paving the way for more personalized healthcare and treatment. The devices also became less bulky and aesthetically pleasing, making them less intrusive and more suitable as everyday wearables. One way that smart wearables such as smartphones are benefiting from the widespread use of other devices is through pairing [30–33]. Some reasons for the success of smart wearables adoption and their success points are listed in Table 1 below.

Table 1. Success reasons and success points of smart wearables.

	Powered By	**Capabilities**
Smart wearables	Low power consumption Compact size Adaptable styles Robustness	Continuous functionality Long-term Monitoring Real-time data sensing Communication with Internet

1.5. Outline and Main Contributions of This Article

In this article, the use of smart wearables in the detection, prediction, and treatment of cardiovascular disease was investigated. For this purpose, a systematic literature review was conducted, following the methodology that is explained in Section 2. Subsequently, in Section 3, the results of the performed search are presented. Later, the obtained results are analyzed in Section 4 and discussed in Section 5. Then, the challenges hindering the progress in the use of smart wearables are discussed, and future perspectives to solve these challenges are presented. Finally, the article is concluded with a concluding section. To the best of our knowledge, there are no systematic reviews addressing the potential of smart wearables for early diagnosis of CVDs. For example, in [34], the authors investigated the application of AI in smart wearables for cardiovascular disease detection. However, the focus of their research was on smart models rather than hardware; the obstacles that have slowed the development of this field are barely addressed, and the same is true for future prospects. Furthermore, in [35,36], the authors explored the use of smart wearables in life course research, but they did not systematically explore the field or provide a complete vision of contextual implementations. Motivated by the large role that smart wearables play in various aspects of daily life and by the lack of a systematic literature review discussing their role in predicting cardiovascular disease, this article therefore attempts to answer the following questions:

- What are the applications of using smart wearables to detect and predict cardiovascular disease?
- What are the different aspects such as hardware and software used in these implementations?
- To what extent are these implementations feasible?
- What are the challenges and limitations in this area?
- What future perspectives can be pursued to improve the use of smart wearables in CVDs management?

Therefore, this article answers the above questions and thus contributes to academic knowledge by:

- Systematically reviewing the use of smart wearables in the treatment of cardiovascular disease;
- Analyzing and discussing the reviewed implementations in a way that facilitates the identification of opportunities for improvement in this area;
- Naming the barriers to progress in this area;
- Proposing solutions that can be used to address these barriers;
- Presenting a collection of research questions and findings that could serve as a starting point for future research.

2. Research Methodology

This section details and explains the methodology used to conduct the systematic review. The steps described here can be used to conduct the same search and review the results or repeat the search in a different time period.

2.1. Eligibility Criteria

In conducting this review, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [37] was used as a guide for preparing a systematic literature review. The structure of this review was based on the latest PRISMA checklist (PRISMA Checklist 2020) [38]. In accordance with PRISMA standards, multiple sources were searched for papers that met the scope of this review. Four variables were used to select these materials. To be considered reputable, a paper must address artificial intelligence or related fields, present a smart wearable solution, address healthcare, and focus on cardiovascular diseases. In addition, the material should have been published in a peer-reviewed journal or as a conference proceedings. In addition, only documents published between January 2010 and

October 2022 were considered. As a final eligibility criterion, an English language filter was applied to eligible papers.

2.2. Information Sources

Several academic abstract and citation databases for peer-reviewed literature were used, including IEEE, PubMed, and Scopus Elsevier, to ensure superior results and to cover the largest number of documents possible. Each of these three databases provides access to millions of documents and has powerful, sophisticated search tools to facilitate thorough literature searches.

2.3. Search Strategy

In order to conduct a thorough search of the above materials, three queries were formulated. While these queries all follow the same logical structure, they use different syntaxes to comply with the different rules imposed by each data source. Targeted articles are found at the intersection of four query blocks, each defining a different topic of interest. AI, health, wearables, and CVDs (or related areas) are the four basic focus areas. The phrase "AND" was used to combine areas for a more effective query, while the term "OR" was used to combine different terms within each area. The three queries used to find what is being searched for are as follows:

- IEEE: ((("ARTIFICIAL INTELLIGENCE" OR "SMART AGENTS" OR "SMART MACHINES" OR "INTELLIGENT" OR "DEEP LEARNING" OR "MACHINE LEARNING" OR "NEURAL NETWORK") AND ("HEALTH*" OR "DISEASE" OR "ILL*" OR"CARE") AND ("WIRELESS SENSORS NETWORK" OR "SMART SENSORS" OR "BODY AREA NETWORK" OR "WEARABLE" OR "SENSOR") AND ("CARDIOLOGY" OR "CARDIOVASCULAR" OR "HEART" OR "CARDI*"))).
- PubMed: ((ARTIFICIAL INTELLIGENCE) OR (SMART AGENTS) OR (SMART MACHINES) OR (INTELLIGENT) OR(DEEP LEARNING) OR (MACHINE LEARNING) OR (NEURAL NETWORK)) AND ((HEALTH) OR (DISEASE) OR (ILL) OR(CARE) OR (HEALTHCARE)) AND ((WIRELESS SENSORS NETWORK) OR (SMART SENSORS) OR(BODY AREA NETWORK) OR (WEARABLE) OR (SENSOR)) AND ((CARDIOLOGY) OR (CARDIOVASCULAR) OR (HEART) OR(CARDIAC)).
- Scopus: TITLE-ABS-KEY(((artificial intelligence) OR (smart agents) OR (smart machines) OR (intelligent) OR (deep learning) OR (machine learning) OR (neural network)) AND ((health*) OR (disease) OR (ill*) OR (care)) AND ((wireless sensors network) OR (smart sensors) OR (body area network) OR (wearable) OR (sensor)) AND ((cardiology) OR (cardiovascular) OR (heart) OR (cardi*))) AND (LIMIT-TO (SRCTYPE, "j") OR LIMIT-TO (SRCTYPE, "p")) AND (LIMIT-TO(DOCTYPE, "cp") OR LIMIT-TO(DOCTYPE, "ar")) AND (LIMIT-TO(LANGUAGE, "English")) AND (LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015) OR LIMIT-TO (PUBYEAR, 2014) OR LIMIT-TO (PUBYEAR, 2013) OR LIMIT-TO (PUBYEAR, 2012) OR LIMIT-TO (PUBYEAR, 2011) OR LIMIT-TO (PUBYEAR, 2010)).

In the case of Scopus, the query was used as described above to retrieve the results. However, in IEEE and PubMed, additional filters were applied through the graphical user interface. In both sources, a "Year" filter was added to limit the selection to articles between 2010 and 2021. However, in IEEE, articles with the types "Conferences" and "Journals" were selected, whereas in PubMed, articles with the types "Clinical Trial" and "Journal Article" were selected. Finally, a filter was performed in the PubMed interface to limit the documents to those published in English. On the other hand, Scopus offers the possibility to limit the search to the title, abstract, or keywords, while the other two sources perform the search in the whole text of the document. It is worth noting that the search was performed in October 2021.

2.4. Selection Process

The data extracted from the records were selected in three steps to determine which files were relevant to this analysis. The first step was to review the titles and abstracts of all documents to determine if they were relevant to the topic of this study. The documents that passed this step were then downloaded. In a second step, we reviewed the downloaded files to quickly verify their content and determine if they were relevant to our evaluation. The documents selected in this phase are the ones that are examined in detail. Finally, the documents were researched and evaluated to extract the data needed to demonstrate the development of smart wearables for CVDs.

3. Results

The steps mentioned in the previous section led to a systematic result. The results of this search are listed in this section.

3.1. Study Selection

Initially, 4002 documents were identified from the three libraries based on the above searches. The search on IEEE yielded 1013 documents, on PubMed 1020, and on Scopus 1969, after which duplicate entries were excluded, removing 1021 and leaving 2981 documents. Then, the aforementioned selection procedure was applied, excluding 2382 documents on the basis of irrelevance and advancing 599 to the next stage. Documents were classified as irrelevant if they met the search criteria or if they contained the search terms specified in the search queries but did not deal with cardiovascular disease or were not wearable systems. In the second phase, full-text screening, the 599 documents were downloaded and skimmed to assess their suitability. In this phase, 512 documents were excluded for various reasons, and 87 documents were deemed suitable for this review. All these details are shown in Figure 3 below that matches the PRISMA diagram (information flow through the different phases of a systematic review) [37].

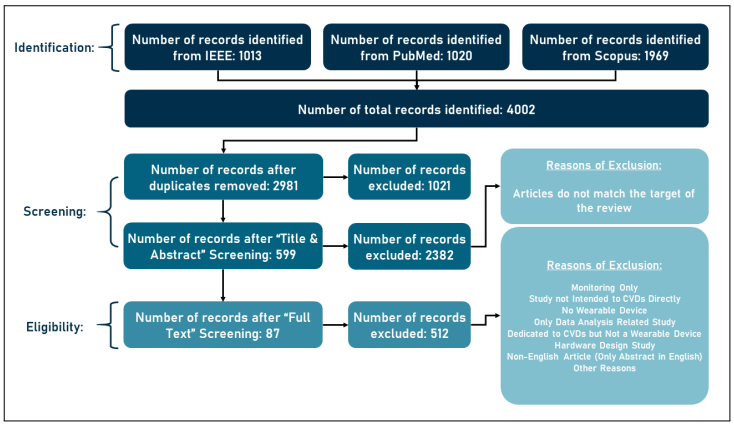

Figure 3. Flow of information through the different phases of a systematic review.

3.2. Study Characteristics

Following the search described above, the 87 papers deemed appropriate were carefully reviewed and examined to extract all relevant information. From each paper, the year of publication, disease(s) treated, vital signs recorded, hardware of the wearable device(s), embedded intelligent models, dataset(s) used, and outcome metrics were extracted. Table 2 below shows all retrieved details from the eligible studies, thus forming one of the main outcomes of this research, listing all implementations of smart wearables in cardiovascular disease management between 2010 and 2022 along with their relevant details.

Table 2. Implementations of smart wearables in detection of CVDs.

Ref#	Year	Disease(s) Targeted	Vital Signs Collected	Hardware Employed	Smart Model(s) Used	Training Dataset(s)	Results Metrics
[39]	2010	Atrial Fibrillation	Electrocardiogram	A wearable vest including dry foam ECG acquisition device A mobile phone (Nokia N85)	Not Identified	PhysioNet MIT-BIH dataset	Sensitivity: 94.56% Positive Predictive Value: 99.22%
[40]	2010	Right Bundle Branch Block Beats Premature Ventricular Contraction Paced Beats Fusion of Paced and Normal Beats Sinus Tachycardia	Electrocardiogram	Plug-In-Based GUI Platform: An Alive Bluetooth ECG heart monitor and Amoi E72 Microsoft Windows Mobile 5 Smartphone Machine-Learning-Based Platform: An Alive Bluetooth ECG heart monitor and an HTC Microsoft Windows Mobile 6 Smartphone	Multilayer Perceptron	PhysioNet MIT-BIH dataset	Accuracy > 90%
[41]	2010	Sinus Bradycardia Cardiac Asystole Atrial Fibrillation Wide QRS Complex	Electrocardiogram	A three-lead ECG device that contain two main parts: NCTU ECG Aquisition tool as the data acquisition (DAQ) unit and a wireless-transmission unit. Medi-Trace 200, Kendall are also used to read the ECG from the body	Not Identified	Dataset collected at MUSE ECG system (GE health care, USA) in China Medical University (CMUH) database	Accuracy> 92%
[42]	2011	Premature Ventricular Contraction Atrial Premature Contraction	Electrocardiogram Electroencephalogram Respiratory Rate Skin Temperature	Wearable Sensor Node and it consists of seven modules: analog front-end circuits for four physiological signals, a radio communication module, a storage module, and MSP430F2618 as microcontroller unit (MCU) Smartphone: HTC HD2 with a 1 GHz CPU and 448 MB RAM (can be replaced with any android, Windows or IOS phone)	Hidden Markov Model Layered Hidden Markov Model	PhysioNet MIT-BIH dataset	Sensitivity: 99.72% Positive Predictive Value: 99.64%
[43]	2011	Congestive Heart Failure Malignant Ventricular Ectopy Ventricular Tachycardia	Electrocardiogram	A wireless ECG sensor S3C6400 mobile phone HBE-ZigBeX motes as a wireless sensor network	Multilayer Perceptron	PhysioNet MIT-BIH dataset	BIDMC Congestive Heart Failure: 100% Malignant Ventricular Ectopy: 90.9% Ventricular Tachyarrhythmia: 83.3%
[44]	2015	Atrial Fibrillation	Electrocardiogram	Rejiva ECG wearable sensor and a smartphone	Support Vector Machines	PhysioNet MIT-BIH dataset	Specificity: 77.25% Sensitivity: 93.13% Accuracy: 95%
[45]	2016	Atrial Fibrillation	Electrocardiogram Photoplethysmogram	Samsung Simband wrist band smart watch	Elastic Net Logistic model	Private Data	Sensitivity: 97% Specificity: 94% AUROC: 99%
[46]	2016	Myocardial Ischemia	Electrocardiogram	A smart cloth composed of four units: Smart cloth unit to measure physiological signal-ECG signal Signal control unit to control and memorize the status of the device by an ultra-low power MCU and SD card to save the signal data Signal sensing unit that has a motion tracking sensor module to capture the accelerometer signal Wireless connection unit to transmit the data A smartphone	Neural Network	PhysioNet MIT-BIH dataset PhysioNet MIT-BIH Normal Sinus Rhythm dataset	Accuracy > 76%
[47]	2017	Atrial Fibrillation	Electrocardiogram Photoplethysmogram	Samsung Simband wrist band smart watch	Convolutional Neural Network Elastic Net Logistic model	Private Data	Accuracy: 91.8%
[48]	2017	Heart Attack	Electrocardiogram Body Temperature	Device composed of pulse sensor, a temperature sensor, an Arduino, and a Low Energy (LE) Bluetooth A smartphone	Not Identified	Private Data	
[49]	2017	Ventricular Premature Complex Atrial Premature Complex Ventricular Fibrillation Atrial Fibrillation	Electrocardiogram	Bio Clothing One, XYZ life BC1	Artificial Neural Networks	PhysioNet American Heart Association database PhysioNet Creighton University Ventricular Tachyarrhythmia database PhysioNet MIT-BIH dataset PhysioNet MIT-BIH Noise Stress Test database	Accuracy > 75%
[50]	2017	Atrial Fibrillation	Electrocardiogram	Wrist bracelet designed for the purpose: based on the ultra low power series Microcontroller STM32L471RG	Support Vector Machines	Private Data	Accuracy: 95%
[51]	2017	Atrial Fibrillation	Audio Signal in Radial Artery	The PAG monitoring device consists of four components audiogram sensor: Panasonic capacitive microphone analog-digital converter: Embedded in Atmega328P microprocessor: Atmega328P chip data storage unit A smartphone	Convolutional Neural Network	Dataset collected at National Cheng Kung University Hospital (NCKUH), Tainan, Taiwan.	Accuracy : 98.92%

Table 2. *Cont.*

Ref#	Year	Disease(s) Targeted	Vital Signs Collected	Hardware Employed	Smart Model(s) Used	Training Dataset(s)	Results Metrics
[52]	2018	Myocardial Infarction	Electrocardiogram	ECG sensor using AD8232 and Espressif ESP-32 Wi-Fi + BLE module	Convolutional Neural Network	PhysioNet PTB Diagnostic ECG Database	Accuracy: 84%
[53]	2018	Ventricular Arrhythmia Junctional Arrhythmia Supraventricular Arrhythmia Arrhythmias	Electrocardiogram	a smart clothing consisting of cloth carrier, biosen sor platform, and smart terminals. In biosensor platform, ADI ECG analog front-end (ADAS1001) is used for obtaining the ECG signals, Microcontroller (STM32) is used to realize the data processing and a Bluetooth module is available for data transfer	Deep Neural Network with a Softmax Regression model	PhysioNet MIT-BIH dataset	Accuracy > 94%
[54]	2018	Hypertension	Heart Rate	A waist belt comprised of three kinds of sensors: three dry electrodes, a 3-axis accelerometer and two pressure sensors with different sensitivities	Logistic Regression Support Vector Machines	Private Data	Accuracy: 93.33%
[55]	2018	Atrial Fibrillation	Electrocardiogram Photoplethysmogram	Samsung gear device wearable device	Convolution–Recurrent Hybrid Model (CRNN)	Private Data	Accuracy > 98%
[56]	2018	Atrial Fibrillation	Electrocardiogram	A smart shirt equipped with ECG sensors A smartphone		Dataset collected at the Dongsan Medical Center in South Korea	Accuracy: 98.2%
[57]	2018	Ventricular Tachycardia Ventricular Bradycardia Premature Atrial Contractions Premature Ventricular Contractions	Electrocardiogram	for ECG Sensing: ECG body sensor with analog conditioning circuit (AD8232), Microcontroller unit (MCU) (PIC12F1822), Bluetooth module (HC-06), and charging controller module for analysis and display: processing and displaying unit of that process the ECG signal and display it on thin film transistor (TFT) liquid crystal display (LCD) consisting of Rpi computer, Bluetooth module, TFT screen, and power supply	Support Vector Machines	PhysioNet MIT-BIH dataset	Accuracy: 96.2%
[58]	2019	Myocardial Infarction Heart Failure Arrhythmias Fusion Beats Supraventricular Ectopic Beats Ventricular Ectopic Beats	Electrocardiogram Heart Rate Respiratory Rate	A patch with electronic circuit is built for the purpose and proposed in the article and an Android smartphone and a cloud server for data storage and further analysis	Convolutional Neural Network	PhysioNet PTB Diagnostic ECG Database St Petersburg INCART 12-lead Arrhythmia Database	Accuracy: 98.7%
[59]	2019	Atrial Fibrillation	Electrocardiogram	A patch with electronic circuit is built for the purpose and proposed in the article and an Android smartphone and a cloud server for data storage and further analysis	Decision Tree	PhysioNet MIT-BIH dataset	Accuracy > 97.18%
[60]	2019	Atrial Fibrillation Atrial Flutter Ventricular Fibrillation	Electrocardiogram	A wearable ECG sensing device and an Android smartphone and a cloud server for data storage and further analysis	Convolutional Neural Network	PhysioNet MIT-BIH dataset	Accuracy > 94%
[61]	2019	Atrial Fibrillation	Electrocardiogram	Smart vest equipped with two ECG sensing units	Long Short-Term Memory	PhysioNet dataset of the 2017 Computing in Cardiology Challenge	Sensitivity: 83.82% Specificity: 97.84% F1-score: 81.43%
[62]	2019	Supraventricular Ectopic Beats Ventricular Ectopic Beats	Electrocardiogram	ECG sensing device with a smartphone or tablet	Long Short-Term Memory	PhysioNet MIT-BIH dataset	Accuracy > 79%
[63]	2019	Atrial Fibrillation	Heart Rate	Commercial HR Sensor	Long Short-Term Memory	PhysioNet Atrial Fibrillation Database (AFDB) PhysioNet MIT-BIH dataset	Accuracy: 98.51%
[64]	2019	Arrhythmias Congestive Heart Failure	Electrocardiogram	One lead ECG sensor	Convolutional Neural Network	PhysioNet MIT-BIH Normal Sinus Rhythm database	Accuracy: 93.75%
[65]	2019	Arrhythmias	Electrocardiogram	A device composed of a single-lead heart rate monitor front end AD8232 chip, Atmel's ATmega128 as a microcontroller and a BLE module A smartphone is also used	Support Vector Machines K-Nearest Neighbors Logistic Regression Random Forest Decision Tree Gradient Boosting Decision Tree	PhysioNet MIT-BIH dataset	Accuracy > 77%
[66]	2019	Atrial Fibrillation	Photoplethysmogram	Wearable wristband device	Support Vector Machines	Private Data	Accuracy: 90%

Table 2. *Cont.*

Ref#	Year	Disease(s) Targeted	Vital Signs Collected	Hardware Employed	Smart Model(s) Used	Training Dataset(s)	Results Metrics
[67]	2020	Atrial Bigeminy Atrial Fibrillation Atrial Flutter Ventricular Bigeminy Heart Block Ventricular Trigeminy Ventricular Flutter Ventricular Tachycardia Supraventricular Tachyarrhythmia Idioventricular Rhythm Paced Beats Nodal (A-V Junctional) Rhythm	Electrocardiogram	SparkFun Single Lead Heart Rate Monitor AD8232 as the data acquisition device Smartphone as a gateway to the server	Convolutional Neural Network	PhysioNet MIT-BIH dataset	Accuracy: 94.13%
[68]	2020	Atrial Fibrillation Left Bundle Branch Block Beats Right Bundle Branch Block Beats	Electrocardiogram Photoplethysmogram	Amazfit Healthband 1S for ECG and PPG sensing smartphone for data reception and analysis	Convolutional Neural Network	Dataset collected at Peking University First Hospital	Sensitivity: 80.00% Specificity: 96.81% Accuracy: 90.52%
[69]	2020	Atrial Premature Contraction Ventricular Premature Contraction Paced Beats Ventricular Escape Beats	Electrocardiogram	A sensing device composed from a single lead heart rate monitor AD8232 and interfaced with NodeMCU development board having ESP8266 microcontroller capable of connecting to internet via WiFi Smartphone for the analysis of the data	Convolutional Neural Network	PhysioNet MIT-BIH dataset	Accuracy > 90%
[70]	2020	Cardiovascular Risk	Electrocardiogram Electroencephalogram Electromyogram Heart Rate Blood Pressure Respiratory Rate Blood Sugar Level Oxygen Saturation Level Cholesterol Levels	Wearable medical sensors and a wearable smart watch	Convolutional Neural Network	UCI Cleveland Heart Diseases Dataset	Accuracy: 98.5%
[71]	2020	Atrial Fibrillation	Electrocardiogram Photoplethysmogram Photoplethysmogram Oxygen saturation Level Body Temperature	The sensing device used is composed of three parts: AD8232r for ECG detection, ADS1115 analog-to-digital converter and SX1276 LoRa chip that transmits the data to the fog device The fog device: a low-cost raspberry pi system integrated with Intel Neural Compute Stick 2 (NCS 2) that is capable of handling deep learning algorithms	Convolutional Neural Network	PhysioNet dataset of the 2017 Computing in Cardiology Challenge	Accuracy: 90%
[72]	2020	Cardiovascular Risk	Electrocardiogram Blood Pressure	An ECG sensing device built with AD8232 unit A smart watch raspberry pi with SX1272 unit to transmit the data for LoRa gateway	Convolutional Neural Network	UCI Cleveland Heart Diseases Dataset	Accuracy: 98.2%
[73]	2020	Aortic Stenosis Mitral Insufficiency Mitral Stenosis Tricuspid Regurgitation	Electrocardiogram Photoplethysmogram Gyrocardiography Seismocardiogram	Shimmer 3 from Shimmer Sensing for ECG detection A three-axis MEMS accelerometer: (Kionix KXRB5-2042, Kionix, Inc.) to measure the SCG signal A three-axis MEMS gyroscope (Invensense MPU9150, Invensense, Inc.) to record the GCG signal An ear-lobe photoplethysmography (PPG) sensor	Decision Tree Random Forest Neural Network	Dataset collected at Columbia University Medical Center (CUMC)	Accuracy > 90%
[74]	2020	Left Bundle Branch Block Beats Right Bundle Branch Block Beats Atrial Escape Beats Nodal (Junctional) Escape Beats Atrial Premature Beats Aberrated Atrial Premature Beats Nodal Premature Beats Supraventricular Premature Beats Premature Ventricular Contractions Ventricular Escape Beats Fusion of Ventricular and Normal Beats Paced Beats Fusion of Paced and Normal Beats	Electrocardiogram	A sensing device composed of AD8232 single-lead three-electrode ECG Heart Rate monitor and a ESP8266 Wi-Fi module used to provide wireless data transmission access to the Arduino Nano and is used to connect it to the cloud	Convolutional Neural Network	PhysioNet MIT-BIH dataset	Accuracy: 99.625% Sensitivity: 97.736% Specificity: 99.713% Precision: 97.835%

Table 2. *Cont.*

Ref#	Year	Disease(s) Targeted	Vital Signs Collected	Hardware Employed	Smart Model(s) Used	Training Dataset(s)	Results Metrics
[75]	2020	Ventricular Ectopic Beats Arrhythmias Premature Atrial Contractions	Electrocardiogram	Sensing device composed of Raspberry Pi for processing, ADS1115 as Analog to Digital Converter and AD8232 as ECG sensor	Convolutional Neural Network	PhysioNet MIT-BIH dataset	Accuracy: 95.76%
[76]	2020	Premature Ventricular Contractions Atrial Fibrillation	Electrocardiogram Photoplethysmogram	7-lead Holter monitor (Rozinn RZ153+ Series, Rozinn Electronics Inc., Glendale, NY, USA) Smartwatch (Simband 2, Samsung Digital Health, San Jose, CA,USA)	Random Forest Support Vector Machines	Dataset collected at the ambulatory cardiovascular clinic at the University of Massachusetts Medical Center (UMMC)	Best Model Accuracy: 94%
[77]	2020	Arrhythmias	Electrocardiogram	Sensing device built using Raspberry Pi 3 model B+ and two ECG sensors AD8232 with a pulse sensor and an analog digital converter ADS1015	Support Vector Machines Naïve Bayes Artificial Neural Networks	PhysioNet MIT-BIH dataset	Best Model Accuracy: 97.8%
[78]	2020	Atrial Fibrillation	Electrocardiogram	the wearable system is composed to work on a prototype developed by Medicaltech srl (Rovereto, Italy)	A Custom model based on Thresholding of Shannon Entropy values	PhysioNet MIT-BIH dataset	Sensitivity: 99.2% Specificity: 97.3%
[79]	2020	Atrial Fibrillation	Electrocardiogram	The sensing device is composed of Raspberry pi 3, Arduino UNO, AD8232 single lead ECG sensor, HC-05 Bluetooth, biomedical sensor pad and battery	Long Short-Term Memory	PhysioNet MIT-BIH dataset	Accuracy: 97.57%
[80]	2020	Atrial Escape Beats Junctional Escape Beats Left Bundle Branch Block Beats Right Bundle Branch Block Beats Atrial Premature Beats Aberrated Atrial Premature Beats Junctional Premature Beats Supraventricular Premature Beats Premature Ventricular Contractions Ventricular Escape Beats Fusion of Ventricular and Normal Beats Paced Beats Fusion of Paced and Normal Beats Supraventricular Arrhythmia	Electrocardiogram	Moto 360 NanoPi Neo Plus2 Raspberry Pi Zero	Long Short-Term Memory	PhysioNet MIT-BIH dataset	Accuracy > 98.6 %
[81]	2020	Atrial Fibrillation Arrhythmias	Electrocardiogram	A wearable sensing device composed of AD8232 as an ECG sensor, MCP3008 ias an ADC and Raspberry Pi as a computing unit	Support Vector Machines	UCI Cleveland Heart Diseases Dataset	Accuracy: 72.41%
[82]	2020	Arrhythmias	Electrocardiogram Body Temperature Heart Rate Blood Oxygen Level	A sensing device composed of: Temperature sensor: MLX90614 Heart rate and blood oxygen sensors: MAX30100 ECG sensor: AD8232 Inter-Integrated Circuit (I2C) communication protocol Microcontroller: Arduino UNO Wireless transmission: Wi-Fi chip ESP8266 A smartphone	Long Short-Term Memory Convolutional Neural Network	PhysioNet MIT-BIH dataset	Accuracy: 99.05%
[83]	2020	Premature Ventricular Contraction	Electrocardiogram	A wireless 3-lead ECG sensor from Shimmer Sensing	Support Vector Machines	PhysioNet MIT-BIH dataset	Sensitivity: 96.51% Predictive Value: 81.92%
[84]	2020	Atrial Fibrillation Syncope	Electrocardiogram	A sensing device composed of: The SparkFun AD8232 ECG sensing unit Arduino Mega 2560 microcontroller Raspberry Pi 3 board ADXL345 triple-axis accelerometer HC-05 Bluetooth sensor	Long Short-Term Memory	PhysioNet MIT-BIH dataset	Accuracy: 97.61%
[85]	2021	Atrial Fibrillation	Pulse Plethysmogram	A smartphone Wrist-type pulse wave monitor (type: Smart TCM-I, product by: Shanghai Asia & Pacific Computer Information System CO, Ltd, Shanghai, China)	Time Synchronous Averaging	Private Data	Accuracy: 98.4%

Table 2. *Cont.*

Ref#	Year	Disease(s) Targeted	Vital Signs Collected	Hardware Employed	Smart Model(s) Used	Training Dataset(s)	Results Metrics
[86]	2021	Cardiovascular Risk	Photoplethysmogram	Pulse rate sensor with ATmega32 microcontroller	Support Vector Machines Naive Bayes Random Forest Decision Tree Logistic Regression Artificial Neural Networks Recurrent Neural Networks	Dataset collected at Framingham University	Accuracy: 94.9%
[87]	2021	Ventricular Ectopic Beats Supraventricular Ectopic Beats Premature Atrial Contractions Premature Ventricular Contractions	Electrocardiogram	Ternary second-order delta modulator circuits	Support Vector Machines	PhysioNet MIT-BIH dataset	Accuracy > 98%
[88]	2021	Atrial Fibrillation Ventricular Tachycardia Sinus Bradycardia Atrial Tachycardia	Electrocardiogram	A custom-built ECG Signal acquisition circuit	Gramian Angular Fields (GAFs) Deep Residual Network (ResNet)	PhysioNet MIT-BIH dataset LTAF database Simulated Data (Prosim2 Vital Sign Simulator)	Accuracy: 98.1% Sensitivity: 97.6% Specificity: 99.7% F1 Score: 97.6%
[89]	2021	Arrhythmias Congestive Heart Failure	Electrocardiogram	ARDUINO UNO ECG SENSOR AD8232 DISPOSABLE ECG ELECTRODES	Support Vector Machines	PhysioNet dataset of the 2016 Computing in Cardiology Challenge	Accuracy: 98%
[90]	2021	Atrial Fibrillation	Electrocardiogram	A consumer-grade, single-lead heart belt (Suunto Movesense, Suunto, Vantaa, Finland)	Not Identified	Private Data	Accuracy 97.8%
[91]	2021	Atrial Fibrillation Atrial Flutter Supraventricular Tachycardia Ventricular Tachycardia	Electrocardiogram	ECG247 Smart Heart Sensor	Not Identified	Private Data	Accuracy > 95%
[92]	2021	Heart Attack	Electrocardiogram Heart Rate Body Temperature Blood Pressure	A device composed of ECG, heart rate, body temperature, and blood pressure sensors	Not Identified	Private Data	Accuracy: 83%
[93]	2021	Atrial Fibrillation Ventricular Bradycardia Ventricular Tachycardia Bundle Branch Block	Electrocardiogram	HealthyPiV3 biosensors	Convolutional Neural Network	PhysioNet MIT-BIH dataset PhysioNet PAF Prediction Challenge Database for AF records PhysioNet PTB Diagnostic ECG Database PhysioNet dataset of the of 2015 bradycardia Challenge PhysioNet Fantasia Database and PAF Prediction Challenge Database for healthy signals	Accuracy > 98.75%
[94]	2021	Heart Attack	Electrocardiogram Electrocardiogram	AD8232 ECG sensor	Sequential Covering Algorithm Support Vector Machines	PhysioNet PTB-XL dataset	F1 Score: 87.8%
[95]	2021	Heart Attack	Body Temperature Activity Parameters Oxygen Saturation Level	Composed of different sensors to collect different vital signs which are: LM35, MPU 6050, MAX30100 and AD8232 respectively	Linear Regression K-Nearest Neighbors Naive Bayes	Private Data	Accuracy: 80%
[96]	2021	Ventricular Premature Beats Supraventricular Premature Beats Atrial Fibrillation	Electrocardiogram	IREALCARE2.0 Wearable ECG Sensor	Time-Span Convolutional Neural Network Recurrent Neural Networks	Private Data	F1 Score: 86.5% Precision: 87.7% Recall: 86.8%
[97]	2021	Cardiovascular Risk	Electrocardiogram Oxygen Saturation Level	Composed of AD8232 (ECG sensor) and MAX30102 (SPO2 sensor)	Convolutional Neural Network Convolutional Neural Network	PhysioNet MIT-BIH dataset	Shallow CNN Accuracy: 96.06% Deep CNN Accuracy: 98.47%
[98]	2021	Heart Failure Hypertension Atrial Fibrillation Peripheral Artery Disease Myocardial Contraction	Heart Rate Activity Parameters	GENEActiv and Activinsights Band (Activinsights Ltd., Kimbolton, UK)	Not Identified	To be collected	To be provided
[99]	2021	Atrial Fibrillation	Heart Rate Respiratory Rate	BioHarness 3.0 by Zephyr	Support Vector Machines	PhysioNet MIT-BIH dataset	Sensitivity: 78% Specificity: 66%
[100]	2021	Atrial Fibrillation Bigeminy Arrhythmias	Electrocardiogram	AD8232	Decision Tree	Private Data	Accuracy > 95%

Table 2. *Cont.*

Ref#	Year	Disease(s) Targeted	Vital Signs Collected	Hardware Employed	Smart Model(s) Used	Training Dataset(s)	Results Metrics
[101]	2021	Atrial Fibrillation Atrial flutter Left Bundle Branch Block Beats Wolff-Parkinson-White Syndrome Atrial Premature Contraction Premature Ventricular Contraction	Electrocardiogram	A smart vest equipped with AD8232 ECG Sensor	Shallow Wavelet Scattering Network (ScatNet)	PhysioNet MIT-BIH dataset	Accuracy > 96%
[102]	2021	Tachycardia	Heart Rate Respiratory Rate Blood Oxygen Level	Medical-grade wearable embedded system (SensEcho, Beijing SensEcho Science & Technology Co Ltd)	Long Short-Term Memory	Medical Information Mart for Intensive Care III (MIMIC-III)	Up to 80% accuracy 2 h before onset of Tachycardia
[103]	2021	Atrial Fibrillation	Photoplethysmogram	Samsung Galaxy Active 2 Watch	Convolutional Neural Network	Private Data	Accuracy 91.6% Specificity 93.0% Sensitivity 90.8%
[104]	2021	Arrhythmias	Electrocardiogram	A chest sticker that is composed from BMD101 ECG sensing device with YJ33 power supply, BQ24072 as a power source and JDY-30 as a Bluetooth module	Convolutional Neural Network	PhysioNet MIT-BIH dataset	Accuracy: 99.83%
[105]	2022	Supraventricular Ectopic Beats Ventricular Ectopic Beats Fusion Beats	Electrocardiogram	Polar H10	Decision Tree Gradient Boosting k-Nearest Neighbors Multilayer Perceptron Random Forest Support Vector Machines	PhysioNet MIT-BIH dataset	Best Model Accuracy: 99.67%
[106]	2022	Supraventricular Ectopic Beats Ventricular Ectopic Beats Fusion Beats	Electrocardiogram	Polar H10	Decision Tree Gradient Boosting k-Nearest Neighbors Multilayer Perceptron Random Forest Support Vector Machines	PhysioNet MIT-BIH dataset	Best Model Accuracy: 99%
[107]	2022	Heart Failure Reduced Ejection Fraction	Electrocardiogram	Galaxy Watch Active & AppleWatch 6	Convolutional Neural Network	Private Data	Area Under Curve 93.4%
[108]	2022	Atrial Fibrillation	Photoplethysmogram Electrocardiogram	Samsung GalaxyWatch Active 2 Chest ECG Patch	Hybrid Decision Model	Private Data	Average: 67.8%
[109]	2022	Atrial Fibrillation	Photoplethysmogram	Custom-built device that contains the PPG sensor MAX30102	Convolutional Neural Network	Data obtained from Kaunas University of Technology	F1-score: 94%
[110]	2022	Atrial Fibrillation	Electrocardiogram	Firstbeat Bodyguard 2, Firstbeat Technologies	Not Identified	Private Data	Accuracy 98.7% Sensitivity 99.6%, Specificity 98.0%
[111]	2022	Supraventricular Ectopic Beats Ventricular Ectopic Beats	Electrocardiogram	Custom-built device that contains the ECG AFE sensor	Artificial Neural Networks Decision Tree K-Nearest Neighbors	PhysioNet MIT-BIH dataset	Accuracy: 98.7%
[112]	2022	Atrial Fibrillation	Photoplethysmogram	Apple Watch	Gradient Boosting Decision Tree	Private Data	Accuracy: 94.16%
[113]	2022	Congestive Heart Failure Atrial Fibrillation	Electrocardiogram	AD8232 sensor	Random Forest	PhysioNet MIT-BIH dataset	Accuracy: 85%
[114]	2022	Cardiovascular Risk	Photoplethysmogram Body Temperature Activity Parameters	Custom-built device with Pulse Sensor, DS18B20 temperature sensor and ADXL 1335 as accelerometer sensor	Naïve Bayes Decision Tree K-Nearest Neighbors Support Vector Machines	Kaggle Human Gait Dataset Kaggle Heart Disease Prediction Dataset	Accuracy: 82%
[115]	2022	Cardiovascular Risk	Heart Rate Respiratory Rate Blood Oxygen Level Electrocardiogram	Not identified (WBAN)	Enhanced version of Recurrent Neural Network named ERNN	Private Data	Accuracy: 96%
[116]	2022	Cardiovascular Risk	Electroencephalogram Body Temperature Blood Oxygen Level Respiratory Rate Blood Sugar Level	A custom-built device equipped with electrocardiogram sensor, electroencephalogram sensor, an electro-mammography sensor, an oxygen level sensor, a temperature sensor, a respiration rate sensor, and a glucose level sensor	Long Short-Term Memory	UCI Cardiac Arrhythmia Dataset	Average Positive Predictive Value: 96.77% Average Negative Predictive Value: 95.12% Average Sensitivity: 95.30%

Table 2. *Cont.*

Ref#	Year	Disease(s) Targeted	Vital Signs Collected	Hardware Employed	Smart Model(s) Used	Training Dataset(s)	Results Metrics
[117]	2022	ST Elevation Myocardial Infarction (STEMI)	Electrocardiogram Motion Data	Custom-built device with 3-axis accelerometer (ADXL335), 3-axis gyroscope (LSM6DS3) and single-lead ECG sensors	Logistic Regression	Private Data	Sensitivity: 73.9% Specificity: 85.7%
[118]	2022	Cardiovascular Risk	Electrocardiogram Motion Data	A custom-built device with accelerometers, Galvanic Skin Response (GSR) and electrocardiograms (ECG) sensors	Mixed Kernel Based Extreme Learning Machine (MKELM)	Private Data	Accuracy: 99.5%
[119]	2022	Cardiovascular Risk	Heart Rate	Wrist Strap & Rohm BH1790GLC-EVK-001 Development board BH1790GLC Optical heart rate sensor	Convolutional Neural Network	Simulated Data	F1-Score: Up to 99%
[120]	2022	Myocardial Infarction Dilated Cardiomyopathy Hypertension	Pulse Plethysmogram	PTN-104 PPG sensor	Support Vector Machines K-Nearest Neighbors Decision Tree	Private Data	Accuracy: 98.4% Sensitivity: 96.7% Specificity: 99.6%
[121]	2022	Cardiovascular Risk	Heart Rate Blood Sugar Level	Heart rate sensor by Sunrom Electronics Glucose monitor by Medtonic	Naïve Bayes K-Nearest Neighbors Support Vector Machines Random Forest Artificial Neural Networks	Private Data	Accuracy: 97.32% Recall: 97.58% Precision: 97.16% F1-Measure: 97.37% Specificity: 96.87% G-Mean: 97.22%
[122]	2022	Cardiovascular Risk	Electrocardiogram	A custom-built device composed of ECG sensor (AD8232) and other components	Random Forest	UCI Cleveland Heart Diseases Dataset	Accuracy: 88%
[123]	2022	Cardiovascular Risk	Heart Rate Oxygen Saturation Level Systolic Pressure Diastolic Pressure	Custom-built soft transducer equipped with MAX30100 SpO2 and HR monitor sensor	Long Short-Term Memory	Kaggle dataset (Not Specified)	Accuracy > 93%
[124]	2022	Cardiovascular Risk	Electrocardiogram Blood Pressure Pulse Plethysmogram Body Temperature	Custom-built device equipped with ECG sensor, TMP117 temperature sensor, Honeywell's 26 PC SMT blood pressure sensor, and a pulse oximeter	Recurrent Neural Networks	UCI Cleveland Heart Diseases Dataset	Accuracy: 99.15% Precision: 98.06% Recall: 98.95% Specificity: 96.32% F1-Score: 99.02%
[125]	2022	Congenital Heart Disease	Electrocardiogram Seismocardiogram	Custom-built chest wearable sensor equipped with ECG sensor (ADS1291; Texas Instruments, Dallas, TX) and seismocardiogram sensor (ADXL355; Analog Devices, Norwood, MA)	Ridge Regression	Private Data	-

3.3. Results of Individual Studies

The systems presented in the eligible studies share common features that allow for easy classification. Contextually, the studies can be divided into three categories, for example, according to whether the measuring devices used are commercially available or not. Systems in the first group use components that are not commonly available; these components were custom-made by the researchers for the study. Systems that use readily available technology and commercially available devices comprise the second category. The final category includes studies that used unspecified devices, making it impossible to determine whether or not they are now available for purchase. The following categories of systems were formed according to the devices used.

3.3.1. Studies Using Custom-Built Devices

Throughout the analysis of the eligible documents, it was shown that 55 studies built their own devices using various vital sign sensors, power resources, storage resources, communications, and other technical components. Within this group, two subgroups stood out, the first of which did not name all the components used, particularly the sensor devices, but, rather, stated that they composed their own wearables from sensor devices. Thus, in [39,41–43,46,48,50,51,54,56,58–62,66,86,88,92,116,118,123], the authors proposed custom-built wearables with unspecified components. These studies were able to detect various cardiovascular diseases such as atrial fibrillation, atrial flutter, atrial premature contraction, atrial tachycardia, cardiac asystole, cardiovascular risk, fusion beats, heart attack, heart failure, hypertension, myocardial infarction, myocardial ischemia, premature atrial contractions, and premature ventricular contractions. Vital signs obtained for this purpose were radial artery audio signal, blood oxygen level, blood pressure, blood sugar level, body temperature, diastolic pressure, electrocardiogram, electroencephalogram, heart rate, motion data, oxygen saturation level, photoplethysmogram, respiratory rate, and skin temperature. In addition, the databases used for training and the performance metrics are detailed in Table 2.

On the other hand, several studies used commercially available sensors to develop their wearable devices. In this context, various sensors such as ECG, accelerometer, and other sensors were used. For example, in [53,73,104,117,125], the ECG sensors ADAS1001, Shimmer 3, BMD101, ADXL355, and ADS1291 were used in combination with other materials to build a wearable device that collects records used to detect or predict cardiovascular disease. In contrast, the authors in [114] used the DS18B20 temperature sensor and ADXL1335 accelerometer to develop the desired wearable system. In addition, the authors in [52,57,65,67,69,71,72,74,75,77,79,81,89,94,100,101,113,122] used the AD8232 ECG sensor to collect vital signs data. In these studies, as discussed in Table 2, different processing units, connector modules, and power sources were used to build the wearable device. Alternatively, in [82,84,95,97], the authors combined different sensor materials with the AD8232 ECG sensor in their wearable device. Specifically, the authors in [82,95] used the MAX30100 blood oxygen sensor in addition to the ECG sensor, whereas the authors in [84] used the ADXL345 triaxial accelerometer, and the authors in [97] used the MAX30102 pulse oximeter sensor. Other studies also used different ECG sensors, with the authors in [87] building their portable devices using the "Ternary Second-Order Delta Modulator Circuits" to acquire ECG data. In the same context, the authors in [111,124] used the sensor ECG AFE and other tools to build a wearable device capable of acquiring the necessary vital signs data. Finally, the authors of [109,119] used MAX30102 photoplethysmography and BH1790GLC optical heart rate sensors in their wearable devices, respectively.

3.3.2. Studies Using Commercially Available Wearable Devices

The other group of studies consists of studies that used commercially available devices. These devices were capable of recording various vital signs such as activity parameters, blood oxygen level, blood pressure, blood glucose level, cholesterol level, electrocardiogram, electroencephalogram, electromyogram, heart rate, oxygen saturation level,

photoplethysmogram, pulse plethysmogram, and respiratory rate. Depending on the type of device used, three categories can be distinguished, namely, wristband devices, belts, and others. For example, in [45,47,55,68,70,80,85,98,103,107,108,112], the authors used smartwatches and smart wristbands to record vital signs. In addition, the authors in [40,44,49,63,64,76,78,83,91,93,96,102,110,120,121] used various wearable ECG devices such as smart vests and patches. In addition, the authors in [90,99,105,106] used smart belts to collect ECG data. Overall, the devices used in all the studies mentioned in this section can be summarized in the following list:

- Alive ECG Heart Monitor;
- Amazfit Health band 1S;
- Apple Smart Watch;
- Bio Clothing One, XYZ life BC1;
- BioHarness 3.0 by Zephyr;
- ECG247 Smart Heart Sensor;
- Firstbeat Bodyguard Chest Patch 2 by Firstbeat Technologies;
- GENEActiv and Activinsights Band by Activinsights Ltd.;
- Glucose Monitor by Medtonic;
- HealthyPiV3 biosensors;
- Heart Rate sensor by Sunrom Electronics;
- IREALCARE2.0 Wearable ECG Sensor;
- Kimbolton, UK;
- Medical-Grade Wearable Embedded System Beijing Sensecho Science & Tech.;
- Wearable device provided by Medicaltech SRL;
- Moto 360;
- NanoPi Neo Plus2;
- Polar H10;
- PTN-104 PPG Sensor;
- Raspberry Pi Zero;
- Rejiva ECG Wearable Sensor;
- Rozinn RZ153+ ECG Monitor;
- Samsung Galaxy Active 2 Smart Watch;
- Samsung Galaxy Active Smart Watch;
- Samsung Gear Wearable Device;
- Samsung Simband 2 Wrist Band Smart Watch;
- Samsung Simband Wrist Band Smart Watch;
- Shimmer ECG Monitor;
- Single-Lead Heart Belt by Suunto Movesense, Suunto, Vantaa, Finland;
- Wrist-Type Pulse Wave Monitor by: Shanghai Asia & Pacific Computer Info. System.

3.3.3. Studies That Did Not Specify the Devices Used

Finally, in a single study, the device used was not specified. The authors in [115] mentioned only that they used a wide body area network (WBAN) to record respiratory rate and blood oxygen levels to detect the presence of cardiovascular risk. Their study achieved 96% accuracy in the classification algorithm, demonstrating high feasibility in detecting CVDs. Unlike some of the studies excluded in the screening phases (see Section 2.4), this study mentioned that a wearable device was used, but did not specify which device was used.

4. Results Analysis

Studies that met the criteria for inclusion in this review, or those that specifically address the use of smart wearables for the diagnosis, prognosis, or treatment of cardiovascular disease, contain a wealth of information worthy of further investigation. In the previous section, the devices used were mentioned. However, to better understand the field, this

part examines factors such as the year of publication, vital signs collected, diseases treated, smart models used, datasets used for training, etc.

4.1. Progress with Years

Once the data are extracted from the papers, it is clear that there has been significant progress in the field of wearables for CVD research over the past four years, with 78% of the publications published in 2019 or later. During those years, a total of 68 studies were published (compared to only 19 in 2010–2018). There are nine in 2019, eighteen in 2020, twenty-one in 2021, and twenty-one in 2022. The number of publications addressing the use of smart wearables for cardiovascular disease management has jumped, reflecting both the growing interest in this area and the widespread acceptance of such devices. The data from this section are shown as a pie chart in Figure 4 below.

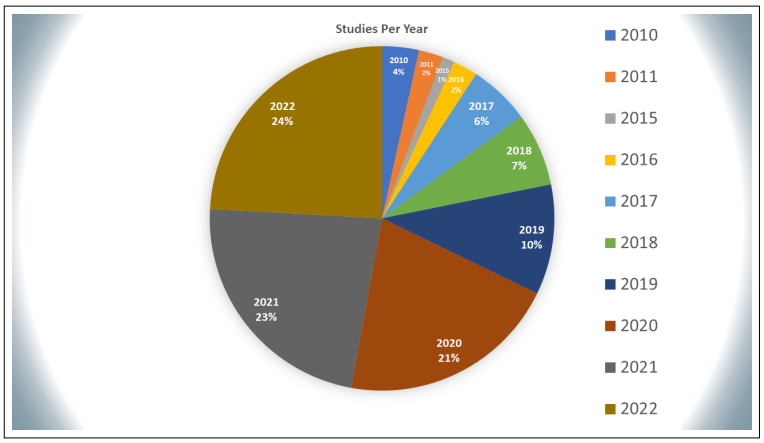

Figure 4. Distribution of studies per year.

4.2. Vital Signs in Use

The electrocardiogram (ECG) is used in 69 of the systems described in the articles to diagnose disease and identify cardiac abnormalities, although many other methods have been offered. Electrocardiograms are routinely performed to check the health of the heart and quickly identify potential problems. An electrocardiogram (ECG) shows the development of the heart's electrical activity over time. When the heart muscle cells are electrically depolarized, the heart muscle contracts. An electrocardiogram records and amplifies this electrical activity over a period of time. Studies have shown that smart watches such as the Samsung Active and Apple Watch have significant efficiency in capturing ECG signals, complementing the accuracy of ECGs performed in a doctor's office, clinic, or hospital room. In addition, the P wave, the QRS complex, and the T wave are the three components of the ECG signal. Figure 3 shows the ECG signal in terms of these components. In a normal electrocardiogram, the heartbeat is detected by [126]:

- PR interval: measured from the beginning of the P wave to the first deflection of the QRS complex with a normal range of 120–200 ms;
- QRS complex: measured from first deflection of QRS complex to end of QRS complex at isoelectric line with a normal range of up to 120 ms;
- QT interval: measured from first deflection of QRS complex to end of T wave at isoelectric line with a normal range of up to 440 ms (though it varies with heart rate and may be slightly longer in females).

Twenty more measures, including photoplethysmogram, heart rate, and others, were also employed in addition to ECG in order to identify CVDs. Table 3 below details the frequency and utilization of these parameters across studies.

Table 3. Vital signs used in studies with count and percentages.

Vital Sign	Count	Percentage
Electrocardiogram	69	79.31%
Photoplethysmogram	15	17.24%
Heart rate	13	14.94%
Body temperature	8	9.20%
Respiratory rate	7	8.05%
Oxygen saturation level	5	5.75%
Blood oxygen level	4	4.60%
Blood pressure	4	4.60%
Activity parameters	3	3.45%
Blood sugar level	3	3.45%
Electroencephalogram	3	3.45%
Pulse plethysmogram	3	3.45%
Motion data	2	2.30%
Seismocardiogram	2	2.30%
Audio signal in radial artery	1	1.15%
Cholesterol levels	1	1.15%
Diastolic pressure	1	1.15%
Electromyogram	1	1.15%
Gyrocardiography	1	1.15%
Skin temperature	1	1.15%
Systolic pressure	1	1.15%

4.3. Diseases Targeted

Because a single document may focus on a single disease or multiple diseases, the number of diseases studied in these publications exceeds 70. Atrial fibrillation (AFib) is the most commonly studied disease, with 39 of 87 studies addressing it. AFib is the leading cause of death and morbidity due to stroke, heart failure, thromboembolism, and reduced quality of life, and accounts for the majority of these cases [127]. Other conditions are also being studied, including premature ventricular contractions (PVCs), ventricular ectopic beats, bradycardia, paced beat (PACE), and many others. Figure 5 is a bar graph showing the number of diseases found in the 87 papers.

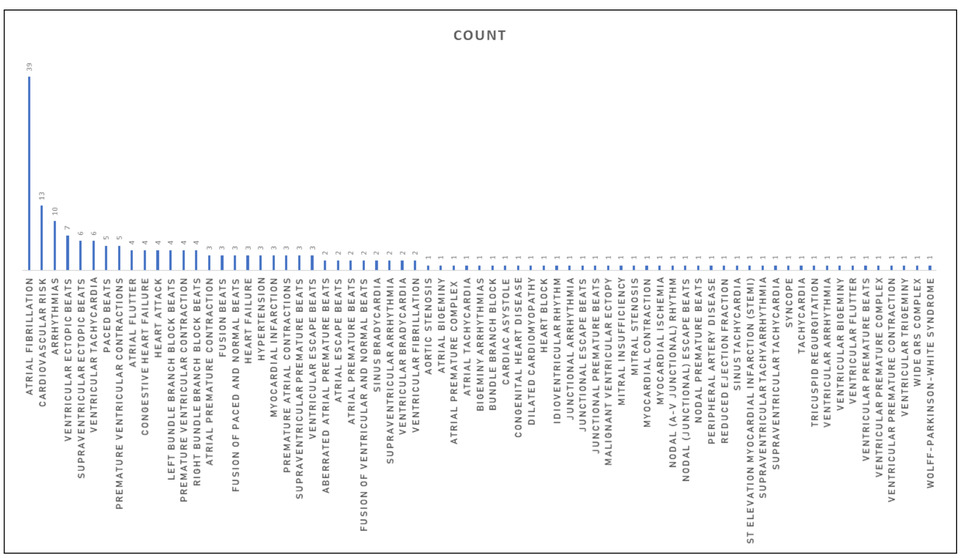

Figure 5. Diseases distribution per studies.

4.4. Smart Models in Use

It is well known that several subfields of artificial intelligence are widely used in different fields. Two of the best-known subfields of AI are machine learning (ML) and deep learning (DL); the former is described as a set of techniques that allow a machine to acquire new information and skills through learning, and the latter is a branch of machine

learning that focuses on algorithms inspired by the structure and function of the brain, called artificial neural networks [128,129]. The relationship between AI, ML, and DL is illustrated in Figure 6. However, in the studies analyzed in this review, many machine learning and deep learning models were used to detect cardiovascular disease. Although each publication proposes a different method to detect the disease(s), all agree that some type of algorithm should be used to classify cardiac abnormalities. Convolutional neural networks, support vector machines, long short-term memories, and decision trees are the most commonly used algorithms.

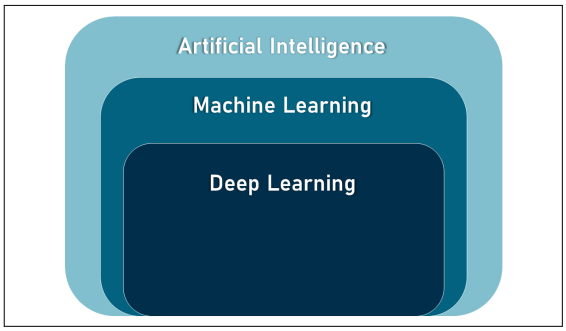

Figure 6. Artificial intelligence, machine learning, and deep learning relation.

The intelligent models of machine learning and deep learning are attracting much attention and are proving to be very practical [130,131] in the healthcare industry. With this in mind, it is of great interest to analyze the efficiency of these models in detecting CVDs. However, this task requires separate studies, as this research focuses on smart wearables as a whole system. This article aims to fill this gap by discussing the four most commonly used smart models, namely:

- Convolutional Neural Network (CNN): CNN is a kind of deep neural network used to analyze visual images. These neural networks are modeled after the neural networks of the human visual system. Neurons are the basic computational unit of a neural network, just as they are the basic functional unit of the human nervous system. In the case of convolutional neural networks, instead of normal matrix multiplication, convolution is used, a special form of mathematical operation. In addition to the input and output layers, a convolutional neural network has numerous hidden layers (a neural layer is a stack of neurons in a single row). A neuron in the input layer receives an input, analyzes it, and performs computations on it, and then transmits a nonlinear function called an activation function to produce the final output of a neuron [132];
- Support Vector Machines (SVMs): SVM is a supervised machine learning model for two-group classification problems that employs classification techniques. An SVM model is able to classify new data after receiving a set of labeled training data for each category [133];
- Long Short-Term Memory (LSTM): LSTM networks are a type of recurrent neural network (RNN) that can learn sequence dependence in sequence predictions. RNNs contain cycles that use network activations from a previous time step as inputs to influence predictions at the current time step. These activations are stored in the internal states of the network, theoretically preserving long-term contextual timing information. This method allows RNNs to use a contextual window that changes dynamically over the course of the input sequence. Complex problem domains such as machine translation, speech recognition, and others require this behavior [134];
- Decision Trees (DTs): A decision tree is a type of supervised machine learning used to make classifications or predictions based on answers to a prior set of questions. The model is a type of supervised learning, meaning that it is trained and evaluated on a dataset that contains the desired classification. Occasionally, the decision tree may not

provide a definitive answer or conclusion. Instead, it may suggest possibilities from which the data scientist can make an informed choice. Because decision trees replicate human thought processes, it is often easy for data scientists to understand and explain the results [135].

The machine learning and deep learning models utilized in the studies analyzed in this review were assessed with different performance metrics such as accuracy, specificity, sensitivity, precision, recall and F1-score. These parameters are explained in detail in the literature, with the authors in [129] providing a detailed explanation in this regard, for example. These parameters can be summarized as follows [129]:

- Accuracy: the fraction of predictions that the model predicted right and is calculated by dividing the number of correct predictions by the total number of predictions.
- Specificity: is the parameter used to calculate model's ability to predict a true negative (no cardiovascular diseases in our case) of each category available.
- Sensitivity: is the parameter used to calculate model's ability to predict the true positives (existence of CVDs in our case) of each category available.
- Precision: is the parameter used to calculate what proportion of positive identifications (existence of CVDs in our case) was actually correct.
- Recall: is the parameter used to calculate what proportion of actual positives (existence of CVDs in our case) was identified correctly.

The performance of models used by each study is detailed in Table 2. Furthermore, the list of smart models used in smart wearables for the detection of cardiovascular diseases is mentioned in Table 4 below, along with the count of use of each model. In this context, and for more details on the potential of machine learning and deep learning models in predicting CVDs, readers are advised to refer the work of Solam Lee and his colleagues [34], which targets these models and discusses their feasibility in this domain.

Table 4. Smart Models Used in Studies.

Smart Model	Count
Convolutional neural network	23
Support vector machines	20
Decision tree	10
Long short-term memory	10
Random forest	9
K-nearest neighbors	8
Artificial neural networks	5
Naïve Bayes	5
Not identified	5
Logistic regression	4
Multilayer perceptron	4
Recurrent neural networks	3
Elastic net logistic model	2
Gradient boosting	2
Gradient boosting decision tree	2
Neural network	2
A custom model based on thresholding of Shannon entropy	1
Convolution–recurrent hybrid model (CRNN)	1
Deep neural network with a softmax regression model	1
Deep residual network (ResNet)	1
Enhanced version of recurrent neural network named ERNN	1
Gramian angular fields (GAFs)	1
Hidden Markov model	1
Hybrid decision model	1
Layered hidden Markov model	1
Linear regression	1
Mixed-kernel-based extreme learning machine (MKELM)	1
Ridge regression	1
Sequential covering algorithm	1
Shallow wavelet scattering network (ScatNet)	1
Time-synchronous averaging	1
Time-span convolutional neural network	1

4.5. Datasets in Use

In all 87 publications examined, at least one dataset was used to train the AI model, and this is consistently the case. In addition, it was noted that certain sources advocate the use of multiple datasets in the development and evaluation of a model. The PhysioNet MIT-BIH dataset, accessible through the PhysioNet library of publicly available medical research data, is the most popular. Of the 87 total studies, 36 used it. Another 25 studies also used researchers' private data. Figure 7 below shows a graphical statistical representation of the frequency of use of datasets. The MIT-BIH Arrhythmia database is the first publicly available collection of standardized test material for the evaluation of arrhythmia detectors. The BIH Arrhythmia Laboratory collected these ambulatory two-channel ECG recordings from 47 patients between 1975 and 1979 and included 48 30-minute samples [136]. The PhysioNet MIT-BIH Atrial Fibrillation Database, the PhysioNet MIT-BIH Noise Stress Test Database (NSTDB), and the PhysioNet MIT-BIH Normal Sinus Rhythm Database were also consulted.

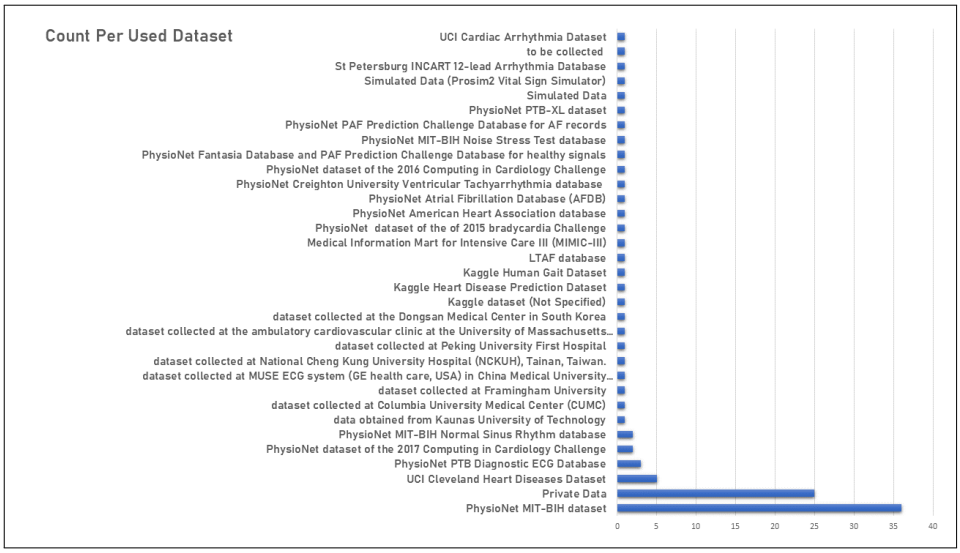

Figure 7. Training datasets in use.

5. Results Discussion

This study systematically collects and analyzes the literature on the use of smart wearables for cardiovascular disease diagnosis and prognosis. However, there is more to be said about the studies discussed so far, especially in terms of their effectiveness and conformity with the latest research areas in artificial intelligence. This topic will be elaborated and explored in this section.

5.1. Performance, Usability, and Feasibility

To predict CVDs, many tools have been used. The wide range of research is due to the wide range of vital signs and devices used to achieve this goal. ECG, BP, HR, and temperature were all reliable predictors of cardiovascular disease. This is evidenced by the fact that the results of several studies (described in Table 2) showing the use of different implementations yielded an accuracy rate of over 99%. However, there are several things to consider when making a final decision on a wearable gadget. The following is a list of features that would make a smart wearable more practical:

- Noninvasive: the gadget should not penetrate or pierce the skin to collect data;
- Compact: the wearable device should not be bulky or large, as its main purpose is to monitor health symptoms without interfering with one's life activities;
- Affordable: the affordability of the device plays a role in how well it fits into everyday life;

- Robust: the device should be durable enough to handle cold, hot, humid, or dry weather, as well as harsh operating conditions such as light scratches or bumps;
- Ease of use: if the hardware used requires little human input, it should have an intuitive interface;
- Durable power source: the portable device must be powered reliably enough to collect meaningful data over an extended period of time.

On the other hand, the electrocardiogram (ECG) is considered the most effective indicator of cardiovascular disease due to its high accuracy in recording the presence of such disease and its practicality and reliability in detecting it. Conventional ECG signal acquisition relies on electrodes, which can be uncomfortable to wear during normal daily activities. Smart watches and wristbands, on the other hand, are quite effective at capturing ECG signals and are also convenient for a number of other reasons. They are available to everyone and are the best option as they combine a variety of useful features with accurate monitoring of heart rate and other vital signs. Commercially available smart watches and wristbands are cheap and have simple user interfaces. They are small, are not in the way, and do not limit people's options. In addition, they are equipped with reliable power sources that allow them to last for a long time. Finally, their ability to record a wide range of biometric data makes them an excellent, if not ideal, option for ECG capture devices and thus for predicting CVD parameters.

5.2. Latest Tech-Trends and Wearables in CVDs

Alternatively, it is interesting to examine whether or not smart wearables used to control CVD are consistent with current machine learning practices. Several subfields of machine learning were identified as current research areas, but "Explainable AI", "Federated Machine Learning", and "Multimodal Machine Learning" were most frequently mentioned. The compliance of smart wearables used to detect CVDs to those topics is discussed in the following sections.

5.2.1. Explainable AI

The more complex AI becomes, the more difficult it becomes for humans to understand and reconstruct the thought process of the algorithm. The entire computational process becomes a so-called "black box", something that cannot be understood by humans. These black box models are created from scratch using nothing but the raw data. They are so complicated that not even the engineers or data scientists who create them can explain how their artificial intelligence algorithms arrive at their conclusions. Insight into the reasoning behind an AI system's results can be very helpful. Being able to explain a decision can be critical in allowing stakeholders to challenge or change the conclusion, in meeting regulatory criteria, or in ensuring that the system works as intended by its creators [137–139].

In this context, and to address the challenges posed by the black-box nature of AI, ML, and DL models, explainable AI (XAI) is proposed as a viable solution. The goal of XAI is to make the results and outputs generated by machine learning algorithms understandable and reliable to human users. The term refers to a method for describing an AI model along with its intended effects and possible biases. In AI-driven decision-making, it helps describe the precision, fairness, transparency, and outcomes of the model. When it comes to bringing AI models into production, a company's ability to explain the rationale behind its decisions is critical to building trust with employees and customers. Companies may take a more ethical approach to AI development if AI can be explained [137–139].

However, it was found that not a single study mentioned above implemented the explainable AI. While the aforementioned studies were able to achieve high accuracy in diagnosing cardiovascular disease, it may be difficult to implement such wearable technologies into the healthcare cycle if people do not know how the models arrive at such results. In other words, if the results are not explained, the medical community and patients will not have confidence in them, or at least be wary of adopting them.

5.2.2. Federated Machine Learning

The importance of protecting sensitive information has been studied for some time, leading to the development of a variety of protocols for encrypting communications between participants. Differential privacy [140], k-order anonymity [141], homomorphic encryption [142], and other approaches have been developed to protect data before they are transmitted. While several attacks have been uncovered in ML, such as the model inversion attack [143] and the affiliation attack [144], none of them are foolproof, as they can infer raw data by accessing the model.

Federated machine learning, often referred to as federated learning (FL), is a novel idea recently introduced by Google in the machine learning field [145]. The main concept behind FL is to eliminate the exchange of user data between peripherals. FL is a collaborative, distributed/decentralized ML privacy-preserving technology that eliminates the need to transfer data from peripherals to a central server in order to train a model. Instead, the models are sent to the peripheral nodes, where they are trained on the local data, and then sent back to the central aggregation node, where the global model is created without the nodes ever seeing the embedded data. Fortunately, federated learning has emerged as a powerful response to user privacy concerns, paving the way for the collection of additional data to train ML models to improve their accuracy and efficiency.

Furthermore, FL enables training models using data from multiple locations that have data with different structure and composition, also known as data islands, and integrating the information into a global trained model, improving the efficiency of the models. In addition, FL enabled "Learning Transfer", where models can share their knowledge without having to transfer users' private data, and made it possible to deal with heterogeneous data scattered in multiple data spaces containing different attributes. The main concept of federated learning is explained in Figure 8 below.

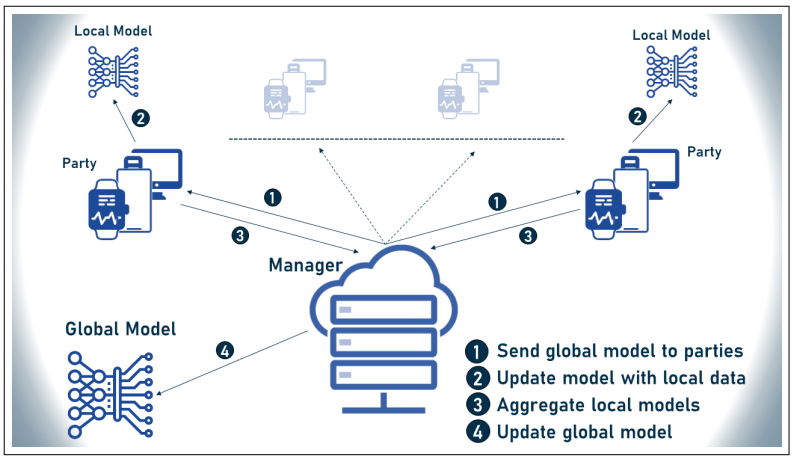

Figure 8. Federated machine learning classical structure.

Federated machine learning has shown promising results in the healthcare industry, as indicated in [146,147]. However, none of the studies included in this review addressed the integration of federated learning into wearable devices to make accurate predictions of cardiovascular disease while maintaining privacy. There could be a few reasons for this. For example, FL is still in its infancy and is still vulnerable to various challenges [148,149]. As a result, these factors may slow down the widespread use of FL in smart wearables in cardiology. However, integrating federated learning into smart wearables may lead to the following outcomes:

- Preserving users' private data, especially health-related data;
- Enabling analysis of data from multiple sources in addition to the vital signs captured by the wearables, such as the patient's medical history derived from electronic health records and the ECG recorded in real time, to provide more accurate results;

- Building users' confidence in smart wearables for cardiovascular disease management and subsequent product adoption.

5.2.3. Multimodal Machine Learning

Multimodal machine learning is concerned with integrating different and divergent data sources to benefit from complementary information in a single computational framework that takes care of a single task, and follows this rule in the context of machine learning (ML), a branch of AI. When it comes to predictive capability, the ability to explore many datasets simultaneously leads to more trustworthy and accurate results, making multimodal machine learning an area of high efficiency and amazing potential. To determine a single goal, multimodal machine learning combines information from many modalities [150].

In this context, data fusion is the process of combining information from many databases. "The process of merging data to improve state estimates and projections" [151] is a more precise definition of data fusion. The Joint Directors of Laboratories (JDL) Data Fusion Subpanel concludes that the method of "data fusion" is essential for dealing with many types of data. This description is supported by the authors in [152], who state that any process that deals with linking, correlating, or combining data retrieved from one or more sources to generate improved information is considered a process that employs data fusion. Because the literature on data fusion is still relatively young, there is no general agreement on the optimal way to merge disparate datasets. This is especially true considering that there are four different methods for performing this [151,152]:

- Early fusion: disparate data sources are merged into a single feature vector before being used by a single machine learning algorithm.
- Intermediate fusion: takes place in the intermediate phase between input and output of a ML architecture, when all data sources have the same representation format.
- Late fusion: defines the aggregation of decisions from multiple ML algorithms, each trained with different data sources.
- Hybrid fusion: defines the use of more than one fusion discipline in a single deep algorithm.

The approaches to data fusion defined above are illustrated in Figure 9 below. In addition, none of the smart wearable CVD detection studies reviewed here explored the use of multimodal ML in their algorithms. However, by using this technology, researchers can evaluate many datasets simultaneously, which greatly improves the accuracy of their results. Multimodal ML allows researchers to analyze medical imaging data such as MRIs, ECGs, and EHR data, giving the public more confidence in the accuracy of our AI models.

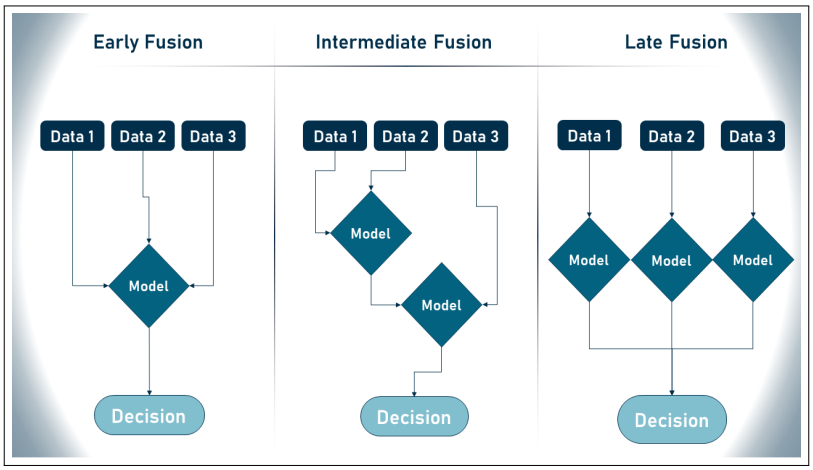

Figure 9. Data fusion different approaches.

6. Challenges and Future Perspectives

Despite the significant role that smart wearables play in the detection of cardiovascular disease, several issues may arise with their use. In addition, the introduction of new artificial intelligence tools and concepts presents many new opportunities to improve the management of heart disease. In this section, challenges and future prospects are discussed to help future studies select starting points for future investigations.

6.1. Challenges

The following are the most common challenges faced by smart wearables in detecting cardiovascular disease. These challenges were identified by analyzing the studies listed in Table 2 and reviewing the literature on smart wearables. Additional information can be obtained from a variety of sources, including, but not limited to [153–157].

6.1.1. Data Privacy and Confidentiality

AI models built into smart wearable technologies work only as well as the information they have access to. While the technical structure of the models themselves—including the cleanliness and suitability of the data—can affect how much data can be used to train AI models, it is generally accepted that more data can lead to more accurate models. In practice, however, there are several obstacles that make data collection the most difficult part of developing AI models. First and foremost is privacy and confidentiality. The security and privacy of personal data are not only strengthened by people, but also by society in general, governments, and companies. Numerous laws and regulations have been enacted to protect personal data, including the European Union's General Data Protection Regulation (GDPR) [158], the Chinese People's Republic of China's Cybersecurity Law [159], the People's Republic of China's General Principles of Civil Law [160], Singapore's PDPA [161], and hundreds of other principles around the world. Although these regulations help protect private information, they pose new challenges to the traditional AI data processing model to varying degrees by making it more difficult to collect data to train models, which in turn makes it more difficult to improve the accuracy of model performance.

6.1.2. Noise and Artifacts

The noninvasive nature of vital signs collection by smart wearables leaves the recordings open to a greater amount of background noise, known as "artifacts". Artifacts are unwanted signals or signal distributions that distort the actual signal and contribute to the noise in the data, degrading the quality of the data and reducing the performance and accuracy of the smart models. Artifacts can be divided into two categories, depending on where they originate: intrinsic artifacts, which come from the monitored body itself, and extrinsic artifacts, which are caused by the monitored person's external environment. The origin of artifacts can be divided into many categories [162,163]:

- Intrinsic artifacts (also known as physiological or internal artifacts):
  - Ocular artifacts: created by ocular motions including blinking, horizontal and vertical eye movement, fluttering of the eyes, etc.;
  - Muscle artifacts: caused by things such as sneezing, swallowing, clenching, talking, lifting the eyebrows, chewing, contracting the scalp, etc.;
  - Respiratory artifacts: resulting from an electrode's movement while breathing, which might manifest as slow, repetitive EEG activity;
  - Sweat artifacts: result of sweat's electrolyte concentration shifts on the electrode's surface after contact with the scalp and are obtained in wearables that collect vital signs that are related to skin.
- Extrinsic artifacts (also known as extra-physiological/external artifacts):
  - Motion artifacts: EEG monitoring systems are susceptible to motion artifacts due to the subject's physical movement;

– Environmental artifacts: these include, but are not limited to, loss of electrode-to-scalp contact, electrode rupture, electromagnetic wave interference from nearby electrical or electronic equipment, etc.

6.1.3. Data Diversity and Heterogeneity

Research in the field of medicine has shown that the use of multiple vital signs may be more helpful in detecting a disease than the use of a single vital sign. Therefore, combining multiple vital signs in the analysis process could allow for more accurate prediction of cardiovascular disease. Combining ECG signals with medical history data from the electronic health record (EHR) and medical images such as magnetic resonance imaging (MRI) is a robust example of multiple vital signs that can be analyzed together to predict cardiovascular disease. However, these data differ in their nature and structure, or even in the devices used to acquire them. More specifically, ECG data are usually stored in the form of real numbers, while EHR data may be in the form of clinical reports, health tests, or other forms, and MRI images are usually stored in different image formats. In this context, classical machine learning models such as support vector machines are usually well suited for linear data, but it is well known that images can be analyzed with deep learning algorithms such as convolutional neural networks. Therefore, it is a difficult task to analyze these data together given their different formats and structures, even if it is more practical for disease detection.

6.1.4. User Technology Adoption and Engagement

One of the major barriers to the use of smart wearables to detect and predict cardiovascular disease is user acceptance, adoption, and participation. Wearing such sensors is received differently by users due to concerns about privacy, discomfort, ethics, and other contextual factors.

Therefore, we may characterize the difficulties as the following set of study questions. In addition, those questions are illustrated in Figure 10 below (the symbol RQ in the list below and in Figure 10 refers to the term "research question"):

- **RQ1**: Disclosure of subject data may be limited by law. If we utilize these records, how can we ensure that no one's privacy will be compromised?
- **RQ2**: There are several potential noise and interference contributors to CVDs detection data. The question is, how should specialists deal with noisy data and artifacts?
- **RQ3**: The identification of CVDs may be enhanced by analyzing a variety of data. Can AI models handle the analysis of diverse datasets?
- **RQ4**: Did smart wearables earn enough confidence in the field despite their excellent accuracy in detecting CVDs, and how can this be improved?

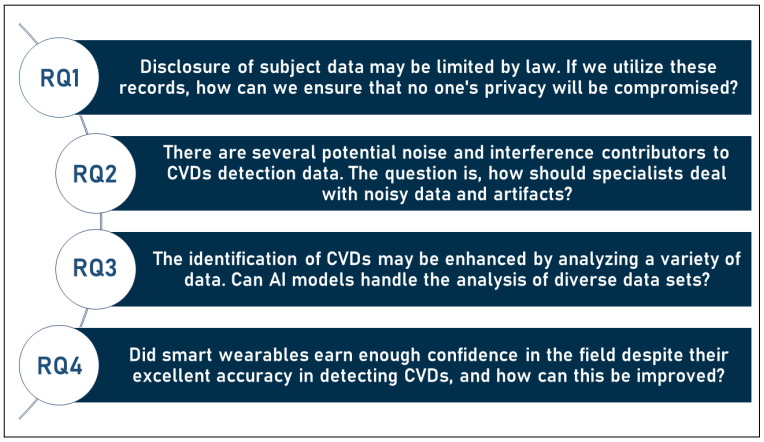

Figure 10. Research questions arising from analysing usage of wearables in CVDs detection.

6.2. Future Perspectives

Cardiovascular disease detection through smart wearables is now a reality. With the global incidence of this disease and the deaths associated with it, there is a growing need to improve the overall process and take more measures for proactive and preventive methods. More research and development on smart wearables is needed to keep up with the increasing demand.

6.2.1. Preserving Data Privacy and Confidentiality

Newer machine learning methods offer new opportunities to protect the privacy and security of user data. One potential technique that can help solve privacy problems is federated learning (FL). Federated learning, a type of collaborative decentralized machine learning that protects user privacy, does not require data to be transported from edge devices to a central server [149,164–167]. It is expected that using FL to identify CVDs will make it easier to collect more data, which in turn will improve detection accuracy.

6.2.2. Artifacts Removal and Data Readiness

Before proceeding with signal processing, it is important to eliminate or reduce all artifacts, both extrinsic and intrinsic, that might interfere with the signals. References [163,168–170] detail some of the existing implementations that perform this function. In order to clean and preprocess the data to improve the accuracy of cardiovascular disease detection, it is necessary to investigate the automation of noise reduction.

6.2.3. Analysis of Heterogeneous and Diverse Data

Multimodal machine learning is a good solution that allows analyzing data with alternative structures and formats. Since current cardiovascular disease detection and prediction implementations usually analyze only one type of data structure (linear, images, etc.), multimodal machine learning allows analyzing multiple types of data simultaneously to improve the overall result of the intelligent model. Learning a complex task by analyzing data from multiple sources and using complementary knowledge are examples of what multimodal machine learning is capable of. In this context, multimodal datasets are described as information with different structures and formats that come from a variety of sources, each of which contributes a unique set of information (or "modality") to the overall dataset. Therefore, using the concept of multimodal ML to analyze different data such as ECG, EHR recordings, and MRI images can help increase the accuracy of CVD detection and prediction.

6.2.4. Raising Trust by Enhancing Accuracy, Privacy, and Explainability

Given the prevalence and devastating impact of cardiovascular disease, there is a growing need for practical and viable solutions that can help detect and even predict the onset of these conditions. Consequently, smart wearables have proven to be viable in this area, providing both continuous and real-time monitoring without interfering with daily life routines. However, there is a great need to improve the prediction of CVDs using smart wearables, whether through increased accuracy, better explainability, or by addressing other issues that hinder their adoption by users, such as privacy and ethical constraints. This is a well-known fact that does not need further explanation, because when it comes to health, users are only willing to use tools that are highly accurate, understandable, private, and reliable. In other words, greater trust and wider use of smart wearables as tools for predicting CVDs will result from improved accuracy, reliability, feasibility, privacy, and explainability of such devices.

For this reason, we may summarize the outlook into the following trending research topics. In addition, those research topics are illustrated in Figure 11 below (the symbol TR in the list below and in Figure 11 refers to the term "trending research topic"):

- **TR1:** To protect user privacy, smart wearables should employ federated learning for CVDs detection;

- **TR2:** The use of automated artifact and noise removal methods to mitigate the effects of interference and background noise;
- **TR3:** Improve the quality of recognition models by analyzing data from numerous modalities and sources using multimodal ML techniques;
- **TR4:** Raising precision, explainability, and adaptability will help build users' confidence in smart wearables.

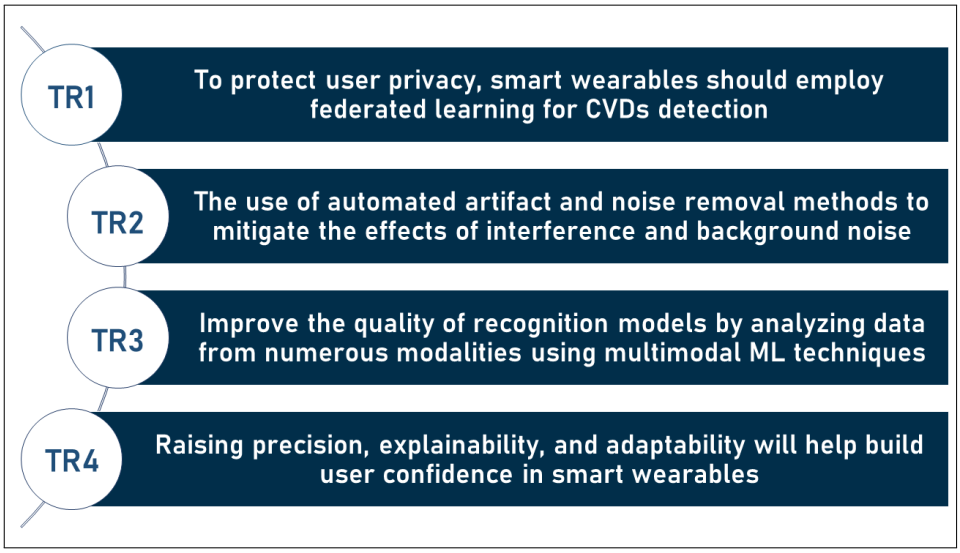

Figure 11. Research topics that may serve as solutions to the challenges in the domain.

Figure 12 below summarizes the challenges–future solutions relationship and illustrates how future views may act as potential solutions in the domain, all of which can assist to enhance research into the use of smart wearables in the detection of CVDs.

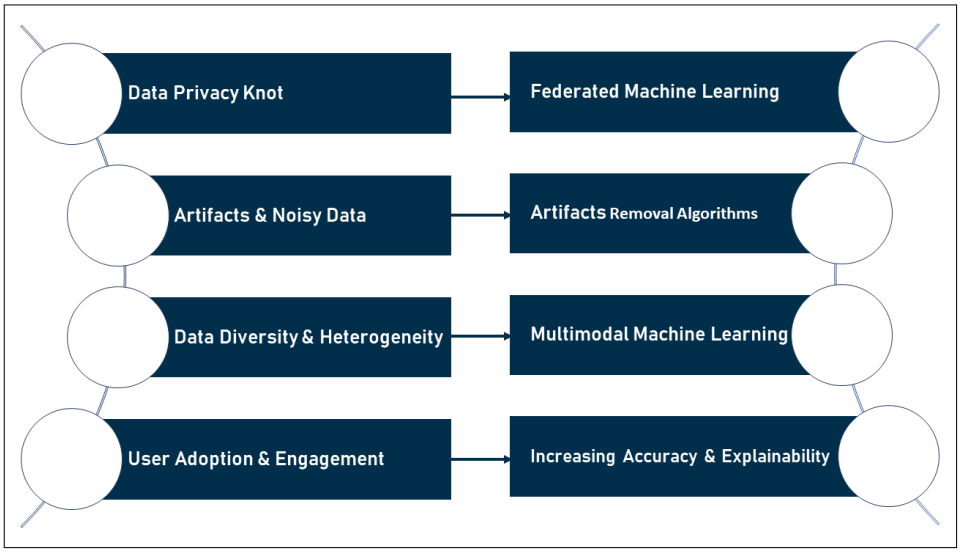

Figure 12. Challenges–future solutions chart.

7. Conclusions

Recently, the use of smart wearables in the diagnosis and prediction of cardiovascular disease has received increasing attention. This is partly due to the technological potential of smart wearables and partly due to the data processing power of artificial intelligence and its derivatives, machine learning and deep learning. In this research, we thoroughly

investigated the use of smart wearables to treat fatal heart diseases. The review of the research area showed the high practicality and effectiveness of such methods, reflecting the growing interest that has surged in recent years. However, given the challenges and limitations discussed in this review, there is a large window for improvement that smart wearables should undergo to prove their feasibility and reliability. Increasing accuracy, automating noise reduction, solving privacy issues, dealing with heterogeneity, and improving explainability are interesting topics that should be considered when trying to promote the use of smart wearables in the management of CVDs. As a result, this review provides a brief overview of a number of relevant topics that can be used as recommendations for further research.

Author Contributions: Conceptualization: M.M. and M.A.; Formal analysis: M.M.; Investigation: M.M.; Methodology: M.M. and M.A.; Supervision: M.A., A.B., H.I. and A.R.; Visualization: M.M.; Writing—original draft: M.M.; Writing—review & editing: M.A., A.B., H.I. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number 06351.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Acknowledgments: We acknowledge the support of Centre d'Entrepreneuriat et de Valorisation des Innovations (CEVI).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Urgent Health Challenges for the Next Decade. 13 January 2020. Available online: https://www.who.int/news-room/photo-story/photo-story-detail/urgent-health-challenges-for-the-next-decade (accessed on 1 April 2021).
2. Holmes, J.; Sacchi, L.; Bellazzi, R. Artificial intelligence in medicine. *Ann. R. Coll. Surg. Engl.* **2004**, *86*, 334–338.
3. Maddox, T.M.; Rumsfeld, J.S.; Payne, P.R. Questions for artificial intelligence in health care. *JAMA* **2019**, *321*, 31–32. [CrossRef] [PubMed]
4. Tran, B.X.; Vu, G.T.; Ha, G.H.; Vuong, Q.H.; Ho, M.T.; Vuong, T.T.; La, V.P.; Ho, M.T.; Nghiem, K.C.P.; Nguyen, H.L.T.; et al. Global evolution of research in artificial intelligence in health and medicine: A bibliometric study. *J. Clin. Med.* **2019**, *8*, 360. [CrossRef]
5. Noorbakhsh-Sabet, N.; Zand, R.; Zhang, Y.; Abedi, V. Artificial intelligence transforms the future of health care. *Am. J. Med.* **2019**, *132*, 795–801. [CrossRef]
6. Roth, G.A.; Mensah, G.A.; Johnson, C.O.; Addolorato, G.; Ammirati, E.; Baddour, L.M.; Barengo, N.C.; Beaton, A.Z.; Benjamin, E.J.; Benziger, C.P.; et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **2020**, *76*, 2982–3021. [CrossRef]
7. Centers for Disease Control and Prevention (CDC). Heart Disease Facts | Cdc.Gov. 14 October 2022. Available online: https://www.cdc.gov/heartdisease/facts.htm (accessed on 10 April 2021).
8. Agency of Canada, Public Health. Heart Disease in Canada—Canada.Ca. August 2022. Available online: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/heart-disease-canada.html (accessed on 15 April 2021).
9. Tsao, C.W.; Aday, A.W.; Almarzooq, Z.I.; Alonso, A.; Beaton, A.Z.; Bittencourt, M.S.; Boehme, A.K.; Buxton, A.E.; Carson, A.P.; Commodore-Mensah, Y.; et al. Heart disease and stroke statistics—2022 update: A report from the American Heart Association. *Circulation* **2022**, *145*, e153–e639. [CrossRef]
10. National Heart Lung and Blood Institute. *Brief: Your Guide to Living Well with Heart Disease (NIH Publication 06-5716)*; National Institutes of Health: Washington, DC, USA, 2005.
11. Johnson, K.W.; Torres Soto, J.; Glicksberg, B.S.; Shameer, K.; Miotto, R.; Ali, M.; Ashley, E.; Dudley, J.T. Artificial intelligence in cardiology. *J. Am. Coll. Cardiol.* **2018**, *71*, 2668–2679. [CrossRef] [PubMed]
12. Prashanth, K.; Manjappa, M.; Srikar, C. The advent of artificial intelligence in cardiology: The current applications and future prospects. *Eur. J. Mol. Clin. Med.* **2020**, *7*, 14–20.
13. Kim, K.J.; Shin, D.H. An acceptance model for smart watches: Implications for the adoption of future wearable technology. *Internet Res.* **2015**, *25*, 527–541. [CrossRef]
14. Perera, C.; Vasilakos, A.V. A knowledge-based resource discovery for Internet of Things. *Knowl.-Based Syst.* **2016**, *109*, 122–136. [CrossRef]

15. Liu, L.; Peng, Y.; Liu, M.; Huang, Z. Sensor-based human activity recognition system with a multilayered model using time series shapelets. *Knowl.-Based Syst.* **2015**, *90*, 138–152. [CrossRef]
16. Park, E.; Kim, K.J.; Kwon, S.J. Understanding the emergence of wearable devices as next-generation tools for health communication. *Inf. Technol. People* **2016**, *29*, 717–732 [CrossRef]
17. Turing, A. Computing machinery and intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
18. Niknejad, N.; Ismail, W.B.; Mardani, A.; Liao, H.; Ghani, I. A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges. *Eng. Appl. Artif. Intell.* **2020**, *90*, 103529. [CrossRef]
19. Kortuem, G.; Kawsar, F.; Sundramoorthy, V.; Fitton, D. Smart objects as building blocks for the internet of things. *IEEE Internet Comput.* **2009**, *14*, 44–51. [CrossRef]
20. Cheng, J.W.; Mitomo, H. The underlying factors of the perceived usefulness of using smart wearable devices for disaster applications. *Telemat. Inform.* **2017**, *34*, 528–539. [CrossRef]
21. Poslad, S. *Ubiquitous Computing: Smart Devices, Environments and Interactions*; John Wiley Sons: Hoboken, NJ, USA, 2011.
22. Jeong, S.C.; Kim, S.H.; Park, J.Y.; Choi, B. Domain-specific innovativeness and new product adoption: A case of wearable devices. *Telemat. Inform.* **2017**, *34*, 399–412. [CrossRef]
23. Fernandez, P. Wearable Technology: Beyond Augmented Reality. *Libr. Hi Tech News* **2014**, *31*. [CrossRef]
24. Thorp, E.O. The invention of the first wearable computer. In Proceedings of the Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215), Pittsburgh, PA, USA, 19–20 October 1998; pp. 4–8.
25. Mann, S. Smart clothing: The shift to wearable computing. *Commun. ACM* **1996**, *39*, 23–24. [CrossRef]
26. Park, S.; Jayaraman, S. Smart textiles: Wearable electronic systems. *MRS Bull.* **2003**, *28*, 585–591. [CrossRef]
27. Wright, R.; Keith, L. Wearable technology: If the tech fits, wear it. *J. Electron. Resour. Med. Libr.* **2014**, *11*, 204–216. [CrossRef]
28. Dimou, E.; Manavis, A.; Papachristou, E.; Kyratsis, P. A conceptual design of intelligent shoes for pregnant women. In Proceedings of the Workshop on Business Models and ICT Technologies for the Fashion Supply Chain, Florence, Italy, 20–22 April 2016; Springer: Cham, Switzerland, 2016; pp. 69–77.
29. Yang, H.; Yu, J.; Zo, H.; Choi, M. User acceptance of wearable devices: An extended perspective of perceived value. *Telemat. Inform.* **2016**, *33*, 256–269. [CrossRef]
30. Sana, F.; Isselbacher, E.M.; Singh, J.P.; Heist, E.K.; Pathik, B.; Armoundas, A.A. Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **2020**, *75*, 1582–1592. [CrossRef] [PubMed]
31. DeVore, A.D.; Wosik, J.; Hernandez, A.F. The future of wearables in heart failure patients. *JACC Heart Fail.* **2019**, *7*, 922–932. [CrossRef]
32. Raja, J.M.; Elsakr, C.; Roman, S.; Cave, B.; Pour-Ghaz, I.; Nanda, A.; Maturana, M.; Khouzam, R.N. Apple watch, wearables, and heart rhythm: Where do we stand? *Ann. Transl. Med.* **2019**, *7*, 417. [CrossRef]
33. Vashistha, R.; Dangi, A.K.; Kumar, A.; Chhabra, D.; Shukla, P. Futuristic biosensors for cardiac health care: An artificial intelligence approach. *3 Biotech* **2018**, *8*, 358. [CrossRef]
34. Lee, S.; Chu, Y.; Ryu, J.; Park, Y.J.; Yang, S.; Koh, S.B. Artificial intelligence for detection of cardiovascular-related diseases from wearable devices: A systematic review and meta-analysis. *Yonsei Med. J.* **2022**, *63*, S93–S107. [CrossRef]
35. Bayoumy, K.; Gaber, M.; Elshafeey, A.; Mhaimeed, O.; Dineen, E.H.; Marvel, F.A.; Martin, S.S.; Muse, E.D.; Turakhia, M.P.; Tarakji, K.G.; et al. Smart wearable devices in cardiovascular care: Where we are and how to move forward. *Nat. Rev. Cardiol.* **2021**, *18*, 581–599. [CrossRef]
36. Kumar, S.; Victoria-Castro, A.M.; Melchinger, H.; O'Connor, K.D.; Psotka, M.; Desai, N.R.; Ahmad, T.; Wilson, F.P. Wearables in Cardiovascular Disease. *J. Cardiovasc. Transl. Res.* **2022**, 1–12. [CrossRef]
37. Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ* **2021**, *372*, n160. [CrossRef]
38. PRISMA. Available online: https://www.prisma-statement.org//PRISMAStatement/Checklist.aspx (accessed on 17 November 2022).
39. Wang, I.-J.; Liao, L.-D.; Wang, Y.-T.; Chen, C.-Y.; Lin, B.-S.; Lu, S.-W.; Lin, C.-T. A wearable mobile electrocardiogram measurement device with novel dry polymer-based electrodes. In Proceedings of the TENCON 2010–2010 IEEE Region 10 Conference, Fukuoka, Japan, 21–24 November 2010; pp. 379–384.
40. Oresko, J.J.; Jin, Z.; Cheng, J.; Huang, S.; Sun, Y.; Duschl, H.; Cheng, A.C. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 734–740. [CrossRef]
41. Lin, C.-T.; Chang, K.-C.; Lin, C.-L.; Chiang, C.-C.; Lu, S.-W.; Chang, S.-S.; Lin, B.-S.; Liang, H.-Y.; Chen, R.-J.; Lee, Y.-T.; et al. An intelligent telecardiology system using a wearable and wireless ECG to detect atrial fibrillation. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 726–733.
42. Hu, S.; Shao, Z.; Tan, J. A real-time cardiac arrhythmia classification system with wearable electrocardiogram. In Proceedings of the 2011 International Conference on Body Sensor Networks, Dallas, TX, USA, 23–25 May 2011; pp. 119–124.
43. Lee, H.-J.; Kim, D.-O.; Kang, B.-J.; Ban, S.-W. Mobile embedded health-care system working on wireless sensor network. In Proceedings of the 2011 Third International Conference on Communications and Mobile Computing, Qingdao, China, 18–20 April 2011; pp. 161–164.

44. Cheng, S.; Tamil, L.S.; Levine, B. A mobile health system to identify the onset of paroxysmal atrial fibrillation. In Proceedings of the 2015 International Conference on Healthcare Informatics, Dallas, TX, USA, 21–23 October 2015; pp. 189–192.
45. Nemati, S.; Ghassemi, M.M.; Ambai, V.; Isakadze, N.; Levantsevych, O.; Shah, A.; Clifford, G.D. Monitoring and detecting atrial fibrillation using wearable technology. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 3394–3397.
46. Wu, T.-K.; Lin, C.-C.; Ku, W.-Y.; Liou, Y.-S.; Yang, C.-Y.; Lee, M.-Y.; Lin, W.-Y.; Tsai, T.-H. The Study of the Enhanced External Counterpulsation System Based on Smart Clothes. In Proceedings of the 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Washington, DC, USA, 27–29 June 2016; pp. 125–129.
47. Shashikumar, S.P.; Shah, A.J.; Li, Q.; Clifford, G.D.; Nemati, S. A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology. In Proceedings of the 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Orlando, FL, USA, 16–19 February 2017; pp. 141–144.
48. ElSaadany, Y.; Majumder, A.J.A.; Ucci, D.R. A wireless early prediction system of cardiac arrest through IoT. In Proceedings of the 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC), Turin, Italy, 4–8 July 2017; Volume 2, pp. 690–695.
49. Sadrawi, M.; Lin, C.-H.; Lin, Y.-T.; Hsieh, Y.; Kuo, C.-C.; Chien, J.-C.; Haraikawa, K.; Abbod, M.F.; Shieh, J.-S. Arrhythmia evaluation in wearable ECG devices. *Sensors* **2017**, *17*, 2445. [CrossRef]
50. Roj, D.; Wrobel, J.; Matonia, A.; Horoba, K.; Henzel, N. Control and signal processing software embedded in smart wristband monitor of silent atrial fibrillation. In Proceedings of the 2017 MIXDES-24th International Conference Mixed Design of Integrated Circuits and Systems, Bydgoszcz, Poland, 22–24 June 2017; pp. 585–590.
51. Lin, C.-W.; Chang, Y.; Lin, C.-C.K.; Tsai, L.-M.; Chen, J.-Y. Development of an AI-based non-invasive Pulse AudioGram monitoring device for arrhythmia screening. In Proceedings of the 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT), Bethesda, MD, USA, 6–8 November 2017; pp. 40–43.
52. Chowdhuryy, M.H.I.; Sultana, M.; Ghosh, R.; Ahamed, J.U.; Mahmood, M.A.I. AI assisted portable ECG for fast and patient specific diagnosis. In Proceedings of the 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 8–9 February 2018; pp. 1–4.
53. Xia, Y.; Zhang, H.; Xu, L.; Gao, Z.; Zhang, H.; Liu, H.; Li, S. An automatic cardiac arrhythmia classification system with wearable electrocardiogram. *IEEE Access* **2018**, *6*, 16529–16538. [CrossRef]
54. Ni, H.; Cho, S.; Mankoff, J.; Yang, J. Automated recognition of hypertension through overnight continuous HRV monitoring. *J. Ambient. Intell. Humaniz. Comput.* **2018**, *9*, 2011–2023. [CrossRef]
55. Aliamiri, A.; Shen, Y. Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor. In Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Las Vegas, NV, USA, 4–7 March 2018; pp. 442–445.
56. Park, E.-B.; Lee, J.-H. The Automated Heart Disease Detection System with Dry-Electrode Based Outdoor Shirts. In Proceedings of the 2018 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), JeJu, Republic of Korea, 24–26 June 2018; pp. 206–212.
57. Alfarhan, K.A.; Mashor, M.Y.; Mohd Saad, A.R.; Omar, M.I. Wireless heart abnormality monitoring kit based on Raspberry Pi. In *Journal of Biomimetics, Biomaterials and Biomedical Engineering*; Trans Tech Publications Ltd.: Stafa-Zurich, Switzerland, 2018; Volume 35, pp. 96–108.
58. Kim, Y.S.; Mahmood, M.; Lee, Y.; Kim, N.K.; Kwon, S.; Herbert, R.; Kim, D.; Cho, H.C.; Yeo, W.-H. All-in-one, wireless, stretchable hybrid electronics for smart, connected, and ambulatory physiological monitoring. *Adv. Sci.* **2019**, *6*, 1900939. [CrossRef] [PubMed]
59. Lin, Y.-J.; Chuang, C.-W.; Yen, C.-Y.; Huang, S.-H.; Chen, J.-Y.; Lee, S.-Y. An AIoT wearable ECG patch with decision tree for arrhythmia analysis. In Proceedings of the 2019 IEEE Biomedical Circuits and Systems Conference (BioCAS), Nara, Japan, 17–19 October 2019; pp. 1–4.
60. Lin, Y.-J.; Chuang, C.-W.; Yen, C.-Y.; Huang, S.-H.; Huang, P.-W.; Chen, J.Y.; Lee, S.Y. Artificial intelligence of things wearable system for cardiac disease detection. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), Hsinchu, Taiwan, 18–20 March 2019; pp. 67–70.
61. Van Zaen, J.; Chételat, O.; Lemay, M.; Calvo, E.M.; Delgado-Gonzalo, R. Classification of cardiac arrhythmias from single lead ECG with a convolutional recurrent neural network. *arXiv* **2019**, arXiv:1907.01513.
62. Scirè, A.; Tropeano, F.; Anagnostopoulos, A.; Chatzigiannakis, I. Fog-computing-based heartbeat detection and arrhythmia classification using machine learning. *Algorithms* **2019**, *12*, 32. [CrossRef]
63. Faust, O.; Ciaccio, E.J.; Majid, A.; Acharya, U.R. Improving the safety of atrial fibrillation monitoring systems through human verification. *Saf. Sci.* **2019**, *118*, 881–886. [CrossRef]
64. Karboub, K.; Tabaa, M.; Dellagi, S.; Dandache, A.; Moutaouakkil, F. Intelligent patient monitoring for arrhythmia and congestive failure patients using internet of things and convolutional neural network. In Proceedings of the 2019 31st International Conference on Microelectronics (ICM), Cairo, Egypt, 15–18 December 2019; pp. 292–295.
65. Zhao, P.; Quan, D.; Yu, W.; Yang, X.; Fu, X. Towards deep learning-based detection scheme with raw ECG signal for wearable telehealth systems. In Proceedings of the 2019 28th International Conference on Computer Communication and Networks (ICCCN), Valencia, Spain, 29 July–1 August 2019; pp. 1–9.

66. Yang, C.; Veiga, C.; Rodriguez-Andina, J.J.; Farina, J.; Iniguez, A.; Yin, S. Using PPG signals and wearable devices for atrial fibrillation screening. *IEEE Trans. Ind. Electron.* **2019**, *66*, 8832–8842. [CrossRef]
67. HHuda, N.; Khan, S.; Abid, R.; Shuvo, S.B.; Labib, M.M.; Hasan, T. A low-cost, low-energy wearable ECG system with cloud-based arrhythmia detection. In Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 5–7 June 2020; pp. 1840–1843.
68. Chen, E.; Jiang, J.; Su, R.; Gao, M.; Zhu, S.; Zhou, J.; Huo, Y. A new smart wristband equipped with an artificial intelligence algorithm to detect atrial fibrillation. *Heart Rhythm* **2020**, *17*, 847–853. [CrossRef] [PubMed]
69. Bhat, T.; Bhat, S.; Manoj, T. A Real-Time IoT Based Arrhythmia Classifier Using Convolutional Neural Networks. In Proceedings of the 2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Udupi, India, 30–31 October 2020; pp. 79–83.
70. Ali, F.; El-Sappagh, S.; Islam, S.R.; Kwak, D.; Ali, A.; Imran, M.; Kwak, K.S. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Inf. Fusion* **2020**, *63*, 208–222. [CrossRef]
71. Rincon, J.A.; Guerra-Ojeda, S.; Carrascosa, C.; Julian, V. An IoT and fog computing-based monitoring system for cardiovascular patients with automatic ECG classification using deep neural networks. *Sensors* **2020**, *20*, 7353. [CrossRef]
72. Khan, M.A. An IoT framework for heart disease prediction based on MDCNN classifier. *IEEE Access* **2020**, *8*, 34717–34727. [CrossRef]
73. Yang, C.; Aranoff, N.D.; Green, P.; Tavassolian, N. Classification of aortic stenosis using time–frequency features from chest cardio-mechanical signals. *IEEE Trans. Biomed. Eng.* **2019**, *67*, 1672–1683. [CrossRef]
74. Reddy, S.; Seshadri, S.B.; Bothra, G.S.; Suhas, T.G.; Thundiyil, S.C. Detection of arrhythmia in real-time using ECG signal analysis and convolutional neural networks. In Proceedings of the 2020 IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE), Online, Poland, 16–19 September 2020; pp. 1–4.
75. Gowtham, A.; Anirudh, L.; Sreeja, B.S.; Aakash, B.A.; Adittya, S. Detection of Arrhythmia using ECG waves with Deep Convolutional Neural Networks. In Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 5–7 November 2020; pp. 1390–1396.
76. Han, D.; Bashar, S.K.; Zieneddin, F.; Ding, E.; Whitcomb, C.; McManus, D.D.; Chon, K.H. Digital image processing features of smartwatch photoplethysmography for cardiac arrhythmia detection. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 4071–4074.
77. Afadar, Y.; Akram, A.; Alkeebali, A.; Majzoub, S. Heart Arrhythmia Detection & Monitoring Using Machine Learning & ECG Wearable Device. In Proceedings of the 2020 Seventh International Conference on Information Technology Trends (ITT), Abu Dhabi, United Arab Emirates, 25–26 November 2020; pp. 107–112.
78. Marsili, I.A.; Biasiolli, L.; Masè, M.; Adami, A.; Andrighetti, A.O.; Ravelli, F.; Nollo, G. Implementation and validation of real-time algorithms for atrial fibrillation detection on a wearable ECG device. *Comput. Biol. Med.* **2020**, *116*, 103540. [CrossRef] [PubMed]
79. Ahsanuzzaman, S.M.; Ahmed, T.; Rahman, M.A. Low cost, portable ECG monitoring and alarming system based on deep learning. In Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 5–7 June 2020; pp. 316–319.
80. Saadatnejad, S.; Oveisi, M.; Hashemi, M. LSTM-based ECG classification for continuous monitoring on personal wearable devices. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 515–523. [CrossRef] [PubMed]
81. Malepati, N.; Fatima, R.; Gupta, S.; Ramsali, V.; Shobha, K.R. Portable ECG Device for Remote Monitoring and Detection of Onset of Arrhythmia. In Proceedings of the 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2–4 July 2020; pp. 1–5.
82. Panganiban, E.B.; Chung, W.-Y.; Tai, W.-C.; Paglinawan, A.C.; Lai, J.-S.; Cheng, R.-W.; Chang, M.-K.; Chang, P.H. Real-time intelligent healthcare monitoring and diagnosis system through deep learning and segmented analysis. In *International Conference on Biomedical and Health Informatics*; Springer: Cham, Switzerland, 2019; pp. 15–25.
83. Bazi, Y.; Al Rahhal, M.M.; AlHichri, H.; Ammour, N.; Alajlan, N.; Zuair, M. Real-time mobile-based electrocardiogram system for remote monitoring of patients with cardiac arrhythmias. *Int. J. Pattern Recognit. Artif. Intell.* **2020**, *34*, 2058013. [CrossRef]
84. Rahman, M.A.; Samin, M.J.A.; Al Hasan, Z. Remote ECG Monitoring and Syncope Detection System Using Deep Learning. In Proceedings of the 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), Dhaka, Bangladesh, 28–29 November 2020; pp. 201–206.
85. Ding, X.; Wang, Y.; Hao, Y.; Lv, Y.; Chen, R.; Yan, H. A New Measure of Pulse Rate Variability and Detection of Atrial Fibrillation Based on Improved Time Synchronous Averaging. *Comput. Math. Methods Med.* **2021**, *2021*, 5597559. [CrossRef]
86. Elwahsh, H.; El-Shafeiy, E.; Alanazi, S.; Tawfeek, M.A. A new smart healthcare framework for real-time heart disease detection based on deep and machine learning. *Peerj Comput. Sci.* **2021**, *7*, e646. [CrossRef]
87. Tang, X.; Tang, W. An ECG Delineation and Arrhythmia Classification System Using Slope Variation Measurement by Ternary Second-Order Delta Modulators for Wearable ECG Sensors. *IEEE Trans. Biomed. Circuits Syst.* **2021**, *15*,1053–1065. [CrossRef] [PubMed]
88. Li, H.; An Z.; Zuo, S.; Zhu, W.; Zhang, Z.; Zhang, S.; Zhang, C.; Song, W.; Mao, Q.; Mu, Y.; et al. Artificial Intelligence-Enabled ECG Algorithm Based on Improved Residual Network for Wearable ECG. *Sensors* **2021**, *21*, 6043. [CrossRef] [PubMed]
89. Mohandas, P.; Aswin, P.R.; John, A.; Madhu, M.; Thomas, G.; Kurupath, V. Automated cardiac condition diagnosis using AI based ECG analysis system for school children. In Proceedings of the 2021 8th International Conference on Smart Computing and Communications (ICSCC), Kochi, India, 1–3 July 2021; pp. 362–366.

90. Santala, O.E.; Halonen, J.; Martikainen, S.; Jäntti, H.; Rissanen, T.T.; Tarvainen, M.P.; Laitinen, T.P.; Laitinen, T.M.; Väliaho, E.S.; Hartikainen, J.E.K.; et al. Automatic Mobile Health Arrhythmia Monitoring for the Detection of Atrial Fibrillation: Prospective Feasibility, Accuracy, and User Experience Study. *JMIR mHealth and uHealth* **2021**, *9*, e29933. [CrossRef]
91. Sandberg, E.L.; Grenne, B.L.; Berge, T.; Grimsmo, J.; Atar, D.; Halvorsen, S.; Fensli, R.; Jortveit, J. Diagnostic accuracy and usability of the ECG247 smart heart sensor compared to conventional Holter technology. *J. Healthc. Eng.* **2021**, *2021*, 5230947. [CrossRef] [PubMed]
92. Mihiranga, A.; Shane, D.; Indeewari, B.; Udana, A.; Nawinna, D.; Attanayaka, B. Digital Tool for Prevention, Identification and Emergency Handling of Heart Attacks. In Proceedings of the 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), Bangalore, India, 30 September–2 October 2021; pp. 1–6.
93. Panganiban, E.B.; Paglinawan, A.C.; Chung, W.Y.; Paa, G.L.S. ECG diagnostic support system (EDSS): A deep learning neural network based classification system for detecting ECG abnormal rhythms from a low-powered wearable biosensors. *Sens. Bio-Sens. Res.* **2021**, *31*, 100398. [CrossRef]
94. Virgeniya, S.C.; Ramaraj, E. IoT and Big Data for ECG Signal Classification-A Quick Decision System. In Proceedings of the 2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4), Bangalore, India, 16–17 December 2021; pp. 1–6.
95. Sahoo, S.; Borthakur, P.; Baruah, N.; Chutia, B.P. IoT and machine learning based health monitoring and heart attack prediction system. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2021; Volume 1950, p. 012056.
96. Meng, L.; Ge, K.; Song, Y.; Yang, D.; Lin, Z. Long-term wearable electrocardiogram signal monitoring and analysis based on convolutional neural network. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [CrossRef]
97. Kadhim, O.R.; Taha, Z.K.; Abd AL-Majeed, S.W. Low-Cost Cloud-based Intelligent Remote Patient Monitoring for Heart Patient. In Proceedings of the 2021 International Conference on Communication & Information Technology (ICICT), Basrah, Iraq, 5–6 June 2021; pp. 31–36.
98. Frodi, D.M.; Kolk, M.Z.; Langford, J.; Andersen, T.O.; Knops, R.E.; Tan, H.L.; Svendsen, J.H.; Tjong, F.V.; Diederichsen, S.Z. Rationale and design of the SafeHeart study: Development and testing of a mHealth tool for the prediction of arrhythmic events and implantable cardioverter-defibrillator therapy. *Cardiovasc. Digit. Health J.* **2021**, *2*, S11–S20. [CrossRef]
99. Pierleoni, P.; Belli, A.; Gentili, A.; Incipini, L.; Palma, L.; Raggiunto, S.; Sbrollini, A.; Burattini, L. Real-time smart monitoring system for atrial fibrillation pathology. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 4461–4469. [CrossRef]
100. Siavashi, A.; Majidi, M. Sensing, Wireless Transmission, and Smart Processing of Heart Signals. In Proceedings of the 2021 5th International Conference on Internet of Things and Applications (IoT), Isfahan, Iran, 19–20 May 2021; pp. 1–6.
101. Feng, H.-Y.; Chen, P.-Y.; Hou, J. SR-ScatNet Algorithm for On-device ECG Time Series Anomaly Detection. In Proceedings of the SoutheastCon 2021, Atlanta, GA, USA, 10–13 March 2021; pp. 1–5.
102. Liu, X.; Liu, T.; Zhang, Z.; Kuo, P.C.; Xu, H.; Yang, Z.; Lan, K.; Li, P.; Ouyang, Z.; Ng, Y.L.; et al. TOP-Net prediction model using bidirectional long short-term memory and medical-grade wearable multisensor system for tachycardia onset: Algorithm development study. *JMIR Med. Inform.* **2021**, *9*, e18803. [CrossRef] [PubMed]
103. Zhang, H.; Zhu, L.; Nathan, V.; Kuang, J.; Kim, J.; Gao, J.A.; Olgin, J. Towards Early Detection and Burden Estimation of Atrial Fibrillation in an Ambulatory Free-living Environment. *Proc. Acm Interact. Mob. Wearable Ubiquitous Technol.* **2021**, *5*, 1–19. [CrossRef]
104. Hui, Y.; Yin, Z.; Wu, M.; Li, D. Wearable devices acquired ECG signals detection method using 1D convolutional neural network. In Proceedings of the 2021 15th International Symposium on Medical Information and Communication Technology (ISMICT), Xiamen, China, 14–16 April 2021; pp. 81–85.
105. Pramukantoro, E.S.; Gofuku, A. A heartbeat classifier for continuous prediction using a wearable device. *Sensors* **2022**, *22*, 5080. [CrossRef] [PubMed]
106. Pramukantoro, E.S.; Gofuku, A. A real-time heartbeat monitoring using wearable device and machine learning. In Proceedings of the 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech), Osaka, Japan, 7–9 March 2022; pp. 270–272.
107. Kwon, J.M.; Jo, Y.Y.; Lee, S.Y.; Kang, S.; Lim, S.Y.; Lee, M.S.; Kim, K.H. Artificial Intelligence-Enhanced Smartwatch ECG for Heart Failure-Reduced Ejection Fraction Detection by Generating 12-Lead ECG. *Diagnostics* **2022**, *12*, *654*. [CrossRef]
108. Zhu, L.; Nathan, V.; Kuang, J.; Kim, J.; Avram, R.; Olgin, J.; Gao, J. Atrial fibrillation detection and atrial fibrillation burden estimation via wearables. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 2063–2074. [CrossRef]
109. Sabbadini, R.; Riccio, M.; Maresca, L.; Irace, A.; Breglio, G. Atrial Fibrillation Detection by Means of Edge Computing on Wearable Device: A Feasibility Assessment. In Proceedings of the 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Messina, Italy, 22–24 June 2022; pp. 1–6.
110. Santala, O.E.; Lipponen, J.A.; Jäntti, H.; Rissanen, T.T.; Tarvainen, M.P.; Laitinen, T.P.; Laitinen, T.M.; Castrén, M.; Väliaho, E.S.; Rantula, O.A.; et al. Continuous mHealth Patch Monitoring for the Algorithm-Based Detection of Atrial Fibrillation: Feasibility and Diagnostic Accuracy Study. *JMIR Cardio* **2022**, *6*, e31230. [CrossRef]
111. Baraeinejad, B.; Shayan, M.F.; Vazifeh, A.R.; Rashidi, D.; Hamedani, M.S.; Tavolinejad, H.; Gorji, P.; Razmara, P.; Vaziri, K.; Vashaee, D.; et al. Design and Implementation of an Ultralow-Power ECG Patch and Smart Cloud-Based Platform. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–11. [CrossRef]

112. Hiraoka, D.; Inui, T.; Kawakami, E.; Oya, M.; Tsuji, A.; Honma, K.; Kawasaki, Y.; Ozawa, Y.; Shiko, Y.; Ueda, H.; et al. Diagnosis of Atrial Fibrillation Using Machine Learning with Wearable Devices After Cardiac Surgery: Algorithm Development Study. *JMIR Form. Res.* **2022**, *6*, e35396. [CrossRef] [PubMed]
113. Shrestha, A.P.; Yu, C.-H. ECG Data Analysis with IoT and Machine Learning. In Proceedings of the 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 26–29 January 2022; pp. 0323–0327.
114. Jenifer, A.; Jeba, G.; Paulraj, L.; Kumar, N.; Yuvaraj, T.; Alen, G.; Rozario, P.; Amoli, R. Edge-based Heart Disease Prediction Device using Internet of Things. In Proceedings of the 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 9–11 May 2022; pp. 1500–1504.
115. Muthu Ganesh, V. Nithiyanantham, J. Heuristic-based channel selection with enhanced deep learning for heart disease prediction under WBAN. *Comput. Methods Biomech. Biomed. Eng.* **2022**, *25*, 1–20. [CrossRef]
116. Dami, S. Internet of things-based health monitoring system for early detection of cardiovascular events during COVID-19 pandemic. *World J. Clin. Cases* **2022**, *10*, 9207. [CrossRef]
117. Koivisto, T.; Lahdenoja, O.; Hurnanen, T.; Vasankari, T.; Jaakkola, S.; Kiviniemi, T.; Airaksinen, K.J. Mechanocardiography in the Detection of Acute ST Elevation Myocardial Infarction: The MECHANO-STEMI Study. Sensors **2022**, *22*, 4384. [CrossRef]
118. Sheeba, A.; Padmakala, S.; Subasini, C.A.; Karuppiah, S.P. MKELM: Mixed Kernel Extreme Learning Machine using BMDA optimization for web services based heart disease prediction in smart healthcare. *Comput. Methods Biomech. Biomed. Eng.* **2022**, *25*, 1–15. [CrossRef]
119. Shafi, J.; Obaidat, M.S.; Krishna, P.V.; Sadoun, B.; Pounambal, M.; Gitanjali, J. Prediction of heart abnormalities using deep learning model and wearabledevices in smart health homes. *Multimed. Tools Appl.* **2022**, *81*, 543–557. [CrossRef]
120. Aziz, S.; Khan, M.U.; Iqtidar, K.; Ali, S.; Remete, A.N.; Javid, M.A. Pulse plethysmograph signal analysis method for classification of heart diseases using novel local spectral ternary patterns. *Expert Syst.* **2022**, *39*, e13011. [CrossRef]
121. Chakraborty, C.; Kishor, A. Real-Time Cloud-Based Patient-Centric Monitoring Using Computational Health Systems. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 1613–1623. [CrossRef]
122. Jansi Rani, S.V.; Chandran, K.R.; Ranganathan, A.; Chandrasekharan, M.; Janani, B.; Deepsheka, G. Smart wearable model for predicting heart disease using machine learning. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *13*, 4321–4332. [CrossRef]
123. Arpaia, P.; Crauso, F.; De Benedetto, E.; Duraccio, L.; Improta, G.; Serino, F. Soft Transducer for Patient's Vitals Telemonitoring with Deep Learning-Based Personalized Anomaly Detection. *Sensors* **2022**, *22*, 536. [CrossRef]
124. Manimurugan, S.; Almutairi, S.; Aborokbah, M.M.; Narmatha, C.; Ganesan, S.; Chilamkurti, N.; Alzaheb, R.A.; Almoamari, H. Two-Stage Classification Model for the Prediction of Heart Disease Using IoMT and Artificial Intelligence. *Sensors* **2022**, *22*, 476. [CrossRef]
125. Ganti, V.G.; Gazi, A.H.; An, S.; Srivatsa, A.V.; Nevius, B.N.; Nichols, C.J.; Carek, A.M.; Fares, M.; Abdulkarim, M.; Hussain, T.; et al. Wearable Seismocardiography-Based Assessment of Stroke Volume in Congenital Heart Disease. *J. Am. Heart Assoc.* **2022**, *11*, e026067. [CrossRef] [PubMed]
126. Price, D. How to read an electrocardiogram (ECG). Part 1: Basic principles of the ECG. The normal ECG. *South Sudan Med J.* **2010**, *3*, 26–31.
127. Lip, G.Y.; Tse, H.F. Management of atrial fibrillation. *Lancet* **2007**, *370*, 604–618. [CrossRef]
128. Bini, S.A. Artificial intelligence, machine learning, deep learning, and cognitive computing: What do these terms mean and how will they impact health care? *J. Arthroplast.* **2018**, *33*, 2358–2361. [CrossRef]
129. Ertel, W. *Introduction to Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2018.
130. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinform.* **2018**, *19*, 1236–1246. [CrossRef]
131. Bhatt, C.; Kumar, I.; Vijayakumar, V.; Singh, K.U.; Kumar, A. The state of the art of deep learning models in medical science and their challenges. *Multimed. Syst.* **2021**, *27*, 599–613. [CrossRef]
132. Aloysius, N.; Geetha, M. A review on deep convolutional neural networks. In Proceedings of the 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 6–8 April 2017; pp. 0588–0592.
133. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [CrossRef]
134. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
135. Myles, A.J.; Feudale, R.N.; Liu, Y.; Woody, N.A.; Brown, S.D. Brown. An introduction to decision tree modeling. *J. Chemom. A J. Chemom. Soc.* **2004**, *18*, 275–285.
136. MIT-BIH Arrhythmia Database v1.0.0. Available online: https://www.physionet.org/content/mitdb/1.0.0/ (accessed on 1 November 2022).
137. Xu, F.; Uszkoreit, H.; Du, Y.; Fan, W.; Zhao, D.; Zhu, J. Explainable AI: A brief survey on history, research areas, approaches and challenges. In *CCF International Conference on Natural Language Processing and Chinese Computing*; Springer: Cham, Switzerland, 2019; pp. 563–574.
138. Doran, D.; Schulz, S.; Besold, T.R. Besold. What does explainable AI really mean? A new conceptualization of perspectives. *arXiv* **2017**, arXiv:1710.00794.
139. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. XAI—Explainable artificial intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [CrossRef]

140. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 265–284.
141. El Emam, K.; Dankar, F.K. Protecting privacy using k-anonymity. *J. Am. Med. Inform. Assoc.* **2008**, *15*, 627–637. [CrossRef]
142. Li, P.; Li, J.; Huang, Z.; Li, T.; Gao, C.Z.; Yiu, S.M.; Chen, K. Multi-key privacy-preserving deep learning in cloud computing. *Future Gener. Comput. Syst.* **2017**, *74*, 76–85. [CrossRef]
143. Fredrikson, M.; Jha, S.; Ristenpart, T. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; pp. 1322–1333.
144. Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership inference attacks against machine learning models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2017; 2017; pp. 3–18.
145. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*; PMLR: Ft. Lauderdale, FL, USA, 2017; pp. 1273–1282.
146. Xu, J.; Glicksberg, B.S.; Su, C.; Walker, P.; Bian, J.; Wang, F. Federated learning for healthcare informatics. *J. Healthc. Informatics Res.* **2021**, *5*, 1–19. [CrossRef] [PubMed]
147. Nguyen, D.C.; Pham, Q.V.; Pathirana, P.N.; Ding, M.; Seneviratne, A.; Lin, Z.; Dobre, O.; Hwang, W.-J. Federated learning for smart healthcare: A survey. *ACM Comput. Surv. (CSUR)* **2022**, *55*, 1–37. [CrossRef]
148. Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H.B.; Patel, S.; Ramage, D.; Segal, A.; Seth, K. Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017; pp. 1175–1191.
149. Zhang, K.; Song, X.; Zhang, C.; Yu, S. Challenges and future directions of secure federated learning: A survey. *Front. Comput. Sci.* **2022**, *16*, 1–8. [CrossRef]
150. Baltrušaitis, T.; Ahuja, C.; Morency, L.P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 423–443. [CrossRef]
151. Kline, A.; Wang, H.; Li, Y.; Dennis, S.; Hutch, M.; Xu, Z.; Wang, F.; Cheng, F.; Luo, Y. Multimodal Machine Learning in Precision Health. *arXiv* **2022**, arXiv:2204.04777.
152. White, F.E. *Data Fusion Lexicon*; Joint Directors of Labs: Washington, DC, USA, 1991.
153. Jiang, H.; Chen, X.; Zhang, S.; Zhang, X.; Kong, W.; Zhang, T. Software for wearable devices: Challenges and opportunities. In Proceedings of the 2015 IEEE 39th Annual Computer Software and Applications Conference, Taichung, Taiwan, 1–5 July 2015; Volume 3, pp. 592–597.
154. Bhushan, D.; Agrawal, R. Security challenges for designing wearable and IoT solutions. In *A Handbook of Internet of Things in Biomedical and Cyber Physical System*; Springer: Cham, Switzerland, pp. 109–138. 2020.
155. Motti, V.G. Assistive wearables: Opportunities and challenges. In *Adjunct Proceedings of the 2019 Acm International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1040–1043.
156. Balsamo, D.; Merrett, G.V.; Zaghari, B.; Wei, Y.; Ramchurn, S.; Stein, S.; Weddell, A.S.; Beeby, S. Wearable and autonomous computing for future smart cities: Open challenges. In Proceedings of the 2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 21–23 September 2017; pp. 1–5.
157. Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review. *Sensors* **2022**, *22*, 7472. [CrossRef]
158. Albrecht, J.P. How the GDPR will change the world. *Eur. Data Prot. Law Rev.* **2016**, *2*, 287. [CrossRef]
159. Parasol, M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams. *Comput. Law Secur. Rev.* **2018**, *34*, 67–98. [CrossRef]
160. Gray, W.; Zheng, H.R. General Principles of Civil Law of the People's Republic of China. *Am. J. Comp. Law* **1986**, *34*, 715–743. [CrossRef]
161. Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li, W.; Gao, Y. A survey on federated learning. *Knowl.-Based Syst.* **2021**, *216*, 106775. [CrossRef]
162. De Luca, C.J.; Gilmore, L.D.; Kuznetsov, M.; Roy, S.H. Filtering the surface EMG signal: Movement artifact and baseline noise contamination. *J. Biomech.* **2010**, *43*, 1573–1579. [CrossRef] [PubMed]
163. Islam, M.K.; Rastegarnia, A.; Sanei, S. Signal Artifacts and Techniques for Artifacts and Noise Removal. In *Signal Processing Techniques for Computational Health Informatics*; Springer: Cham, Switzerland, 2021; pp. 23–79
164. Mammen, P.M. Federated learning: Opportunities and challenges. *arXiv* **2021**, arXiv:2101.05428.
165. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol. (TIST)* **2019**, *10*, 1–19. [CrossRef]
166. Li, Q.; Wen, Z.; Wu, Z.; Hu, S.; Wang, N.; Li, Y.; He, B. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Trans. Knowl. Data Eng.* **2021**, 1. [CrossRef]
167. Li, L.; Fan, Y.; Tse, M.; Lin, K.Y. A review of applications in federated learning. *Comput. Ind. Eng.* **2020**, *149*, 106854. [CrossRef]
168. Iriarte, J.; Urrestarazu, E.; Valencia, M.; Alegre, M.; Malanda, A.; Viteri, C.; Artieda, J. Independent component analysis as a tool to eliminate artifacts in EEG: A quantitative study. *J. Clin. Neurophysiol.* **2003**, *20*, 249–257. [CrossRef]

169. Ram, M.R.; Madhav, K.V.; Krishna, E.H.; Komalla, N.R.; Reddy, K.A. A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. *IEEE Trans. Instrum. Meas.* **2011**, *61*, 1445–1457. [CrossRef]
170. Daly, I.; Billinger, M.; Scherer, R.; Müller-Putz, G. On the automated removal of artifacts related to head movement from the EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 427–434. [CrossRef] [PubMed]

sensors

MDPI

Review

Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review

Mohammad Moshawrab [1,*], Mehdi Adda [1], Abdenour Bouzouane [2], Hussein Ibrahim [3] and Ali Raad [4]

[1] Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC G5L 3A1, Canada
[2] Département d'Informatique et de Mathématique, Université du Québec à Chicoutimi, 555 Boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada
[3] Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, QC G4R 5B7, Canada
[4] Faculty of Arts & Sciences, Islamic University of Lebanon, Wardaniyeh P.O. Box 30014, Lebanon
* Correspondence: mohammad.moshawrab@uqar.ca; Tel.: +1-(581)624-9394

Abstract: Today's world is changing dramatically due to the influence of various factors. Whether due to the rapid development of technological tools, advances in telecommunication methods, global economic and social events, or other reasons, almost everything is changing. As a result, the concepts of a "job" or work have changed as well, with new work shifts being introduced and the office no longer being the only place where work is done. In addition, our non-stop active society has increased the stress and pressure at work, causing fatigue to spread worldwide and becoming a global problem. Moreover, it is medically proven that persistent fatigue is a cause of serious diseases and health problems. Therefore, monitoring and detecting fatigue in the workplace is essential to improve worker safety in the long term. In this paper, we provide an overview of the use of smart wearable devices to monitor and detect occupational physical fatigue. In addition, we present and discuss the challenges that hinder this field and highlight what can be done to advance the use of smart wearables in workplace fatigue detection.

Keywords: smart wearables; occupational fatigue; fatigue detection; smart health; productivity management; heart rate variability; diseases prediction

Citation: Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review. *Sensors* **2022**, *22*, 7472. https://doi.org/10.3390/s22197472

Academic Editors: Maria Linden, Andrea Cataldo, Mia Folke, Mats Björkman and Ning Xiong

Received: 19 August 2022
Accepted: 29 September 2022
Published: 2 October 2022

1. Introduction

Our world has recently been changing at a fast pace. Several global events have clearly impacted many areas of our lives. For example, the improvement of information and communication technologies (ICT) has changed many of our concepts, such as educational habits, business processes, entertainment methods, health services, and much more. Nevertheless, some events have had a negative impact on the global economy and labour market, such as the 11 September attacks, the 2008 economic crisis, and more recently, the COVID-19 pandemic. Whether it is due to technology having increased the pace of work or economic stress forcing people to work more to adapt, or that working life has changed, the pace of business has increased, or work has become more intense and faster, is yet to be determined [1–8]. In addition, the concept of the "24/7 society" has also increased time pressure. The need to increase productivity requires the working hours to be extended, which has lengthened the average working day and shortened the average recovery times [9]. In addition, the introduction of rotating shifts has contributed to disrupting the biological clock and circadian rhythms of workers. Therefore, fatigue, sleep deprivation, and psychosocial stress are considered the main consequences of this increased work intensity and time pressure [10].

1.1. Fatigue Definition(s)

Despite its severity and health significance, and although the term fatigue has been extensively studied recently, it is used in many different meanings and there is currently no single accepted definition [1]. For example, the authors of Ref. [11] defined it as "a reduction in physical and/or mental performance as a result of physical, mental, or emotional exertion that may affect virtually all physical abilities, including strength, speed, reaction time, coordination, decision making, or balance". However, Ref. [12] described it as a state that fluctuates between alertness and drowsiness, whereas Ref. [13] defined it as a state of the muscles and central nervous system in which prolonged physical activity or mental processing, in the absence of adequate rest, results in insufficient capacity or energy to maintain the initial level of activity and/or processing. In addition, Ref. [14] defines fatigue as a decreased capacity or motivation to work that is accompanied by feelings of tiredness and sleepiness. Despite the differences between definitions, all agree that fatigue is associated with or is itself a lack of activity and motivation. Researchers often distinguish between acute and chronic fatigue [15]. Acute fatigue is clearly due to a single cause, occurs in healthy people, is considered normal, sets in quickly, and lasts only a short time. Chronic fatigue, on the other hand, is known to have multiple, additive, or unknown causes, occurs regardless of activity or exertion, and, according to the author, usually cannot be eliminated by common means [16]. In addition, researchers distinguish between different types of acute fatigue, such as: Occupational physical fatigue, occupational mental fatigue, occupational heat stress, occupational noise stress, and others [17]. Occupational physical fatigue, which is the subject of this article, is thus defined as the work-related physical fatigue due to various causes that can be divided into two groups: work-related and person-related causes and contributors [18].

1.2. Fatigue Is Silent—Never Underestimate It

Fatigue has become a commonplace and almost universal feature of our modern lives. Increasing fatigue has led to sleep problems and has gradually entered standard disease patterns [1]. Although acute fatigue has identifiable causes and is considered normal, it can become pathological if it persists. The consequences of fatigue can range from mild, infrequent symptoms to severe, disabling symptoms, and even lead to chronic fatigue syndrome [19]. Consequently, it is important to track fatigue, not only because of its potential consequences, but also because individuals may not accurately assess their fatigue level, which requires immediate or real-time measurement [20]. Moreover, this real-time measurement and assessment is necessary because physicians may erroneously conclude during routine field examinations that fatigue measured in the field is not severe and will not lead to certain illnesses [14].

1.2.1. Health Consequences

Studies and research have shown that fatigue is not only widespread in almost all sectors of the economy, but that there is also a direct relationship between occupational physical fatigue and various diseases. For example, it has been demonstrated in Refs. [21,22] that prolonged physical fatigue can weaken the immune system and cause chronic fatigue syndrome. In addition, studies have shown that 33% of all work-related musculoskeletal injuries and illnesses in the construction industry in the United States are due to fatigue and overexertion [23]. Similarly, Refs. [24–26] have also found that physical fatigue is a leading cause of work-related injuries in the oil, gas, and construction industries. In addition, fatigue is considered particularly dangerous where work safety is of outermost importance, such as in public transportation, health care, and other fields. In addition, numerous studies have found a direct relationship between occupational physical fatigue and disease. For example, in Refs. [27,28], the authors mentioned that fatigue can lead to one or more serious, critical, and fatal diseases. Figure 1 below shows some of the diseases that can be caused by the accumulation and persistence of occupational physical fatigue [10,14,24–31].

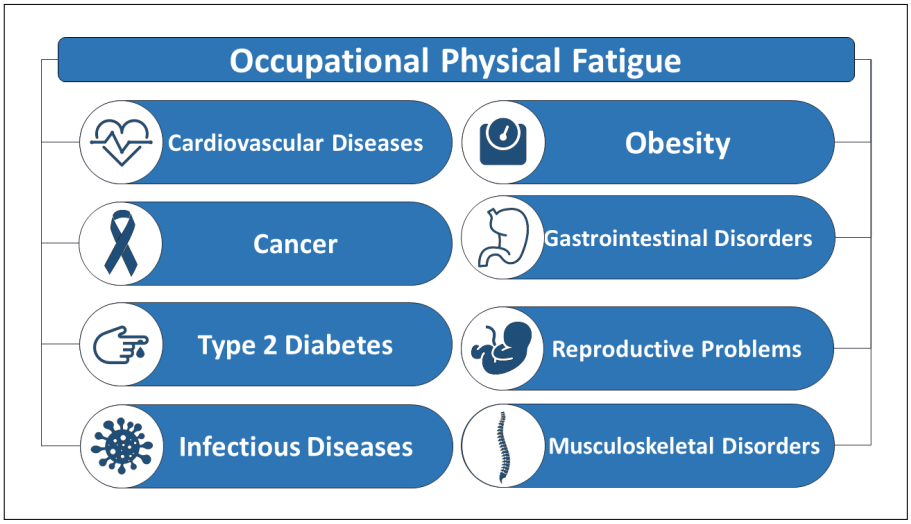

Figure 1. Diseases caused by occupational physical fatigue.

1.2.2. Fatigue and Cardiovascular Diseases

The most critical concept lies in the fact that studies have proven that there is a direct relationship between occupational physical fatigue and heart disease. This relationship reaches a level of causality, as persistent fatigue is confirmed as a direct cause of future heart diseases, or so-called Cardiovascular Diseases (CVDs) [29–31]. More disturbingly, CVDs are considered the most deadly diseases, causing the most deaths and disability-adjusted life years (DALYs) worldwide [32,33]. In this context, numerous studies have discussed the relationship between fatigue and the cardiac system, showing that haemodynamic correlates, decreased indices of stroke volume and cardiac output, hypertension, myocardial infarction, cardiac arrest, and acute myocardial infarction are all consequences of prolonged acute fatigue [10,14,28,34–52]. This causal relationship makes occupational physical fatigue in the workplace intolerable as it causes one of the most dangerous diseases—cardiovascular diseases. Therefore, solutions are needed to control fatigue and avoid deterioration of the health of the workers.

1.3. Detection of Occupational Physical Fatigue

Fatigue is a health symptom to watch out for, and its presence should not be underestimated. As mentioned previously, the presence of fatigue can be considered normal, but its persistence is a dangerous alarm signal for critical health situations. For this reason, instruments for measuring fatigue are not new concepts, as numerous attempts to detect fatigue have already been developed and used [53]. For example, subjective questionnaires were developed in the early 1990s to quantify physical fatigue in the general population, as proposed in Refs. [54,55], and, later, similar attempts were made with the same goal. However, because no standardized scale was developed to assess physical fatigue, different scales were used to measure fatigue, which made it impossible to compare the results of different studies. In addition, the subjective questionnaire technique, although considered a low-cost instrument, is subject to recall errors, is considered intrusive because it takes up workers' time and attention, and, most importantly, is unable to capture fatigue or its consequences in real time. To overcome all the above limitations of the questionnaire, researchers have attempted to collect and analyse various vital signs to detect the presence of fatigue.

Detection by Vital Signs

The need to accurately detect physical fatigue in real time requires monitoring and tracking of some vital signs and biomarkers such as heart rate, heart rate variability (HRV), skin temperature, electroencephalogram (EEG), electromyography (EMG), jerk metrics, and others [17,18,53,56,57]. However, some studies have shown that fatigue has no significant effects on simple measures such as heart rate or blood pressure [14]. Therefore, EEG is the most commonly-used signal to analyse a person's level of relaxation and fatigue. However, EEG is measured with equipment that restricts the worker's activity and is therefore considered invasive. Accordingly, other alternatives are crucial to detect physical fatigue using vital signs without restricting the worker's activity and movement, such as the nocturnal autonomic nervous system (ANS) activity. ANS activity is detected using heart rate variability, motion, and sleep data [19,20,58–60]:

- Heart rate variability (HRV): is an analysis of milliseconds variations in the intervals between heartbeats and reflects the build-up of self-regulatory forces in the body while performing a stressful task [19];
- Motion data: consists of the number of steps, acceleration, rotation, and other parameters and is necessary to improve the accuracy of fatigue detection [20];
- Sleep data: it is proved that there is a bidirectional relationship between fatigue and sleep, where the lack of sleep increases the feeling of fatigue and increasing fatigue leads to sleep problems [20].

In this context, analysis of heart rate variability data is an efficient method to detect fatigue in different populations. In particular, low parasympathetic activity has been associated with the diagnosis of fatigue and burnout [61]. This is possible because HRV mimics the build-up of self-regulatory forces in the body during stressful activities with high mental or physical workload. Parameters extracted from HRV data and analysed to detect fatigue are divided into three main groups: time domain parameters, frequency domain parameters, and non-linear parameters [19,62–71]. These parameters are presented and explained in Table 1 below.

Time-domain parameters are used to calculate the amount of variance in measurements of the interbeat interval (IBI), which is the period between successive heartbeats. Time domain parameters can be expressed in original units, or as the natural logarithm (Ln) of the original units. On the other hand, the frequency domain parameters evaluate the absolute or relative power distribution in the frequency bands: very low frequency (VLF), low frequency (LF), and high frequency (HF). Finally, the nonlinear parameters allow to measure the unpredictability of a time series [19,71];

In addition, In Table 1, the two terms NN Intervals and RR Intervals are used. The RR interval signifies the time between two successive heartbeats, measured from peak (R) to peak (R) on the QRS complex, which is the combination that represents ventricular depolarization of the heart and is composed of Q wave, R wave, and S wave. However, the NN interval denotes the RR interval data but with added filtering to eliminate the artefacts and noise that make some RR intervals unreliable [19,71].

Table 1. Heart rate variability parameters.

Group	Parameter	Unit	Description
Time domain parameters	Mean NN	(ms)	Mean NN ms Mean of NN interval
	SDNN	(ms)	Standard deviation of NN intervals
	RMSSD	(ms)	Square root of the mean squared differences of successive NN intervals
	pNN50	(ms)	Proportion of interval differences of successive NN intervals greater than 50 ms
Frequency domain parameters	VLF	(ms^2)	Power in very low frequency range (0–0.04 Hz)
	LF	(ms^2)	Power in low frequency range (0.04–0.15 Hz)
	HF	(ms^2)	HF ms2 Power in high frequency range (0.15–0.4 Hz)
	LF/HF	(ratio)	Ratio of LF over HF
Non-linear parameters	SD1	(ms)	Standard deviation of points perpendicular to the axis of line of identity or standard deviation of the successive intervals scaled by $\sqrt{\frac{1}{2}}$ $\sqrt{\frac{1}{2}var(RR_{\mathrm{n}} - RR_{\mathrm{n+1}})}$
	SD2	(ms)	Standard deviation of points along the axis of line of identity, or $\sqrt{2SDNN^2 - \frac{1}{2}SD1^2}$
	SD1/SD2	(ratio)	Ratio of SD1 over SD2

1.4. Main Contributions of This Article

This article addresses the use of smart wearables in the detection of occupational physical fatigue. Since there are already several reviews on the use of smart wearables for fatigue detection, the topic presented here is a new one. To our knowledge, previous articles either discussed the use of wearables to detect fatigue in general without distinguishing between categories, or addressed other categories such as mental or cognitive fatigue, so the topic of this review is new. Therefore, the main contributions in this article can be summarized by:

- Discussing the use of smart wearables to detect and monitor occupational physical fatigue, which is a new topic, as indicated by:
  - Presentation of different devices/models used in this field;
  - Listing the current state-of-the-art of implementation of smart wearables for occupational physical fatigue detection, classified by the type of device used (custom-built vs. commercially available devices), and the vital signs collected;
  - Naming the artificial intelligence smart models that were embedded in the smart wearable systems and that were used to analyse the subjects' data;
- Investigating the use of smart wearables to predict cardiovascular diseases in the workplace and how these devices can be used to help maintain both worker health and company productivity;
- Comprehensively indicating the challenges that may hinder progress in the use of smart wearables in the workplace and what future prospects can be targeted to overcome these issues.

Throughout the article, Section 2 discusses the definition, history, and classification of smart wearables. Then, Section 3 explains the use of smart wearables to detect physical fatigue in the workplace. Section 4 presents the challenges that hinder the progress of smart wearables in this area and identifies future directions that can be pursued to overcome these issues. A concluding section briefly summarizes the entire article.

2. Smart Wearables: A New Computing Concept

The rapid development of information and communication technologies along with the improvement of electronics, especially microprocessors, has given rise to a new generation of tiny, robust, and efficient computing devices, such as smart wearables, which can also be referred to as smart wearable technology or wearable devices. These devices provide access to data anytime and anywhere and are heralded as the next generation of ubiquitous technologies after smartphones [72–75]. Smart wearables are a broad technological field that now has applications in many areas of our lives. In the following, we define the term "smart wearables" and provide an overview of the history of wearables. In addition, some classifications of smart wearables are mentioned below.

2.1. Term Definition

The concept of "Smart Machines" was originally launched by Alan Turing in 1950 when he asked his famous question, "Can machines think?" [76]. This question inspired the translation of the concept into reality, where researchers around the world worked to turn computers into intelligent machines. However, the term "Smart" is not uniformly defined in the literature and is introduced in various ways by different researchers [77]. For example, in Ref. [78], the authors define smart objects by their independence, with the embedded sensors, processors, and network devices giving them the ability to act according to their own knowledge. The tools embedded in the smart object allow it to collect data, analyse it, make decisions based on the results, and even interact with humans. In this sense, smart wearables can be defined as computers embedded in anything that covers the human body [79]. Other definitions of smart wearables describe their functionality. The authors in Refs. [80,81] define smart wearables as devices that are equipped with tools to collect, store, and even analyse human data, and can be worn by the user at any time to measure parameters such as personal data, vital signs, locations, environments, movements, and more.

2.2. Smart Wearables; A Brief History

Smart wearables are defined as a subset of the Internet of Things. The term IoT was coined in 1999 by Kevin Ashton, who proposed a vision of a fantasy world in which all devices are equipped with sensors and actuators and connected via the Internet so that they can interact with each other and with people [82]. However, the entire concept of smart wearables was known decades before Ashton's statement. In 1961, Edward Thorp and Claude Shannon developed a small computer that fit inside a shoe and helped them cheat at a roulette game. This is considered the first wearable computing device ever known [83,84]. In the 1980s, Steve Mann designed and built the "EyeTap glasses", a device that could project computer-generated images onto one eye and support the user's visual perceptions with text information [85]. In addition, in 1996, the U.S. Navy Department of Defense invested in a project to monitor the vital signs of its soldiers, which is also considered an important milestone in the development of smart wearables [86,87]. Since then, researchers have expanded their projects in this field to different areas of life such as health, fitness, sports, fashion, and even other sectors, and smart wearables have gradually evolved from invasive, heavy, and huge technologies to more adaptable, compact, and weightless devices [77].

2.3. Classification of Smart Wearables

Over the past few decades, more than a thousand smart wearables have been researched. Nevertheless, there is no specific standard classification of smart wearables. Therefore, the authors in Ref. [88], classified smart wearables into six categories, which are:

- Entertainment: used for Augmented Reality (AR), control devices, and smart gloves;
- Lifestyle: used for video and voice calls or gesture controls;
- Fitness: used for measuring step count, acceleration, heart rate, and body temperature;
- Medical: used for hearing aids, heart monitoring, remote patient monitoring, and much more;
- Industrial: used for remote and hands-free operations related to industrial and business goals;
- Gaming: used for gaming, such as AR devices.

In contrast, the authors in Ref. [89] classified smart wearables by their type rather than functionality. They illustrated their classification in three groups, which are:

- Watch-type: devices that can receive notifications from smartphones such as text messages and emails;
- Necklace or Wristband-type: devices that are used to monitor people's health data in real time;
- Headmount Display-type: devices that can be used for Virtual Reality (VR) and three-dimensional gaming.

However, this classification may miss some devices such as electronic patches, health clothing, and others. Figure 2 below shows some smart wearables that are currently in use in different medical fields.

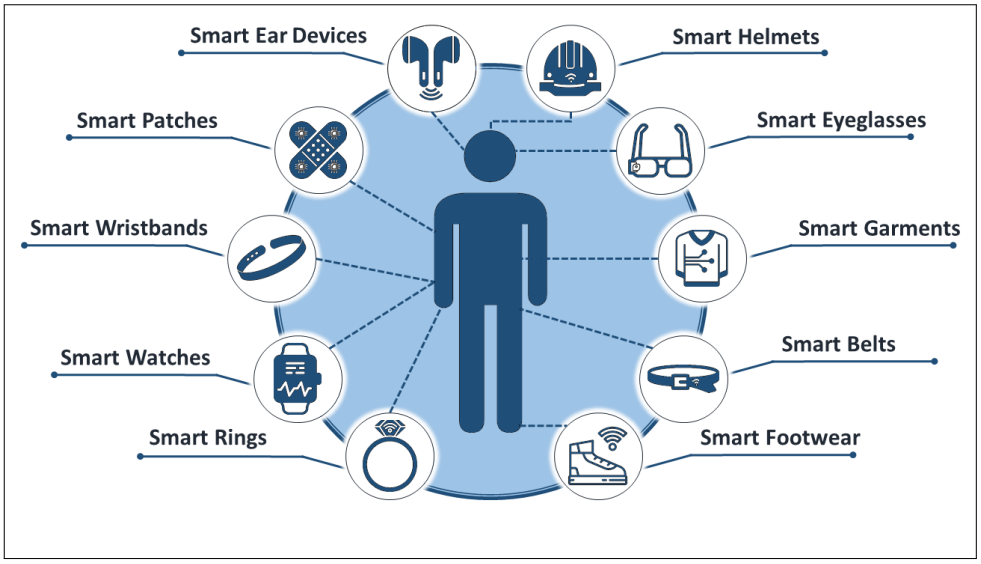

Figure 2. Some of the currently available smart wearables.

3. Smart Wearables and Occupational Physical Fatigue Detection

Given its serious consequences, occupational physical fatigue requires consistent and effective medical intervention, regardless of its causes, burdens, costs, and effects. Artificial intelligence (AI), such as machine learning (ML), the internet of things (IoT), and other vital signs measurement and analysis tools promise to increase the effectiveness of occupational physical fatigue detection devices. Improving the performance of microprocessors, combined with their miniaturization, will help improve fatigue detection to enhance clinical services and meet the growing demand for healthcare services. This is because, on the one hand, patients demand faster and more personalized care, and, on the other hand,

physicians are inundated with data that they need to interpret better, while at the same time they are expected to be more efficient [90,91].

3.1. Smart Wearables and Fatigue: State of the Art

The growing need for real-time fatigue assessment tools has encouraged researchers to work on the appropriate solutions. Over the past decade, several smart wearable systems have been developed to detect occupational physical fatigue. These systems can be divided into three groups in terms of the devices used. The first group includes implementations that use purpose-built devices, the second group includes implementations that use commercially available devices, and the third group consists of implementations where the used devices are not specified.

In the first group, researchers built a variety of devices to monitor vital signs to detect fatigue. The variety of devices stems from the variety of vital signs that can be tracked to detect fatigue. For example, in Refs. [92–94], heart rate data were collected. In Ref. [92], the authors developed a system that can detect different types and degrees of fatigue. The proposed system consists of a smart vest with integrated textrodes, ECG and motion sensors, and a real-time mobile application. The vest collected ECG and thoracic electroimpedance data for this purpose. The system proved to be functional and user-friendly for fatigue risk assessment. In addition, Ref. [93] presented the use of a smart vest with four inertial measurement devices (IMU) and a Shimmer-3 ECG sensor was presented to detect physical fatigue and estimate fatigue levels over time. Once the data were collected, they were analysed using models based on penalized logistic regression and penalized regression, respectively. Similarly, in Ref. [94], the authors proposed the development of a smart vest equipped with a SparkFun heart rate monitor, a Grove Galvanic Skin Response (GSR) sensor, and an MPU-6050 accelerometer/temperature sensor. The vest collects heart rate data to detect workers' level of physical fatigue.

In contrast, the authors in Refs. [95,96] used motion data, proposing a novel, non-intrusive method for monitoring the physical fatigue of construction workers using computer vision technology. Motion data were collected using a 3D motion capture algorithm and IMU sensors. The sensors are attached to a smart vest worn by the test subjects and are monitored by the 3D motion cameras placed in the work area. The captured data was then analysed using Deep Learning algorithms to detect the presence of occupational physical fatigue. In addition, Ref. [96] used time series methods to predict physical fatigue. To achieve their goal, they used ratings of perceived exertion (RPE) and gait data. Data were collected during simulated manual material handling in the laboratory (Lab Study 1) and during a fatiguing squat with intermittent walking (Lab Study 2). The devices used for data collection were IMU, which was strapped around the right ankle, and a smartphone-based IMU sensor strapped around the left lower leg in each study. Data were then analysed using five time series models: Naïve Method, Autoregression (AR), Autoregressive Integrated Moving Average (ARIMA), Vector Autoregression (VAR), and the Vector Error Correction Model (VECM). Those models are explained later in Section 3.3.

Moreover, eye blinks have also been used as a fatigue indicator. In Ref. [97], the authors demonstrated an electronic patch consisting of a flexible strain sensor based on a morphologically modulated laser-patterned film of reduced graphene oxide (LPG) fabricated in a one-step process. The strain sensor was used to monitor human fatigue by analysing the frequency and duration of eye blinks to determine the fatigue level. Similarly, in Ref. [98], the authors proposed a system capable of assessing fatigue based on eye blinks. The device used to monitor the eyes consists of two photovoltaic dye cells. The sensors were attached to the temple of the glasses and positioned on the side of the eye so that they do not interfere with the user's vision. The device records several parameters, including the frequency, duration, and speed of eye blinking, and then analyses the collected data to detect fatigue.

Besides, in Ref. [99], the authors presented a custom-built Smart Safety Helmet (SSH) that can track a worker's head movements and brain activity to detect abnormal behaviour.

The helmet consists of an inertial measurement unit and dry EEG electrodes, and is capable of tracking and analysing a worker's movements to detect fatigue, high stress, or errors to prevent and reduce workplace injuries and accidents. In addition, the helmet is equipped with a small motor that vibrates when risk limits are reached.

In the second group, Ref. [100] offered a new application designed to work with data collected by a Samsung Gear S smartwatch to detect drowsiness in drivers. The smartwatch collects ECG data and analyses it using an intelligent fast Fourier transform (FFT) model to detect drowsiness. The application has two main functions: It reminds drivers to rest every few hours, and it alerts them to nervousness, which can lead to a risky condition.

Finally, in the third group, the authors proposed in Ref. [101] a novel method for detecting physical fatigue in the workplace using heart rate signals. The authors did not specify which device was used to collect the subject's vital signs. However, the model used to analyse the data was built using the k-nearest neighbours (KNN) method. The proposed model provided good results with accuracy, sensitivity, and specificity rates of 78.18%, 60.96%, and 82.15%, respectively.

Alternatively, the implementations of smart wearables for monitoring and detection of occupational physical fatigue in the workplace can be classified based on the collected vital signs. In this context, heart rate, motion, eye blinks, and electroencephalogram were the main biometrics tracked by the existing implementations. Figure 3 shows a classification of these implementations in terms of the vital signs captured and the devices used.

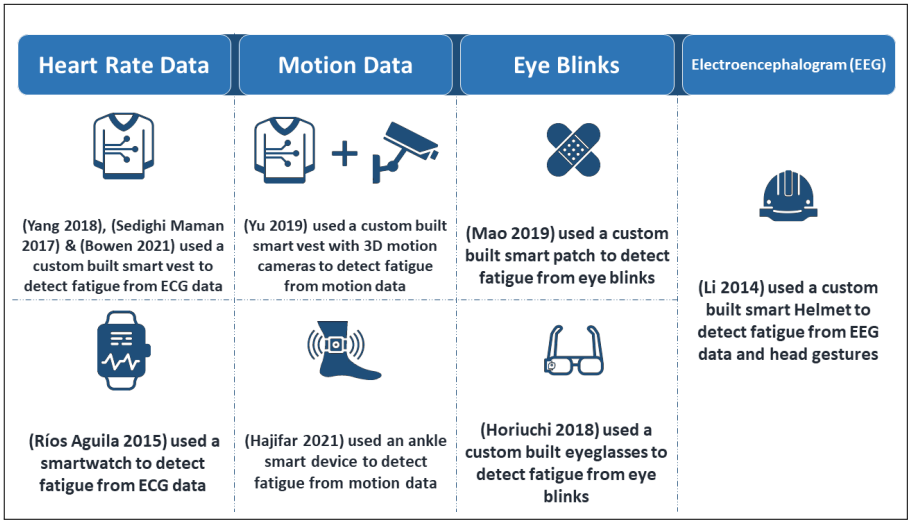

Figure 3. Occupational physical fatigue detection implementations in terms of the vital sign(s) tracked and the device(s) used [92–100].

3.2. Smart Wearables in Fatigue; A Brief Discussion

Several devices have been used in the literature to determine physical fatigue in the workplace. The variety of devices stems from the variety of health biomarkers recorded. Electroencephalogram, electrocardiogram, exercise, eye blinks, and others are good indicators of the presence of fatigue. This is shown by the good results obtained with the different implementations that use these indicators. However, there are some elements that should be considered when selecting hardware for detecting fatigue in the workplace. Below is a list of features that a smart wearable should have for better feasibility:

- Non-invasive: the device should collect data without breaking the subject's skin or invading the body;
- Compact: the wearable should be lightweight and small so that it can be used in the workplace without obstructing the user's activities and movements;
- Affordable: the price of the device affects its adaptation at the workplace;

- Robust: the device should be robust to endure mild, hot, wet, or dry environments and must even withstand harsh working conditions such as minimal scratches or shocks;
- Ease of use: the hardware used should include an easy-to-use interface if it requires minimal user intervention;
- Durable power source: the wearable should have a durable power source to ensure usability for at least one complete work shift to collect significant data.

Knowing that EEG signals are collected by placing small metal discs, also known as EEG electrodes, on the scalp of the subject, devices that use EEG as a vital sign to detect occupational physical fatigue are the least practical among the other devices. The device worn on the head may be heavier, immobilizing, and even more expensive compared to other devices. On the other hand, eyeglasses that record the blinking of the eyes are considered to be lighter and more comfortable in terms of movement and activity of the worker. Moreover, devices that are attached to the body, such as vests, smartwatches, wristbands, ankle bands, or even electronic patches, are considered the most convenient, portable, compact, and lightweight devices that can be used in the workplace to detect physical fatigue.

However, it seems necessary to identify different physical activities associated with the vital signs studied in order to improve the accuracy and robustness of fatigue detection. The reviewed literature showed that the methods that examined motion with other vital signs were promising in terms of accurate fatigue detection. However, it is worth noting that motion data is best captured at the wrist, hip, or feet, while heart rate data is best captured at the wrist or chest, as they are in close proximity to the major blood vessels to check the pulse.

All in all, the smart wearable devices for the wrist, such as the smartbands or smartwatches available on the market, are the best choice for combining the necessary functions and efficiency in measuring the required vital parameters, such as HR and motion. Smart watches and wristbands are commercially available at affordable prices and have easy-to-use interfaces. They are also compact and non-invasive and do not restrict workers in their activities. In addition, they come with acceptable power sources, so they can last for at least an entire work shift. Finally, the ability to capture various vital signs provides them with great efficiency to act as occupational physical fatigue detection devices in the workplace, and they are even the best choice.

3.3. Artificial Intelligence and Fatigue: Smart Models and Data Analysis

Artificial intelligence has been widely used in health area recently [102]. The term AI is explained as a technique that allows a machine to mimic human behaviour and design a working model of the human brain that has the ability to make decisions based on its learning [102–111]. In addition, machine learning (ML) is a subfield of AI that uses statistical techniques to allow a machine to improve itself through learning and experience [102–111]. In addition, deep learning (DL) is a special class of machine learning that has led to the idea of neural networks by simulating how our brain cells, or neurons, work [102–111]. Figure 4 below shows the logical relationship between deep learning, machine learning, and artificial intelligence. It is worth noting that DL has recently attracted more and more attention from health researchers due to its high accuracy, sometimes surpassing human diagnoses [103–111]. The development of AI smart models has helped to develop accurate and efficient systems that can detect fatigue in the workplace.

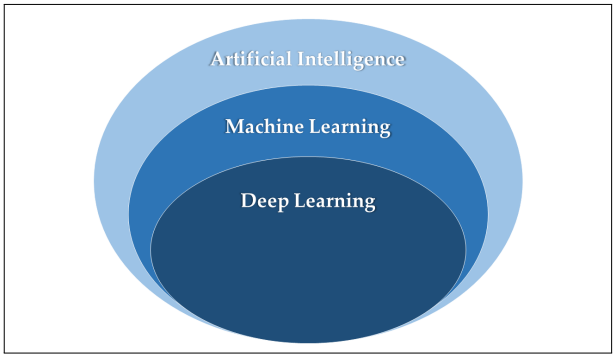

Figure 4. The relationship between AI, ML, and DL.

The authors in Ref. [93] used penalized logistics and multiple linear regression models to detect and estimate physical fatigue over time. In addition, they used least absolute shrinkage and selection operator (LASSO) for the feature selection method. Furthermore, in Ref. [96], the authors compared the use of five time series algorithms for detecting fatigue at work. For this purpose, they applied naïve method, autoregression (AR), autoregressive integrated moving average (ARIMA), vector autoregression (VAR), and vector error correction model (VECM). Similarly, in Ref. [100], the authors used the fast fourier transform (FFT) time series algorithm for fatigue detection. However, in Ref. [101], the k-nearest neighbours method was used as an intelligent model for physical fatigue detection. Table 2 below provides a brief definition of each model.

The information presented in Table 2 shows that it is possible to use vital signs not only to detect occupational physical fatigue, but also to predict its occurrence and estimate its magnitude in the near future. While the use of classification algorithms is suitable for detecting physical fatigue, as in Refs. [93,100,101], the implementation of time series algorithms is suitable for predicting the fatigue state of workers based on past fatigue data, as the authors did in Refs. [93,96]. However, the information provided shows that robust classical machine learning algorithms and the latest deep learning models such as support vector machines (SVMs), deep convolutional neural networks (DCNNs), long short term memory networks (LSTMs), and others that promise higher accuracy have not yet been used for fatigue detection in the literature.

On the other hand, the capability of smart wearables may allow researchers to predict the productivity of future companies or work trends based on current and past fatigue data of their workers. Such an estimation requires that productivity data is collected along with fatigue-related vital signs for further analysis and evaluation. To our knowledge, there are no artificial intelligence or machine learning models that predict productivity based on fatigue monitoring and detection.

Table 2. Artificial intelligence models used in occupational physical fatigue detection.

Ref.	Algorithm(s) Used	Description	Used For	Performance
[93]	Penalized Logistic	Logistic regression is a predictive analysis used to describe data and to explain the association among one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables. However, penalized logistic regression requires a penalty to the logistic model for having too many variables, which leads to shrinking the coefficients of the less contributive variables toward zero and is also recognized as regularization [112,113]	Physical Fatigue Detection: Classification Physical Fatigue Estimation: Forecasting	Best Model Results: Sensitivity: 0.96 Specificity: 0.88
	Multiple Linear Regression Models	Multiple linear regression or known as multiple regression is a method used in statistics to predict the likely outcome based on several variables, plotting the association between these multiple independent variables and single dependent variables [114]		
[96]	Naïve Method	A method that involves using the previous observation directly as the forecast without any change and it can be adjusted slightly for seasonal data [115,116]	Forecast Physical Fatigue : Forecasting	Best model: VECM Mean Absolute Scaled Error (MASE): 0.43 for a 6-steps ahead fatigue forecasting
	Autoregression (AR)	A time series model that uses observations from previous time steps as input to a regression equation to predict the value at the next time step [116]		
	Autoregressive Integrated Moving Average (ARIMA)	A time series forecasting model that uses time series data to either better understand the data set or to predict future trends based on past values. It is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables [116]		
	Vector Autoregression (VAR)	A time series multivariate forecasting algorithm that is used when two or more time series influence each other [116]		
	Vector Error Correction Model (VECM)	A restricted vector autoregression model intended for usage with no stationary series that are to be co-integrated [117]		
[100]	Fast Fourier Transform	A computational tool that simplifies signal analysis by computing the discrete Fourier transform (DFT) and its inverse. It works by sampling a signal over a period of time and dividing it into its frequency components used to improve the computational efficiency [118]	Detection of Drowsiness: Classification	-
[101]	K-Nearest Neighbours	A data classification method that guesses how likely a data point relates to a group depending on what group the data points nearest to it are [119]	Physical Fatigue Detection: Classification	Accuracy: 78.18% Sensitivity: 60.96% Specificity: 82.15%

3.4. Occupational Physical Fatigue as a CVD Prediction Parameter

Occupational physical fatigue at the workplace is a normal phenomenon. Its causes are well known, such as repetitive movements and physical exertion. However, it can become pathological when it becomes chronic and leads to various diseases, which in some cases can lead to death [21–52]. However, since there are no static medical formulas that link occupational physical fatigue to disease, there are no applications to date that can

predict the future occurrence of disease due to the presence and persistence of fatigue in the workplace. However, several attempts have been made by researchers to identify cardiovascular risks based on heart rate variability analysis, using time domain, frequency domain, and non-linear HRV parameters for this purpose. Those implementations are discussed below and are also summarized in Table 3 below.

For example, the authors in Ref. [63] used multilayer perceptron (MLP), radial basis function (RBF), and support vector machines (SVM) to analyse HRV series in conjunction with classification schemes to predict cardiovascular risks. The created solution was trained with data collected by the authors and achieved a maximum accuracy of 96.67%. In addition, Ref. [65] proposed a solution to help physicians predict sudden cardiac death (SCD) using smart models based on the k-nearest neighbour (k-NN) and multilayer perceptron neural network algorithms. The models created were based on the PhysioNet databases "Sudden Cardiac Death Holter" [120] and "MIT-BIH Normal Sinus Rhythm" databases [121]. The proposed solution has a high accuracy of 99.73%, 96.52%, 90.37%, and 83.96% for the first, second, third, and fourth one-minute intervals, respectively. Similarly, Ref. [66] proposed an instrument to predict SCD two minutes before its occurrence. The smart models were built using SVM and probabilistic neural network (PNN) and trained with PhysioNet databases "Sudden Cardiac Death Holter" and "MIT-BIH Normal Sinus Rhythm". The presented solution proved its efficiency, with SVM and PNN, achieving a maximum mean SCA prediction rate of 96.36% and 93.64%, respectively.

Table 3. Implementations of cardiovascular risk prediction using HRV.

Ref.	Diseases(s) Detected	Model(s) Used	Dataset(s)	Results
[63]	Cardiovascular Risk	Multilayer Perceptron (MLP) Radial Basis Function (RBF) Support Vector Machines (SVM)	-	Accuracy: 96.67%
[65]	Sudden Cardiac Death (SCD)	k-Nearest Neighbor (k-NN) Multilayer Perceptron Neural Network	"Sudden Cardiac Death Holter" [120] "MIT-BIH Normal Sinus Rhythm" [121]	Accuracy: 99.73%
[66]	Sudden Cardiac Death (SCD)	Support Vector Machines Probabilistic Neural Network (PNN)	Sudden Cardiac Death Holter" "MIT-BIH Normal Sinus Rhythm"	Mean SCA prediction rate: 96.36%
[67]	Cardiovascular Risk	Support Vector Machine (SVM) Trees Based Classifier Artificial Neural Networks (ANN) Random Forest	"Smart Health for Assessing the Risk of Events via ECG" [122]	Sensitivity: 71.4% Specificity: 87.8%
[68]	Ventricular Tachycardia (VT)	Artificial Neural Network (ANN)	-	Accuracy: 82%
[69]	Hypertension	Statistical model called MIL	-	Accuracy: 92.73%
[70]	Arterial Hypertension (AH)	-	World Health Organization's (WHO) MONICA project data [123]	-

Moreover, in Ref. [67], the authors developed novel models to predict cardiovascular risk in hypertensive patients. The models are based on data mining algorithms such as Support Vector Machines, Trees Based Classifier, Artificial Neural Networks (ANN), and Random Forest to provide an automated tool for risk stratification. The models were built using the "Smart Health for Assessing the Risk of Events via ECG" database [122], available on the PhysioNet data repository and achieved a sensitivity of 71.4% and a specificity of 87.8% in risk prediction. In addition, in Ref. [68], the authors proposed a solution to predict

ventricular tachycardia (VT) one hour before its occurrence by using an artificial neural network (ANN) created with 14 parameters from HRV and respiratory rate variability (RRV) analysis. The solution created was trained using data collected by the authors and was accurate in its results up to 82%. Besides, Ref. [69] used photoplethysmography (PPG) to estimate HRV and predict the occurrence of hypertension in the studied subjects. A statistical model called MIL was used for the solution, which was trained using the data collected by the authors and achieved an accuracy of 92.73%. Finally, Ref. [70] further provided a solution to determine the effects of workplace stress on the risk of developing arterial hypertension (AH) in the population. The study used data from the World Health Organization's (WHO) MONICA project data [123] and was able to establish an association between workplace stress and the development of AH.

3.4.1. Cvds Prediction: A Brief Discussion

The state of the art in using HRV to predict cardiovascular diseases or cardiovascular risk is promising, as it serves as an obvious indication that HRV can be collected and analysed in the workplace to detect not only the presence of fatigue but also the possibility of risk for developing CVD in the future. However, because there is no clear formula that can be relied upon to predict health risk due to fatigue, the relationship between the prevalence of fatigue and the presence of cardiovascular risk is an area that requires in-depth investigation. However, this area of investigation may be complicated by several issues, such as the reliability of the results from a medical perspective. In addition, the debate about the possibility of biased reasoning in predicting the ability of developing a cardiovascular risk based on fatigue is a research question that should be studied and analysed in depth to find an appropriate way to link fatigue and CVDs. However, the question here is: **why CVDs, when it has been proven that fatigue can cause many other diseases**?

3.4.2. Why to Predict CVDs at Workplace

Cardiovascular diseases are known as the most deadly diseases worldwide. The number of deaths caused by these diseases is the highest in the world, and these numbers are increasing rapidly. According to a study by the World Health Organization (WHO), the number of deaths caused by CVDs have increased from 12.1 million to 18.6 million between 1990 and 2019 [33]. In addition, the burden of CVDs are also being studied from an economic perspective. For example, the "Medical Expenditure Panel Survey" noted in a report that costs due to CVDs in the United States alone were an estimated USD 378.0 billion between 2017 and 2018. These costs are not limited to expenditures, which were estimated at USD 226.0 billion, but also include an estimate of USD 151.8 billion in lost future productivity, which is considered an extremely high number in governments economics [124]. These facts encourage working on solutions to predict future CVDs in the workplace, not only to protect workers' lives, which are the most sacred, but also to avoid future productivity losses that will impact the national economy and therefore, in turn, have negative public health consequences.

4. Challenges and Future Limitations

Despite the large role smart wearables are expected to play in detecting occupational physical fatigue, several challenges may arise during their implementation. In addition, the emergence of new tools and concepts in artificial intelligence opens up many ideas that can be used to improve fatigue monitoring and detection in the workplace.

4.1. Challenges

The following are the most common obstacles encountered when using smart wearables to detect occupational physical fatigue [53,125].

4.1.1. Data Privacy and Confidentiality

The performance of AI models embedded in smart wearable systems depends on the availability of data. Although achieving highly accurate models depends on the technical structure of the models themselves—the cleanliness and readiness of the data, and other aspects—it is common that the availability of more data to train AI models increases their accuracy. However, in the real world, collecting data is the biggest challenge in developing AI models for several reasons: most importantly, privacy and confidentiality. Not only individuals, but also society, governments, and organizations are strengthening the protection of data privacy and security. In this regard, several regulations and laws have been enacted, such as the European Union's General Data Protection Regulation (GDPR) [126], China's Cyber Security Law of the People's Republic of China [127], the General Principles of the Civil Law of the People's Republic of China [128], the PDPA in Singapore [129], and hundreds of principles that have been legislated around the world. Although these regulations help protect private information, they pose new challenges to the traditional AI data processing model to varying degrees by making it more difficult to collect data to train models, which in turn makes it more difficult to improve the accuracy of model performance [130–134].

4.1.2. Noise and Artefacts

Smart wearables collect vital signs data in a non-invasive way, which makes the records more susceptible to many external sources of noise. These noisy data are called artefacts", which are unwanted signals or signal distributions that interfere with the actual signal. Artefacts are divided into two main groups depending on their origin: intrinsic artefacts, which originate from the monitored body, and extrinsic artefacts, which are caused by the monitored person's environment. There are different sources of artefacts that can be grouped according to their origin [135,136]:

- **Intrinsic artefacts** (also called physiological or internal artefacts)
  - **Ocular artefacts**: any artefact caused by the movement of the eyeball that interferes with EEG recording, such as eye blinks, horizontal and vertical eye movements, eye flutter, etc.;
  - **Muscle artefacts**: arise from activities such as sniffing, swallowing, clenching, talking, eyebrow raising, chewing, scalp contraction, etc.;
  - **Cardiac artefacts**: slow waves that are not recorded on the ECG and are due to the electrical activity of the heart;
  - **Respiratory artefacts**: caused by the movement of an electrode during inhalation or exhalation and may take the form of slow, rhythmic EEG activity;
  - **Sweat artefacts**: caused by changes in the electrolyte concentration of the electrode due to sweat secretion on the scalp.
- **Extrinsic artefacts** (also called extra-physiological/external artefacts)
  - **Motion artefact**: The motion of the monitored body in an EEG monitoring system produces a lot of motion artefacts;
  - **Environmental artefact**: These can occur when contact is lost between the electrode and the scalp, when the electrode bursts, or when electrical or electronic devices in the environment that generate electromagnetic waves cause interference, etc.

Artefacts and noise affects the quality of data, which therefore reduces the performance and precision of detecting and predicting occupational physical fatigue.

4.1.3. Data Heterogeneity

As mentioned earlier, fatigue in the workplace can be monitored and tracked with smart wearables. However, accurate and reliable measurement of fatigue requires the collection of more than one vital sign, such as heart rate and motion, as discussed in Section 3.2. In addition, embedding other health data, such as some medical tests extracted

from electronic health records (EHRs), can improve monitoring results. Nevertheless, it is not easy to analyse data with heterogeneous structures, especially when they are scattered in more than one data space. Therefore, integrating data from different modalities or different measurement devices and merging them to monitor and detect occupational physical fatigue in the workplace is a challenging task.

4.1.4. Some Vital Signs Limitations

Vital signs are considered the most important indicators for detecting physical fatigue. However, some studies have shown that there is no significant effect of fatigue on simple signs such as heart rate or blood pressure [14]. This limits the selection to EEG and HRV or eye-blinks. Since EEG limits the activity of the worker and eye-blinks cannot be readily detected in some work environments, HRV is considered to be almost the only biomarker that can detect occupational physical fatigue without affecting the activity of the worker.

4.1.5. Lack of Standard and Unified Fatigue Classification Scale

Although the preliminary results of using smart wearables to detect work-related physical fatigue are promising, there is no clear standard or unified scale to refer to when classifying fatigue. Although some questionnaire-based assessment methods have succeeded in classifying fatigued individuals into different groups, as in Ref. [137], there is no unified scale that can be used to measure fatigue when using smart wearables. Therefore, almost all implementations that use smart wearables to detect fatigue look for a binary result of whether fatigue is present or not. Furthermore, to our knowledge, no study has validated the use of physiological measures versus the gold standard for assessing physical fatigue (i.e., blood lactate levels).

4.1.6. Lack of Knowledge about Clear Thresholds of Vital Signs for Severe Physical Fatigue

One of the major challenges in analysing vital signs data obtained from smart wearables is the lack of information on the unique thresholds of individual vital signs for severe physical fatigue. Although it is clear that accumulation of physical fatigue over a long period of time can lead to various health problems, there are no clear thresholds for various vital signs that indicate extreme fatigue. Furthermore, to our knowledge, there are no formulas that can be used to predict disease based on fatigue data.

4.1.7. Difficulty Going beyond Fatigue Detection toward Diseases Prediction

In the absence of a unified fatigue scale, clear thresholds for severe fatigue, and unambiguous formulas that can link accumulation of fatigue symptoms to disease, smart wearables are being used almost as detectors of physical fatigue in the workplace. Researchers are trying to explore what role smart wearables can play in predicting diseases caused by persistent fatigue. However, as far as we know, there are no such implementations, as most applications that predict diseases analyse HRV or other vital signs independent of fatigue status.

4.1.8. User Technology Adoption and Engagement

One of the most common challenges hindering the use of smart wearables to detect physical fatigue at work is user acceptance, adoption, and engagement. User acceptance of wearing such sensors varies due to issues of privacy, comfort, or other social circumstances.

We can therefore summarize the challenges and obstacles as the research questions mentioned in the following list. In addition, those questions are illustrated in Figure 5 below (the symbol RQ in the list below and in Figure 5 refers for the term research question):

- **RQ1**: Subject data are private, and laws may restrict their disclosure. How can these data be used without violating privacy?
- **RQ2**: Data collected in the workplace are exposed to various sources of noise and interference. How should noisy data and artefacts be handled?

- **RQ3**: Analysing diverse data can improve fatigue detection. Is it possible to analyse heterogeneous data with AI models?
- **RQ4**: There are several biometric parameters that can be used to detect occupational physical fatigue in the workplace. Which one(s) is/are most appropriate and how can health characteristics be associated with fatigue duration?
- **RQ5**: Proactive fatigue prediction can help maintain both worker health and organizational productivity. Is it possible to use smart wearables to predict illness in the workplace?

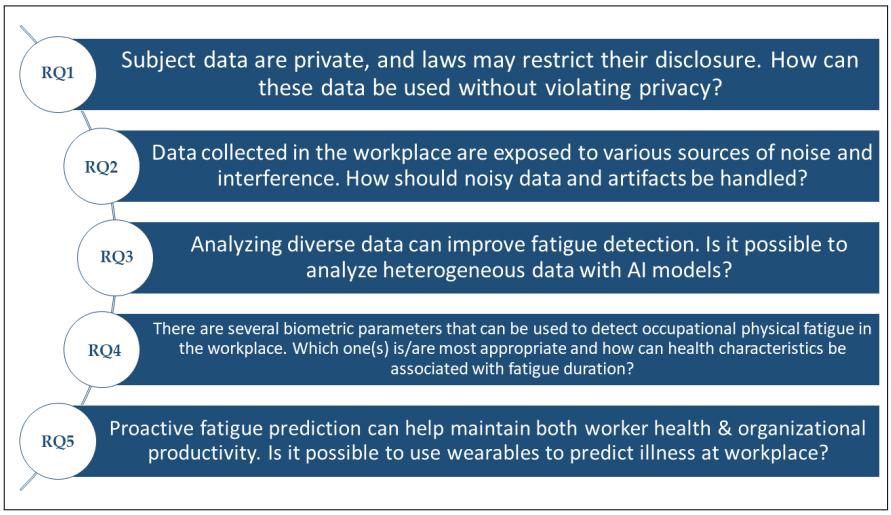

Figure 5. Research questions arising from analysing usage of wearables in fatigue detection.

4.2. Future Perspectives and Research Trends

Smart wearables are already being used successfully to detect and monitor fatigue. However, given the global prevalence of occupational physical fatigue due to changing work patterns, such as varied and rotating shifts, there is a growing need to improve the entire process and take further steps toward proactive and preventive approaches. This growing need requires additional efforts in the development of smart wearables that go beyond simple fatigue detection.

4.2.1. Preserving Data Privacy

Regulations, laws, user disapproval, and other factors limit the collection of worker health data. Traditionally, data collected from subjects should be collected on a local centralized server or distributed to various decentralized storage and processing devices to create and train AI models that are then able to detect fatigue. Therefore, the model created has full access to the subject's data, whether anonymous or labelled by the subject. Consequently, the data are not private. However, later machine learning approaches propose new privacy alternatives. For example, federated learning (FL) is a promising technology that can help solve privacy problems. Federated learning is defined as a type of collaborative distributed/decentralized machine learning privacy-preserving technology in which a model is trained without the need to transfer data from edge devices to a central server. Instead, the trained models are shared between the edge devices and the central server, which acts as an aggregation station to build the global model without knowing the embedded data [130–134]. The use of FL in occupational physical fatigue detection and monitoring is expected to help overcome the privacy issue and therefore facilitates the collection of more data, which helps improve accuracy.

4.2.2. Removing Artefacts and Noisy Data

Extrinsic and intrinsic signal artefacts that obscure the signals should be removed or minimized before processing the signals. Several implementations have already been made for this purpose, such as those mentioned in Refs. [136,138–140]. Therefore, automation of noise reduction is an area that should be investigated to clean and preprocess the data to improve the accuracy of physical fatigue detection in the workplace.

4.2.3. Analysing Diverse and Heterogeneous Data

Medical studies have shown that precise and accurate assessment of occupational physical fatigue at work requires the use of multiple vital signs rather than a single indicator. However, with the advent of multimodal machine learning technology, it becomes possible to analyse data read from or collected by multiple devices. Multimodal machine learning is defined as the ability to analyse data from multimodal datasets, observe a common phenomenon, and use complementary information to learn a complex task. Here, multimodal datasets are defined as data observed with multiple sensors, where the output of each sensor is called a modality and can be associated with a dataset [141]. Multimodal ML is based on the concept of "data fusion", which is defined as "the process of combining data to refine state estimates and predictions". According to the Joint Directors of Laboratories Data Fusion Subpanel (JDL), the technique referred to as "data fusion" is a must for processing more than one type of data [142]. In this context, data fusion is divided into the following three categories:

- Early fusion: can be referred to as a multiple data, single smart model;
- Intermediate fusion: occurs in the intermediate phase between input and output of a ML architecture when all data sources have the same representation format. In this phase, features are combined to perform various tasks such as feature selection, decision making, or predictions based on historical data;
- Late fusion: defines the aggregation of decisions from multiple ML algorithms, each of which has been trained with different data sources.

Therefore, embedding multimodal ML into smart wearables is crucial to analyse heterogeneous data and thus enhancing the accuracy and precision of detection and monitoring.

4.2.4. Raising Accuracy, Increasing Explainability, and Gaining Trust

In the workplace, it is becoming increasingly important to monitor the health of workers, especially as work pressures increase due to the changing concepts of work around the world. Given the need to keep an eye on health without hindering workers in their work, smart wearables are considered as one of the most practical tools that can be used. However, there is a need to increase the accuracy of fatigue detection with wearables, improve the explainability of these tools, and eliminate the black box characteristics of the models embedded in these smart wearables as much as possible. Increased accuracy and better explainability will help these devices gain trust and, as a result, be used as health monitoring devices in the workplace.

4.2.5. Using Smart Wearables as Predictive Tool

Smart wearables have demonstrated their high efficiency in monitoring workers' vital signs, such as heart rate and other metrics such as movement and activity data. The ability to capture such parameters in the workplace and in real time, as well as the high accuracy with which AI and ML models can analyse this data, opens the door to using all of these capabilities in predicting health problems based on fatigue data. This will help maintain the long-term health of the working population.

4.2.6. Monitoring Workers Productivity Linked to Fatigue

Furthermore, the use of smart wearables can be extended to productivity management in companies. This can be achieved by identifying the relationship between worker fatigue and productivity. To the best of our knowledge, all previous implementations of smart

wearables for fatigue detection have not considered worker productivity and have not addressed the identification of the relationship between fatigue and productivity. Detecting such a link would also help companies increase revenue by improving work processes while maintaining the health of their employees.

Therefore, we can summarize the future perspectives into the trending research topics mentioned in the following list. In addition, those research topics are illustrated in Figure 6 below (the symbol TR in the list below and in Figure 6 refers for the term trending research topic):

- **TR1:** Integrate federated learning into smart wearables implementations for fatigue detection to preserve subject privacy;
- **TR2:** Automate artefact and noise removal algorithms to reduce the impact of interference and noise;
- **TR3:** Use multimodal ML algorithms to analyse data from multiple modalities and sources to improve the precision and accuracy of recognition models;
- **TR4:** Use the multimodal ML to step for analysis of more than one vital sign when possible, rather than limiting analysis to just one biometric parameter;
- **TR5:** Increase efforts to build predictive models to predict workplace illnesses for a win-win for both workers and commercial enterprises.

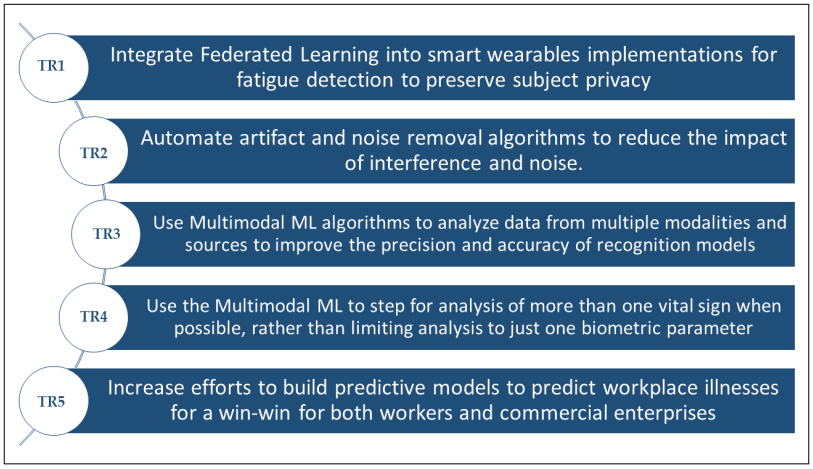

Figure 6. Research topics that may serve as solutions to the challenges in the domain.

To summarize the challenges-future-solutions, and to help boost the research of the usage of smart wearables in the detection of occupational physical fatigue, Figure 7 below presents a link between the current top challenging issues and future perspectives that can serve as possible solutions in the domain.

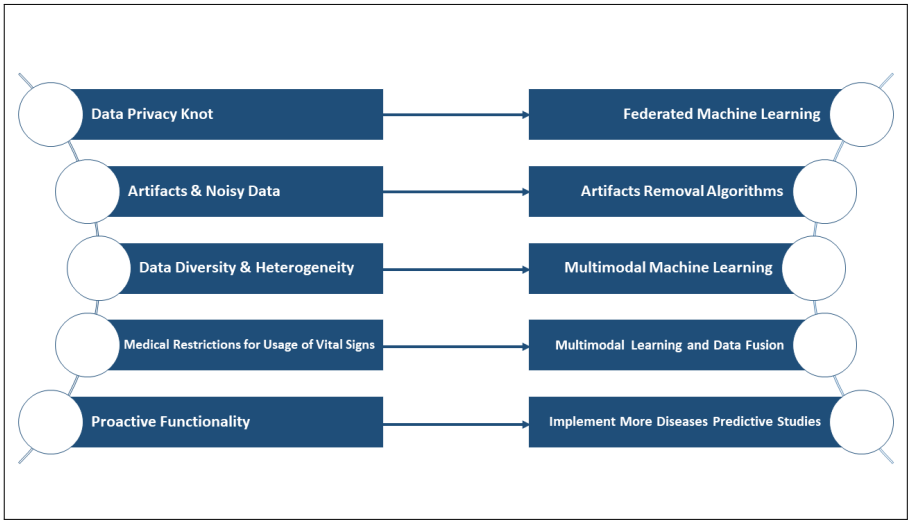

Figure 7. Challenges-future-solutions chart.

5. Conclusions

Our world is changing at an accelerating pace, with almost our entire environment changing within years and sometimes months. The "when", "where", and "how" to work are also concepts that have changed for various reasons, such as the COVID-19 pandemic, which may not be the last to change our notions of work or increase work pressure. Consequently, work-related fatigue, also known as occupational physical fatigue, is spreading and becoming more common worldwide. This increases the need for solutions that can monitor workplace fatigue to prevent workers' health from deteriorating, especially because the accumulation of fatigue can seriously affect workers' health and even lead to death, according to some studies. However, smart wearables associated with artificial intelligence and machine learning technologies have proven their effectiveness in detecting and monitoring fatigue in the workplace, especially when the relevant challenges can be addressed with the latest and most advanced technologies. They also promise to act as predictive tools that can limit the serious impact of fatigue on workers' health.

Author Contributions: Conceptualization: M.M. and M.A.; Formal analysis: M.M.; Investigation: M.M.; Methodology: M.M. and M.A.; Supervision: M.A., A.B., H.I. and A.R.; Visualization: M.M.; Writing—original draf: M.M.; Writing—review & editing: M.A., A.B., H.I. and A.R.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number 06351.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dawson, D.; Noy, Y.I.; Härmä, M.; Åkerstedt, T.; Belenky, G. Modelling fatigue and the use of fatigue models in work settings. *Accid. Anal. Prev.* **2011**, *43*, 549–564. [CrossRef]
2. Peric, M.; Vitezic, V. Impact of global economic crisis on firm growth. *Small Bus. Econ.* **2016**, *46*, 1–12. [CrossRef]
3. Chang, S.S.; Stuckler, D.; Yip, P.; Gunnell, D. Impact of 2008 global economic crisis on suicide: Time trend study in 54 countries. *BMJ* **2013**, *347*, f5239. [CrossRef] [PubMed]
4. Gupta, M.; Abdelmaksoud, A.; Jafferany, M.; Lotti, T.; Sadoughifar, R.; Goldust, M. COVID-19 and economy. *Dermatol. Ther.* **2020**, *33*, e13329. [CrossRef]

5. Jason, L.A.; Corradi, K.; Gress, S.; Williams, S.; Torres-Harding, S. Causes of death among patients with chronic fatigue syndrome. *Health Care Women Int.* **2006**, *27*, 615–626. [CrossRef]
6. Irvine, J.; Basinski, A.; Baker, B.; Jandciu, S.; Paquette, M.; Cairns, J.; Dorian, P. Depression and risk of sudden cardiac death after acute myocardial infarction: Testing for the confounding effects of fatigue. *Psychosom. Med.* **1999**, *61*, 729–737. [CrossRef] [PubMed]
7. Paoli, P.; Merllié, D. Third European Survey on Working Conditions 2000. European Foundation for the Improvement of Living and Working Conditions. 2001. Available online: https://www.eurofound.europa.eu/publications/report/2001/working-conditions/third-european-survey-on-working-conditions-2000 (accessed on 10 October 2021).
8. Figart, D.M.; Golden, L. (Eds.) *Working Time: International Trends, Theory and Policy Perspectives*, 1st ed.; Routledge: London, UK, 2000.
9. Åkerstedt, T.; Folkard, S. Validation of the S and C components of the three-process model of alertness regulation. *Sleep* **1995**, *18*, 1–6. [CrossRef] [PubMed]
10. Härmä, M. Workhours in relation to work stress, recovery and health. *Scand. J. Work. Environ. Health* **2006**, *32*, 502–514. [CrossRef] [PubMed]
11. IMO. *Guidelines on Fatigue*; International Maritime Organization: London, UK, 2019
12. Desmond, P.A.; Hancock, P.A. Active and passive fatigue states. In *Stress, Workload, and Fatigue*; CRC Press: Boca Raton, FL, USA, 2000; pp. 455–465.
13. Job, R.S.; Dalziel, J. Defining fatigue as a condition of the organism and distinguishing it from habituation, adaptation, and boredom. In *Stress, Workload, and Fatigue*; CRC Press: Boca Raton, FL, USA, 2000; pp. 466–476.
14. Nelesen, R.; Dar, Y.; Thomas, K.; Dimsdale, J.E. The relationship between fatigue and cardiac functioning. *Arch. Intern. Med.* **2008**, 168, 943–949. [CrossRef]
15. Mohren, D.C.L.; Jansen, N.W.H.; van Amelsvoort, L.G.P.M.; Kant, I.A. *Epidemiological Approach of Fatigue and Work: Experiences from the Maastricht Cohort Study*; Wilco: Amersfoort, The Netherlands, 2007.
16. Piper, B.F. Fatigue: Current bases for practice. In *Management of Pain, Fatigue and Nausea*; Macmillan Education UK: London, Greater London; 1989; pp. 187–198.
17. Spook, S.M.; Koolhaas, W.; Bültmann, U.; Brouwer, S. Implementing sensor technology applications for workplace health promotion: A needs assessment among workers with physically demanding work. *BMC Public Health* **2019**, *19*, 1100. [CrossRef]
18. Williamson, A.; Friswell, R. Fatigue in the workplace: Causes and countermeasures. *Fatigue Biomed. Health Behav.* **2013**, 1, 81–98. [CrossRef]
19. Besson, C.; Saubade, M.; Gremeaux, V.; Millet, G.P.; Schmitt, L. Heart rate variability: Methods, limitations and clinical examples. *Rev. Medicale Suisse* **2020**, *16*, 1432–1437. [CrossRef]
20. Gonzalez, K.; Sasangohar, F.; Mehta, R.K.; Lawley, M.; Erraguntla, M. Measuring fatigue through Heart Rate Variability and activity recognition: A scoping literature review of machine learning techniques. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Rome, Italy, 28–30 September 2017; SAGE Publications: Los Angeles, CA, USA; Volume 61, pp. 1748–1752.
21. Afari, N.; Buchwald, D. Chronic fatigue syndrome: A review. *Am. J. Psychiatry* **2003**, *160*, 221–236. [CrossRef] [PubMed]
22. Evengard, B.; Kuratsune, H.; Jason, L.A.; Natelson, B.H. *Fatigue Science for Human Health*; Watanabe, Y., Ed.; Springer: New York, NY, USA, 2008; pp. V–XI.
23. BLS. Nonfatal Occupational Injuries and Illnesses Requiring Days Away from Work in 2015. 2016. Available online: https://www.bls.gov/news.release/osh2.toc.htm (accessed on 25 September 2021).
24. Chan, M. Fatigue: The most critical accident risk in oil and gas construction. *Constr. Manag. Econ.* **2011**, *29*, 341–353. [CrossRef]
25. Adane, M.M.; Gelaye, K.A.; Beyera, G.K.; Sharma, H.R.; Yalew, W.W. Occupational Injuries among Building Construction Workers in Gondar City, Ethiopia. *Occup. Med. Health Aff.* **2013**, *1*, 125. [CrossRef]
26. Tadesse, S.; Israel, D. Occupational injuries among building construction workers in Addis Ababa, Ethiopia. *J. Occup. Med. Toxicol.* **2016**, *11*, 1–6. [CrossRef] [PubMed]
27. Caruso, C.C.; Waters, T.R. A review of work schedule issues and musculoskeletal disorders with an emphasis on the healthcare sector. *Ind. Health* **2008**, *46*, 523–534. [CrossRef] [PubMed]
28. Van der Hulst, M. Long workhours and health. *Scand. J. Work. Environ. Health* **2003**, *29*, 171–188. [CrossRef] [PubMed]
29. Collins, S. Occupational factors, fatigue, and cardiovascular disease. *Cardiopulm. Phys. Ther. J.* **2009**, *20*, 28. [CrossRef] [PubMed]
30. Melamed, S.; Shirom, A.; Toker, S.; Berliner, S.; Shapira, I. Burnout and risk of cardiovascular disease: Evidence, possible causal paths, and promising research directions. *Psychol. Bull.* **2006**, *132*, 327. [CrossRef] [PubMed]
31. Melamed, S.; Kushnir, T.; Shirom, A. Burnout and risk factors for cardiovascular diseases. *Behav. Med.* **1992**, *18*, 53–60. [CrossRef]
32. WHO Reveals Leading Causes of Death and Disability Worldwide: 2000–2019. 2020. Available online: https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019 (accessed on 10 March 2021).
33. Roth, G.A.; Mensah, G.A.; Johnson, C.O.; Addolorato, G.; Ammirati, E.; Baddour, L.M.; Barengo, N.C.; Beaton, A.Z.; Benjamin, E.J.; Benziger, C.P.; et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. *J. Am. Coll. Cardiol.* **2020**, *76*, 2982–3021. [CrossRef]

34. George Citroner. Vital Exhaustion' Increases Heart Attack Risk in Men: What to Know. 16 March 2021. Available online: https://www.healthline.com/health-news/vital-exhaustion-increases-heart-attack-risk-in-men-what-to-know (accessed on 15 August 2021).
35. Cai, W. Fatigue Can Cause a Heart Attack, Doctors Say. 23 April 2021. Available online: https://www.shine.cn/news/metro/2104237896/ (accessed on 15 August 2021).
36. Gafarov, V.; Gromova, E.; Panov, D.; Gagulin, I.; Gafarova, A. Vital exhaustion and risk of myocardial infarction in male population aged 25–64 years in RussiaSiberia. Epidemiological program WHO Monica-psychosocial. *Eur. Heart J. Acute Cardiovasc. Care* **2021**, *10*, zuab020-224. [CrossRef]
37. Fairclough, S.H.; Houston, K. A metabolic measure of mental effort. *Biol. Psychol.* **2004**, *66*, 177–190. [CrossRef] [PubMed]
38. Dworkin, H.J.; Lawrie, C.; Bohdiewicz, P.; Lerner, A.M. Abnormal left ventricular myocardial dynamics in eleven patients with chronic fatigue syndrome. *Clin. Nucl. Med.* **1994**, *19*, 675–677. [CrossRef] [PubMed]
39. Lerner, A.M.; Zervos, M.; Chang, C.H.; Beqaj, S.; Goldstein, J.; O'Neill, W.; Dworkin, H.; Fitgerald, T.; Deeter, R.G. A small, randomized, placebo-controlled trial of the use of antiviral therapy for patients with chronic fatigue syndrome. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **2001**, *32*, 1657–1658. [CrossRef] [PubMed]
40. Peckerman, A.; LaManca, J.J.; Dahl, K.A.; Chemitiganti, R.; Qureishi, B.; Natelson, B.H. Abnormal impedance cardiography predicts symptom severity in chronic fatigue syndrome. *Am. J. Med. Sci.* **2003**, *326*, 55–60. [CrossRef] [PubMed]
41. Peckerman, A.; LaManca, J.J.; Smith, S.L.; Taylor, A.; Tiersky, L.; Pollet, C.; Korn, L.R.; Hurwitz, B.E.; Ottenweller, J.E.; Natelson, B.H. Cardiovascular stress responses and their relation to symptoms in Gulf War veterans with fatiguing illness. *Psychosom. Med.* **2000**, *62*, 509–516. [CrossRef] [PubMed]
42. Choi, J.B.; Nelesen, R.; Loredo, J.S.; Mills, P.J.; Ancoli-Israel, S.; Ziegler, M.G.; Dimsdale, J.E. Sleepiness in obstructive sleep apnea: A harbinger of impaired cardiac function? *Sleep* **2006**, *29*, 1531–1536. [CrossRef] [PubMed]
43. European Society of Cardiology. Exhaustion linked with increased risk of heart attack in men. ScienceDaily. 13 March 2021. Available online: www.sciencedaily.com/releases/2021/03/210313151926.htm (accessed on 10 June 2021).
44. Caruso, C.C.; Hitchcock, E.M.; Dick, R.B.; Russo, J.M.; Schmit, J.M. *Overtime and Extended Work Shifts; Recent Findings on Illnesses, Injuries, and Health Behaviors*; Department of Health and Human Services National Institute for Occupational Safety and Health: Washington, DC, USA, 2004.
45. Liu, Y.; Tanaka, H. Overtime work, insufficient sleep, and risk of non-fatal acute myocardial infarction in Japanese men. *Occup. Environ. Med.* **2002**, *59*, 447–451. [CrossRef]
46. Hayashi, T.; Kobayashi, Y.; Yamaoka, K.; Yano, E. Effect of overtime work on 24-hour ambulatory blood pressure. *J. Occup. Environ. Med.* **1996**, *38*, 1007–1011. [CrossRef]
47. Iwasaki, K.; Sasaki, T.; Oka, T.; Hisanaga, N. Effect of working hours on biological functions related to cardiovascular system among salesmen in a machinery manufacturing company. *Ind. Health* **1998**, *36*, 361–367. [CrossRef] [PubMed]
48. Rhoads, J.M. Overwork. *JAMA* **1977**, *237*, 2615–2618. [CrossRef] [PubMed]
49. Kageyama, T.; Nishikido, N.; Kobayashi, T.; Kurokawa, Y.; Kabuto, M. Commuting, overtime, and cardiac autonomic activity in Tokyo. *Lancet* **1997**, *350*, 639. [CrossRef]
50. Tochikubo, O.; Ikeda, A.; Miyajima, E.; Ishii, M. Effects of insufficient sleep on blood pressure monitored by a new multibiomedical recorder. *Hypertension* **1996**, *27*, 1318–1324. [CrossRef] [PubMed]
51. Sesoko, S.; Akema, N.; Matsukawa, T.; Kaneko, Y. Predisposing factors for the development of malignant essential hypertension. *Arch. Intern. Med.* **1987**, *147*, 1721–1724. [CrossRef]
52. Theorell, T.; Karasek, R.A. Current issues relating to psychosocial job strain and cardiovascular disease research. *J. Occup. Health Psychol.* **1996**, *1*, 9. [CrossRef] [PubMed]
53. Anwer, S.; Li, H.; Antwi-Afari, M.F.; Umer, W.; Wong, A.Y. Evaluation of Physiological Metrics as Real-Time Measurement of Physical Fatigue in Construction Workers: State-of-the-Art Review. *J. Constr. Eng.-Manag.-Asce* **2021**, *147*, 03121001. [CrossRef]
54. Lee, K.A.; Hicks, G.; Nino-Murcia, G. Validity and reliability of a scale to assess fatigue. *Psychiatry Res.* **1991**, *36*, 291–298. [CrossRef]
55. Chalder, T.; Berelowitz, G.; Pawlikowska, T.; Watts, L.; Wessely, S.; Wright, D.; Wallace, E.P. Development of a fatigue scale. *J. Psychosom. Res.* **1993**, *37*, 147–153. [CrossRef]
56. Williamson, A.; Lombardi, D.A.; Folkard, S.; Stutts, J.; Courtney, T.K.; Connor, J.L. The link between fatigue and safety. *Accid Anal. Prev.* **2011**, *43*, 498–515. [CrossRef]
57. Spencer, M.B.; Robertson, K.A.; Folkard, S. *The Development of a Fatigue Risk Index for Shiftworkers*; Research Report 446; Health Safety Executive: London, UK, 2006.
58. Pichot, V.; Bourin, E.; Roche, F.; Garet, M.; Gaspoz, J.M.; Duverney, D.; Barthélémy, J.C. Quantification of cumulated physical fatigue at the workplace. *PflüGers Arch.* **2002**, *445*, 267–272. [CrossRef] [PubMed]
59. Vicente, J.; Laguna, P.; Bartra, A.; Bailón, R. Drowsiness detection using heart rate variability. *Med. Biol. Eng. Comput.* **2016**, *54*, 927–937. [CrossRef] [PubMed]
60. Desmond, P.A.; Neubauer, M.C.; Matthews, G.; Hancock, P.A. (Eds.) *The Handbook of Operator Fatigue*; Ashgate Publishing, Ltd.: London, UK, 2012.
61. Lennartsson, A.K.; Jonsdottir, I.; Sjörs, A. Low heart rate variability in patients with clinical burnout. *Int. J. Psychophysiol.* **2016**, *110*, 171–178. [CrossRef] [PubMed]

62. Joo, S.; Choi, K.J.; Huh, S.J. Prediction of ventricular tachycardia by a neural network using parameters of heart rate variability. In Proceedings of the 2010 Computing in Cardiology, Belfast, UK, 26–29 September 2010; pp. 585–588.
63. Ramirez-Villegas, J.F.; Lam-Espinosa, E.; Ramirez-Moreno, D.F.; Calvo-Echeverry, P.C.; Agredo-Rodriguez, W. Heart rate variability dynamics for the prognosis of cardiovascular risk. *PLoS ONE* **2011**, *6*, e17060.
64. Song, T.; Qu, X.F.; Zhang, Y.T.; Cao, W.; Han, B.H.; Li, Y.; Da Cheng, H. Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. *BMC Cardiovasc. Disord.* **2014**, *14*, 1–8. [CrossRef] [PubMed]
65. Ebrahimzadeh, E.; Pooyan, M.; Bijar, A. A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. *PLoS ONE* **2014**, *9*, e81896. [CrossRef] [PubMed]
66. Murukesan, L.; Murugappan, M.; Iqbal, M.; Saravanan, K. Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features. *J. Med. Imaging Health Inform.* **2014**, *4*, 521–532. [CrossRef]
67. Melillo, P.; Izzo, R.; Orrico, A.; Scala, P.; Attanasio, M.; Mirra, M.; Pecchia, L. Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. *PLoS ONE* **2015**, *10*, e0118504. [CrossRef] [PubMed]
68. Lee, H.; Shin, S.Y.; Seo, M.; Nam, G.B.; Joo, S. Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks. *Sci. Rep.* **2016**, *6*, 1–7. [CrossRef]
69. Lan, K.C.; Raknim, P.; Kao, W.F.; Huang, J.H. Toward hypertension prediction based on PPG-derived HRV signals: A feasibility study. *J. Med. Syst.* **2018**, *42*, 1–7. [CrossRef] [PubMed]
70. Gafarov, V.; Gromova, E.; Panov, D.; Gagulin, I.; Gafarova, A.; Krymov, E. Stress at work enhances risk of arterial hypertension in general population. who monica-psychosocial program. *J. Hypertens.* **2021**, *39*, e160. [CrossRef]
71. Shaffer, F.; Ginsberg, J.P. An overview of heart rate variability metrics and norms. *Front. Public Health* **2017**, *5*, 258. [CrossRef]
72. Kim, K.J.; Shin, D.H. An acceptance model for smart watches: Implications for the adoption of future wearable technology. *Internet Res.* **2015**, *25*, 527–541. [CrossRef]
73. Perera, C.; Vasilakos, A.V. A knowledge-based resource discovery for Internet of Things. *Knowl.-Based Syst.* **2016**, *109*, 122–136. [CrossRef]
74. Liu, L.; Peng, Y.; Liu, M.; Huang, Z. Sensor-based human activity recognition system with a multilayered model using time series shapelets. *Knowl.-Based Syst.* **2015**, *90*, 138–152. [CrossRef]
75. Park, E.; Kim, K.J.; Kwon, S.J. Understanding the emergence of wearable devices as next-generation tools for health communication. *Inf. Technol. People* **2016**, *29*, 717–732. [CrossRef]
76. Turing, A. Computing machinery and intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
77. Niknejad, N.; Ismail, W.B.; Mardani, A.; Liao, H.; Ghani, I. A comprehensive overview of smart wearables: The state of the art literature, recent advances, and future challenges. *Eng. Appl. Artif. Intell.* **2020**, *90*, 103529. [CrossRef]
78. Kortuem, G.; Kawsar, F.; Sundramoorthy, V.; Fitton, D. Smart objects as building blocks for the internet of things. *IEEE Internet Comput.* **2009**, *14*, 44–51. [CrossRef]
79. Cheng, J.W.; Mitomo, H. The underlying factors of the perceived usefulness of using smart wearable devices for disaster applications. *Telemat. Inform.* **2017**, *34*, 528–539. [CrossRef]
80. Poslad, S. *Ubiquitous Computing: Smart Devices, Environments and Interactions*; John Wiley Sons: Hoboken, NJ, USA, 2011.
81. Jeong, S.C.; Kim, S.H.; Park, J.Y.; Choi, B. Domain-specific innovativeness and new product adoption: A case of wearable devices. *Telemat. Inform.* **2017**, *34*, 399–412. [CrossRef]
82. Ashton, K. That 'internet of things' thing. *RFID J.* **2009**, *22*, 97–114.
83. Fernandez, P. *Wearable Technology: Beyond Augmented Reality*; Library Hi Tech News: Bingley, UK, 2014.
84. Thorp, E.O. The invention of the first wearable computer. In Digest of Papers. In Proceedings of the Second International Symposium on Wearable Computers (Cat. No. 98EX215), Pittsburgh, PA, USA, 19–20 October 1998; pp. 4–8.
85. Mann, S. Smart clothing: The shift to wearable computing. *Commun. ACM* **1996**,*39*, 23–24. [CrossRef]
86. Park, S.; Jayaraman, S. Smart textiles: Wearable electronic systems. *MRS Bull.* **2003**, *28*, 585–591. [CrossRef]
87. Wright, R.; Keith, L. Wearable technology: If the tech fits, wear it. *J. Electron. Resour. Med. Libr.* **2014**, *11*, 204–216. [CrossRef]
88. Dimou, E.; Manavis, A.; Papachristou, E.; Kyratsis, P. A conceptual design of intelligent shoes for pregnant women. In Proceedings of the Workshop on Business Models and ICT Technologies for the Fashion Supply Chain, Florence, Italy, 20–22 April 2016; Springer: Cham, Switzerland; pp. 69–77.
89. Yang, H.; Yu, J.; Zo, H.; Choi, M. User acceptance of wearable devices: An extended perspective of perceived value. *Telemat. Inform.* **2016**, *33*, 256–269. [CrossRef]
90. Steinhubl, S.R.; Topol, E.J. Moving From Digitalization to Digitization in Cardiovascular Care: Why Is it Important, and What Could it Mean for Patients and Providers? *J. Am. Coll. Cardiol.* **2015**, *66*, 1489–1496. [CrossRef]
91. Boeldt, D.L.; Wineinger, N.E.; Waalen, J.; Gollamudi, S.; Grossberg, A.; Steinhubl, S.R.; McCollister-Slipp, A.; Rogers, M.A.; Silvers, C.; Topol, E.J. How Consumers and Physicians View New Medical Technology: Comparative Survey. *J. Med. Internet Res.* **2015**, *17*, e215. [CrossRef]
92. Yang, L.; Lu, K.; Diaz-Olivares, J.A.; Seoane, F.; Lindecrantz, K.; Forsman, M.; Eklund, J.A. Towards smart work clothing for automatic risk assessment of physical workload. *IEEE Access* **2018**, *6*, 40059–40072. [CrossRef]
93. Sedighi Maman, Z.; Alamdar Yazdi, M.A.; Cavuoto, L.A.; Megahed, F.M. A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Appl. Ergon.* **2017**, *65*, 515–529. [CrossRef] [PubMed]

94. Bowen, J.; Hinze, A.; König, J.; Exton, D. Supporting safer work practice through the use of wearable technology. In *Ergonomics and Human Factors*; CIEHF: Birmingham, UK, 2021; Volume 117.
95. Yu, Y.; Li, H.; Yang, X.; Kong, L.; Luo, X.; Wong, A.Y. An automatic and non-invasive physical fatigue assessment method for construction workers. *Autom. Constr.* **2019**, *103*, 1–12. [CrossRef]
96. Hajifar, S.; Sun, H.; Megahed, F.M.; Jones-Farmer, L.A.; Rashedi, E.; Cavuoto, L.A. A forecasting framework for predicting perceived fatigue: Using time series methods to forecast ratings of perceived exertion with features from wearable sensors. *Appl. Ergon.* **2021**, *90*, 103262. [CrossRef] [PubMed]
97. Mao, L.; Gong, T.; Ai, Q.; Hong, Y.; Guo, J.; He, Y.; Yu, B. Morphologically modulated laser-patterned reduced graphene oxide strain sensors for human fatigue recognition. *Smart Mater. Struct.* **2019**, *29*, 015009. [CrossRef]
98. Horiuchi, R.; Ogasawara, T.; Miki, N. Fatigue assessment by blink detected with attachable optical sensors of dye-sensitized photovoltaic cells. *Micromachines* **2018**, *9*, 310. [CrossRef] [PubMed]
99. Li, P.; Meziane, R.; Otis, M.J.D.; Ezzaidi, H.; Cardou, P. A Smart Safety Helmet using IMU and EEG sensors for worker fatigue detection. In Proceedings of the 2014 IEEE International Symposium on Robotic and Sensors Environments (ROSE) Proceedings, Timisoara, Romania, 16–18 October 2014; pp. 55–60.
100. Ríos Aguilar, S.; Miguel Merino, J.L.; Millan Sanchez, A.; Sanchez Valdivieso, A. Variation of the Heartbeat and Activity as an Indicator of Drowsiness at the Wheel Using a Smartwatch. *Int. J. Interact. Multimed. Artif. Intell.* **2015**, *3*, 96. [CrossRef]
101. Darbandy, M.T.; Rostamnezhad, M.; Hussain, S.; Khosravi, A.; Nahavandi, S.; Sani, Z.A. A new approach to detect the physical fatigue utilizing heart rate signals. *Res. Cardiovasc. Med.* **2020**, *9*, 23.
102. Bhatt, C.; Kumar, I.; Vijayakumar, V.; Singh, K.U.; Kumar, A. The state of the art of deep learning models in medical science and their challenges. *Multimed. Syst.* **2021**, *27*, 599–613. [CrossRef]
103. Bini, S.A. Artificial intelligence, machine learning, deep learning, and cognitive computing: What do these terms mean and how will they impact health care? *J. Arthroplast.* **2018**, *33*, 2358–2361. [CrossRef]
104. Collobert, Ronan, and Samy Bengio. SVMTorch: Support vector machines for large-scale regression problems.*J. Mach. Learn. Res.* **2001**, *1*, 143–160.
105. Salakhutdinov, R.; Hinton, G. Deep boltzmann machines. In Proceedings of the Artificial Intelligence and Statistics, Clearwater Beach, FL, USA, 16–18 April 2009; pp. 448–455.
106. Arel, I.; Rose, D.C.; Karnowski, T.P. Deep machine learning-a new frontier in artificial intelligence research [research frontier]. *IEEE Comput. Intell. Mag.* **2010**, *5*, 13–18. [CrossRef]
107. Sukittanon, S.; Surendran, A.C.; Platt, J.C.; Burges, C.J. Convolutional networks for speech detection. In Proceedings of the Eighth International Conference on Spoken Language Processing, Jeju Island, Korea, 4–8 October 2004.
108. Rizk, Y.; Hajj, N.; Mitri, N.; Awad, M. Deep belief networks and cortical algorithms: A comparative study for supervised classification. *Appl. Comput. Inform.* **2019**, *15*, 81–93. [CrossRef]
109. Dauphin, G.M.Y.; Glorot, X.; Rifai, S.; Bengio, Y.; Goodfellow, I.; Lavoie, E.; Bergstra, J. Unsupervised and transfer learning challenge: A deep learning approach. In Proceedings of the ICML Workshop on Unsupervised and Transfer Learning, Edinburgh, Scotland, 26 June–1 July 2012; pp. 97–110.
110. Bengio, Y.; Courville, A.C.; Vincent, P. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR* **2012**, *1*, abs/1206.5538.
111. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: Review, opportunities and challenges. *Briefings Bioinform.* **2018**, *19*, 1236–1246. [CrossRef]
112. Peng, C.Y.J.; Lee, K.L.; Ingersoll, G.M. An introduction to logistic regression analysis and reporting. *J. Educ. Res.* **2002**, *96*, 3–14. [CrossRef]
113. Algamal, Z.Y.; Lee, M.H. Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Syst. Appl.* **2015**, *42*, 9326–9332. [CrossRef]
114. Olive, D.J. Multiple linear regression. In *Linear Regression*; Springer: Cham, Switzerland, 2017; pp. 17–83.
115. Domingos, P.; Pazzani, M.On the optimality of the simple Bayesian classifier under zero-one loss. *Mach. Learn.* **1997**, *29*, 103–130. [CrossRef]
116. Chatfield, C. *Time-Series Forecasting*; CRC Press: Boca Raton, FL, USA, 2000.
117. Enders, W. *Applied Econometric Time Series*; John Wiley Sons: Hoboken, NJ, USA, 2008.
118. Cochran, W.T.; Cooley, J.W.; Favin, D.L.; Helms, H.D.; Kaenel, R.A.; Lang, W.W.; Welch, P.D. What is the fast Fourier transform? *Proc. IEEE* **1967**, *55*, 1664–1674. [CrossRef]
119. Kramer, O. K-nearest neighbors. In *Dimensionality Reduction with Unsupervised Nearest Neighbors*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 13–23.
120. Sudden Cardiac Death Holter Database v1.0.0. *PhysioNet*, 2 July 2004. Available online: https://physionet.org/content/sddb/1.0.0/ (accessed on 1 November 2021).
121. MIT-BIH Normal Sinus Rhythm Database v1.0.0. *PhysioNet*, 2 August 1999. Available online: https://physionet.org/content/nsrdb/1.0.0/ (accessed on 1 November 2021).
122. Smart Health for Assessing the Risk of Events via ECG Database v1.0.0. *PhysioNet*, 19 May 2015. Available online: https://physionet.org/content/shareedb/1.0.0/ (accessed on 1 November 2021).

123. Tunstall-Pedoe, H. (Ed.) *MONICA, Monograph and Multimedia Sourcebook: World's Largest Study of Heart Disease, Stroke, Risk Factors, and Population Trends 1979–2002*; World Health Organization: Geneva, Switzerland, 2003.
124. Tsao, C.W.; Aday, A.W.; Almarzooq, Z.I.; Alonso, A.; Beaton, A.Z.; Bittencourt, M.S.; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics—2022 Update: A Report From the American Heart Association. *Circulation* **2022**, *145*, e153–e639. [CrossRef]
125. Ahn, C.R.; Lee, S.; Sun, C.; Jebelli, H.; Yang, K.; Choi, B. Wearable sensing technology applications in construction safety and health. *J. Constr. Eng. Manag.* **2019**, *145*, 03119007. [CrossRef]
126. Albrecht, J.P. How the GDPR will change the world. *Eur. Data Prot. Law Rev.* **2016**, *2*, 287. [CrossRef]
127. Parasol, M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams. *Comput. Law Secur. Rev.* **2018**, *34*, 67–98. [CrossRef]
128. Gray, W.; Zheng, H.R. General Principles of Civil Law of the People's Republic of China. *Am. J. Comp. Law* **1986**, *34*, 715–743. [CrossRef]
129. Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li, W.; Gao, Y. A survey on federated learning. *Knowl.-Based Syst.* **2021**, *216*, 106775. [CrossRef]
130. Mammen, P.M. Federated learning: Opportunities and challenges. *arXiv* **2021**, arXiv:2101.05428.
131. Zhang, K.; Song, X.; Zhang, C.; Yu, S. Challenges and future directions of secure federated learning: A survey. *Front. Comput. Sci.* **2022**, *16*, 1–8. [CrossRef]
132. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol. (TIST)* **2019**, *10*, 1–19. [CrossRef]
133. Li, Q.; Wen, Z.; Wu, Z.; Hu, S.; Wang, N.; Li, Y.; He, B. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. *IEEE Trans. Knowl. Data Eng.* **2021**. [CrossRef]
134. Li, L.; Fan, Y.; Tse, M.; Lin, K.Y. A review of applications in federated learning. *Comput. Ind. Eng.* **2020**, *149*, 106854. [CrossRef]
135. De Luca, C.J.; Gilmore, L.D.; Kuznetsov, M.; Roy, S.H. Filtering the surface EMG signal: Movement artifact and baseline noise contamination. *J. Biomech.* **2010**, *43*, 1573–1579. [CrossRef]
136. Islam, M.K.; Rastegarnia, A.; Sanei, S. Signal Artifacts and Techniques for Artifacts and Noise Removal. In *Signal Processing Techniques for Computational Health Informatics*; Springer: Cham, Switzerland, 2021; pp. 23–79.
137. Gafarov, V.V.; Panov, D.O.; Gromova, E.A.; Gagulin, I.V.; Gafarova, A.V.; Krymov, E.A. *Sex Differences in Long-Term Trends of Psychosocial Factors and Gender Effect on Risk of Cardiovascular Diseases: Arterial Hypertension, Myocardial Infarction and Stroke*; IntechOpen: London, UK, 2021.
138. Iriarte, J.; Urrestarazu, E.; Valencia, M.; Alegre, M.; Malanda, A.; Viteri, C.; Artieda, J. Independent component analysis as a tool to eliminate artifacts in EEG: A quantitative study. *J. Clin. Neurophysiol.* **2003**, *20*, 249–257. [CrossRef]
139. Ram, M.R.; Madhav, K.V.; Krishna, E.H.; Komalla, N.R.; Reddy, K.A. A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. *IEEE Trans. Instrum. Meas.* **2011**, *61*, 1445–1457. [CrossRef]
140. Daly, I.; Billinger, M.; Scherer, R.; Müller-Putz, G. On the automated removal of artifacts related to head movement from the EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 427–434. [CrossRef] [PubMed]
141. Ramachandram, D.; Taylor, G.W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process. Mag.* **2017**, *34*, 96–108. [CrossRef]
142. Kline, A.; Wang, H.; Li, Y.; Dennis, S.; Hutch, M.; Xu, Z.; Luo, Y. Multimodal Machine Learning in Precision Health. *arXiv* **2022**, arXiv:2204.04777.

A.2 Reviewing Digital Parameters Used in Fatigue Detection

During my research journey into the realm of Occupational Physical Fatigue detection, a unique opportunity emerged to contribute further to the field. I observed a conspicuous gap in the existing literature—an absence of comprehensive reviews pertaining to the digital parameters and devices employed in tracking and predicting fatigue. Recognizing the significance of this uncharted territory, I was motivated to craft a succinct yet informative review that addresses this crucial aspect of fatigue management. This additional review serves as an invaluable resource, shedding light on the digital parameters and devices so far overlooked, and thereby enriching the discourse on Occupational Physical Fatigue detection. Later, this article was published in the proceedings of the IHSH'2022 conference:

- **Detection of Occupational Fatigue in Digital Era; Parameters In Use** *(IHSH'2022 Conference)*

2022 3rd International Conference on Human-Centric Smart Environments for Health and Well-being (IHSH) | 978-1-6654-6321-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IHSH57076.2022.10092165

Detection of Occupational Fatigue in Digital Era; Parameters In Use

Mohammad Moshawrab
Département de mathématiques, informatique et génie
Université du Québec à Rimouski
Rimouski, Canada
mohammad.moshawrab@uqar.ca

Mehdi Adda
Département de mathématiques, informatique et génie
Université du Québec à Rimouski
Rimouski, Canada
mehdi_adda@uqar.ca

Abdenour Bouzouane
département D'informatique et de Mathématique
Université du Québec à Chicoutimi
Chicoutimi, Canada
abdenour_bouzouane@uqac.ca

Hussein Ibrahim
Institut Technologique de Maintenance Industrielle
Sept-Îles, Canada
hussein.ibrahim@itmi.ca

Ali Raad
Faculty of Arts & Sciences
Islamic University of Lebanon
Wardaniyeh, Lebanon
ali.raad@iul.edu.lb

***Abstract*—Workplace fatigue is common and widespread worldwide and in various occupations. This fact is a result of changing working conditions such as changing shifts, increasing work pressure, circadian rhythm disturbances, and other factors that are common in many industries. Since fatigue has serious implications for workers' health on the one hand and for companies' productivity on the other, its detection is of great importance to all stakeholders. Accordingly, the literature presented in this paper has been studied in depth to investigate the different parameters that can be used to detect fatigue in the workplace. In addition, the tools used to detect these parameters are discussed to help researchers select the best combination of parameters and tools to detect fatigue in the workplace.**

***Index Terms*—Smart Health, Workplace Fatigue, Fatigue Detection, Diseases Prevention, Parameters, Smart Wearables**

I. Introduction

The development and growth of the tools and techniques that surround us have greatly changed our lives. Indeed, some habits have changed accordingly, such as educational routines, medical services, work and job concepts, entertainment and much more. Nevertheless, new concepts were also introduced into our knowledge, such as the "24/7" active society, which increased the demand for higher productivity, thus increasing time pressure and work intensity. Increased productivity led to longer working hours, which lengthened the average workday and shortened the average recovery time. In addition, increased productivity led to higher work demands, circadian rhythm disruptions, social and societal demands, and inadequate sleep. All of these changes have resulted in fatigue becoming a significant problem in modern society. Psychosocial stress and sleep deprivation associated with fatigue are the main consequences of increased work intensity [1:3].

A. Fatigue Definition(s)

Despite its prevalence, severity, and intense research, there is no single definition for the term fatigue [4]. For example, the authors in [5] defined fatigue as a state that fluctuates between wakefulness and sleepiness. In addition, in [6], fatigue is defined as a state of the muscles and central nervous system in which prolonged physical activity, in the absence of adequate rest, results in insufficient ability or energy to maintain the original level of activity. Regardless of the different definitions, all agree that fatigue is related to or is itself a lack of activity and motivation. Moreover, researchers distinguish between acute and chronic fatigue and classify acute fatigue into different types, such as occupational physical fatigue, occupational mental fatigue, occupational heat stress, occupational noise stress and others. Thus, occupational fatigue is described as work-related fatigue due to various causes, which can be divided into two groups: work-related and individual-related causes and contributors [6:8]. In this article, the parameters used to detect or predict fatigue are listed and discussed.

B. Fatigue Consequences

Acute fatigue is considered normal, but it can become pathological if it persists and leads to deterioration of health. Therefore, it is important to track fatigue because people may not correctly assess their level of fatigue, and even physicians

may erroneously conclude during routine examinations that fatigue is not severe and cannot lead to a specific disease [6].

1) Health Risks: Fatigue in the workplace is common in almost all sectors of the economy. In addition, studies have shown that persistent fatigue has various health consequences, such as weakening of the immune system, musculoskeletal injuries, and the development of chronic fatigue syndrome [9:11]. In addition, the health consequences may also result in more serious diseases, such as: Cardiovascular diseases, cancer, type 2 diabetes, infectious diseases, obesity, gastrointestinal diseases and reproductive problems [12,13].

2) Productivity Minimizing: In addition, fatigue has an impact on worker productivity. Numerous studies have examined that fatigue is negatively related to productivity levels, showing that a reduction in fatigue is associated with improved daily activity and work productivity. In addition, early detection and treatment of fatigue has been shown to improve productivity on the one hand. On the other hand, improving working conditions is a key to reducing musculoskeletal problems and fatigue and thus increasing productivity [14,15]. Figure 1 below shows a brief summary of the costs incurred by companies due to fatigue. The data in the figure were provided by [16].

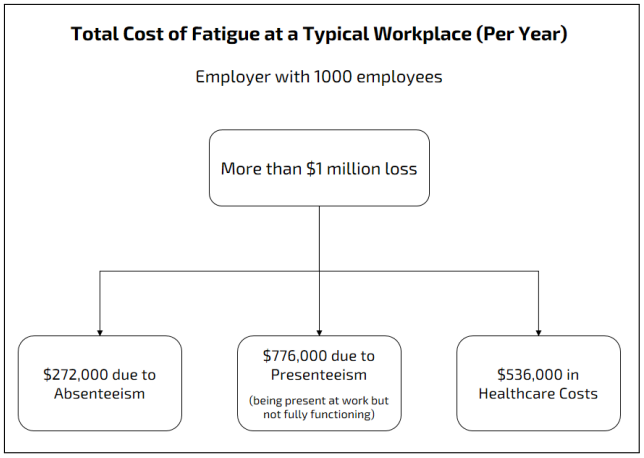

Fig. 1. Diseases caused by Occupational Fatigue

II. Fatigue Detection in Digital Era

Fatigue in the workplace is considered normal, but its persistence can be a dangerous alarm signal of a critical health condition. Therefore, tools and methods to detect and measure fatigue are not new concepts, and several attempts have been made for this purpose, starting with subjective questionnaires developed in the early 1990s to quantify physical fatigue, such as: McCorkle and Young's Symptom Distress Scale and Rhoten's Fatigue Scale and Fatigue Observation Checklist [17,18]. However, since there is no standard scale to assess fatigue, it is impossible to compare the results of different studies. In addition, subjective questionnaires are considered inexpensive and efficient tools for detecting and quantifying fatigue, but they are subject to recall errors, are considered intrusive, and cannot detect fatigue in real time. To overcome all the above limitations of questionnaires, researchers have attempted to record and analyze various vital signs to detect the presence of fatigue.

A. From Subjective Questionnaires to Digitization

The need to monitor occupational fatigue in real time requires monitoring of some vital signs and biomarkers. Given the serious consequences of occupational fatigue, rigorous and effective medical intervention is needed, regardless of the causes, burdens, costs, and effects. In addition, the rapid development of information and communication technologies combined with the improvement of microprocessors has given rise to a new generation of tiny, robust, and efficient computing devices, such as smart wearables, also known as smart wearable technology or wearable devices. These devices provide anytime, anywhere access to data and are being heralded as the next generation of ubiquitous technology after smartphones [19]. The improvement of smart wearables along with Artificial Intelligence (AI) and Machine Learning (ML) models has promoted their use in tracking and analyzing vital signs and thus in detecting fatigue.

B. Fatigue Detection Parameters

Literature indicates that several parameters can be used to detect fatigue in the workplace. Nevertheless, there are no references that list all the parameters that can be used in the detection of fatigue. In this section, the parameters used in the literature to detect fatigue are presented, discussed, and explained.

1) Common Fatigue Indicators: The need for accurate detection of fatigue at work requires monitoring of some vital signs such as Heart Rate, Heart Rate Variability (HRV), Electroencephalogram electroencephalogram (EEG), Jerk metrics, and others [20]. However, some studies discuss the effect of fatigue on simple biomarkers such as heart rate, necessitating the use of other vital signs for accurate results [6]. Because EEG is measured with equipment that can be considered invasive and immobilizing, other alternatives are needed to detect fatigue without restricting worker movement, such as Nocturnal Autonomic Nervous System Activity (ANS) monitored from heart rate variability, movement, and sleep data [21]. Subsequently, heart rate variability and motion are the most commonly used indicators of fatigue in the literature.

2) Further Fatigue Detection Parameters: Fatigue in the workplace has been studied extensively by researchers recently, and numerous implementations have been made to detect it. Whether using smart wearables or other devices, many parameters can be used to detect or even predict fatigue. Therefore, an in-depth analysis of workplace fatigue detection implementations provides a clear idea of almost all parameters that can be used to detect fatigue. A thorough literature review shows that fatigue detection parameters can be classified into four main categories, namely body movement, work-related information, personal information, and vital signs [3,5,6,20:63]. However, further details are discussed below, with each category broken down and explained with relevant examples.

- Body Movement
  - Activity
    - Body Move Acceleration [23,59]
    - Manual Material Handling [25]
    - Rating of Perceived Exertion [25,42]
  - Gait
    - Changes in posture/gait [29,30]
    - Force variability [30]
    - Gait Parameters [25,60]
    - Head Gestures [28]
    - Higher risk of slips, trips, and falls [31,32]
    - Joint Angles Movement [33,34,36]
    - Posture and Exerted Force [35]
    - Stride Length, Height, Width and Duration [36]
    - Tremor [37]
    - Working Postures (standing up, back bending, squatting) [38]
  - Motion
    - Accelerations and inclination angles [23]
    - Jerk Metrics [43]
    - Leg Motion [22]
    - Body Motion [21,24]
    - Movement Variability [23]
    - Walking - Squating - Walking Cycle [25]
- Job Related
  - Working Environment
    - Duration of Continuous Work on One Task [45]
    - Equipment Malfunction [45]
    - Hazardous Materials [45]
    - Noise and other Distractions [45]
    - Team Make-Up and Group Size [45]
    - Environment Temperature & Humidity [40,42,44,45]
    - Tools & Standard Operating Procedures [45]
    - Type of Work [5,46]
    - Worker Qualifications and Training [45]
    - Workload and Time Pressure [45]
  - Working Shifts
    - Breaks During Work [45]
    - Circadian Adjustment [47]
    - Consecutive Shifts [48]
    - Overtime Work [3]
    - Shift Rotation [46,47]
    - Shift Type [46,48]
    - Total Work Time [42,46]
    - Recovery Time [42,46]
- Personal Information
  - Gender
    - Gender [49]
  - Lifestyle
    - Care Giving Duties [45]
    - Hobbies [45]
    - Disengagement From Work [45]
    - Parents with Infant Children [45]
    - Social Factors [45]
    - Traffic and Commuting Times [45]
  - Medical History: Acute/Chronic Diseases [50]
- Vital Signs
  - Autonomic Nervous System (ANS)
    - ANS activity [51]
  - Brain Signals
    - Electroencephalography (EEG) [21,28,52,53]
  - Circadian Process
    - Circadian rhythm [61]
    - A Dark and Quiet Bedroom [45]
    - Age-Related Changes in Circadian Function [54]
    - Circadian Fluctuations Amplitude [54]
    - Circadian Adaptability to Altered External Time Cues [54]
    - Circadian & Homeostatic Process [45,54]
    - Multiple Sleep Latency Test (MLST) [21]
    - Obstructive Sleep Apnea (OSA) [55]
    - Sleep Need [54]
    - Sleep Recovery Speed [54]
  - Eye Signals
    - Eyeblinks [26,27]
  - Heart Rate & HRV
    - Heart Rate [20,22:24,29,30,40,41,42,44,56,62]
    - Heart Rate Variability [20,21,24,57]
  - Inner Vital Signs
    - Body Temperature [56]
    - Breathing Rate [41]
    - Thermoregulatory measures [20]
  - Muscles Signals
    - Electromyography (EMG) [20,34,53]
    - Surface Electromyographic (sEMG) [39,43]
  - Other Heart Parameters
    - Cardiac Index [6]
    - Stroke Index [6]
  - Skin Signals
    - Galvanic Skin Response (GSR) [24,63]
    - Skin Humidity [56]
    - Skin Temperature [20,41]

The variety of parameters listed above reflects the variety of possibilities and the variety of ways that can be used to detect fatigue at work, which makes its detection more feasible and accurate, provided that the best parameters are chosen. The information shows that vital signs, especially heart rate and HRV, and body movement are the most commonly used parameters to detect fatigue. However, it also shows that some parameters that are considered unimportant can lead to fatigue at work. Having a young child or sleeping in a bright and noisy bedroom may not seem that significant at first glance when it comes to fatigue. However, the results shown here are contradictory, as several negligible parameters can be tracked to detect or even predict fatigue.

III. Parameters Classification

In the previous section, the parameters were grouped based on their types. However, the classification can be made according to different aspects and points of view. For example, the equipment used to detect the parameters, detectability, and other concepts are possible classification criteria for these parameters. In addition, the parameters can be categorized according to their criticality, where some can be considered crucial for fatigue detection and others secondary or complementary. This classification can help researchers in selecting parameters when developing real-world fatigue detection applications. In the following sections, some classification perspectives are listed and discussed.

A. Parameters Classification in Terms of Devices Used

Advances in electronics and microprocessors have led to the emergence of new generations of devices that can collect data. The precise, powerful, intelligent and sensory devices available today are used in many areas of our daily lives. Considering the variety of parameters that can be used to detect fatigue in the workplace, many devices and tools can be used for this purpose. The analysis of studies on fatigue and the parameters used to detect and predict fatigue provided a descriptive overview of the devices that can be used in this context. These devices are listed in Table 1 below, which shows where each device was used according to the studies mentioned therein.

The information in the table shows that smartwatches are the devices that can detect and record most parameters such as motion, gait, activity, circadian rhythm, Heart Rate and Heart Rate Variability, internal vital signs, and skin signals. This result shows the great potential of smartwatches for fatigue detection in the workplace. In addition, Inertial Measurement Unit (IMU) sensors and Heart Rate monitors are also widely used to detect motion and heart rate, respectively. This fact also favors the development of wearable devices for workplace fatigue detection, which is the case in most applications that combine both IMUs and heart rate monitors to develop customized smartwatches or other tools for detecting or predicting workplace fatigue, as the authors did in [21:23].

TABLE I
Detection of Parameters Per Device

Device	Body Movement	Job Related	Personal Info	Vital Signs
Impedance Cardiography	-	-	-	[6]
Smartphone Sensors	-	[21,38]	-	-
Inertial Measurement Unit (IMU) Sensors	[22,23,25, 26,28]	-	-	-
Electrocardiogram Electrodes (ECG)	-	-	-	[22,53]
Heart Rate Monitors (Shimmer Device, SparkFun Heart Rate Monitor, POLAR S810i™, Polar Vantage NV™)	-	-	-	[23,24,29,30 40,42,44]
GROVE Galvanic Skine Response Sensor	-	-	-	[24]
Accelerometer Sensors	[24,36]	-	-	-
RGB Cameras	[58]	-	-	-
Laser-Patterned Reduced Graphene Oxide (LPG) Sensor	-	-	-	[26]
Optical Sensors of Dye-Sensitized Photovoltaic Cells	-	-	-	[27]
Electroencephalogram Electordes (EEG)	-	-	-	[28]
Smartwatch	[60]	-	-	[61:63]
Video Cameras	[29]	-	-	-
Interviews & Questionnaires	[29]	[30,42]	[30,42]	[30,55]
MYOTON-3 Device	-	-	-	[29]
Gait Mat (GAITRite Platinum)	[32]	-	-	-
Inclinometers	[33]	-	-	-
Magnetic Field-Based Motion Tracking System	[34]	-	-	[34]
Motion Capture System	[43]	-	-	-
Physiological Status Monitors (PSMs)	-	-	-	[41]
Wireless Surface Electromyography (sEMG) System	-	-	-	[43]
Wet Bulb Globe Temperature (WBGT) Monitor	-	[44]	-	-
Polysomnographic Studies (PSGs)	-	-	-	[51]
MEG Recording Devices	-	-	-	[52]
Electromyography Electrodes (EMG)	-	-	-	[53]
Photoplethysmography (PPG) Sensor	-	-	-	[57]
RadioFrequency Sensors & WiFi Devices	[59]	-	-	-
Raspberry Pi Sensors	[60]	-	-	-

B. Classification in Terms of Criticality and Usefulness

Furthermore, the parameters can be classified according to their criticality and usefulness in detecting fatigue. The information in Tables 1 and 2 confirms that fatigue is mainly detected by vital signs, especially Heart Rate and Heart Rate Variability or by movement-related data such as gait, posture, and motion. However, some studies suggest combining both HRV and movement data for accurate and feasible detection [22].

In addition, other parameters, such as circadian rhythm data, appear to be important for fatigue detection but cannot be used alone as a fatigue indicator. Therefore, parameters can be divided into three groups in terms of criticality: critical, important, and secondary. The first group contains the parameters that can be used alone to detect fatigue, the second group contains parameters that are important for accurate detection but cannot be used alone for detection, and the third group contains the parameters that can increase the accuracy of detection without affecting the results if not used. The following list classifies the parameters from such a perspective:

- Critical Parameters
  - Body Movement
    - Action
    - Gait & Posture
    - Motion
  - Vital Signs
    - Autonomic Nervous System
    - Brain Signals
    - Eye Signals
    - Heart Rate & HRV
    - Muscles Signals
    - Other Heart Parameters
- Important Parameters
  - Job Related
    - Working Environments
    - Working Shifts
  - Vital Signs
    - Circadian Process
    - Inner Vital Signs
    - Skin Signals
- Secondary Parameters
  - Personal Information
    - Gender
    - Lifestyle
    - Medical History

C. Classification in Terms of detectability

In addition, parameters can be classified based on their detectability. For example, some parameters can be detected in real time, such as heart rate and body temperature, while others can be detected asynchronously, such as the occurrence of social or economic events. In addition, some parameters can be collected with digital devices, while others require questionnaires or surveys. Below is a list in which the parameters are ordered according to their recognizability:

- Real Time vs. Asynchronous Parameters:
  - Real Time Parameters
    - Body Movement
      - Activity
      - Gait & Postures
      - Motion
    - Vital Signs
      - Autonomic Nervous System
      - Brain Signals
      - Circadian Process
      - Eye Signals
      - Heart Rate & HRV
      - Inner Vital Signs
      - Muscles Signals
      - Other Heart Parameters
      - Skin Signals
  - Asynchronous Parameters
    - Job Related
      - Working Environment
      - Working Shifts
    - Personal Information
      - Gender
      - Lifestyle
      - Medical History
- Classification By detectability:
  - Collected by Digital Devices
    - Body Movement
      - Activity
      - Gait & Postures
      - Motion
    - Job Related
      - Working Environment
      - Working Shifts
    - Vital Signs
      - Autonomic Nervous System
      - Brain Signals
      - Circadian Process
      - Eye Signals
      - Heart Rate & HRV
      - Inner Vital Signs
      - Muscles Signals
      - Other Heart Parameters
      - Skin Signals
  - Collected by Questionnaires or Surveys
    - Job Related
      - Working Environment
      - Working Shifts
    - Personal Information
      - Gender
      - Lifestyle
      - Medical History

IV. Challenges & Future Opportunities

The implementation of fatigue detection in the workplace has made significant progress recently. However, there are still many challenges in this area. These obstacles are the subject of study and the interest of researchers to increase the level of tools and applications used. There are also great opportunities to improve fatigue detection, whether through the tools used or through predictive and proactive measures.

A. Challenges and Limitations

Although there are several parameters that can be used to detect fatigue, challenges can arise, either from the application of methods and tools for detection or from the nature of the parameters themselves. The following are the most common obstacles encountered in the detection of occupational fatigue [20,64]:

1) Parameters Selection: The wide variety of fatigue detection parameters increases the chances of tracking and monitoring fatigue in the workplace. However, it is not an easy task to collect all parameters at once as it requires different devices and may not be feasible due to irrelevance of data. Therefore, a combination of the best parameters should be selected considering the relationship between the data collected and the minimum number of devices to be used.

Based on the literature reviewed in this article, a combination of activity, gait, motion, and HRV would be a feasible and accurate combination to capture and analyze fatigue detection [22]. In addition, the collection of circadian process data along with work environment parameters would increase the accuracy of detection when added to the previous combination. Moreover, motion data, HRV, and circadian process data can be collected with a smartwatch, which minimizes the number of devices needed to collect such a set of parameters. Consequently, choosing the best combination of parameters is critical to improve the feasibility of fatigue detection solutions.

2) Data Usability; Artifacts and Noise: Some devices record fatigue parameters noninvasively, which makes the recordings more susceptible to many external sources of noise. Noisy data are also referred to as "Signal Artifacts". These are unwanted signals or signal distributions that interfere with the actual signal and are classified into two main groups: intrinsic and extrinsic artifacts [20,64]:

- **Intrinsic artifacts:** originate from the body being monitored
  - **Ocular Artifacts**: caused by the movement of the eyeball and interferes with EEG recording
  - **Muscle Artifacts**: arise from activities such as sniffing and swallowing
  - **Cardiac Artifacts**: noise due to the electrical activity of the heart
  - **Respiratory Artifacts**: caused by inhalation or exhalation while recording vital signs
- **Extrinsic Artifacts:** originate from the environment surrounding the body being monitored
  - **Motion Artifact**: motion while recording vital signs
  - **Environmental Artifact**: like interference with waves from other electronic devices

3) Lack of Knowledge about Clear Vital Signs Thresholds: Another obstacle that complicates the use of these parameters is the lack of information on unique thresholds of each vital sign for severe fatigue. And although it is obvious that prolonged accumulation of fatigue can lead to various health problems, there are no clear thresholds for various vital signs that indicate extreme fatigue.

4) Lack of Medical Formulas That Links Fatigue to Diseases: Moreover, despite the fact that medical evidence suggests that fatigue can have serious consequences for worker health, there are no clear medical formulas that directly link fatigue to illness. The lack of such a formula hinders the use of fatigue detection tools that can predict the presence of future illness based on fatigue data.

5) Stepping Further than Detection Towards Prediction: Because there is no standard scale for measuring fatigue, no clear thresholds for severe fatigue are known, and no clear formulas exist for linking fatigue accumulation to disease, fatigue-related applications are almost used as indicators of physical fatigue in the workplace. Researchers are trying to explore how technological devices such as smart wearables and Artificial Intelligence can be used to predict illness caused by persistent fatigue. However, as far as we know, there are no such implementations and most available implementations that predict diseases analyze HRV or other vital signs independent of fatigue status.

B. Future Opportunities

Many measures are implemented to detect and monitor fatigue. However, as fatigue becomes more prevalent in the workplace, the need to monitor fatigue and draw conclusions about its existence and persistence increases. Therefore, additional efforts are needed beyond simple detection.

1) Raise Accuracy by Widening Range of Collected Data: Work stress is increasing worldwide for various reasons and requires fatigue monitoring to avoid its negative effects. On the other hand, the development of sensors is advancing rapidly, making it easier to collect the data needed to detect fatigue in the workplace. Accordingly, the ability to collect data is becoming easier, which provides the opportunity to collect and analyze the necessary data and select the most appropriate parameters according to the results.

2) Merge Fatigue with Both Health and Productivity; A Win-Win Situation: Monitoring work stress can be beneficial for both workers and companies, because the relationship between fatigue and productivity is inverse: when fatigue increases, worker productivity decreases and vice versa. Therefore, it seems very important to link information on fatigue with health and productivity indicators and their analysis in order to increase companies' productivity while maintaining workers' health.

CONCLUSION

Work fatigue has become a feature of our times and a common characteristic of workers in various sectors around the world. Whether it is due to constant changes in working conditions or other global events, it is worth monitoring and analysing fatigue in order to maintain worker health while increasing company productivity. This article provides an overview of the different parameters that can be used to detect fatigue in the workplace. The aim is to allow those interested in this area to choose the best parameters or combine some of them to achieve the best results in detecting fatigue. In addition, tools that can be used to expand expectations in this area were discussed, as well as challenges and future opportunities.

ACKNOWLEDGMENTS

We acknowledge the support of Natural Sciences and Engineering Research Council of Canada (NSERC), grant number 06351, Fonds Québécois de la Recherche sur la Nature et les Technologies (FRQNT) and Institut Technologique de Maintenance Industrielle (ITMI).

REFERENCES

[1] Chang, S. S., Stuckler, D., Yip, P., & Gunnell, D. (2013). Impact of 2008 global economic crisis on suicide: time trend study in 54 countries. Bmj, 347.
[2] Gupta, M., Abdelmaksoud, A., Jafferany, M., Lotti, T., Sadoughifar, R., & Goldust, M. (2020). COVID-19 and economy. Dermatologic therapy, 33(4), e13329-e13329.
[3] Härmä, M. (2006). Workhours in relation to work stress, recovery and health. Scandinavian journal of work, environment & health, 502-514.
[4] Eidelman, D. (1980). Fatigue: towards an analysis and a unified definition. Medical hypotheses, 6(5), 517-526.
[5] Desmond, P. A., & Hancock, P. A. (2000). Active and passive fatigue states. In Stress, workload, and fatigue (pp. 455-465). CRC Press
[6] Nelesen, R., Dar, Y., Thomas, K., & Dimsdale, J. E. (2008). The relationship between fatigue and cardiac functioning. Archives of internal medicine, 168(9), 943–949
[7] Spook, S.M., Koolhaas, W., B¨ultmann, U. et al. (2019) Implementing sensor technology applications for workplace health promotion: a needs assessment among workers with physically demanding work. BMC Public Health 19, 1100
[8] Ann Williamson, Rena Friswell. (2013). Fatigue in the workplace: causes and countermeasures. Fatigue: Biomedicine, Health & Behavior, 1:1-2, 81-98.
[9] Afari, N., & Buchwald, D. (2003). Chronic fatigue syndrome: a review. American Journal of Psychiatry, 160(2), 221-236.
[10] Evengard, B., Kuratsune, H., Jason, L. A., & Natelson, B. H. (2008). Fatigue science for human health (pp. V-XI). Y. Watanabe (Ed.). New York, NY, USA:: Springer.
[11] BLS. (2016). Nonfatal Occupational Injuries and Illnesses Requiring Days Away From Work in 2015. Retreieved in September 25, 2021 from https://www.bls.gov/news.release/osh2.toc.htm.
[12] Caruso, C. C., & Waters, T. R. (2008). A review of work schedule issues and musculoskeletal disorders with an emphasis on the healthcare sector. Industrial health, 46(6), 523–534
[13] Collins, S. (2009). Occupational factors, fatigue, and cardiovascular disease. Cardiopulmonary physical therapy journal, 20(2), 28.
[14] Michaud, K., Pope, J. E., Emery, P., Zhu, B., Gaich, C. L., DeLozier, A. M., ... & Smolen, J. S. (2019). Relative impact of pain and fatigue on work productivity in patients with rheumatoid arthritis from the RA-BEAM baricitinib trial. Rheumatology and therapy, 6(3), 409-419.
[15] Daneshmandi, H., Choobineh, A. R., Ghaem, H., Alhamd, M., & Fakherpour, A. (2017). The effect of musculoskeletal problems on fatigue and productivity of office personnel: a cross-sectional study. Journal of preventive medicine and hygiene, 58(3), E252.
[16] What is Fatigue Costing Your Company? - National Safety Council. (n.d.). National Safety Council. Retrieved December 15, 2021, from https://www.nsc.org/workplace/safety-topics/fatigue/what-is-fatigue-costing-your-company
[17] McCorkle, R. U. T. H., & Young, K. (1978). Development of a symptom distress scale. Cancer nursing, 1(5), 373-378.
[18] Rhoten, D. (1982). Fatigue and the postsurgical patient. In C.M. Norris (Ed.), Concept clarification in nursing (277-300). Rockville, MD: Aspen.
[19] Park, E., Kim, K. J., & Kwon, S. J. (2016). Understanding the emergence of wearable devices as next-generation tools for health communication. Information Technology & People.
[20] Anwer, S., Li, H., Antwi-Afari, M. F., Umer, W., & Wong, A. Y. L. (2021). Evaluation of physiological metrics as real-time measurement of physical fatigue in construction workers: state-of-the-art review. Journal of Construction Engineering and Management, 147(5), 03121001.

[21] Gonzalez, K., Sasangohar, F., Mehta, R. K., Lawley, M., & Erraguntla, M. (2017, September). Measuring fatigue through Heart Rate Variability and activity recognition: A scoping literature review of machine learning techniques. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 61, No. 1, pp. 1748-1752). Sage CA: Los Angeles, CA: SAGE Publications.

[22] Yang, L., Lu, K., Diaz-Olivares, J. A., Seoane, F., Lindecrantz, K., Forsman, M., ... & Eklund, J. A. (2018). Towards smart work clothing for automatic risk assessment of physical workload. Ieee Access, 6, 40059-40072.

[23] Sedighi Maman, Z., Alamdar Yazdi, M. A., Cavuoto, L. A., & Megahed, F. M. (2017). A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. Applied ergonomics, 65, 515–529.

[24] Bowen, J., Hinze, A., König, J., & Exton, D. (2021). Supporting safer work practice through the use of wearable technology. Ergonomics and Human Factors, 117.

[25] Hajifar, S., Sun, H., Megahed, F. M., Jones-Farmer, L. A., Rashedi, E., & Cavuoto, L. A. (2021). A forecasting framework for predicting perceived fatigue: Using time series methods to forecast ratings of perceived exertion with features from wearable sensors. Applied Ergonomics, 90, 103262.

[26] Mao, L., Gong, T., Ai, Q., Hong, Y., Guo, J., He, Y., ... & Yu, B. (2019). Morphologically modulated laser-patterned reduced graphene oxide strain sensors for human fatigue recognition. Smart Materials and Structures, 29(1), 015009.

[27] Horiuchi, R., Ogasawara, T., & Miki, N. (2018). Fatigue assessment by blink detected with attachable optical sensors of dye-sensitized photovoltaic cells. Micromachines, 9(6), 310.

[28] Li, P., Meziane, R., Otis, M. J. D., Ezzaidi, H., & Cardou, P. (2014). A Smart Safety Helmet using IMU and EEG sensors for worker fatigue detection. In 2014 IEEE International Symposium on Robotic and Sensors Environments (ROSE) Proceedings (pp. 55-60). IEEE.

[29] Roja, Z., Kalkis, V., Vain, A., Kalkis, H., & Eglite, M. (2006). Assessment of skeletal muscle fatigue of road maintenance workers based on heart rate monitoring and myotonometry. Journal of Occupational Medicine and Toxicology, 1(1), 1-9.

[30] Chang, F. L., Sun, Y. M., Chuang, K. H., & Hsu, D. J. (2009). Work fatigue and physiological symptoms in different occupations of high-elevation construction workers. Applied ergonomics, 40(4), 591-596.

[31] Rosengren, K. S., Hsiao-Wecksler, E. T., & Horn, G. (2014). Fighting fires without falling: Effects of equipment design and fatigue on firefighter's balance and gait. Ecological Psychology, 26(1-2), 167-175.

[32] Park, K., Sy, J. F., Horn, G. P., Kesler, R. M., Petrucci, M. N., Rosengren, K. S., & Hsiao-Wecksler, E. T. (2018). Assessing gait changes in firefighters after firefighting activities and while carrying asymmetric loads. Applied ergonomics, 70, 44-50.

[33] Moriguchi, C. S., Carnaz, L., Veiersted, K. B., Hanvold, T. N., Hæg, L. B., Hansson, G. Å., & Coury, H. J. C. G. (2013). Occupational posture exposure among construction electricians. Applied ergonomics, 44(1), 86-92.

[34] Hu, B., & Ning, X. (2015). The influence of lumbar extensor muscle fatigue on lumbar–pelvic coordination during weightlifting. Ergonomics, 58(8), 1424-1432.

[35] Occhipinti, E. (1998). OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs. Ergonomics, 41(9), 1290-1311.

[36] Helbostad, J. L., Leirfall, S., Moe-Nilssen, R., & Sletvold, O. (2007). Physical fatigue affects gait characteristics in older persons. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 62(9), 1010-1015.

[37] Yung, M., Bigelow, P. L., Hastings, D. M., & Wells, R. P. (2014). Detecting within- and between-day manifestations of neuromuscular fatigue at work: an exploratory study. Ergonomics, 57(10), 1562-1573.

[38] Nath, N. D., Akhavian, R., & Behzadan, A. H. (2017). Ergonomic analysis of construction worker's body postures using wearable mobile sensors. Applied ergonomics, 62, 107-117.

[39] Cifrek, M., Medved, V., Tonković, S., & Ostojić, S. (2009). Surface EMG based muscle fatigue evaluation in biomechanics. Clinical biomechanics, 24(4), 327-340.

[40] Chan, A. P., Wong, F. K., Wong, D. P., Lam, E. W., & Yi, W. (2012). Determining an optimal recovery time after exercising to exhaustion in a controlled climatic environment: Application to construction works. Building and environment, 56, 28-37.

[41] Gatti, U. C., Schneider, S., & Migliaccio, G. C. (2014). Physiological condition monitoring of construction workers. Automation in Construction, 44, 227-233.

[42] Yi, W., Chan, A. P., Wang, X., & Wang, J. (2016). Development of an early-warning system for site work in hot and humid environments: A case study. Automation in Construction, 62, 101-113.

[43] Umer, W., Li, H., Szeto, G. P. Y., & Wong, A. Y. (2017). Low-cost ergonomic intervention for mitigating physical and subjective discomfort during manual rebar tying. Journal of Construction Engineering and Management, 143(10), 04017075.

[44] Ueno, S., Sakakibara, Y., Hisanaga, N., Oka, T., & Yamaguchi-Sekino, S. (2018). Heat strain and hydration of Japanese construction workers during work in summer. Annals of work exposures and health, 62(5), 571-582.

[45] Satterfield, B. C., & Van Dongen, H. P. (2013). Occupational fatigue, underlying sleep and circadian mechanisms, and approaches to fatigue risk management. Fatigue: Biomedicine, Health & Behavior, 1(3), 118-136.

[46] Sallinen, M., & Kecklund, G. (2010). Shift work, sleep, and sleepiness—differences between shift schedules and systems. Scandinavian journal of work, environment & health, 121-133.

[47] Folkard, S. (2008). Do permanent night workers show circadian adjustment? A review based on the endogenous melatonin rhythm. Chronobiology international, 25(2-3), 215-224.

[48] Folkard, S., & Lombardi, D. A. (2004). Toward a "risk index" to assess work schedules. Chronobiology international, 21(6), 1063-1072.

[49] Di Milia, L., Smolensky, M. H., Costa, G., Howarth, H. D., Ohayon, M. M., & Philip, P. (2011). Demographic factors, fatigue, and driving accidents: An examination of the published literature. Accident Analysis & Prevention, 43(2), 516-532.

[50] Smolensky, M. H., Di Milia, L., Ohayon, M. M., & Philip, P. (2011). Sleep disorders, medical conditions, and road accident risk. Accident Analysis & Prevention, 43(2), 533-548.

[51] Baharav, A., Kotagal, S., Gibbons, V., Rubin, B. K., Pratt, G., Karin, J., & Akselrod, S. (1995). Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability. Neurology, 45(6), 1183-1187.

[52] Tanaka, M., Ishii, A., & Watanabe, Y. (2015). Effects of mental fatigue on brain activity and cognitive performance: a magnetoencephalography study. Anat Physiol, 4, 1-5.

[53] Rodríguez-Ibáñez, N., García-González, M. A., Fernández-Chimeno, M., & Ramos-Castro, J. (2011, August). Drowsiness detection by thoracic effort signal analysis in real driving environments. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 6055-6058). IEEE.

[54] Williamson, A., & Friswell, R. (2013). Fatigue in the workplace: causes and countermeasures. Fatigue: Biomedicine, Health & Behavior, 1(1-2), 81-98.

[55] Howard, M. E., Desai, A. V., Grunstein, R. R., Hukins, C., Armstrong, J. G., Joffe, D., ... & Pierce, R. J. (2004). Sleepiness, sleep-disordered breathing, and accident risk factors in commercial vehicle drivers. American journal of respiratory and critical care medicine, 170(9), 1014-1021.

[56] Surangsrirat, D., Dumnin, S., & Samphanyuth, S. (2019, April). Heart Rate, Skin Temperature and Skin Humidity and their Relationship to Accumulated Fatigue. In 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART) (pp. 1-4). IEEE.

[57] Li, G., & Chung, W. Y. (2013). Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier. Sensors, 13(12), 16494-16511.

[58] Chaaraoui, A. A., Padilla-López, J. R., & Flórez-Revuelta, F. (2015, May). Abnormal gait detection with RGB-D devices using joint motion history features. In 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG) (Vol. 7, pp. 1-6). IEEE.

[59] Gu, Y., Zhan, J., Ji, Y., Li, J., Ren, F., & Gao, S. (2017). MoSense: An RF-based motion detection system via off-the-shelf WiFi devices. IEEE Internet of Things Journal, 4(6), 2326-2341.

[60] Weiss, G. M., Yoneda, K., & Hayajneh, T. (2019). Smartphone and smartwatch-based biometrics using activities of daily living. IEEE Access, 7, 133190-133202.

[61] Castaldo, R., Prati, M., Montesinos, L., Kulkarni, V., Chappell, M., Byrne, H., ... & Pecchia, L. (2019, April). Investigating the use of wearables for monitoring circadian rhythms: A feasibility study. In International Conference on Biomedical and Health Informatics (pp. 275-280). Springer, Cham.

[62] Takeshita, R., Shoji, A., Hossain, T., Yokokubo, A., & Lopez, G. (2021, November). Emotion Recognition from Heart Rate Variability Data of Smartwatch While Watching a Video. In 2021 Thirteenth International Conference on Mobile Computing and Ubiquitous Network (ICMU) (pp. 1-6). IEEE.

[63] Ciabattoni, L., Ferracuti, F., Longhi, S., Pepa, L., Romeo, L., & Verdini, F. (2017, January). Real-time mental stress detection based on smartwatch. In 2017 IEEE International Conference on Consumer Electronics (ICCE) (pp. 110-111). IEEE.

[64] Ahn, C. R., Lee, S., Sun, C., Jebelli, H., Yang, K., & Choi, B. (2019). Wearable sensing technology applications in construction safety and health. Journal of Construction Engineering and Management, 145(11), 03119007.

A.3 Improved ML Models for Prediction of Cardiovascular Diseases

Following my exploration of smart wearables in disease management, I transitioned into a critical area of research aimed at enhancing the predictive performance of Machine Learning models in Cardiovascular Diseases (CVDs). The gravity of CVDs, recognized as the leading global cause of mortality, served as a compelling impetus behind this phase of my research. Leveraging cutting-edge technologies, I meticulously crafted a suite of eight distinct Machine Learning models designed to predict CVDs. In some instances, these models not only met but exceeded prevailing state-of-the-art benchmarks, achieving classification accuracy rates surpassing 91.80

The culmination of this effort was the development and subsequent publication of three comprehensive articles. Each of these articles delves into the nuances and outcomes of the respective Machine Learning models, which collectively constitute a significant stride in the field of cardiovascular health prediction. These articles found their place in different esteemed events, reflecting the diverse contexts and audiences that have benefited from the insights and innovations arising from my research. These events are as below:

- **Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability** *(MobiSPC2022 Conference **"Best Paper Award" winner**)*
- **Predicting Cardiovascular Events with Machine Learning Models and Heart Rate Variability** *(IASKS-JUSPN Journal / published as an extension for the MobiSPC2022 article)*
- **Machine Learning Models to Predict Cardiovascular Events from Heart Rate Variability Data** *(IHSHS'2022 Conference)*


MobiSPC
The 19th International Conference on Mobile Systems and Pervasive Computing
Niagara Falls, Canada, August 9-11, 2022
ELSEVIER
BEST PAPER AWARD
Awarded to:
Mohammad Moshawrab, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim and Ali Raad
For the Paper:
Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability
August 10, 2022
Date
Dr. Elhadi Shakshuki
SEIT Steering Chair and Founder

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 203 (2022) 231–238

Procedia Computer Science

www.elsevier.com/locate/procedia

The 19th International Conference on Mobile Systems and Pervasive Computing (MobiSPC)
August 9-11, 2022, Niagara Falls, Canada

Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability

Mohammad Moshawrab[a,c,*], Mehdi Adda[a], Abdenour Bouzouane[b], Hussein Ibrahim[c], Ali Raad[d]

[a] *Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, G5L 3A1, Québec, Canada*

[b] *Département D'informatique et de Mathématique, Université du Québec à Chicoutimi, Chicoutimi, 555 boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada*

[c] *Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, G4R 5B7, Québec, Canada*

[d] *Dean of the Faculty of Science and Arts, Islamic University of Lebanon, Wardaniyeh, Lebanon*

Abstract

Artificial Intelligence is exponentially evolving into a solution to many of humanity's complex problems. In this context, healthcare is benefiting from this technology and all its branches to improve the level of services offered, including cardiac health services. Cardiovascular diseases have always been among the most common and deadly diseases around the world, as studies have consistently shown. However, Artificial Intelligence services offer several tools to improve the diagnosis of these diseases and even predict their occurrence. In this study, four models are created and trained with "PhsyioNet Smart Health for Assessing the Risk of Events via ECG Database" to analyze the characteristics of heart rate variability and predict the occurrence of heart diseases and cerebrovascular events. The results obtained support the confidence in the use of Artificial Intelligence in cardiology, where Support Vector Machines, Deep Neural Networks, and XGBoost achieved an accuracy of 91.80%, 90.19%, and 89.10%, respectively.

Keywords: Artificial Intelligence, Machine Learning, Diseases Prediction, Cardiovascular Diseases, Heart Rate Variability

1. Introduction

Cardiovascular Diseases (CVDs) cause the most deaths and are therefore known as the most dangerous disease worldwide. According to the latest figures from the World Health Organization (WHO) in the field of cardiovascular diseases, the number of deaths caused by them increased from 12.1 million to 18.6 million between 1990 and 2019, with deaths accounting for 32% of global mortality in 2019. Moreover, cardiovascular diseases are not only a major cause of health conflict, but also of economic burden. According to "Medical Expenditure Panel Survey," the costs

* Corresponding author; Tel.: +1-581-624-9394
E-mail address: mohammad.moshawrab@uqar.ca

1877-0509
10.1016/j.procs.2022.07.030

due to CVDs were estimated to be $378.0 billion in the United States alone between 2017 and 2018, including $226.0 billion in expenditures and $151.8 billion in lost future productivity [1,2].

1.1. Artificial Intelligence in Healthcare: A New Cardiology Era

The deadly cardiovascular diseases urge to find efficient solutions that can help in the early diagnosis of these diseases and, if possible, even predict their occurrence. Traditional methods to detect these diseases include electrocardiogram, echocardiography, coronary angiography, stress test, magnetic resonance imaging or intracoronary ultrasound. However, technological developments, especially Information and Communication Technologies (ICT), and the rise of Artificial Intelligence (AI) and its variants are helping to improve the quality of healthcare services and thus facilitate the diagnosis of CVDs. Moreover, AI tools are considered the next revolution in cardiology as they help provide faster and more accurate patient care outcomes. Moreover, AI will soon transform the science of heart health, as its tools could outperform experts in diagnosing or even predicting CVDs [3,4].

1.2. Heart Rate Variability as a CVD Indicator

Recently, interest in the use of heart rate variability (HRV) as an indicator of cardiovascular diseases has increased, especially with the development of AI and the data analysis capabilities offered by its branches: Machine Learning and Deep Learning. Moreover, HRV is known as the beat-to-beat variation in heart rate or the duration of the RR peak interval, where R is a wave of the QRS complex extracted from a cardiac ECG signal. Knowing that changes in the autonomic regulation of the heart can be read from the temporal variations in heart rate, the parameters extracted from HRV data are divided into three main categories: Time domain, Frequency Domain and Non-Linear parameters. These categories are listed in Table 1 below [5].

Table 1. Heart Rate Variability Parameters.

Group	Parameter	Unit	Description
Time Domain Features	Mean NN	(ms)	Mean of NN interval
	SDNN	(ms)	Standard deviation of NN intervals
	RMSSD	(ms)	Square root of the mean squared differences of successive NN intervals
	pNN50	(ms)	Proportion of interval differences of successive NN intervals greater than 50 ms
Frequency Domain Parameters	VLF	(ms^2)	Power in very low frequency range (0–0.04 Hz)
	LF	(ms^2)	Power in low frequency range (0.04–0.15 Hz)
	HF	(ms^2)	HF ms2 Power in high frequency range (0.15–0.4 Hz)
	LF/HF	(ratio)	Ratio of LF over HF
Non-Linear Parameters	SD1	(ms)	Standard deviation of points perpendicular to the axis of line of identity or the successive intervals scaled by $\sqrt{\frac{1}{2}}\ \sqrt{\frac{1}{2}\,var(RR_{\mathrm{n}} - RR_{\mathrm{n+1}})}$
	SD2	(ms)	Standard deviation of points along the axis of line of identity, or $\sqrt{2SDNN^2 - \frac{1}{2}SD1^2}$
	SD1/SD2	(ratio)	Ratio of SD1 over SD2

1.3. Prediction of CVDs with HRV; State of the Art

The number of studies on CVDs detection using HRV parameters is increasing rapidly. Researchers are using different AI models to analyze various HRV parameters, and AI has proven its efficiency and accuracy in this domain.
For example, in[6], the authors used the Fast Fourier Transform (FFT) with the Blackman Harris window algorithm to build a model that analyzes various HRV features to predict the occurrence of Ventricular Tachycardia (VT) in the short term. In addition, the authors developed an Artificial Neural Networks (ANN) classifier in [7] and trained it with the "PhysioNet Spontaneous Ventricular Tachyarrhythmia Database" [8] to predict the occurrence of VT. They measured performance using several metrics, recording 76.60% 82.9% and 71.4% for accuracy, sensitivity, and specificity, respectively. In addition, the authors in [9] used Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Support Vector Machines (SVM) to predict cardiovascular risk. Their best model achieved 96.67% accuracy. In addition, the authors in [10] used SVM to develop a predictive model to predict cardiovascular risk after Myocardial Infarction, and their model accuracy was 89%.
Moreover, the authors of [11] created models to predict Sudden Cardiac Death (SCD) using the k-Nearest Neighbor and Multilayer Perceptron Neural Network (MLP) algorithms. They trained their models using the "PhysioNet Sudden Cardiac Death Holter database" [12] and the "PhysioNet Normal Sinus Rhythm database" [13], and their recorded performance measures were 99.73%, 96.52%, 90.37%, and 83.96% accuracy for the first, second, third, and fourth one-minute intervals, respectively. Furthermore, in [14], the authors did the same with SVM and Probabilistic Neural Network (PNN) to predict SCD two minutes before its onset. Similarly, the authors trained their models using the "PhysioNet Sudden Cardiac Death (SCD) Holter database" [12] and the "PhysioNet MIT Normal Sinus Rhythm database" [13] and SVM and PNN recorded prediction rates of 96.36% and 93.64%, respectively.

On the other hand, in [15], the authors targeted hypertensive patients and developed novel SVM, Tree-Based Classifier, Artificial Neural Network and Random Forest models to create an automated cardiovascular risk stratification model. The authors trained their data using the "Smart Health for Assessing the Risk of Events via ECG database" [16] and achieved a sensitivity of 71.4% and a specificity of 87.8%. Furthermore, in [17], the authors developed an Artificial Neural Networks model that analyzes respiratory rate in addition to HRV features to detect Ventricular Tachycardia one hour before its onset. The performance metrics of their model were 88%, 82%, and 93% for sensitivity, specificity, and area under the curve, respectively. In addition, the authors in [18] used a statistical model called MIL, to predict CVDs based on features of heart rate variability. Their model achieved high accuracy, as they mentioned. Finally, in [19], the authors created K Nearest Neighbor (k-NN), Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, Neural Network, and Vote and trained them with the "UCI Heart Diseases Repository" [20]. The models created were able to predict CVDs with 87.4% accuracy.

In this article, several artificial intelligence models were created to predict Cardiovascular Diseases and events. The models used are: Support Vector Machine (SVM), Deep Neural Networks (DNN), XGBoost, and Neural Oblivious Decision Ensembles (NODE). Section 2 below explains the dataset used in this study and the preprocessing steps used to prepare the data for the models. Section 3 explains the models created and the results obtained with these models are listed and discussed in Section 4.

2. Materials & Methods

2.1. Dataset

The dataset used in this study is the "PhysioNet Smart Health for Assessing the Risk of Events via ECG Database" (SHAREEDB) [16] that is offered by the PhysioNet online data repository. This dataset was collected to investigate the efficiency of classifying hypertensive patients at higher risk for cardiac and cerebrovascular events using heart rate variability characteristics. It consists of 139 records of 24-hour Electrocardiographic (ECG) Holter recordings. Each recording contains three ECG signals sampled at a rate of 128 samples per second with a precision of 8 bits. The population in which the data were collected consisted of 49 women and 90 men aged 55 years and older. They were followed up for 12 months to record the occurrence of serious cardiovascular and cerebrovascular events such as Coronary Revascularization, fatal or nonfatal Acute Coronary Syndromes, syncopal events, Myocardial Infarctions, fatal or nonfatal strokes, and Transient Ischemic Attacks. During the follow-up period, 17 patients experienced a cardiovascular event, including 11 Myocardial Infarctions, 3 strokes, and 3 syncopal events. In addition, the dataset contains some demographic and clinical information about the subjects, such as their age, sex, any vascular events, values of systolic and diastolic arterial pressure, and others.

2.2. Data Filtering & Preprocessing

The ECG signals provided by the SHAREEDB dataset are collected in laboratories and may be susceptible to a lot of noise that needs to be removed before the data is passed to the AI models. It is very important to clean the data and remove the noise to obtain high-quality ECG signals that are then analyzed by the models. The data cleaning and preparation steps used in this study are summarized below:

- **Filtering & Artifacts Removal [21]:** ECG recordings are susceptible to noise or interference from various signals, which can be divided into high and low-frequency noise sources. For example, noise can be caused by electrode interference, muscle motion interference, channel interference, baseline drift, or power line interference. Therefore, ECG signals can be cleaned by using the following filters:
  - **IIR Notch Filters:** remove motion artifacts and/or power line interference
  - **FIR Filters:** clean ECG data and are act on the range of ECG data that is between 1 and 100 hertz
- **R Peaks Detection [22,23]:** The ECG signal reflects the electrical activity of the myocardium and is divided into three distinct parts: the P wave, the QRS complex, and the T wave. However, the QRS complex is composed of Q-wave, R-wave and S-wave. The R-peak is the interval between the onset of the QRS complex and the peak of the R-wave and can be determined using various algorithms such as Hamilton, Christov, Engelse and Zeelenberg, Pan and Tompkins, Stationary Wavelet Transform and Two Moving Average. According to the results in [23], Engelse and Zeelenberg provided the best results in detecting R-peaks, which is why they were used in this study
- **Calculation of RR Intervals:** Heart Rate Variability is defined as the RR intervals or the difference between two consecutive R peaks, which are then calculated using the required equations

- **Outliers Removal:** After the RR intervals are detected, the outliers, defined as points that are extremely far from the mean, are removed and replaced with the mean value
- **Extract HRV features:** Finally, the HRV features were calculated using the appropriate mathematical formulas. In this study, 26 HRV features were calculated, and despite the high number of features calculated, the use of all features gave good results

2.3. Artificial Intelligence Models

Cardiology is defined as the healthcare sector that takes care of heart health, and the use of AI in this field is growing briskly. AI has demonstrated high accuracy and efficiency in detecting CVDs, and sometimes it can go beyond professional diagnosis and even be used in predicting cardiovascular diseases instead of detecting them due to its high ability to analyze cardiac data [24,25]. In addition, AI is known for its various branches that are used in different areas of life around the world. For example, Machine Learning, Ensemble Learning and Deep Convolutional Neural Networks are AI branches that were used in this study:

- **Classical Machine Learning Algorithms[26]:** are algorithms that give computers learning potential by training them with experimental data and generating models based on these data, enabling them to make decisions in new situations such as: Support Vector Machines, Naïve Bayes, Logistic and Linear Regression and others.
- **Ensemble Learning [27]:** is a special branch of ML where its algorithms are based on merging predictions from different models. Some of these models are XGBoost, AdaBoost, GradientBoosting, LightGBM and others.
- **Deep Convolutional Neural Network (DCNNs) [26]:** are a type of Neural Networks that are used to analyze data with a grid-like structure. However, these networks are intended for analyzing multidimensional data such as images and videos. Using these networks to analyze tabular data may require transforming the data used. Nevertheless, there are several models that offer transformation of tabular data for use in DCNNs, such as TabNet, GrowNet, TreeEnsemble Layers, TabTransformers, Self Normalizing Neural Networks, Neural Oblivious Decision Ensembles (NODE), AutoInt, and Deep & Cross Neural Networks (DCNs) [28].

3. Construction of AI Models

In this study, different AI models were used to analyze HRV features to detect heart diseases and events. However, before passing the extracted features to the models, some data fitting steps should be performed, as explained below.

3.1. Data Adjustment

Considering that of the study population, 139 patients, only 17 developed a cardiovascular event in the 12-month follow up period, the extracted HRV features show an unbalanced identity, with the majority falling into the ”no cardiovascular event” class. Because the proportion of this class is 122 of 139, the performance of the prediction models may be negatively affected, suggesting the application of some data adjustments such as balancing and scaling:

- **Synthetic Minority Over-sampling Technique (SMOTE):** a data expansion in which new samples are drawn from existing ones to oversample the minority class
- **Preprocessing Standard Scaling:** the standardization of characteristics is achieved by removing the mean of the data and scaling it to a unit variance

3.2. Building the Models; hyperparameters to be considered

After applying the necessary data fitting steps to the extracted HRV features, they are then passed to the models created for fitting with the thresholds listed below:

3.2.1. Classical ML: Support Vector Machines

SVM is a supervised Machine Learning algorithm that is fed labeled training data to learn how to assign labels to objects based on examples, and then gain the ability to predict the category of new example(s) [26]. The performance of the SVM model is affected by the following hyperparameters [29]:

- **Kernel:** the function that converts the input data into the required form
- **Regularization:** denotes the misclassification or error term and is expressed as hyperparameter ”C”.
- **gamma:** interpret how far the effect of a single training sample extends
- **class weight:** used for imbalanced datasets and defines the weight of the classes to be predicted

3.2.2. Deep Learning: Deep Neural Networks

These networks are algorithms that mimic human brain cells called neurons. In general, these networks use brain simulations to improve their learning and increase the accuracy of the models. The structure of DNNs consists of more than two interconnected layers and is affected by the following hyperparameters [30]:

- **Number of layers:** input, output, and the hidden layers that define the structure of the network.
- **Units:** denotes the output of each layer.
- **Activation function:** also known as the ”transfer function”, which defines how the weighted sum of the input is converted into an output from one or more nodes in a layer of the network
- **Number of epochs:** a complete pass through all rows of the training data
- **Batch size:** samples that the model examines within each epoch before updating the weights.
- **Learning rates:** a variable that controls how the optimizer's learning rate changes over time
- **Momentum:** is the ”delay” in learning the mean and variance

3.2.3. Ensemble Learning Algorithms: XGBoost

XGBoost is an Ensemble Learning algorithm that also belongs also to the Machine Learning AI Branch.

- **XGBoost [31]:** eXtreme Gradient Boosting package is a scalable implementation of the gradient boosting framework built with an efficient linear model solver and a tree learning algorithm with hyperparameters:
  - **Booster:** the type of model to run at each iteration
  - **Learning Rate:** is the step size shrinkage used during the update to prevent overfitting
  - **Gamma:** specifies the minimum loss reduction required to perform splitting
  - **Max Depth:** the parameter used to control overfitting
  - **Min Child Weight:** defines the minimum sum of weights of all observations required in a child
  - **Max Delta Step:** helps to make the update step more conservative
  - **Sub Sample:** denotes the fraction of observations that are randomly selected for each tree
  - **Lambdas:** is used to handle the regularization part
  - **Alpha:**is used in case of very high dimensionality to make the algorithm run faster during implementation
  - **Tree Method:** Algorithm for tree construction
  - **Scale Weight:** controls the balance of positive and negative weights
  - **Objective:** defines the loss function to be minimized

3.2.4. Deep Convolutional Neural Networks: Neural Oblivious Decision Ensembles

In this study, the following model was used to apply DCNN to the SHAREEDB tabular data:

- **Neural Oblivious Decision Ensembles (NODE)[32]:** a model with a layered structure built from differentiable oblivious trees, which are decision tables that decompose the data along dd-splitting features and compare each feature to a learned threshold. It was trained in an end-to-end manner using backpropagation and is affected by the following hyperparameters:
  - **Number of Layers:** Number of layers forming the Neural Network
  - **Number of Trees:** Number of trees in each layer
  - **Depth:** Depth of the tree
  - **Learning Rate:** is the shrinkage step size used in the update to prevent overfitting.

3.3. Wrapping Up, Training, Prediction, and Optimization

Once the models were created, they were trained with the fitted version of the extracted HRV features. The obtained results are explained and discussed in detail in Section 4. Figure 1 shows the overall architecture of the data preparation steps and the models created in this study.

4. Results & Discussion

The created models were trained with the HRV features. The SVM, DNN, XGBoost, and NODE models were evaluated with the metrics of Accuracy, Precision, Recall, Specificity, Negative Predictive Value NPV, and F1 Score. For better measurement, Repeated K-fold Cross Validation [33] was implemented with 10 folds and repeated 5 times. The results are shown in Table 2 below, and the values of accuracy, precision, recall, specificity, negative predictive value, and F1 score are denoted as AC, PR, RE, SP, NPV, and F1 , respectively. In addition, the values of the hyperparameters used are listed in the table. Figure 2 below also shows a graphical representation of the performance of the models created in this study.

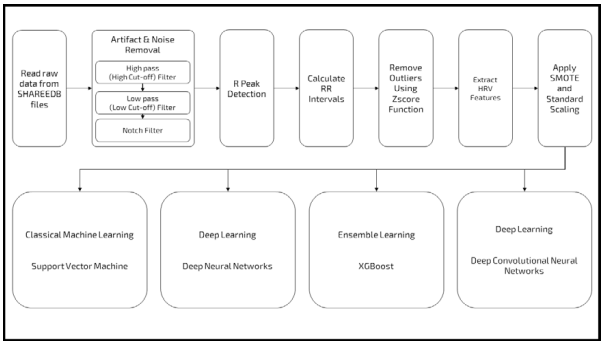

Fig. 1. Overall Architecture Followed in this Study

Table 2. AI Models Evaluation Metrics.

Model	Hyper Parameters		Evaluation Metrics					
	Parameter	Value	AC	PR	RE	SP	NPV	F1
SVM	Training(Testing)	0.75(0.25)	91.80%	87.87%	96.66%	87.09%	96.42%	92.06%
	Kernel	rbf						
	Regularization(C)	2.66						
	Gamma	0.141						
DNN	Training(Testing)	0.79(0.21)	90.19%	85.18%	95.83%	85.18%	95.83%	90.19%
	Layers	Input/3 Hidden/Output						
	Units	512/256/128/64/1						
	Activation Function	tanh/tanh/tanh/sigmoid						
	Dropout	Before Output Layer 0.2						
	Optimizer	SGD						
	Epochs	6850						
	Batch Size	250						
	Learning Rate	0.005						
	Momentum	default						
XGBoost	Training(Testing)	0.79(0.21)	89.10%	86.00%	93.80%	85.10%	92.50%	89.10%
	Booster	gbtree						
	Learning Rate	0.01						
	Gamma	0.1						
	Maximum Depth	10						
	Minimum Child Weight	0.01						
	Max Delta Step	0						
	Sub Sample	0.75						
	Lambda	1						
	Alpa	0.01						
	Tree Method	Auto						
NODE	Training(Testing)	0.71(0.29)	76.92%	77.77%	73.68%	80%	76.19%	75.67%
	Number of Layers	5						
	Depth	10						
	Number of Trees	1						
	Learning Rate	0.1						
	Batch Size	26						

4.1. Discussion

In this study, several models were created to analyze HRV characteristics to detect cardiovascular risks. The results obtained demonstrate the high efficiency of AI models in predicting cardiovascular disease. However, the results obtained in this study outperformed previous implementations.

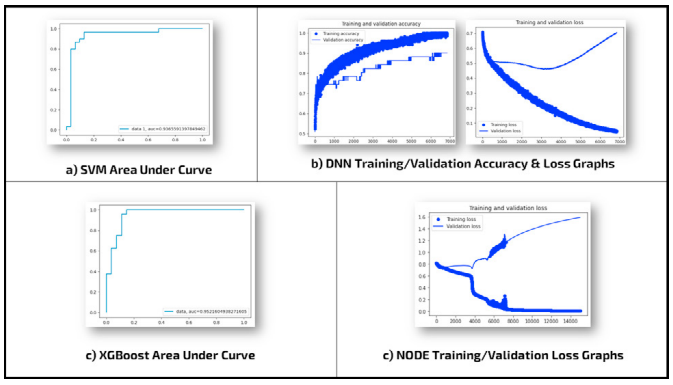

Fig. 2. Models Performance Graphical Representation

First, the authors in [15]applied similar models to the same dataset. Nevertheless, the results obtained in this study exceeded their results. For example, their SVM model recorded accuracy, recall and specificity results were 89.00%, 86.30% and 91.80% respectively, whereas our results are 91.80%, 96.66% and 87.09% for the same performance metrics. In addition, the performance metrics of their Multi Layer Perceptron (MLP) model were Accuracy: 78.10%, Recall: 86.30%, Specificity: 69.90% and our model recorded 90.19%, 95.83% and 85.18% for the same metrics.

In addition, our SVM model achieved 91.80% accuracy, the highest performance among all previous implementations.

For example, the SVM models in [9] recorded an accuracy of 88.64%, 82.95% and 82.58% for the Linear, Polynomial and RBF kernels, respectively. Moreover, the accuracy of SVM in [10,15,19] was 79.81%, 89.00% and 85.19%, respectively. Even though the accuracy is close, other metrics such as Precision and Recall clearly outperform the previous results by a large margin. Knowing that Recall measures how a model correctly classifies True Positives, the models presented in this study are more accurate in predicting whether a person will have a cardiovascular disease in future. The high recall for SVM, DNN, and XGBoost, which are 96.66%, 95.83% and 93.80%, respectively, reflects the highest ability of all implementations to correctly predict that a person is in the cardiovascular risk zone.
Likewise, the DNN model presented in this article also outperforms all previous implementations. The accuracy of this model is 90.19%, whereas the multilayer perceptron in [9,15] is 86.67% and 78.10%, and the accuracy of artificial neural networks in [7,17] is 76.60% and 85.30%. Moreover, precision and recall are significantly higher than the previous implementations, which also reflects a higher capability in cardiovascular risk detection. Table 3 provides a detailed comparison between the results of the models presented in this article and the previous implementations. The sybmoles of the performance metrics used in this table are similar to those in Table 2, and a ”NA” symbol indicates that the corresponding metric was not mentioned in the associated study.
On the other hand, none of the previous implementations used XGBoost , which also outperformed the previous implementations with an accuracy of 98.10% and a recall of 94.60%, reflecting high efficiency in predicting cardiovascular risk, in contrast to the implementation of NODE, which achieved an accuracy of 76.92%, which is not comparable with the previous implementations.
Finally, the SVM, DNN, and XGBoost models discussed in this study can be considered the most accurate models for predicting cardiac disease and events. Even the implementations in [11,14] had higher accuracy and relatively higher recall, but their models were developed to detect sudden cardiac death (SCD) only minutes before its occurrence. For example, the model mentioned in [11] achieved 99.73% accuracy in predicting sudden cardiac death one minute before its onset, but the performance drops to 83.93% when the event is predicted four minutes before its occurrence. However, the models presented here are able to predict cardiovascular disease 12 months before its onset, demonstrating high efficiency in predicting cardiovascular disease and cardiac events long before their onset, thus increasing confidence in the use of AI in detecting and predicting cardiac disease and related events.

Table 3. Comparison with Previous Implementations.

Study	Model	AC	PR	RE	SP	NPV	F1
Our Study	Support Vector Machines	91.80%	87.87%	96.66%	87.09%	96.42%	92.06%
	DNN	90.19%	85.18%	95.83%	85.18%	95.83%	90.19%
	XGBoost	89.10%	86.00%	93.80%	85.10%	92.50%	89.10%
	NODE	76.92%	77.77%	73.68%	80.00%	76.00%	75.67%
[7]	Artificial Neural Network	76.60%	70.70%	82.90%	71.40%	NA	NA
[9]	Support Vector Machines (Linear Kernel)	88.64%	90.84%	86.36%	90.91%	86.96%	NA
	Support Vector Machines (Polynomial Kernel)	82.95%	80.85%	79.55%	86.36%	85.37%	NA
	Support Vector Machines (RBF Kernel)	82.58%	79.45%	77.27%	87.88%	86.44%	NA
	Multi Layer Perceptron (Top 15 Features)	86.67%	100%	73.33%	100%	78.95%	NA
[10]	Support Vector Machines	79.81%	21.15%	91.67%	79.08%	99.36%	NA
[11]	MLP (A Minute Before the SCD Event)	99.73%	NA	NA	NA	NA	NA
	K-NN (A Minute Before the SCD Event)	98.32%	NA	NA	NA	NA	NA
[14]	SVM (2 minutes before VF Event)	96.36%	NA	NA	NA	NA	NA
	Penalized Neural Network	93.64%	NA	NA	NA	NA	NA
[15]	Support Vector Machines	89.00%	NA	86.30%	91.80%	NA	NA
	Multi Layer Perceptron	78.10%	NA	86.30%	69.90%	NA	NA
[17]	Artificial Neural Network	85.30%	83.30%	88.20%	82.40%	87.50%	NA
[18]	MIL Statisitcs Algorithm	85.47%	92.11%	86.42%	83.33%	NA	NA
[19]	Vote	87.41%	NA	NA	NA	NA	NA
	Naïve Bayes	84.81%	NA	NA	NA	NA	NA
	Support Vector Machines	85.19%	NA	NA	NA	NA	NA

5. Conclusion

AI will one day be destiny, some have said. But what we are witnessing today through the use of these technologies in many areas of life confirms that they have become a reality and that their use is increasing day by day. Moreover, AI is expected to evolve the concepts of cardiology and the mechanisms to diagnose its diseases, and even use them to predict the occurrence of these diseases in the future. In this study, we have presented a group of models capable of predicting the occurrence of heart diseases or events with high accuracy, which increases the confidence in AI and its branches in the health field. Furthermore, adapting these models to work in real time will certainly help create personalized and continuous monitoring that can be used to track patients' heart health or even monitor the health of workers who work in stressful environments or for extremely long periods of time.

References

[1] Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., ... & GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. Journal of the American College of Cardiology, 76(25), 2982-3021.

[2] Tsao, C. W., Aday, A. W., Almarzooq, Z. I., Alonso, A., Beaton, A. Z., Bittencourt, M. S., ... & American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2022). Heart Disease and Stroke Statistics—2022 Update: A Report From the American Heart Association. Circulation, 145(8), e153-e639.

[3] National Heart Lung and Blood Institute. (2005). In Brief: Your guide to living well with heart disease (NIH Publication 06-5716). Washington, DC: National Institutes of Health.

[4] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. Journal of the American College of Cardiology, 71(23), 2668-2679.

[5] Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: a review. Medical and biological engineering and computing, 44(12), 1031-1051.

[6] Baumert, M., Wessel, N., Schirdewan, A., Voss, A., & Abbott, D. (2007). Forecasting of ventricular tachycardia using scaling characteristics and entropy of heart rate time series. In World Congress on Medical Physics and Biomedical Engineering 2006 (pp. 1001-1004). Springer, Berlin, Heidelberg.

[7] Joo, S., Choi, K. J., & Huh, S. J. (2010, September). Prediction of ventricular tachycardia by a neural network using parameters of heart rate variability. In 2010 Computing in Cardiology (pp. 585-588). IEEE.

[8] Spontaneous Ventricular Tachyarrhythmia Database v1.0. (2007, May 2). PhysioNet. https://physionet.org/content/mvtdb/1.0/

[9] Ramirez-Villegas, J. F., Lam-Espinosa, E., Ramirez-Moreno, D. F., Calvo-Echeverry, P. C., & Agredo-Rodriguez, W. (2011). Heart rate variability dynamics for the prognosis of cardiovascular risk. PloS one, 6(2), e17060.

[10] Song, T., Qu, X. F., Zhang, Y. T., Cao, W., Han, B. H., Li, Y., ... & Da Cheng, H. (2014). Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. BMC cardiovascular disorders, 14(1), 1-8.

[11] Ebrahimzadeh, E., Pooyan, M., & Bijar, A. (2014). A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. PloS one, 9(2), e81896.

[12] Sudden Cardiac Death Holter Database v1.0.0. (2004, July 2). PhysioNet. https://physionet.org/content/sddb/1.0.0/

[13] MIT-BIH Normal Sinus Rhythm Database v1.0.0. (1999, August 3). PhysioNet. https://physionet.org/content/nsrdb/1.0.0/

[14] Murukesan, L., Murugappan, M., Iqbal, M., & Saravanan, K. (2014). Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features. Journal of Medical Imaging and Health Informatics, 4(4), 521-532.

[15] Melillo, P., Izzo, R., Orrico, A., Scala, P., Attanasio, M., Mirra, M., ... & Pecchia, L. (2015). Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. PloS one, 10(3), e0118504.

[16] Smart Health for Assessing the Risk of Events via ECG Database. (2015, May 19). PhysioNet. Retrieved November 1, 2021, from https://physionet.org/content/shareedb/1.0.0/

[17] Lee, H., Shin, S. Y., Seo, M., Nam, G. B., & Joo, S. (2016). Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks. Scientific reports, 6(1), 1-7.

[18] Lan, K. C., Raknim, P., Kao, W. F., & Huang, J. H. (2018). Toward hypertension prediction based on PPG-derived HRV signals: A feasibility study. Journal of medical systems, 42(6), 1-7.

[19] Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, 82-93.

[20] UCI Machine Learning Repository: Heart Disease Data Set. (n.d.-a). UCI Machine Learning Repository: Heart Disease Data Set. https://archive.ics.uci.edu/ml/datasets/heart+disease.

[21] Luo, S., & Johnston, P. (2010). A review of electrocardiogram filtering. Journal of electrocardiology, 43(6), 486-496.

[22] Ashley, E. A., & Niebauer, J. (2004). Cardiology explained.

[23] Eilers, J., Chromik, J., & Arnrich, B. (2021). Choosing the Appropriate QRS Detector. In BIOSIGNALS (pp. 50-59).

[24] Mathur, P., Srivastava, S., Xu, X., & Mehta, J. L. (2020). Artificial intelligence, machine learning, and cardiovascular disease. Clinical Medicine Insights: Cardiology, 14, 1179546820927404.

[25] Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet?. Heart, 104(14), 1156-1164.

[26] Ertel, W. (2018). Introduction to artificial intelligence. Springer.

[27] Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. Frontiers of Computer Science, 14(2), 241-258.

[28] Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. Advances in Neural Information Processing Systems, 34.

[29] Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

[30] Anderson, J. A. (1995). An introduction to neural networks. MIT press.

[31] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.

[32] Popov, S., Morozov, S., & Babenko, A. (2019). Neural oblivious decision ensembles for deep learning on tabular data. arXiv preprint arXiv:1909.06312.

[33] Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, 21(2), 137-146.

Journal of Ubiquitous Systems & Pervasive Networks
Volume 18, No. 2 (2023) pp. 49-59

Predicting Cardiovascular Events with Machine Learning Models and Heart Rate Variability

Mohammad Moshawrab [a*], Mehdi Adda [a], Abdenour Bouzouane [b], Hussein Ibrahim [c], Ali Raad [d]

[a] *Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, G5L 3A1, Québec, Canada*
[b] *Département D'informatique et de Mathématique, Université du Québec à Chicoutimi, Chicoutimi, 555 boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada*
[c] *Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, G4R 5B7, Québec, Canada*
[d] *Dean of the Faculty of Science and Arts, Islamic University of Lebanon, Wardaniyeh, Lebanon*

Abstract

Artificial Intelligence (AI) is increasingly becoming a potential answer to many of science's most challenging problems. In this context, healthcare is using this technology and its advancement to improve the quality of services provided, including cardiac healthcare services. According to studies, Cardiovascular Diseases (CVDs) are among the most common and deadly diseases in the world. However, Artificial Intelligence and its branches such as Machine Learning (ML) and Deep Learning (DL) offer tremendous potential to improve disease diagnosis and even predict its occurrence. In this study, eight Machine Learning and Deep Learning models are created and trained with "PhsyioNet Smart Health for Assessing the Risk of Events via ECG Database" to analyze the characteristics of Heart Rate Variability and predict the occurrence of heart disease and cerebrovascular events. The results support the use of Artificial Intelligence in cardiology, with five of the proposed models outperforming previous implementations. Specifically, Support Vector Machines, TabTransformers, Deep Neural Networks, AdaBoost, and XGBoost achieved accuracy rates of 91.80%, 90.38%, 90.19%, 89.50%, and 89.10%, respectively. Further performance metrics are presented throught the article such as precision, recall and others.

Keywords: *Artificial Intelligence, Machine Learning, Deep Learning, Cardiovascular Diseases, Heart Rate Variability*

1. Introduction

Cardiovascular Disease causes the most deaths and is therefore considered the most dangerous disease in the world. According to the latest data from the World Health Organization (WHO) in the field of heart disease, the number of deaths caused by these diseases has increased from 12.1 million in 1990 to 18.6 million in 2019, accounting for 32% of global mortality in 2019. In addition, CVDs is a significant source of health conflict and economic hardship. Based on the Medical Expenditure Panel Survey, the cost of CVDs in the United States between 2017 and 2018 was estimated at $378.0 billion, including $226.0 billion in expenditures and $151.8 billion in lost future productivity [1, 2].

1.1. AI in Healthcare: A New Cardiology Era

The potential for AI to automate processes, enhance decision-making, and enable new discoveries has broad implications, with possible applications in healthcare [3],transportation [4], industry [5], luxury [6] and more. Smart health, for instance, is the use of computational methods, data analysis, and artificial intelligence to the healthcare industry with the goal of enhancing patient care, administrative efficiency, and clinical results [3] and in enabling diseases prediction. However, the deadly nature of Cardiovascular Diseases necessitates the development of effective solutions that can help in the early detection of these diseases and, if possible, even predict their development. Electrocardiogram, Echocardiogram, Coronary Angiography, stress test, Magnetic Resonance Imaging or Intracoronary Ultrasound are traditional methods to

*Corresponding author. Tel.: +1234567890
Fax: +9876543210; E-mail: Author.Name@iasks.org

DOI: 10.5383/JUSPN.18.02.001

detect these diseases. However, technological advances, particularly Information and Communication Technologies (ICTs) and the growth of Artificial Intelligence and its derivatives, are improving the quality of healthcare services and facilitating the detection of CVDs. In addition, AI technologies are considered the next revolution in cardiology because they can accelerate and improve patient care outcomes. Moreover, AI will soon transform the field of cardiovascular health, as its tools could outperform specialists in detecting or even predicting CVDs [7, 8].

1.2. Heart Rate Variability as a CVD Indicator

Recently, there has been a rise in interest in using Heart Rate Variability (HRV) as a predictor of CVDs, particularly with the advent of AI and the data analysis capabilities afforded by its branches: Machine Learning and Deep Learning. Furthermore, HRV is defined as the beat-to-beat variation in heart rate or the length of the RR peak interval, where R is a QRS complex wave taken from a cardiac ECG signal. Because changes in the autonomic control of the heart may be interpreted from temporal fluctuations in heart rate, the parameters retrieved from HRV data are classified into three types: time domain, frequency domain, and non-linear parameters [9]. Table 1 below lists these categories:

Table 1. Heart Rate Variability Parameters.

Group	Parameter	Unit	Description
Time Domain Parameters	Mean NN	(ms)	Mean of NN interval
	SDNN	(ms)	Standard deviation of NN intervals
	RMSSD	(ms)	Square root of the mean squared differences of successive NN intervals
	pNN50	(ms)	Proportion of interval differences of successive NN intervals greater than 50 ms
Frequency Domain Parameters	VLF	(ms^2)	Power in very low frequency range (0–0.04 Hz)
	LF	(ms^2)	Power in low frequency range (0.04–0.15 Hz)
	HF	(ms^2)	HF ms2 Power in high frequency range (0.15–0.4 Hz)
	LF/HF	(ratio)	Ratio of LF over HF
Non-Linear Parameters	SD1	(ms)	Standard deviation of points perpendicular to the axis of line of identity or the successive intervals scaled by $\sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}var(RR_{\mathrm{n}} - RR_{\mathrm{n+1}})}$
	SD2	(ms)	Standard deviation of points along the axis of line of identity, or $\sqrt{2SDNN^2 - \frac{1}{2}SD1^2}$
	SD1/SD2	(ratio)	Ratio of SD1 over SD2

1.3. Prediction of CVDs with HRV; State of the Art

There has been a surge in recent years in the number of researches looking at the ability to diagnose CVDs by measuring HRV characteristics. AI has demonstrated its efficacy and precision in this field, and researchers are increasingly turning to AI models to examine a wide range of HRV data.

In [10], for instance, the authors constructed a model to assess several HRV variables and predict the onset of ventricular tachycardia (VT) using the Fast Fourier Transform (FFT) and the Blackman Harris window technique. Additionally, the authors in [11] created an Artificial Neural Networks (ANN) classifier to predict the incidence of VT and trained it using the "PhysioNet Spontaneous Ventricular Tachyarrhythmia Database" [12]. They used a number of different criteria to assess performance, recording rates of 76.60% for accuracy, 82.9% for sensitivity, and 71.4% for specificity. In addition, the authors of [13] employed Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Support Vector Machines (SVM) to make predictions about cardiovascular risk. Accuracy of their best model was 96.67%. Additionally, authors in [14] employed SVM to create a prediction model to predict cardiovascular risk following Myocardial Infarction, and the model accuracy was 89%.

In addition, the authors of [15] used the k-Nearest Neighbor and Multilayer Perceptron Neural Network algorithms to develop models that predict Sudden Cardiac Death (SCD). Their models were trained using the "PhysioNet Sudden Cardiac Death Holter database" [16] and the "PhysioNet Normal Sinus Rhythm database" [17], and their results showed an accuracy of 99.73% for the first minute, 96.52% for the second minute, 90.37% for the third minute, and 83.96% for the fourth minute. And in [18], authors performed the same study using SVM and Probabilistic Neural Network (PNN) to predict SCD two minutes beforehand. SVM and PNN achieved 96.36% and 93.64% accuracy in predicting sudden cardiac death using the "PhysioNet Sudden Cardiac Death Holter database" [16] and the "PhysioNet MIT Normal Sinus Rhythm database" [17].

Besides, in [19], the authors developed a novel SVM, Tree-Based Classifier, Artificial Neural Network, and Random Forest models to automate cardiovascular risk classification for hypertension patients. Using the "Smart Health for Assessing the Risk of Events through ECG database" [20], the authors were able to train their data with a sensitivity of 71.4% and a specificity of 87.8%. In addition, authors in [21] created an Artificial Neural Networks model that examines respiratory rate in addition to HRV data to identify ventricular tachycardia an hour before it manifests. Their model has a sensitivity of 88%, a specificity of 82%, and an area under the curve of 93%. The authors in [22] also employed a statistical model called MIL to predict CVDs using HRV characteristics. As they noted, their model was quite accurate. In addition, the authors of [23] developed and trained a variety of classification methods, including K Nearest Neighbor, Decision Tree, Naive Bayes, Logistic Regression, Support Vector Machine, Neural Network, and Vote. They used the "UCI Heart Diseases Repository" [24], to train their models. It was shown that the models had an accuracy of 87.4% in predicting CVDs.

1.4. Outline & Main Contributions

Several AI models were developed in this study to predict CVDs and related events where eight different models were implemented. The dataset and the preparation processes that were performed to get the data ready for the models are described in Section 2. below. A description of the models developed may be found in Section 3., while Section 4. contains a listing and discussion of the results.

Despite the fact that several Machine Learning implementations have been performed in CVD detection and prediction, this article aims to propose ML models that have either never been used in this field or to propose models already in use and improve their performance. Therefore, this article aims to propose ML models capable of predicting CVDs with improved performance that outperforms previous implementations. The result obtained by the models is a binary result, stating whether a CVD is detected or not. The article thus contributes to the ML field in predicting CVDs:

- Proposing use of new models in the prediction of CVDs
- Enhancing and boosting the performance of ML in CVDs

2. Materials & Methods

2.1. Dataset

The dataset used in this study is the "PhysioNet Smart Health for Assessing the Risk of Events via ECG Database" (SHAREEDB) [20] that is offered by the PhysioNet online data repository. This dataset was collected to investigate the efficiency of classifying hypertensive patients at higher risk for cardiac and cerebrovascular events using HRV characteristics. It consists of 139 records of 24-hour Electrocardiographic (ECG) Holter recordings, each containing three ECG signals sampled at a rate of 128 samples per second with a precision of 8 bits. The population from which the data were gathered included 49 women and 90 men aged 55 and up. They were followed up for 12 months to record the occurrence of cardiovascular and cerebrovascular events. During the follow-up period, 17 patients experienced such event, including 11 Myocardial Infarctions, 3 strokes, and 3 syncopal events. The dataset also includes some demographic and clinical information about the subjects, such as their age, sex, any vascular events, and others. Figure 1 below describes the specifications of the dataset in use: SHAREEDB.

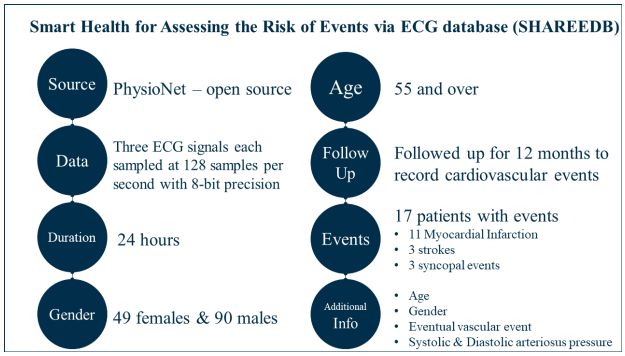

Fig. 1. SHAREEDB Description and Specs.

2.2. Data Filtering & Preprocessing

Since the SHAREEDB dataset contains Electrocardiogram (ECG) signals gathered in a laboratory, there may be substantial background noise that must be eliminated before the data is fed into the AI models. Before feeding the data into the models, it is crucial to clean the data and eliminate the noise in order to produce high-quality ECG signals. Briefly described below are the procedures used to clean and prepare the data for this study:

2.2.1. Filtering & Artifacts Removal

The 3-channel ECG Holter device was used to record the data included in the dataset files. A normal ECG signal has a frequency range of 0.05 Hz to 100 Hz. However, there are a number of signals interreferences that may affect ECG recordings, including baseline drift, channel interference, power line interference, muscle movement interference, and electrode contact interference. Raw ECG readings often include two forms of noise [25]:

- High frequency noise: current conduction noise, white Gaussian noise, Electromyogram or motion noise.
- Low-frequency noise: baseline drift and electrode contact loss

We can successfully identify the kind of noise and then pick the approaches to employ to decrease the noise or eradicate the artifact if we have a thorough grasp of each noise artifact. Various sounds are caused by various things, including:

- Power Line Interference: is caused by harmonics of electromagnetic interference through the power line and the electromagnetic field of nearby electrical equipment and is between 50 Hz and 60 Hz.
- White Gaussian Noise: is similar to channel noise in nature but is difficult to identify its sources because they occur at different levels and are random in nature.
- Electromyogram/Motion Noise: generated by the electrical activity of the muscles or the change in the electrode-skin impedance due to changes in skin temperature, humidity, etc.
- Baseline Drift: low-frequency noise, typically around 1 Hz and caused by respiration and rapid body movements
- Electrode Contact Loss: is caused by loss of contact between the electrode and the skin

Because of this, the following filters are effective in getting rid of both low-frequency and high-frequency artifacts and has been adopted, in this study, to clean the data before being used:

- IIR Notch Filters: are used to remove power line interference and/or motion artifacts in a specific frequency spectrum
- FIR Filters: are very stable filters and operate in the range of 1 Hz to 100 Hz making them suitable for ECG data cleaning

2.2.2. R Peaks Detection

The electrical activity of the heart muscle may be seen in an ECG signal throughout time. The ECG represents the amplified sum of the electrical depolarization of muscle cells that causes the heart muscle to contract during a certain time period. Three components make up the electrocardiogram signal: The P-Wave, the QRS complex, and the T-wave. The ventricular depolarization represented by the QRS complex is the electrical impulse as it travels through the ventricles. Immediately following each other in rapid succession are the Q wave, the R wave, and the S wave. Because HRV is defined as the difference between two successive RR periods, the R Peaks are the peaks to be discovered in this investigation. The R Peaks may be found using any of the available detection methods [26, 27]. These algorithms include:

- Hamilton
- Christov
- Engelse and Zeelenberg
- Pan and Tompkins
- Stationary Wavelet Transform
- Two Moving Average

According to [27], Engelse and Zeelenberg was selected as the most accurate peak detection algorithm. Although the tests were performed on a different data set, Engelse and Zeelenberg was selected for R peak detection in this study based on the recommendation of authors.

2.2.3. Calculation of RR Intervals

Heart Rate Variability is defined as the RR intervals or the difference between two consecutive R peaks, which are then calculated using the required equations.

2.2.4. Outliers Removal

After the RR intervals are detected, the outliers, defined as points that are extremely far from the mean, are removed and replaced with the mean value.

2.2.5. Extract HRV features

Finally, the HRV features were calculated using the appropriate mathematical formulas. In this study, 26 HRV features were calculated, and despite the high number of features calculated, the use of all features gave good results.

2.3. Artificial Intelligence Models

Cardiology is defined as the healthcare sector concerned with heart health, and the usage of AI in this discipline is rapidly expanding. AI has showed excellent accuracy and efficiency in identifying CVDs, and owing to its strong capacity to evaluate cardiac data, it may sometimes go beyond professional diagnosis and even be utilized in predicting CVDs rather than detecting them [28, 29]. Furthermore, AI is notable for its diverse branches that are applied in various aspects of life all over the globe. Figure 2 below shows the different branches of AI. In this research, AI branches such as Machine Learning, Ensemble Learning, and Deep Convolutional Neural Networks were applied:

- **Classical Machine Learning Algorithms [30]:** are algorithms that give computers learning potential by training them with experimental data and generating models based on these data, enabling them to make decisions in new situations such as: Support Vector Machines, Naïve Bayes, Logistic and Linear Regression and others.
- **Ensemble Learning [31]:** is a special branch of ML where its algorithms are based on merging predictions from different models. Some of these models are XGBoost, AdaBoost, Gradient Boosting, LightGBM and others.
- **Deep Convolutional Neural Networks (DCNNs) [30]:** are a type of Neural Networks that are used to analyze data with a grid-like structure. However, these networks are intended for analyzing multidimensional data such as images and videos. Using these networks to analyze tabular data may require transforming the data used. Nevertheless, there are several models that offer transformation of tabular data for use in DCNNs, such as TabNet, GrowNet, TreeEnsemble Layers, TabTransformers, Self-Normalizing Neural Networks, Neural Oblivious Decision Ensembles (NODE), AutoInt, and Deep & Cross Neural Networks (DCNs) [32].

3. Construction of AI Models

In this study, different AI models were used to analyze HRV features to detect heart diseases and events. However, before passing the extracted features to the models, some data fitting steps were performed, as explained below.

3.1. Data Adjustment

Given that only 17 of the 139 patients in the research suffered a cardiovascular event throughout the 12-month follow-up period,

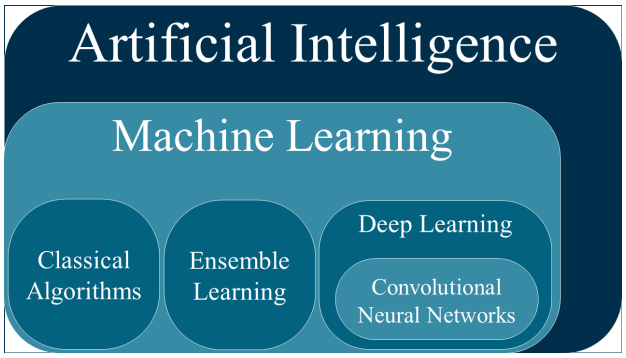

Fig. 2. SHAREEDB Description and Specs.

the retrieved HRV features show an unbalanced identity, with the majority falling into the "no cardiovascular event" class. Because the percentage of non-defected subjects is 122 of 139, the performance of the prediction models may be harmed, implying the usage of data modifications such as balancing and scaling:

- **Synthetic Minority Over-sampling Technique (SMOTE):** a data expansion in which new samples are drawn from existing ones to oversample the minority class
- **Preprocessing Standard Scaling:** the standardization of characteristics is achieved by removing the mean of the data and scaling it to a unit variance

3.2. Building the Models; Hyperparameters to be Considered

After applying the necessary data fitting steps to the extracted HRV features, they are then passed to the models created for fitting with the thresholds listed below:

3.2.1. Support Vector Machines

SVM is a supervised Machine Learning algorithm that is fed labeled training data to learn how to assign labels to objects based on examples, and then gain the ability to predict the category of new example(s) [30]. The performance of the SVM model is affected by the following hyperparameters [33]:

- **Kernel:** the function that converts the input data into the required form such as linear, polynomial and radial basis function (RBF).
- **Regularization:** denotes the misclassification or error term and is expressed as hyperparameter "C".
- **Gamma:** interpret how far the effect of a single training sample extends
- **Class Weight:** used for imbalanced datasets and defines the weight of the classes to be predicted

3.2.2. TabTransformers

TabTransformers is a model based on transformers whose layers convert categorical feature embeddings into robust contextual embeddings to achieve higher prediction accuracy, and is affected by the following hyperparameters [34]:

- **Activation Function:** defines how the weighted sum of the input is converted into an output of a node in a network layer
- **Number of Heads:** specifies the number of heads of attention

- **Dropout:** regularization to reduce overfitting and improve generalization of deep neural networks
- **MLP Hidden Units Factors:** MLP hidden layer units, as factors of the number of inputs.
- **Learning Rate:** is the shrinkage step size used in updating to prevent overfitting

3.2.3. Deep Neural Networks

These networks are algorithms that mimic human brain cells called neurons. In general, these networks use brain simulations to improve their learning and increase the accuracy of the models. The structure of DNNs consists of more than two interconnected layers and is affected by the following hyperparameters [35]:

- **Number of Layers:** input, output, and the hidden layers that define the structure of the network.
- **Units:** denotes the output of each layer.
- **Activation Function:** also known as the "transfer function", which defines how the weighted sum of the input is converted into an output from one or more nodes in a layer of the network
- **Number of Epochs:** a complete pass through all rows of the training data
- **Batch Size:** samples that the model examines within each epoch before updating the weights
- **Learning Rates:** a variable that controls how the optimizer's learning rate changes over time
- **Momentum:** is the "delay" in learning the mean and variance

3.2.4. AdaBoost

AdaBoost is a meta-estimator that first fits a classifier to the original data and then fits additional copies of the classifier to the same data, changing the weights of misclassified instances so that subsequent classifiers examine them extensively, leading to an improved result [36]:

- **Number of Estimators:** the number of base estimators or weak learners to be used in the dataset
- **Learning Rate:** is the step size used in the update to prevent overfitting

3.2.5. XGBoost

XGBoost is an Ensemble Learning algorithm that also belongs also to the Machine Learning AI Branch. XGBoost, eXtreme Gradient Boosting package, is a scalable implementation of the gradient boosting framework built with an efficient linear model solver and a tree learning algorithm with hyperparameters [37]:

- **Booster:** the type of model to run at each iteration
- **Learning Rate:** is the step size shrinkage used during the update to prevent overfitting
- **Gamma:** specifies the minimum loss reduction required to perform splitting
- **Max Depth:** the parameter used to control overfitting, as a higher depth allows the model to learn relationships that are very specific to a given sample
- **Min Child Weight:** defines the minimum sum of weights of all observations required in a child
- **Max Delta Step:** makes updating more conservative
- **Sub Sample:** denotes the fraction of observations that are randomly selected for each tree
- **Lambdas:** is used to handle the regularization part
- **Alpha:** is used in case of very high dimensionality to make the algorithm run faster during implementation
- **Tree Method:** Algorithm for tree construction
- **Scale Weight:** control the weight of positive-negative classes
- **Objective:** defines the loss function to be minimized

3.2.6. Logistic Regression

Logistic Regression is a Machine Learning algorithm that analyzes data for classification and is a supervised algorithm that sorts data into two categories. The algorithm is named after the function that is at the core of the method, the logistic function. There are several forms for LR and in this article we will use binary logistic regression, where the target variable has only two possible outcomes. The performance of LR is affected by three important hyperparameters [38]:

- **Solver:** uses a Coordinate Descent (CD) algorithm that solves optimization problems by successively performing approximate minimization along coordinate directions or coordinate hyperplanes
- **Penalty (Regularization):** is any modification of a learning algorithm that aims to reduce its generalization error, but not its training error
- **C:** the inverse of the regularization strength in Logistic Regression
- **Class Weight:** weight of the classes to be predicted

3.2.7. TabNet

TabNet is a model that uses sequential attention to select which features to infer at each decision step, and is influenced by the following hyperparameters [39]:

- **Optimizer:** an algorithm that modifies the neural network attributes, such as weights and learning rate.
- **Learning Rate:** is the step size used in updating to prevent overfitting
- **Batch Size:** number of examples per batch.

3.2.8. Deep Convolutional Neural Networks: Neural Oblivious Decision Ensembles (NODE)

Neural Oblivious Decision Ensembles is a model with a layered structure built from differentiable oblivious trees, which are decision tables that decompose the data along dd-splitting features and compare each feature to a learned threshold. It was trained in an end-to-end manner using backpropagation and is affected by the following hyperparameters [40]:

- Number of Layers: Number of layers forming the Neural Network
- Number of Trees: Number of trees in each layer
- Depth: Depth of the tree
- Learning Rate: is the shrinkage step size used in the update to prevent overfitting

3.3. Technical Environment Specifications

To implement this study, the computer used carried the below mentioned specifications:

- **Hardware Specs:**
  - **CPU:** Intel(R) Core i7-7500U CPU @ 2.70GHz
  - **RAM:** 16.0 GB DDR4
- **Operating System:**Windows 10 Home
- **Programming Language Used:** python 3.9
- **Libraries Used:**
  - **wfdb:** used to read data from the PhysioNet binary files [41]
  - **Scipy Signal Library:** provides efficient functions for both IIR Notch and FIR filters [42]
  - **py-ecg-detectors:**provide R Peaks detection algorithms [43]
  - **Scipy Zscore:** used for outliers' removal [44]
  - **SMOTE:** to apply Synthetic Minority Over-sampling [45]
  - **SKLearn Preprocessing Standard Scaling:** to apply standard scaling [46]

3.4. Wrapping Up, Training, Prediction, and Optimization

Once the models were created, they were trained using the extracted HRV features. The models were then evaluated using several performance metrics, namely accuracy, precision, recall, F1 score, specificity, and negative predictive value. The results obtained are detailed and discussed in the next section. Figure 3 below describes the overall structure of the implemented system.

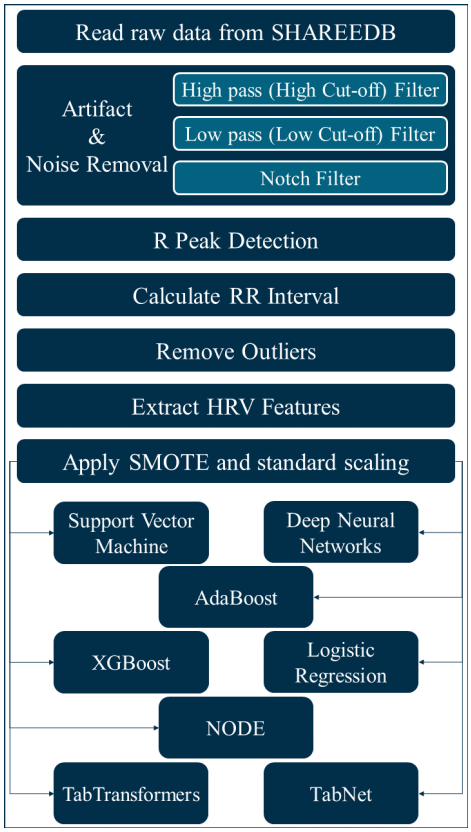

Fig. 3. Overall Architecture Followed in this Study.

4. Results & Discussion

The created models were trained with the HRV features. The eight models were evaluated with the metrics of Accuracy, Precision, Recall, Specificity, Negative Predictive Value NPV, and F1 Score. For better measurement, and to be aware of overfitting, Repeated K-fold Cross Validation [47] was implemented with 10 folds and repeated 5 times. Beside detection of overfitting, the use of K-fold cross validation ensure that the recorded results are not obtained from an optimistic execution. Consequently, the performance graphs are illustrated in Figure 4, 5 & 6 respectively, where the first shows the graphs for classical ML models, the second shows the graphs related to Ensemble ML models and the third shows the graphs of the DL models. In addition, the results are shown in Table 2 below, and the values of accuracy, precision, recall, specificity, negative predictive value, and F1 score are denoted as AC, PR, RE, SP, NPV, and F1, respectively. In addition, the values of the hyperparameters used are listed in the table.

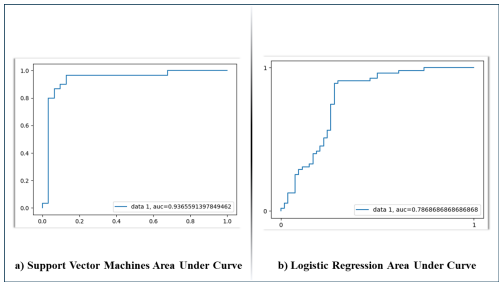

Fig. 4. Classical ML Models Performance Graphs.

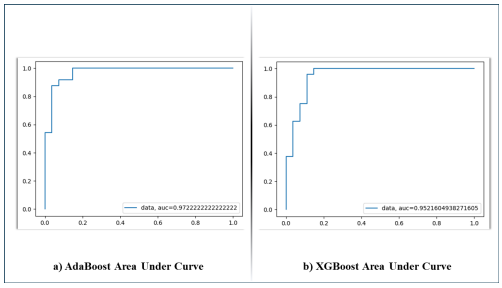

Fig. 5. Ensemble Learning ML Models Performance Graphs.

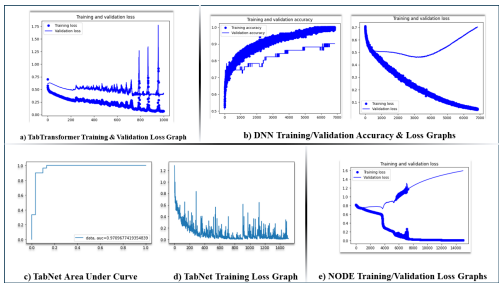

Fig. 6. Deep Learning Models Performance Graphs.

Table 2. AI Models Evaluation Metrics & Hyperparameters Used.

#	Model	Hyperparameters Used		AC	PR	RE	SP	NPV	F1
		Parameter	Value						
1	Support Vector Machines	Train/Test Split	0.75(0.25)	91.80%	87.87%	96.66%	87.09%	96.42%	92.06%
		Kernel	rbf						
		Regularization (C)	2.66						
		Gamma	0.141						
2	TabTransformers	Train/Test Split	0.72(0.28)	90.38%	86.66%	96.29%	91.22%	84%	95.45%
		Activation Function	Sigmoid						
		Number of Transformer Blocks	1024						
		Number of Attention Heads	1024						
		Dropout Rate	0.25						
		MLP Hidden Units	[1024,512]						
		Learning Rate	0.05						
		Epochs	1000						
3	Deep Neural Networks	Train/Test Split	0.79(0.21)	90.19%	85.18%	95.83%	85.18%	95.83%	90.19%
		Layers	Input/3 Hidden/Output						
		Units	512/256/128/64/1						
		Activation Function	tanh/tanh/tanh/sigmoid						
		Dropout	Before Output Layer 0.2						
		Optimizer	SGD						
		Epochs	6850						
		Batch Size	250						
		Learning Rate	0.005						
		Momentum	default						
4	AdaBoost	Train/Test Split	0.79(0.21)	89.50%	87.20%	94.60%	84.90%	93.60%	90.80%
		Number of Estimators	200						
		Learning Rate	1						
5	XGBoost	Train/Test Split	0.79(0.21)	89.10%	86.00%	93.80%	85.10%	92.50%	89.10%
		Booster	gbtree						
		Learning Rate	0.01						
		Gamma	0.1						
		Maximum Depth	10						
		Minimum Child Weight	0.01						
		Max Delta Step	0						
		Sub Sample	0.75						
		Lambda	1						
		Alpha	0.01						
		Tree Method	Auto						
6	Logistic Regression	Train/Test Split	0.71(0.29)	80.73%	76.56%	89.09%	72.22%	86.66%	82.35%
		Solver	newton-cg						
		Regularization (Penalty)	none						
		C	3.1						
7	TabNet	Train/Test Split	0.79(0.21)	76%	74.70%	82.60%	74.70%	80.50%	76.50%
		Learning Rate	0.9						
		Batch Size	1024						
		Virtual Batch Size	1024						
8	NODE	Train/Test Split	0.71(0.29)	76.92%	77.77%	73.68%	80%	76.19%	75.67%
		Number of Layers	5						
		Depth	10						
		Number of Trees (per layer)	1						
		Learning Rate	0.1						
		Batch Size	26						
		Epochs	9000						

4.1. Discussion

In this study, several models were created to analyze HRV characteristics to detect cardiovascular risks. The results obtained demonstrate the high efficiency of AI models in predicting cardiovascular disease. However, the results obtained in this study outperformed previous implementations.

First, the authors in [19] applied similar models to the same dataset. Nevertheless, the results obtained in this study exceeded their results. For example, their SVM model recorded accuracy, recall and specificity results were 89.00%, 86.30% and 91.80% respectively, whereas our results are 91.80%, 96.66% and 87.09% for the same performance metrics. In addition, the performance metrics of their Multi-Layer Perceptron model were Accuracy: 78.10%, Recall: 86.30%, Specificity: 69.90% and our model recorded 90.19%, 95.83% and 85.18% for the same metrics.

In addition, our SVM model achieved 91.80% accuracy, the highest performance among all previous implementations. For example, the SVM models in [13] recorded an accuracy of 88.64%, 82.95% and 82.58% for the Linear, Polynomial and RBF kernels, respectively. Moreover, the accuracy of SVM in [14, 19, 23] was 79.81%, 89.00% and 85.19%, respectively. Even though the accuracy is close, other metrics such as Precision and Recall clearly outperform the previous results by a large margin. Knowing that Recall measures how a model correctly classifies True Positives, the models presented in this study are more accurate in predicting whether a person will have a CVD in future. The high recall for SVM, DNN, and XGBoost, which are 96.66%, 95.83% and 93.80%, respectively, reflects the highest ability of all implementations to correctly predict that a person is in the cardiovascular risk zone.

Likewise, the DNN model presented in this article also outperforms all previous implementations. The accuracy of this model is 90.19%, whereas the multilayer perceptron in [13, 19] is 86.67% and 78.10%, and the accuracy of artificial neural networks in [11, 21] is 76.60% and 85.30%. Moreover, precision and recall are significantly higher than the previous implementations, which also reflects a higher capability in cardiovascular risk detection. Table 3 provides a detailed comparison between the results of the models presented in this article and the previous implementations. The symbols of the performance metrics used in this table are similar to those in Table 2, and an "NA" symbol indicates that the corresponding metric was not mentioned in the associated study.

On the other hand, none of the previous implementations used XGBoost, which also outperformed the previous implementations with an accuracy of 98.10% and a recall of 94.60%, reflecting high efficiency in predicting cardiovascular risk, in contrast to the implementation of NODE, which achieved an accuracy of 76.92%, which is not comparable with the previous implementations.

Finally, the SVM, DNN, and XGBoost models discussed in this study can be considered the most accurate models for predicting cardiac disease and events. Even the implementations in [15, 19] had higher accuracy and relatively higher recall, but their models were developed to detect Sudden Cardiac Death only minutes before its occurrence. For example, the model mentioned in [15] achieved 99.73% accuracy in predicting sudden cardiac death one minute before its onset, but the performance drops to 83.93% when the event is predicted four minutes before its occurrence. However, the models presented here are able to predict cardiovascular disease 12 months before its onset, demonstrating high efficiency in predicting cardiovascular disease and cardiac events long before their onset, thus increasing confidence in the use of AI in detecting and predicting cardiac disease and related events.

4.2. Challenges & Future Recommendations

Although Machine Learning are ready to play a significant role in predicting CVDs, there are a number of potential obstacles that might occur in the course of their deployment. What follows are some of the most typical problems that arise in such a setting:

- Data Readiness and Availability: Data determines machine learning model performance. The availability of more data will help in improving the performance of the smart models and therefore increase their accuracy in predicting CVDs. However, the availability of data is prone to different problems such as the legal or ethical restrictions. However, assumed available and accessible, the data to be used may be noisy since digital ECG recordings are more vulnerable to environmental noise. Artefacts—unwanted signals or signal distributions—interfere with the signal in noisy data. In this context, Intrinsic Artefacts come from the monitored body, whereas Extrinsic Artefacts come from their surroundings [48, 49]
- Data Privacy and Confidentiality: Although the technical structure of the models, data cleanliness and readiness, and other factors affect model accuracy, more data to train AI models usually improves their accuracy. For privacy and secrecy considerations, gathering data is the largest hurdle in constructing AI models in the real world. Society, governments, and organizations are enhancing data privacy and security. The European Union's General Data Protection Regulation (GDPR) [50], China's Cyber Security Law [51] and hundreds of other principles have been legislated worldwide. These restrictions safeguard private data, but also make it harder to gather data to train models, which makes it harder to increase model performance [49]
- Users Acceptability: User acceptability, adoption, and engagement are of the most significant obstacles to using AI and its branches to identify CVDs. Using those technologies to predict illnesses has met with mixed reception from users owing to concerns about privacy, discomfort, and other contextual factors
- Additional Computation-Cost: Due to the additional computing imposed by the added tasks such as data balancing and noise removal, an increase in computation time is obtained, and thus this imposes additional slowdowns that may impair the overall performance of the models

However, several approaches have been made to resolve those challenges in the attempt to enhance the feasibility of using AI and its descendants to predict cardiac illness. Those solutions are considered as hot topics that are being studied carefully nowadays:

- Automating Noise Removal: Before processing the signals, artifacts, both extrinsic and intrinsic, that obfuscate the signals should be eliminated or greatly reduced. This goal has already been accomplished by a number of existing solutions, some of which are discussed in Refs. [52]. Thus, research into automated noise reduction to clean and preprocess the data to enhance the precision of physical tiredness detection in the workplace is warranted
- Privacy Preserving: Data used in Machine Learning models training should be stored on a local server or distributed to decentralized storage and processing devices to construct and train the models. Thus, the model has complete access to the subject's data, whether anonymous or labelled by the subject. Federated learning (FL) may address this issue. Federated learning is a defined as collaborative distributed/decentralized machine learning privacy-preserving method that trains models without transferring data from edge devices to a central server. Instead, edge devices communicate learned models with the central server, which works as an aggregation station to create the global model without understanding the embedded data [53, 54]. The use of Federated Learning into CVDs prediction would help resolve privacy issues and therefore resolve the challenges in this regard
- Increase Accuracy, Explainability and Trust: Predicting the onset of cardiovascular disease is crucial in light of the growing health burden caused by this condition. The black box nature of the models used, however, must be reduced as much as possible, and the accuracy of AI tools and procedures in this area must be increased. Devices that are more accurate and easier to explain will be more likely to be employed as a CVDs prediction device. In this context, several technologies can be adopted such as the one mentioned in [55] that automates assessing the quality of a smart model

Table 3. Comparison with Previous Implementations.

Study	Model	AC	PR	RE	SP	NPV	F1
Our Study	Support Vector Machines	91.80%	87.87%	96.66%	87.09%	96.42%	92.06%
	TabTransformers	90.38%	86.66%	96.29%	91.22%	84.00%	95.45%
	Deep Neural Network	90.19%	85.18%	95.83%	85.18%	95.83%	90.19%
	AdaBoost	89.50%	87.20%	94.60%	84.90%	93.60%	90.80%
	XGBoost	89.10%	86.00%	93.80%	85.10%	92.50%	89.10%
	Logistic Regression	80.73%	76.56%	89.09%	72.22%	86.66%	82.35%
	TabNet	76.00%	74.70%	82.60%	74.70%	80.50%	76.50%
	NODE	76.92%	77.77%	73.68%	80.00%	76.00%	75.67%
[11]	Artificial Neural Network	76.60%	70.70%	82.90%	71.40%	NA	NA
[13]	Support Vector Machines (Linear Kernel)	88.64%	90.84%	86.36%	90.91%	86.96%	NA
	Support Vector Machines(Polynomial Kernel)	82.95%	80.85%	79.55%	86.36%	85.37%	NA
	Support Vector Machines (RBFKernel)	82.58%	79.45%	77.27%	87.88%	86.44%	NA
	Multi Layer Perceptron (Top15 Features)	86.67%	100%	73.33%	100%	78.95%	NA
[14]	Support Vector Machines	79.81%	21.15%	91.67%	79.08%	99.36%	NA
[15]	MLP (A Minute Before the SCD Event)	99.73%	NA	NA	NA	NA	NA
	K-NN (A Minute Before the SCDEvent)	98.32%	NA	NA	NA	NA	NA
[18]	SVM (2 minutes before VF Event)	96.36%	NA	NA	NA	NA	NA
	Penalized Neural Network	93.64%	NA	NA	NA	NA	NA
[19]	Support Vector Machines	89.00%	NA	86.30%	91.80%	NA	NA
	Multi Layer Perceptron	78.10%	NA	86.30%	69.90%	NA	NA
[21]	Artificial Neural Network	85.30%	83.30%	88.20%	82.40%	87.50%	NA
[22]	MIL Statistics Algorithm	85.47%	92.11%	86.42%	83.33%	NA	NA
[23]	Vote	87.41%	NA	NA	NA	NA	NA
	Naïve Bayes	84.81%	NA	NA	NA	NA	NA
	Support Vector Machines	85.19%	NA	NA	NA	NA	NA

5. Conclusion

Ultimately, AI will determine the fate of humans, they say. However, the widespread adoption and use of these technologies today proves that they are no longer science fiction. The field of cardiology, as well as methods for diagnosing and treating Cardiovascular Disease, will benefit from the development of AI, which could one day enable accurate prediction of disease. Research has produced a number of models that can accurately predict the occurrence of cardiac problems or events, boosting confidence in AI and its applications in medicine. If these models are operational in real time, this will undoubtedly contribute to the development of personalized and continuous monitoring that can be used to monitor the heart health of patients or even the health of workers who work in stressful environments or for extremely long periods of time.

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number 06351], Fonds Québécois de la Recherche sur la Nature et les Technologies (FRQNT), and Centre d'Entrepreneuriat et de Valorisation des Innovations (CEVI).

References

[1] Roth, Gregory A., et al. "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study." Journal of the American College of Cardiology 76.25 (2020): 2982-3021.

[2] Tsao, Connie W., et al. "Heart disease and stroke statistics—2022 update: a report from the American Heart Association." Circulation 145.8 (2022): e153-e639.

[3] Patel, Vishv, Devansh Shah, and Nishant Doshi. "Emerging Technologies and Applications for Smart Cities." J. Ubiquitous Syst. Pervasive Networks 15, no. 02 (2021): 19-24.

[4] Karakostas, B. (2023). Control of autonomous UAV using an onboard LSTM neural network. Journal of Ubiquitous Systems & Pervasive Networks, 18(1), 09-14.

[5] Haricha, Karim, Azeddine Khiat, Yassine Issaoui, Ayoub Bahnasse, and Hassan Ouajji. "Towards smart manufacturing: Implementation and benefits." J. Ubiquitous Syst. Pervasive Networks 15, no. 02 (2021): 25-31.

[6] Zaki, Marquos, Ali Alquraini, and Tarek R. Sheltami. "Home Automation using EMOTIV: Controlling TV by Brainwaves." J. Ubiquitous Syst. Pervasive Networks 10, no. 1 (2018): 27-32.

[7] National Heart Lung and Blood Institute. "In Brief: Your guide to living well with heart disease (NIH Publication 06-5716)." Washington, DC: National Institutes of Health (2005).

[8] Johnson, Kipp W., et al. "Artificial intelligence in cardiology." Journal of the American College of Cardiology 71.23 (2018): 2668-2679.

[9] Rajendra Acharya, U., et al. "Heart rate variability: a review." Medical and biological engineering and computing 44.12 (2006): 1031-1051.

[10] Baumert, Mathias, et al. "Forecasting of ventricular tachycardia using scaling characteristics and entropy of heart rate time series." World Congress on Medical Physics and Biomedical Engineering 2006. Springer, Berlin, Heidelberg, 2007.

[11] Joo, Segyeong, Kee-Joon Choi, and Soo-Jin Huh. "Prediction of ventricular tachycardia by a neural network using parameters of heart rate variability." 2010 Computing in Cardiology. IEEE, 2010.

[12] "Spontaneous Ventricular Tachyarrhythmia Database v1.0." Spontaneous Ventricular Tachyarrhythmia Database v1.0, 2 May 2022, https://physionet.org/content/mvtdb/1.0. Accessed on 01 March 2022

[13] Ramirez-Villegas, Juan F., et al. "Heart rate variability dynamics for the prognosis of cardiovascular risk." PloS one 6.2 (2011): e17060.

[14] Song, T., Qu, X. F., Zhang, Y. T., Cao, W., Han, B. H., Li, Y., ... & Da Cheng, H. (2014). Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. BMC cardiovascular disorders, 14(1), 1-8.

[15] Ebrahimzadeh, E., Pooyan, M., & Bijar, A. (2014). A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. PloS one, 9(2), e81896.

[16] "Sudden Cardiac Death Holter Database v1.0.0." Sudden Cardiac Death Holter Database v1.0.0, 2 July 2004, https://physionet.org/content/sddb/1.0.0. Accessed 01 March 2022.

[17] "MIT-BIH Normal Sinus Rhythm Database v1.0.0." MIT-BIH Normal Sinus Rhythm Database v1.0.0, 3 Aug. 1999, https://physionet.org/content/nsrdb/1.0.0. Accessed 01 March 2022.

[18] Murukesan, L., et al. "Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features." Journal of Medical Imaging and Health Informatics 4.4 (2014): 521-532.

[19] Melillo, Paolo, et al. "Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis." PloS one 10.3 (2015): e0118504.

[20] "Smart Health for Assessing the Risk of Events via ECG Database v1.0.0." Smart Health for Assessing the Risk of Events via ECG Database v1.0.0, 19 May 2015, https://physionet.org/content/shareedb/1.0.0. Accessed 01 March 2022.

[21] Lee, Hyojeong, et al. "Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks." Scientific reports 6.1 (2016): 1-7.

[22] Lan, Kun-chan, et al. "Toward hypertension prediction based on PPG-derived HRV signals: A feasibility study." Journal of medical systems 42.6 (2018): 1-7.

[23] Amin, Mohammad Shafenoor, Yin Kia Chiam, and Kasturi Dewi Varathan. "Identification of significant features and data mining techniques in predicting heart disease." Telematics and Informatics 36 (2019): 82-93.

[24] "UCI Machine Learning Repository: Data Set." UCI Machine Learning Repository: Data Set, https://archive.ics.uci.edu/ml/datasets/heart+disease. Accessed 01 March 2022.

[25] Luo, Shen, and Paul Johnston. "A review of electrocardiogram filtering." Journal of electrocardiology 43.6 (2010): 486-496.

[26] Ashley, Euan A., and Josef Niebauer. "Cardiology explained." (2004).

[27] Eilers, Justus, Jonas Chromik, and Bert Arnrich. "Choosing the Appropriate QRS Detector." BIOSIGNALS. 2021.

[28] Mathur, Pankaj, et al. "Artificial intelligence, machine learning, and cardiovascular disease." Clinical Medicine Insights: Cardiology 14 (2020): 1179546820927404.

[29] Mathur, Pankaj, et al. "Artificial intelligence, machine learning, and cardiovascular disease." Clinical Medicine Insights: Cardiology 14 (2020): 1179546820927404.

[30] Ertel, Wolfgang. Introduction to artificial intelligence. Springer, 2018.

[31] Dong, Xibin, et al. "A survey on ensemble learning." Frontiers of Computer Science 14.2 (2020): 241-258.

[32] Gorishniy, Yury, et al. "Revisiting deep learning models for tabular data." Advances in Neural Information Processing Systems 34 (2021): 18932-18943.

[33] Noble, William S. "What is a support vector machine?." Nature biotechnology 24.12 (2006): 1565-1567.

[34] Huang, Xin, et al. "Tabtransformer: Tabular data modeling using contextual embeddings." arXiv preprint arXiv:2012.06678 (2020).

[35] Anderson, James A. An introduction to neural networks. MIT press, 1995.

[36] Schapire, Robert E. "Explaining adaboost." Empirical inference. Springer, Berlin, Heidelberg, 2013. 37-52.

[37] Chen, Tianqi, et al. "Xgboost: extreme gradient boosting." R package version 0.4-2 1.4 (2015): 1-4.

[38] Kleinbaum, David G., et al. Logistic regression. New York: Springer-Verlag, 2002.

[39] Arik, Sercan Ö., and Tomas Pfister. "Tabnet: Attentive interpretable tabular learning." Proceedings of the AAAI

Conference on Artificial Intelligence. Vol. 35. No. 8. 2021.

[40] Popov, Sergei, Stanislav Morozov, and Artem Babenko. "Neural oblivious decision ensembles for deep learning on tabular data." arXiv preprint arXiv:1909.06312 (2019).

[41] "Wfdb." PyPI, 27 June 2022, https://pypi.org/project/wfdb. Accessed on 10 March 2022

[42] Signal Processing (scipy.signal) — SciPy v1.9.3 Manual, 1 June 2022, https://docs.scipy.org/doc/scipy/reference/signal.html. Accessed on 05 June 2022

[43] "Py-ecg-detectors." PyPI, 1 June 2022, pypi.org/project/py-ecg-detectors.

[44] Scipy.stats.zscore — SciPy v1.9.3 Manual, 1 June 2022, https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.zscore.html. Accessed on 15 March 2022.

[45] "SMOTE — Version 0.9.1." SMOTE — Version 0.9.1, 1 Sept.2000, https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html. Accessed on 15 March 2022.

[46] "Sklearn.Preprocessing.StandardScaler." Scikit-learn, 1 Jan.2000, https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html. Accessed on 15 March 2022.

[47] Fushiki, Tadayoshi. "Estimation of prediction error by using K-fold cross-validation." Statistics and Computing 21.2 (2011): 137-146.

[48] Islam, Md, Amir Rastegarnia, and Saeid Sanei. "Signal artifacts and techniques for artifacts and noise removal." Signal processing techniques for computational health informatics. Springer, Cham, 2021. 23-79.

[49] Talha, Muhammad, Nabil Elmarzouqi, and Anas Abou El Kalam. "Quality and Security in Big Data: Challenges as opportunities to build a powerful wrap-up solution." J. Ubiquitous Syst. Pervasive Networks 12, no. 1 (2020): 9-15.

[50] Albrecht, Jan Philipp. "How the GDPR will change the world." Eur. Data Prot. L. Rev. 2 (2016): 287.

[51] Parasol, Max. "The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams." Computer law & security review 34.1 (2018): 67-98.

[52] Daly, Ian, et al. "On the automated removal of artifacts related to head movement from the EEG." IEEE Transactions on neural systems and rehabilitation engineering 21.3 (2013): 427-434.

[53] Mammen, Priyanka Mary. "Federated learning: opportunities and challenges." arXiv preprint arXiv:2101.05428 (2021).

[54] Yang, Qiang, et al. "Federated machine learning: Concept and applications." ACM Transactions on Intelligent Systems and Technology (TIST) 10.2 (2019): 1-19

[55] Kara, Madjid, Olfa Lamouchi, and Amar Ramdane-Cherif. "Software Quality Assessment Algorithm Based on Fuzzy Logic." J. Ubiquitous Syst. Pervasive Networks 8, no. 1 (2017): 1-9.

2022 3rd International Conference on Human-Centric Smart Environments for Health and Well-being (IHSH) | 978-1-6654-6321-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/IHSH57076.2022.10092060

Machine Learning Models to Predict Cardiovascular Events from Heart Rate Variability Data

Mohammad Moshawrab
Département de mathématiques, informatique et génie
Université du Québec à Rimouski
Rimouski, Canada
mohammad.moshawrab@uqar.ca

Mehdi Adda
Département de mathématiques, informatique et génie
Université du Québec à Rimouski
Rimouski, Canada
mehdi_adda@uqar.ca

Abdenour Bouzouane
département D'informatique et de Mathématique
Université du Québec à Chicoutimi
Chicoutimi, Canada
abdenour_bouzouane@uqac.ca

Hussein Ibrahim
Institut Technologique de Maintenance Industrielle
Sept-Îles, Canada
hussein.ibrahim@itmi.ca

Ali Raad
Dean of Faculty of Arts & Sciences
Islamic University of Lebanon
Wardaniyeh, Lebanon
ali.raad@iul.edu.lb

***Abstract*—Among the diseases known to mankind, cardiovascular diseases remain the deadliest and most expensive. However, Artificial Intelligence offers new solutions that can help diagnose these diseases and even predict their occurrence with high accuracy. In this study, we present several AI models that attempt to predict cardiovascular diseases. The models created are Support Vector Machines, AdaBoost, TabNet and TabTransformers. They were trained using Heart Rate Variability features extracted from the PhysioNet Smart Health for Assessing the Risk of Events via ECG Database. The models achieved high accuracies, which were 91.80%, 89.50%, 76.00% and 90.38% for the Support Vector Machines, AdaBoost, TabNet, and TabTransformers models, respectively.**

***Index Terms*—Cardiovascular Diseases, Artificial Intelligence, Machine Learning, Ensemble Learning, Deep Convolutional Neural Networks, Heart Rate Variability, Predictive Models**

I. Introduction

Cardiovascular Diseases (CVDs) are considered the world's deadliest disease, with most deaths resulting from heart disease or heart-related medical events. The World Health Organization (WHO) has mentioned in its reports that the number of CVD patients increased from 271 to 523 million worldwide between 1990 and 2019. Moreover, the number of deaths caused by CVDs increased from 12.1 to 18.6 million during the same period. In 2019, deaths due to cardiovascular diseases accounted for 32% of all global deaths. These figures demonstrate that heart disease is not only the leading cause of deaths but also of health burdens worldwide [1:3].

A. Artificial Intelligence and CVDs

Because of the severity of CVDs, it is important to detect them as early as possible so that treatment with counselling and medications can begin. The classical routines for diagnosing CVDs vary between electrocardiogram, stress test, magnetic resonance imaging, echocardiography, coronary angiography, or intracoronary ultrasound [4]. However, the rise of Information and Communication Technologies (ICT) combined with the rapid development of Artificial Intelligence (AI) and its offshoots such as Machine Learning (ML) and Deep Learning (DL), is changing several concepts that surround us. As a result, healthcare services are also being transformed by this technological revolution and have recently emerged as one of the largest areas of interest for technology research. AI tools are proving to be highly viable in improving disease diagnosis and treatment, Ambient Assisted Living (AAI), clinical robotics activities, medical research, and many other areas of healthcare. For example, the management of cardiovascular disease has been improved through the use of the latest technologies, and AI tools are even considered the next revolution in heart disease science by enabling faster, more accurate, and less error-prone patient care. Moreover, AI is expected to transform cardiology in the near future as its tools match and sometimes surpass expert performance on real-world data [5,6].

B. Heart Rate Variability for CVDs Diagnosis

Heart rate variability (HRV) is defined as the variation in heart rate from beat to beat or the duration of the peak interval. It is generally believed that the temporal variations in heart rate reflect changes in cardiac autonomic regulation. Detection of cardiovascular disease based on heart rate variability (HRV) characteristics has increased recently, especially with

the development of data analysis techniques offered by Machine Learning and Deep Learning. Parameters extracted from HRV data and analyzed for cardiovascular event detection are classified into three main groups: time-domain parameters, frequency-domain parameters, and nonlinear parameters [7]:

- Time Domain Features
  - Mean NN (ms): Mean of NN interval
  - SDNN (ms): Standard deviation of NN intervals
  - RMSSD (ms): Square root of the mean squared differences of successive NN intervals
  - pNN50 (ms): Proportion of interval differences of successive NN intervals greater than 50 ms
- Frequency Domain Features
  - VLF (ms^2): Power in very low frequency range (0–0.04 Hz)
  - LF (ms^2): Power in low frequency range (0.04–0.15 Hz)
  - HF (ms^2): Power in high frequency range (0.15–0.4 Hz)
  - LF/HF (ratio): Ratio of LF over HF
- Non-Linear Parameters
  - SD1 (ms): Standard deviation of points perpendicular to the axis of line of identity or Standard deviation of the successive intervals scaled by $\sqrt{\frac{1}{2}}$
  - SD2 (ms): Standard deviation of points along the axis of line of identity, or $\sqrt{2SDNN^2 - \frac{1}{2}SD1^2}$
  - SD1/SD2 (ratio): Ratio of SD1 over SD2

C. Prediction of CVDs with HRV; State of the Art

Several attempts have been made in the attempt to detect cardiovascular events using HRV. To this end, researchers have applied various artificial intelligence models to HRV in the time domain, frequency domain, and nonlinear features.

For example, the authors in [8] applied the Fast Fourier Transform (FFT) with Blackman Harris windowing to HRV features in the time domain. The authors created their model to predict the occurrence of Ventricular Tachycardia (VT) in the short term. Similarly, the authors in [9] used an Artificial Neural Networks (ANNs) classifier with the "PhysioNet Spontaneous Ventricular Tachyarrhythmia Database." [10] to predict the future occurrence of VT events. Model performance metrics were 76.60% 82.9% and 71.4% for accuracy, sensitivity, and specificity, respectively. In addition, in [11], in an attempt to classify cardiovascular risk prediction schemes, the researchers analyzed various HRV features using Multilayer Perceptron (MLP), Radial Basis Function (RBF), and Support Vector Machines (SVM), with the best model achieving 96.67% accuracy. Furthermore, in [12], the authors developed predictive models based on SVM to improve risk stratification after acute Myocardial Infarction with their model achieving 89% accuracy.

In addition, the authors developed a k-Nearest Neighbor (k-NN) and a Multilayer Perceptron Neural Network (MLP) as sudden cardiac death (SCD) prediction models in [13]. The models were compared with the "PhysioNet Sudden Cardiac Death Holter database." [14] and the "PhysioNet Normal Sinus Rhythm database" [15] and achieved an accuracy of 99.73%, 96.52%, 90.37%, and 83.96% for the first, second, third, and fourth one-minute intervals, respectively. Similarly, the authors in [16] used SVM and Probabilistic Neural Network (PNN) models to predict SCD two minutes before its occurrence. They trained the models using the "PhysioNet Sudden Cardiac Death (SCD) Holter database" [14] and the "PhysioNet MIT Normal Sinus Rhythm database" [15]. SVM and PNN achieved a maximum mean SCD prediction rate of 96.36% and 93.64%, respectively.

Furthermore, the authors developed in [17] models for automatic risk stratification of hypertensive patients using SVM, tree-based classifiers, Artificial Neural Network and Random Forest. The models were run with the "Smart Health for Assessing the Risk of Events via ECG database" [18] and achieved a sensitivity of 71.4% and a specificity of 87.8%. In [19], the authors also detected Ventricular Tachycardia one hour before its onset using an Artificial Neural Networks model that analyzed not only HRV features but also respiratory rate variability (RRV). The performance of the model was reported as sensitivity, specificity, and area under the curve as 88%, 82%, and 93%, respectively. In addition, the authors analyzed in [20] HRV characteristics using a statistical model called MIL, to predict CVDs, and their models achieved high accuracy. Finally, the authors used in [21] Data mining algorithms to predict CVDs such as K Nearest Neighbor (k-NN), decision tree, Naive Bayes, logistic regression (LR), support vector machine, neural network, and vote. Models were run with the "UCI Heart Diseases Repository" [22] and achieved an accuracy of 87.4%.

In this article, several AI models were used to predict cardiovascular diseases, namely, Support Vector Machines (machine learning), AdaBoost (ensemble learning), and TabNet and TabTransformers (deep convolutional neural networks). In Section 2, the dataset used and the preprocessing steps applied to it are discussed along with a brief explanation of the different AI branches. Section 3 explains the models used, while the results obtained are presented in Section 4.

II. Materials & Methods

A. Dataset

This study used the PhysioNet Smart Health for Assessing the Risk of Events via ECG Database (SHAREEDB) [18] database, which was collected to investigate the feasibility of classifying hypertensive patients at higher risk for cardiac events using the Heart Rate Variability features. This dataset consists of 139 records, each representing a 24-hour electrocardiographic (ECG) Holter recording. The 139 subjects who contributed to this dataset are 49 women and 90 men aged 55 years and older. Subjects were followed up for up to 12 months to record major cardiovascular and cerebrovascular events such as fatal or nonfatal acute coronary syndromes, including Myocardial Infarctions, syncopal events, Coronary Revascularizations, fatal or nonfatal strokes, and transient ischemic attacks. During the 12-month period, 17 cardiovascular events were recorded in 17 different subjects, namely 11 Myocardial Infarctions, 3 strokes, and 3 syncopal events. In addition, the 24-hour recordings include three ECG signals, each sampled at 128 samples per second with 8-bit accuracy. In addition, demographic and clinical information is provided with the

dataset such as: age, gender, any vascular event, systolic and diastolic arterial pressure values, and others.

B. Data Filtering & Preprocessing

It is well known that electrocardiography data are noisy for various reasons, and filtering of these data is necessary to obtain high-quality ECG signals for examination. In this study, the data is filtered and preprocessed before being fed into the AI models. The steps applied to prepare the data for the models are:

- **Filtering & Artifacts Removal [23]:** knowing that the frequency of a clean ECG signal is between 0.05 Hertz and 100 Hertz, the collected ECG signals can be interfered by various external factors. These interferences may result in high or low frequency noise, such as baseline drift, channel interference, current conduction interference, muscle motion interference, or electrode contact interference, which then require the application of various filtering methods, such as:
  - **IIR Notch Filters:** are effective in removing current conduction interference and/or motion artifacts.
  - **FIR Filters:** are filters that operate in the range of 1 hertz to 100 hertz
- **R Peaks Detection[24,25]:** The ECG signal represents the electrical activity of the heart muscle over time and is characterized by three parts: P-Wave, QRS complex, and T-Wave. The QRS complex includes the Q wave, R wave, and S wave, which occur in rapid succession. R-peaks are detected using several common algorithms such as Hamilton, Christov, Engelse and Zeelenberg, Pan and Tompkins, Stationary Wavelet Transform and Two Moving Average. According to [25], Engelse and Zeelenbergm have proven to be the most accurate algorithm for detecting R-peaks, which is why it was used in this study.
- **Calculation of RR Intervals:** RR intervals are defined as the difference between two consecutive R peaks and represent the plots of the Heart Rate Variability.
- **Outliers Removal:** After the RR intervals are calculated, they are corrected by replacing outliers, i.e. points that are extremely far from the mean, with the mean value
- **Extracting HRV Features:** Finally, 26 HRV features were extracted using appropriate mathematical calculations. Despite the high number of features, the obtained results prove the high efficiency of using all features.

C. Artificial Intelligence Smart Models

The use of artificial intelligence in healthcare has increased dramatically recently, especially in cardiology, the healthcare field that deals with heart disease. Due to the high accuracy of AI and its branches in detecting cardiovascular diseases, sometimes surpassing even human diagnosis, researchers' interest in this field is also increasing. The development of AI models has helped to develop accurate and efficient systems that can diagnose or even predict CVDs by analyzing HRV characteristics [26,27]. In addition, AI is currently known with its various branches that are used in almost all areas of life worldwide. For example, Machine Learning, Ensemble Learning and Deep Convolutional Neural Networks are branches of AI that were used in this study. Figure 1 shows the relationship between AI and these models, and a brief explanation of each model is provided below:

- **Classical Machine Learning Algorithms [28]:** give computers the ability to learn without being explicitly programmed by feeding them experimental data to generate models that enable them to evaluate new situations. There are many classical algorithms such as Support Vector Machines (SVM), Linear Regression, Naïve Bayes and others.
- **Ensemble Learning [29]:** is a branch of machine learning that aims to improve predictive performance by combining the predictions of multiple models. Its implementations are mainly divided into bagging, stacking, and boosting algorithms. Some ensemble learning algorithms are XGBoost, AdaBoost, GradientBoosting, LightGBM, and other
- **Deep Convolutional Neural Network [28,30]:** are networks that apply filters to input data to create a feature map that embodies the presence of detected features in the input. However, Deep Convolutional Neural Networks are designed to process images and videos that represent multidimensional input. Therefore, tabular data, such as that used in this study, must be transformed into a multidimensional form. However, this conversion is offered with different models such as: TabNet, GrowNet, TreeEnsemble Layers, TabTransformers, Self Normalizing Neural Networks, Neural Oblivious Decision Ensembles (NODE), AutoInt, and Deep & Cross Neural Networks (DCNs).

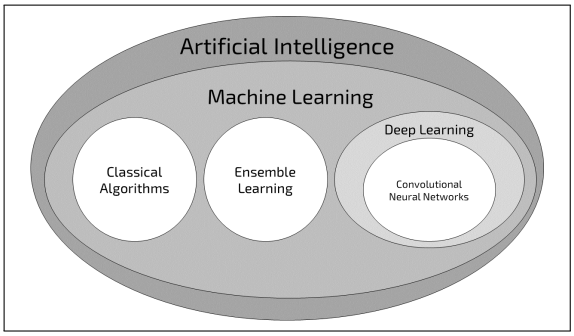

Fig. 1. Artificial Intelligence and its Branches

III. Applied Models

The models presented in the previous section were used in this study. However, before the extracted HRV features were passed to the models, they were treated with some data fitting steps.

A. Data Adjusting

A detailed examination of the architecture of the extracted HRV features shows that the majority of samples were clas-

sified without cardiovascular events (122 of 139 samples). Knowing that unbalanced data can negatively affect the performance of the models, it is necessary to balance the data before using them. In addition, the data are scaled by normalizing or standardizing the real-valued input and class variables. These steps are achieved by performing:

- **Synthetic Minority Over-sampling Technique (SMOTE):** a type of data expansion in which new samples are synthesized from existing ones to oversample the minority class
- **Preprocessing Standard Scaling:** Standardization of characteristics by removing the mean and scaling to a unit variance.

B. Building and Training Models

Therefore, after the extraction of HRV features and the necessary data fitting steps, the treated data are passed to different models built based on the following approaches.

1) Classical ML: Support Vector Machines: Support Vector Machines (SVM) is a supervised Machine Learning algorithm that learns to assign labels to objects based on examples, feeding it training data labeled with the appropriate category to predict the category of new examples [28]. The performance of the SVM is affected by the following important hyperparameters [31]:

- **Kernel:**. transforms the input data into the desired form, such as linear, polynomial and radial basis function (RBF).
- **Regularization:** represents the misclassification or error term
- **Gamma:** defines how far the influence of a single training sample extends, with low values meaning ”far” and high values meaning ”close”
- **Class weight:** defines the weighting of the classes to be predicted and is effective for imbalanced datasets, which includes the case studied in this article

2) Ensemble Learning Algorithms: AdaBoost: AdaBoost is an Ensemble Learning algorithm that falls under the domain of machine learning.

- **AdaBoost [32]:** is a meta-estimator that first fits a classifier to the original data and then fits additional copies of the classifier to the same data, changing the weights of misclassified instances so that subsequent classifiers examine them extensively, leading to an improved result
  - Number of Estimators: the number of base estimators or weak learners to be used in the dataset.
  - Learning Rate: is the step size used in the update to prevent overfitting

3) Deep Convolutional Neural Networks: TabNet & TabTransformers: In this study, the following models were used to implement DCNN on the tabular data of SHAREEDB.

- **TabNet [33]:** a model that uses sequential attention to select which features to infer at each decision step, and is influenced by the following hyperparameters:
  - Optimizer: an algorithm that modifies the neural network attributes, such as weights and learning rate.
  - learning rate: is the step size used in updating to prevent overfitting
  - Batch size: number of examples per batch.
- **TabTransformers [34]:** is a model based on transformers whose layers convert categorical feature embeddings into robust contextual embeddings to achieve higher prediction accuracy, and is affected by the following hyperparameters:
  - Activation function: defines how the weighted sum of the input is converted into an output of a node in a network layer.
  - Number of Heads: specifies the number of heads of attention
  - Dropout: regularization to reduce overfitting and improve generalization of deep neural networks
  - MLP Hidden Units Factors: MLP hidden layer units, as factors of the number of inputs.
  - Learning Rate: is the shrinkage step size used in updating to prevent overfitting

C. Wrapping Up, Training and Predictions

Once the models were created, they were trained using the extracted HRV features. The models were then evaluated using several performance metrics, namely accuracy, precision, recall, F1 score, specificity, and negative predictive value. The results obtained are detailed and discussed in the next section.

IV. Results & Discussion

The HRV features extracted by the steps described above are then passed to the models created: SVM, AdaBoost, TabNet, and TabTransformers. The models were trained with the data and their performance was measured with the performance metrics: Accuracy, Precision (also referred to as Positive Predictive Value PPV), Recall (also referred to as Sensitivity), Specificity (also referred to as True Negative Rate), Negative Predictive Value NPV, and F1 Score. To accurately assess the performance of the models, Repeated K-Fold Cross Validation [35] was applied to the models. A 10-fold cross validation was performed and repeated 5 times. The results are shown in Table 1 below, where the performance measures Accuracy, Precision, Recall, Specificity, Negative Predictive Value, and F1 Score are denoted as AC, PR, RE, SP, NPV, and F1, respectively. The table also lists the values of the hyperparameters used.

In addition, Figure 2 graphs the performance of the models, with parts a, b, and c showing the Area Under Curve for the SVM, AdaBoost, and TabNet models, respectively, while part d shows the TabNet training loss graph and part e shows the training and validation loss for the TabTransformer model.

1) Discussion: In this study, heart rate variability features were extracted from the 24-hour data set SHAREEDB ECG and analyzed with different AI models. The models presented in this study showed high feasibility in predicting cardiovascular events. Despite the high performance metrics recorded by the different models, our results were comparable to those

TABLE I
MACHINE LEARNING MODELS EVALUATION METRICS

Model	Hyper Parameters		Evaluation Metrics					
	Parameter	Value	AC	PR	RE	SP	NPV	F1
SVM	Training(Testing)	0.75(0.25)	91.80%	87.87%	96.66%	87.09%	96.42%	92.06%
	Kernel	rbf						
	Regularization(C)	2.66						
	Gamma	0.141						
AdaBoost	Training(Testing)	0.79(0.21)	89.50%	87.20%	94.60%	84.90%	93.60%	90.80%
	Estimators	200						
	Learning Rate							
TabNet	Training(Testing)	0.79(0.21)	76%	74.70%	82.60%	74.70%	80.50%	76.50%
	Learning Rate	0.9						
	Batch Size	1024						
	Virtual Batch Size	1024						
TabTransformers	Training(Testing)	0.72(0.28)	90.38%	86.66%	96.29%	91.22%	84%	95.45%
	Activation Function	Sigmoid						
	Transformer Blocks	1024						
	Attention Heads	1024						
	Dropout	0.25						
	MLP Hidden unit	[1024,512]						
	Learning Rate	0.05						
	Epochs	1000						

TABLE II
COMPARISON WITH PREVIOUS IMPLEMENTATIONS

Study	Model	AC	PR	RE	SP	NPV	F1
Our Study	Support Vector Machines	91.80%	87.87%	96.66%	87.09%	96.42%	92.06%
	AdaBoost	89.50%	87.20%	94.60%	84.90%	93.60%	90.80%
	TabNet	76.00%	74.70%	82.60%	74.70%	80.50%	76.50%
	TabTransformers	90.38%	86.66%	96.29%	91.22%	84.00%	95.45%
[9]	Artificial Neural Network	76.60%	70.70%	82.90%	71.40%	NA	NA
[11]	Support Vector Machines (Linear Kernel)	88.64%	90.84%	86.36%	90.91%	86.96%	NA
	Support Vector Machines (Polynomial Kernel)	82.95%	80.85%	79.55%	86.36%	85.37%	NA
	Support Vector Machines (RBF Kernel)	82.58%	79.45%	77.27%	87.88%	86.44%	NA
	Multi Layer Perceptron (Top 15 Features)	86.67%	100%	73.33%	100%	78.95%	NA
[12]	Support Vector Machines	79.81%	21.15%	91.67%	79.08%	99.36%	NA
[13]	MLP (A Minute Before the SCD Event)	99.73%	NA	NA	NA	NA	NA
	K-NN (A Minute Before the SCD Event)	98.32%	NA	NA	NA	NA	NA
[16]	SVM (2 minutes before VF Event)	96.36%	NA	NA	NA	NA	NA
	Penalized Neural Network	93.64%	NA	NA	NA	NA	NA
[17]	Support Vector Machines	89.00%	NA	86.30%	91.80%	NA	NA
	Multi Layer Perceptron	78.10%	NA	86.30%	69.90%	NA	NA
[19]	Artificial Neural Network	85.30%	83.30%	88.20%	82.40%	87.50%	NA
[20]	MIL Statisitcs Algorithm	85.47%	92.11%	86.42%	83.33%	NA	NA
[21]	Vote	87.41%	NA	NA	NA	NA	NA
	Naïve Bayes	84.81%	NA	NA	NA	NA	NA
	Support Vector Machines	85.19%	NA	NA	NA	NA	NA

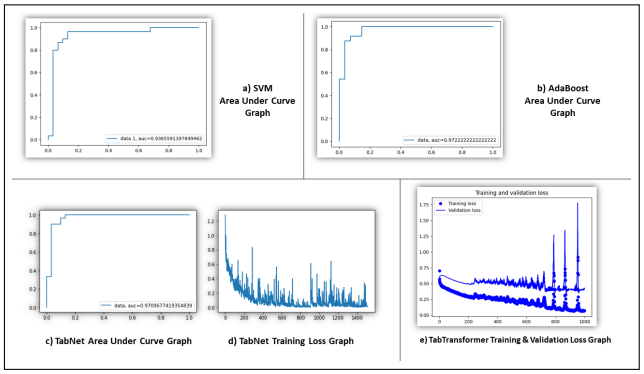

Fig. 2. Models Performance Graphical Representation

of previous implementations.

For example, in [17], the authors used the same dataset and also employed the SVM model. However, our results were better in terms of the performance metrics captured. They stated that their SVM model was considered the most accurate model in their study, with performance metrics such as Accuracy: 89.00%, Recall: 86.30%, and Specificity: 91.80%, while our SVM model metrics were: 91.80%, 96.66%, and 87.09% for the same parameters.

On the other hand, the accuracy of the SVM model presented in this study is better than all other models in previous implementations, where the accuracy is 91.80%. In previous studies, SVMs with linear, polynomial, and RBF kernels in [11] have achieved an accuracy of 88.64%, 82.95%, and 82.58%, respectively. The same is true for [12,21], where the accuracy was 79.81%, 89.00%, and 85.19%, respectively.

Similarly, the AdaBoost and TabTransformers models show better performance compared to other implementations, achieving an accuracy of 89.50% and 90.38%, respectively. Even though the results are close in accuracy, the improvement in precision and recall is very illustrative. Because recall is the measure of a model's correct classification of true positives, our model is better able to correctly predict a subject as potentially having cardiovascular event in the future. With a recall of 96.04%, 94.60%, and 96.29% achieved by our SVM, AdaBoost, and TabTransformers models, respectively, this is the highest ability among all implementations to correctly predict that a subject is in a cardiovascular risk zone. However, the TabNet model did not show competitive results, as shown in the table.

Finally, the SVM, AdaBoost, and TabTransformer models developed in this study are considered the most accurate models for predicting cardiovascular events with high accuracy and relatively high recall. Nevertheless, the performance metrics in [13] and [16] are higher than our results, but they are not comparable with our models because their models were developed for the prediction of sudden cardiac death (SDC) only minutes before its occurrence. A further look into their implementations shows a dramatic decrease in prediction accuracy as the prediction time interval increases by minutes. For example, in [13], the accuracy was 99.73% when the SDC was predicted one minute before its occurrence and drops to 83.93% when the event is predicted four minutes before its occurrence. This is in contrast to the models offered in this study, which are able to predict cardiovascular events 12 months before their occurrence with relatively high performance. Consequently, this study highlights the feasibility of using artificial intelligence models with heart rate variability features in predicting cardiovascular events, even 12 months before their occurrence. Therefore, the results reinforce confidence in the use of AI and its ramifications in the prediction of cardiovascular disease and related events, or at least in its use as a diagnostic assistant for this deadly disease.

CONCLUSION

Artificial intelligence and its ramifications are increasingly entering all aspects of our lives. Moreover, the increase in communication tools combined with AI capabilities is expected to help cardiologists reach a new level of disease diagnosis and prediction. The models developed and presented in this article show high efficiency in predicting cardiovascular events one year before their occurrence. However, future efforts are needed to adapt such models with real-time data processing to improve the personalization of health services and provide immediate insight into the health status of the population. Real-time management, for example, will be key to monitoring people in the workplace to improve their health and increase their productivity, especially for those working under stressful conditions or for long periods of time.

ACKNOWLEDGMENTS

We acknowledge the support of Natural Sciences and Engineering Research Council of Canada (NSERC), grant number 06351, Fonds Québécois de la Recherche sur la Nature et les Technologies (FRQNT) and Institut Technologique de Maintenance Industrielle (ITMI).

REFERENCES

[1] WHO reveals leading causes of death and disability worldwide: 2000–2019. (2020, December 9). World Health Organization. Retrieved June 1, 2021, from https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019

[2] Cardiovascular diseases (CVDs). (2021, June 11). World Health Organization. Retrieved January 15, 2022, from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[3] Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., ... & GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. Journal of the American College of Cardiology, 76(25), 2982-3021.

[4] National Heart Lung and Blood Institute. (2005). In Brief: Your guide to living well with heart disease (NIH Publication 06-5716). Washington, DC: National Institutes of Health.

[5] Besson, C., Saubade, M., Gremeaux, V., Millet, G. P., & Schmitt, L. (2020). Heart rate variability: methods, limitations and clinical examples. Revue Medicale Suisse, 16(701), 1432-1437.

[6] Gonzalez, K., Sasangohar, F., Mehta, R. K., Lawley, M., & Erraguntla, M. (2017, September). Measuring fatigue through Heart Rate Variability and activity recognition: A scoping literature review of machine learning techniques. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 61, No. 1, pp. 1748-1752). Sage CA: Los Angeles, CA: SAGE Publications.

[7] Billman, G. E. (2011). Heart rate variability–a historical perspective. Frontiers in physiology, 2, 86.

[8] Baumert, M., Wessel, N., Schirdewan, A., Voss, A., & Abbott, D. (2007). Forecasting of ventricular tachycardia using scaling characteristics and entropy of heart rate time series. In World Congress on Medical Physics and Biomedical Engineering 2006 (pp. 1001-1004). Springer, Berlin, Heidelberg.

[9] Joo, S., Choi, K. J., & Huh, S. J. (2010, September). Prediction of ventricular tachycardia by a neural network using parameters of heart rate variability. In 2010 Computing in Cardiology (pp. 585-588). IEEE.

[10] Spontaneous Ventricular Tachyarrhythmia Database v1.0. (2007, May 2). PhysioNet. https://physionet.org/content/mvtdb/1.0/

[11] Ramirez-Villegas, J. F., Lam-Espinosa, E., Ramirez-Moreno, D. F., Calvo-Echeverry, P. C., & Agredo-Rodriguez, W. (2011). Heart rate variability dynamics for the prognosis of cardiovascular risk. PloS one, 6(2), e17060.

[12] Song, T., Qu, X. F., Zhang, Y. T., Cao, W., Han, B. H., Li, Y., ... & Da Cheng, H. (2014). Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. BMC cardiovascular disorders, 14(1), 1-8.

[13] Ebrahimzadeh, E., Pooyan, M., & Bijar, A. (2014). A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals. PloS one, 9(2), e81896.

[14] Sudden Cardiac Death Holter Database v1.0.0. (2004, July 2). PhysioNet. https://physionet.org/content/sddb/1.0.0/

[15] MIT-BIH Normal Sinus Rhythm Database v1.0.0. (1999, August 3). PhysioNet. https://physionet.org/content/nsrdb/1.0.0/

[16] Murukesan, L., Murugappan, M., Iqbal, M., & Saravanan, K. (2014). Machine learning approach for sudden cardiac arrest prediction based on optimal heart rate variability features. Journal of Medical Imaging and Health Informatics, 4(4), 521-532.

[17] Melillo, P., Izzo, R., Orrico, A., Scala, P., Attanasio, M., Mirra, M., ... & Pecchia, L. (2015). Automatic prediction of cardiovascular and cerebrovascular events using heart rate variability analysis. PloS one, 10(3), e0118504.

[18] Smart Health for Assessing the Risk of Events via ECG Database. (2015, May 19). PhysioNet. Retrieved November 1, 2021, from https://physionet.org/content/shareedb/1.0.0/

[19] Lee, H., Shin, S. Y., Seo, M., Nam, G. B., & Joo, S. (2016). Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks. Scientific reports, 6(1), 1-7.

[20] Lan, K. C., Raknim, P., Kao, W. F., & Huang, J. H. (2018). Toward hypertension prediction based on PPG-derived HRV signals: A feasibility study. Journal of medical systems, 42(6), 1-7.

[21] Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, 82-93.

[22] UCI Machine Learning Repository: Heart Disease Data Set. (n.d.-a). UCI Machine Learning Repository: Heart Disease Data Set. https://archive.ics.uci.edu/ml/datasets/heart+disease

[23] Luo, S., & Johnston, P. (2010). A review of electrocardiogram filtering. Journal of electrocardiology, 43(6), 486-496.

[24] Ashley, E. A., & Niebauer, J. (2004). Cardiology explained.

[25] Eilers, J., Chromik, J., & Arnrich, B. (2021). Choosing the Appropriate QRS Detector. In BIOSIGNALS (pp. 50-59).

[26] Mathur, P., Srivastava, S., Xu, X., & Mehta, J. L. (2020). Artificial intelligence, machine learning, and cardiovascular disease. Clinical Medicine Insights: Cardiology, 14, 1179546820927404.

[27] Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet?. Heart, 104(14), 1156-1164.

[28] Ertel, W. (2018). Introduction to artificial intelligence. Springer.

[29] Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. Frontiers of Computer Science, 14(2), 241-258.

[30] Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2021). Revisiting deep learning models for tabular data. Advances in Neural Information Processing Systems, 34.

[31] Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565-1567.

[32] Schapire, R. E. (2013). Explaining adaboost. In Empirical inference (pp. 37-52). Springer, Berlin, Heidelberg.

[33] Arık, S. O., & Pfister, T. (2021). Tabnet: Attentive interpretable tabular learning. In AAAI (Vol. 35, pp. 6679-6687).

[34] Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). Tabtransformer: Tabular data modeling using contextual embeddings. arXiv preprint arXiv:2012.06678.

[35] Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. Encyclopedia of database systems, 5, 532-538.

A.4 Toward The Goal; Narrowing Ideas Down

After immersing myself in research and making a notable contribution to publications, I turned my attention to addressing challenges within the field of Machine Learning (ML). I honed in on two particularly intriguing research areas: Multimodal Machine Learning and Federated Learning.

Beside the interest in Federated Learning domain, Multimodal Machine Learning, aimed at tackling the heterogeneity in ML, captured my interest. This research effort culminated in a comprehensive long-form review, dedicated to exploring this topic and its applications in health domain. This review were later published with MDPI:

- **Reviewing Multimodal Machine Learning and Its Use in Cardiovascular Diseases Detection** *(MDPI-Electronics / Impact Factor 2.9)*

electronics

MDPI

Review

Reviewing Multimodal Machine Learning and Its Use in Cardiovascular Diseases Detection

Mohammad Moshawrab [1,*], Mehdi Adda [1], Abdenour Bouzouane [2], Hussein Ibrahim [3,*] and Ali Raad [4]

[1] Département de Mathématiques, Informatique et Génie, Université du Québec à Rimouski, 300 Allée des Ursulines, Rimouski, QC G5L 3A1, Canada
[2] Département d'Informatique et de Mathématique, Université du Québec à Chicoutimi, 555 Boulevard de l'Université, Chicoutimi, QC G7H 2B1, Canada
[3] Institut Technologique de Maintenance Industrielle, 175 Rue de la Vérendrye, Sept-Îles, QC G4R 5B7, Canada
[4] Faculty of Arts & Sciences, Islamic University of Lebanon, Wardaniyeh P.O. Box 30014, Lebanon
* Correspondence: mohammad.moshawrab@uqar.ca (M.M.); hussein.ibrahim@itmi.ca (H.I.); Tel.: +1-(581)624-9394 (M.M.)

Abstract: Machine Learning (ML) and Deep Learning (DL) are derivatives of Artificial Intelligence (AI) that have already demonstrated their effectiveness in a variety of domains, including healthcare, where they are now routinely integrated into patients' daily activities. On the other hand, data heterogeneity has long been a key obstacle in AI, ML and DL. Here, Multimodal Machine Learning (Multimodal ML) has emerged as a method that enables the training of complex ML and DL models that use heterogeneous data in their learning process. In addition, Multimodal ML enables the integration of multiple models in the search for a single, comprehensive solution to a complex problem. In this review, the technical aspects of Multimodal ML are discussed, including a definition of the technology and its technical underpinnings, especially data fusion. It also outlines the differences between this technology and others, such as Ensemble Learning, as well as the various workflows that can be followed in Multimodal ML. In addition, this article examines in depth the use of Multimodal ML in the detection and prediction of Cardiovascular Diseases, highlighting the results obtained so far and the possible starting points for improving its use in the aforementioned field. Finally, a number of the most common problems hindering the development of this technology and potential solutions that could be pursued in future studies are outlined.

Keywords: multimodal machine learning; multimodal learning; data heterogeneity; data fusion; model heterogeneity; model fusion; diseases prediction; cardiovascular diseases; Internet of Things; smart wearables

Citation: Moshawrab, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Reviewing Multimodal Machine Learning and Its Use in Cardiovascular Diseases Detection. *Electronics* **2023**, *12*, 1558. https://doi.org/10.3390/electronics12071558

Academic Editors: Francisco Luna-Perejón, Lourdes Miró Amarante and Francisco Gómez-Rodríguez

Received: 2 March 2023
Revised: 14 March 2023
Accepted: 25 March 2023
Published: 26 March 2023

1. Introduction

Artificial Intelligence (AI) has experienced rapid growth over the past two decades. The concept of AI has been around since 1950, and the term itself was coined in 1965 at the Dartmouth Summer Workshop, which is considered the founding event of AI as a field [1]. However, the growth in Information and Communication Technologies (ICTs) and the increasing power of computers have contributed significantly to the increasing feasibility and adoption of AI [2]. AI technologies are becoming more advanced and are capable of analyzing enormous amounts of data, learning from past experiences, and making predictions based on patterns and trends [3]. Despite the popularity of AI, there is no single definition for this technology. Researchers in [4], for example, defined it as a set of tools and techniques that use principles and devices from various fields, such as computation, mathematics, logic, and biology, to address the problem of realizing, modeling, and mimicking human intelligence and cognitive processes. Furthermore, the authors define in [5] AI as the study of an "Intelligent Agent", i.e., machines that are able to recognize and understand their environment and consequently

take appropriate actions to increase their chances of achieving their goals. In an attempt to unify definitions, the authors defined in [6] AI as a program that can cope in an arbitrary world no worse than a human. These different definitions reflect the different competencies of AI, which explains the diversity of AI implementations in our daily lives.

Machine Learning (ML) [7], Deep Learning (DL) [8], Federated Machine Learning (FL) [9], and Multimodal Machine Learning [10] are all well-known and popular derivatives of the AI concept that have been adopted by users and applied in various aspects of our daily lives. These different branches of AI are depicted in Figure 1. In this context, Machine Learning is defined as a field of study that focuses on the development of algorithms and statistical models that enable computer systems to learn from data and make predictions or decisions without being explicitly programmed. It involves the application of various approaches, such as supervised and unsupervised learning, Reinforcement Learning, and Deep Learning, that allow computers to automatically improve their performance on a given task through experience [7].

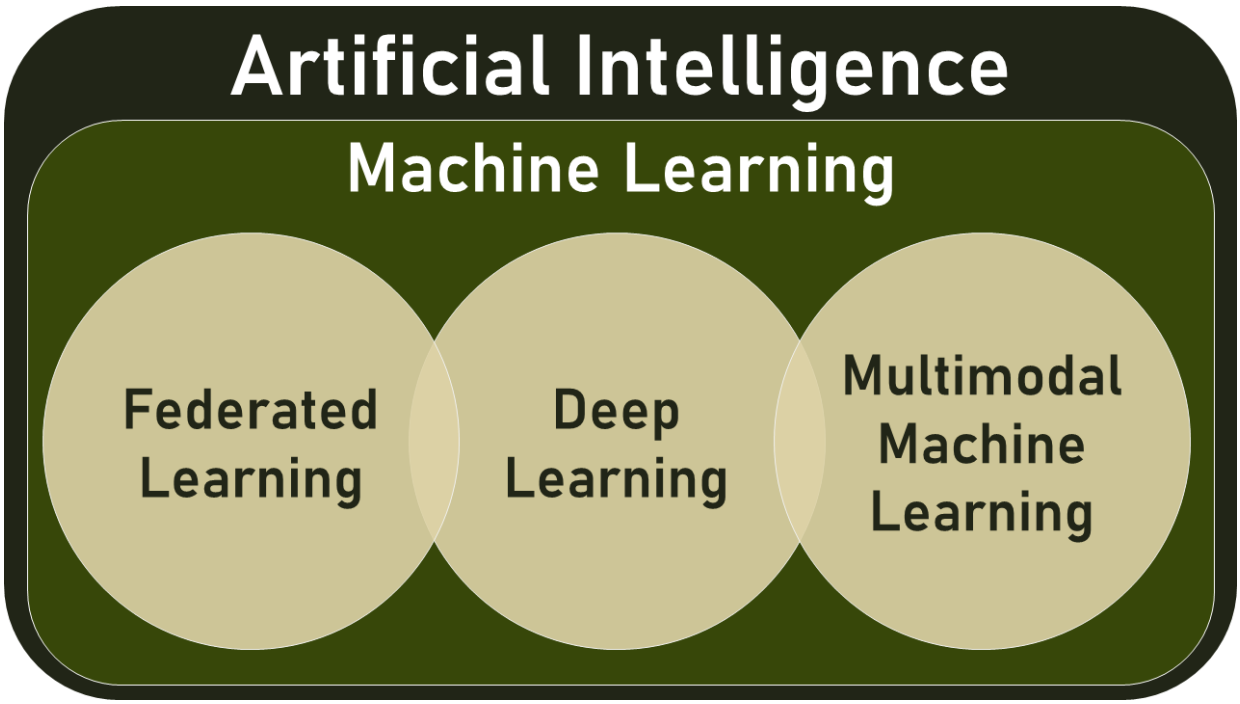

Figure 1. Artificial intelligence branches.

On the other hand, Machine Learning has demonstrated high efficiency in solving classification and regression problems. Machine Learning's ability to extract meaningful insights and patterns from vast and complicated datasets and use this knowledge to make accurate predictions, automate decision making, and enable intelligent systems to learn and adapt in real-time is fundamental to its success. This success has led researchers from different fields to implement ML algorithms, and their efficiency can be observed in various fields, such as:

- Healthcare services [11–13];
- Image, speech and pattern recognition [14,15];
- Internet of Things (IoT) and smart cities [14,16];
- Cybersecurity and threat intelligence [17];
- Natural language processing and sentiment analysis [18];
- User behavior analytics and context-aware smartphone applications [14,15];
- E-commerce and product recommendations [14,15];

- Sustainable agriculture [19];
- Industrial applications [20].

1.1. Machine Learning Domain Challenges

The great success of Machine Learning is not magic but the result of its ability to analyze large amounts of data at high speed and with high accuracy. However, the field of ML still suffers from various challenges and obstacles arising from different problems. Table 1 below summarizes the Machine Learning challenges and categorizes them based on their source. These challenges have been extensively studied in the literature, and more details can be found in several articles, such as [9,21–23].

Table 1. Machine Learning domain common challenges.

Group	Challenges	
Data-Related Challenges [21,22]	Data Availability and Accessibility [23] Data Locality [16]	
	Data Readiness [23]	Data Heterogeneity Noise and Signal Artifacts Missing Data Classes Imbalance
	Data Volume	Course of Dimensionality Bonferroni principle [24]
	Feature Representation and Selection	
Models Related Challenges [25,26]	Accuracy and Performance Model Evaluation Variance and Bias Explainability	
Implementation-Related Challenges [23,27]	Real-Time Processing Model Selection Execution Time and Complexity	
General Challenges [25,26]	User Data Privacy and Confidentiality User Technology Adoption and Engagement Ethical Constraints	

1.2. Heterogeneity: Motivation(s) Behind Multimodal ML

Advances in sensor technologies, storage concepts, communication networks, and other tools have driven data collection [28]. According to recent figures from Statista [29], the total amount of data generated worldwide will reach 64.2 zettabytes or 6.42×10^{16} Megabytes in 2020. This increase exceeded predictions due to increasing demand as a result of the COVID-19 pandemic, as more individuals worked and studied from home and increasingly used utilized home entertainment alternatives. For the above reasons, data volumes are expected to reach 180 zettabytes in the next five years by 2025.

However, these data differ in type, structure, format, usability, lifespan, and other aspects. This heterogeneity poses several challenges in Machine Learning, as it can make it difficult to use data from different resources to gain useful insights or build accurate models. There are many types of heterogeneity, the most common of which are listed below [21,30,31]:

- Structured vs. unstructured data: structured data are highly organized and usually follow a specific schema, while unstructured data have no predefined structure or organization;
- Numeric vs. categorical data: Numeric data are quantitative and can be expressed as numbers, while categorical data are qualitative and represent discrete values, such as colors, types, or labels;

- Temporal data: This type of data contains time-stamped information and can be used to analyze patterns and trends over time;
- Multimodal data: This type of data combines different types of information, such as text, audio, images, and videos.

Thus, dealing with heterogeneous data requires careful processing and feature engineering to put the data into the form required for a single Machine Learning model [31]. In addition, multiple preprocessing steps may be required to analyze heterogeneous data, such as normalization, scaling, or other steps. In some cases, however, it may seem impossible to analyze heterogeneous data, even though training the model with this variety of resources improves its feasibility and increases confidence in its predictions.

For example, Magnetic Resonance Imaging (MRI) analysis using ML models has shown high efficiency in predicting Cardiovascular Diseases (CVDs), as shown in [32]. In addition, smart wearables equipped with ML models are also highly feasible in predicting cardiac disease, as shown in [33]. In addition, the use of Electronic Health Records (EHRs) collected from various health centers such as clinics, hospitals, or smart homes is also a good source for Cardiovascular Disease prediction using ML algorithms [34]. However, trying to merge these three types of data seems to be technically impossible because the first data source, namely MRI images, are stored in the form of medical electronic image files, and the data collected by wearables are structured data, while EHRs can be a collection of both structured and unstructured data, free text reports, medical examination data, or other formats. In the real world, a physician may analyze all of these data to make a more accurate diagnosis, though it is not easy to analyze these data sets simultaneously using the same model. This case is illustrated in Figure 2 below.

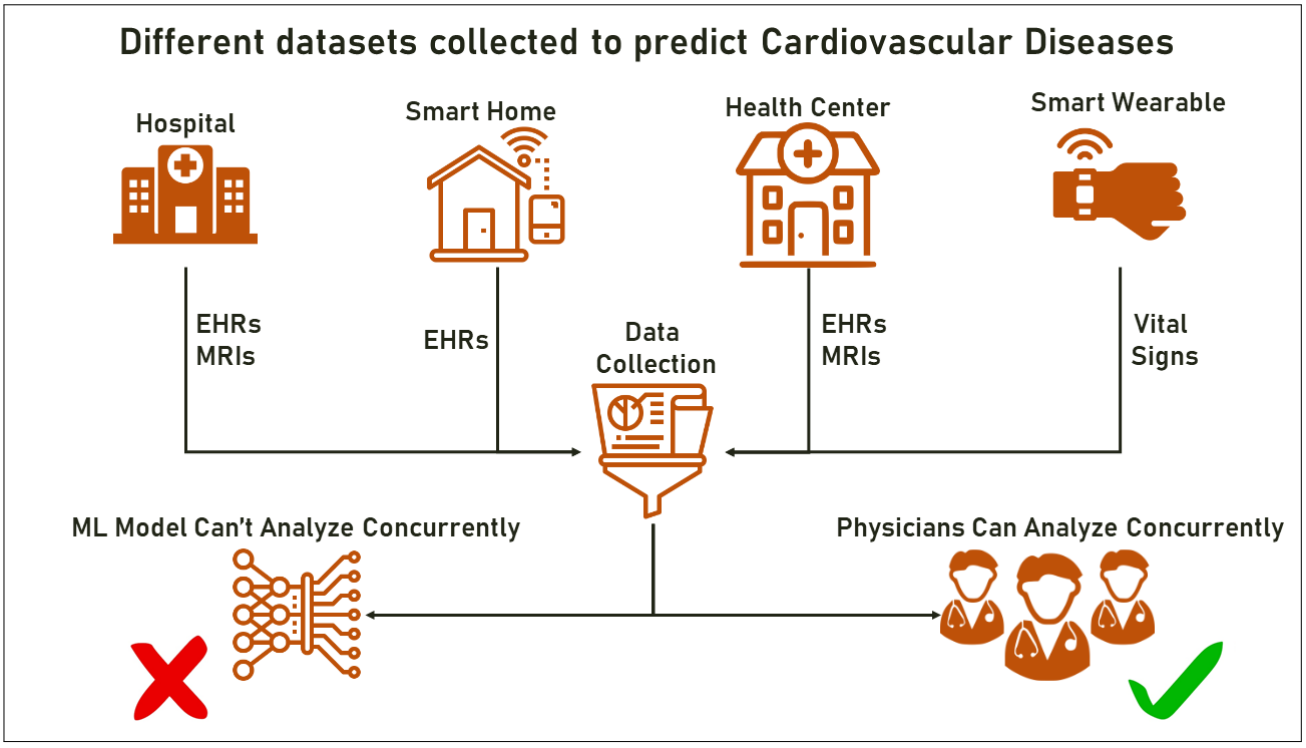

Figure 2. Prediction of CVDs with heterogeneous data—a showcase.

In this context, Multimodal Machine Learning is proposed as a solution to the challenge of data heterogeneity in ML. Multimodal ML gives models the ability to analyze different data within the same ML workflow, whether by merging different datasets, by merging different models, or both, to arrive at a single result, such as the diagnosis of CVDs in the showcase mentioned above [10]. The ability to analyze these heterogeneous data with

multiple views can be of varying importance to a learning task. Therefore, merging all of these data sets and treating them with equal importance is unlikely to lead to optimal learning outcomes [30].

1.3. Machine Learning and Healthcare

The importance of health to human life cannot be overstated, as it is essential for meeting basic needs, pursuing goals, maintaining relationships, and having an adequate quality of life, and poor health can have significant financial and societal consequences. Therefore, researchers are constantly striving to improve the quality of healthcare services. In this context, Artificial Intelligence and its branches, such as Machine Learning and Deep Learning, have been incorporated into healthcare services due to their high feasibility and usability in this field. Machine learning, in particular, is a powerful tool that has the potential to revolutionize healthcare in many ways [35]. ML has made remarkable progress in healthcare, not because of any mystical powers, but because of its superior data processing capabilities compared to those of humans. Because of its speed and precision, thousands of AI applications have already been developed for healthcare, making it a potentially revolutionary tool for solving a wide range of healthcare problems [36].

Machine Learning has been used in various areas of healthcare. Whether diagnosing diseases or even predicting diseases, it has proven to be very useful. Moreover, the development of communication tools, such as smart wearables equipped with Machine Learning and Deep Learning models, has opened the door to real-time continuous monitoring. In this context, smart wearables have shown high feasibility in predicting various diseases such as Cardiovascular Diseases [33], diabetes [37], liver disease [38], fatigue and stress [39], mental illness [40], and many other diseases [41]. In addition, ML models have been used to increase the efficiency of healthcare decision systems [42]. In addition, ML has also been used in the field of genomic medicine [43]. Overall, ML has succeeded in transforming health services and creating personalized digital health services that support physicians and improve the overall quality of public health [44].

Therefore, considering the importance of healthcare, it is urgent to improve the efficiency of ML. The use of state-of-the-art methods and the removal of obstacles to progress are essential to improving performance. The challenges described previously are reflected in the barriers to expanding the use of ML in healthcare, which are common to all ML implementations across all diseases. With this in mind, new solutions that could help promote the use of ML will lead to improved applications in a variety of settings.

1. Define the scope of the review: Clearly define the scope and objective of the review article. What is the main topic or research question that the review aims to address? What specific subtopics or themes will be covered? 2. Identify the key concepts and themes: Based on the scope and objective of the review, identify the key concepts and themes that will be discussed in the article. These should be organized in a logical and coherent manner that supports the overall objective of the review. 3. Develop a framework for presenting the review: Once the key concepts and themes have been identified, develop a framework for presenting the review. This could involve organizing the content chronologically, thematically, or conceptually, depending on the nature of the review and the key concepts and themes identified. 4. Clearly articulate the review framework: Finally, clearly articulate the review framework in the introduction or early sections of the review article. This will help to orient readers to the overall structure and organization of the review and make it easier for them to follow the content. Overall, the goal is to provide a clear and structured overview of the review article that highlights the key concepts and themes and guides the reader through the content in a logical and coherent manner.

1.4. Review Framework: Scope, Outline and Main Contributions

In this article, Multimodal Machine Learning is explored, and its role as a solution to the challenge of heterogeneity is detailed. In addition, the use of Multimodal ML in Cardiovascular Disease detection and prediction is technically reviewed to support its use in this field.

1.4.1. Scope of Research

To achieve the objectives of the study, Multimodal Machine Learning has been explored, along with the data fusion concept, which is the basis of the technology under study. In addition, the technical perspectives of Multimodal ML are studied, and the workflows related to it are examined. Furthermore, a comparison between Multimodal ML and other known techniques is made in order to distinguish between these different techniques. On the other hand, distinct areas where Multimodal ML is used are inspected, and a comprehensive overview of its application in Cardiovascular Diseases, including the state of the art, is therefore obtained. In addition, these implementations were analyzed from different perspectives to understand the limitations and future areas of research. Finally, the challenges and future recommendations associated with advancing this field are reviewed.

1.4.2. Research Questions

The scope of the article defined in the previous section is summarized by the research questions mentioned in the list below:

- What is Multimodal Machine Learning?
- What are the motivations for this technology?
- What are the technical perspectives on which Multimodal ML is based?
- What are the differences between Multimodal ML, classical ML, Multimodal datasets, ensemble ML and other techniques?
- What are the existing Multimodal ML frameworks, and what contributions do each make?
- What is the state of the art in the use of Multimodal ML in CVD prediction, and what technical summaries can be derived?
- What challenges still impede progress in this area, and what approaches could be taken to overcome these issues?

1.4.3. Outline

To answer the above questions, the article is outlined as follows. In Section 2, Multimodal ML is reviewed from various angles, including technical definition(s), differences from other domains, such as classical ML, ensemble ML and others, available frameworks, and other details. Then, in Section 3, the use of Multimodal ML technology in CVD detection and prediction is presented by listing the state of the art in this field and discussing the technical details of the implementations mentioned in the literature. Later, in Section 4, the challenges that hinder progress in this field are discussed, and therefore, some future perspectives that could help in overcoming these challenges are proposed. This article attempts to answer the following questions:

1.4.4. Comparison with Previous Review Frameworks

The topic of Multimodal ML has been a hot and trending topic in recent years. As a result, numerous studies have already addressed this topic, with a large proportion of these studies reviewing Multimodal ML. However, this article proposes several new ideas that add to the knowledge of Multimodal ML. First, this study proposes a technical study for Multimodal ML that, on the one hand, helps to understand this technology and distinguish it from other existing AI techniques. Moreover, none of the previous studies proposed a technical review for the use of this technology in CVD detection and prediction. Moreover, this review discusses in detail the challenges and future ideas in

this field to help future researchers select the most relevant ideas on which to build their future work.

1.4.5. Key Findings and Contributions

Consequently, this article is rich in various new points that contribute to the body of knowledge on Multimodal ML:

- Discuss fusion and its fundamental role in defining the structure of Multimodal ML;
- Establish clear and precise boundaries to distinguish between Multimodal ML, traditional ML, multimodal datasets, multilabel models, and ensemble learning;
- Propose a new description for the different workflows that can be followed in the implementation of Multimodal ML algorithms;
- Discuss existing frameworks in the area of Multimodal ML and evaluate the contributions to this area;
- Review and discuss the state of the art of Multimodal ML in the diagnosis of CVDs;
- Examine the technical details associated with these implementations;
- Present completely and in detail the challenges that hinder Multimodal ML and the possible future perspectives that can be pursued to increase the efficiency of the technology.

2. Materials and Methods: What is Multimodal ML?

The human mind processes information from multiple senses simultaneously. Sometimes it is not enough to just hear about a problem; individuals need to see it for themselves in order to make an informed judgment. For Artificial Intelligence to expand its knowledge of the world, it must be able to process a variety of information sources that may contradict each other. This principle also applies to the field of AI known as Machine Learning (ML), where Multimodal Machine Learning focuses on using numerous data sources to achieve a single goal by leveraging complementary information in a unified computational framework. The ability to explore diverse data increases predictive power and leads to more accurate and reliable results, making Multimodal Machine Learning a multidisciplinary topic with tremendous efficiency and amazing potential [5,10].

2.1. Overview and Definition(s)

Despite the fact that Multimodal Machine Learning is a popular and young research area that has received much attention, it is still in its infancy [4–6,45]. As a result, there is no single and universally accepted definition. Nevertheless, all definitions lead to the same concept: the ability to analyze different data sets to reach a single conclusion. For example, the authors describe in [4] Multimodal ML as the ability to evaluate data from Multimodal datasets, identify a common phenomenon, and use complementary knowledge to learn a complex task. Multimodal datasets are described in this way as data seen with many sensors, where the output of each sensor is called a modality and can be associated with a dataset. Similarly, the authors of [5] describe Multimodal ML as the integration of multiple data sources collected by different instruments, devices, or techniques, followed by the analysis of these merged data using different ML architectures. In addition, Multimodal Machine Learning is described in [10] as an area that aims to develop intelligent models that can process and link data from many sources.

2.2. Multimodal ML and Data Fusion

Multimodal ML brings together data from multiple and disparate modalities to identify a single task. The discipline behind merging data from multiple sources is called data fusion. More specifically, data fusion is defined as "the process of combining data to refine state estimates and predictions" [5]. According to the Joint Directors of Laboratories Data Fusion Subpanel (JDL), the technique of "data fusion" is a must for processing more than one type of data [46]. The authors in [46] support this definition by explaining that any process that deals with associating, correlating, or combining data from one or more sources

to obtain enriched information is called a process that uses data fusion. In data fusion, given the novelty of the literature, there is no consensus on how best to combine different data, especially since there are four different techniques for implementing data fusion, which may have many names depending on the context and research area [5,46,47]. These different approaches are illustrated in Figure 3:

- Early Fusion: also called Low-Level Fusion, is the simplest form of data fusion in which disparate data sources are merged into a single feature vector before being used by a single Machine Learning algorithm. Therefore, it can be referred to as a multiple-data, single-algorithm technique.
- Intermediate Fusion: is also referred to as Medium-Level Fusion, joint fusion, or Feature-Level Fusion, and occurs in the intermediate phase between the input and output of a ML architecture when all data sources have the same representation format. In this phase, features are combined to perform various tasks such as feature selection, decision-making, or predictions based on historical data.
- Late Fusion: also known as decision-level fusion, defines the aggregation of decisions from multiple ML algorithms, each trained with different data sources. In addition, various rules are used to determine how decisions from different classifiers are combined, e.g.,:
  - Max-fusion
  - Averaged-fusion
  - Bayes' rule-based
  - Even rules learned using a metaclassifier
- Hybrid Fusion: defines the use of more than one fusion discipline in a single deep algorithm.

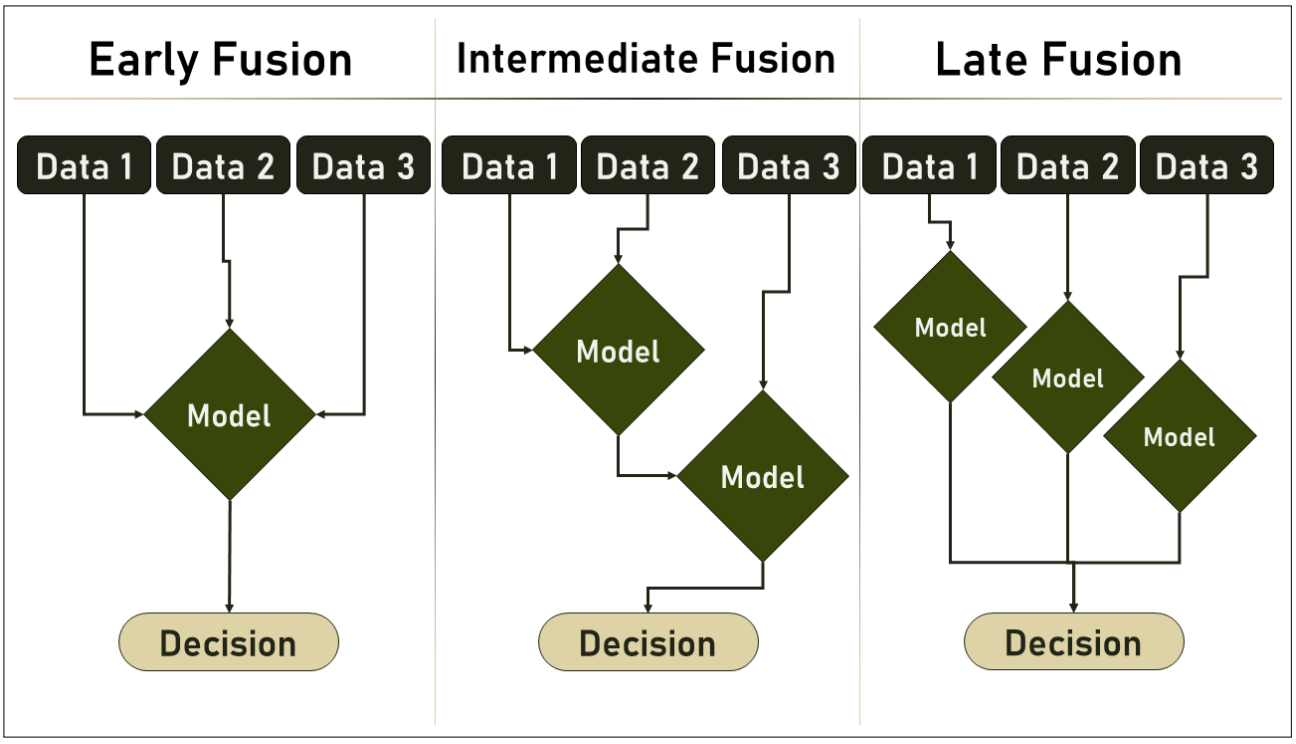

Figure 3. Data fusion approaches.

Based on the information in [4,5], early fusion is the most common form of fusion, which has the advantage of converting all data into the same representation that can be classified using robust classical models, such as Support Vector Machines or Logistic Regression. However, when the input modalities are particularly uncorrelated and have

widely varying dimensionality and sampling rate, it is easier to use a late fusion approach. In addition, both early and late fusion offer the most flexibility in terms of the number of models that can be used to analyze the data, but there is no conclusive evidence that late fusion is better than early fusion because its performance is highly problem dependent. Alternatively, intermediate fusion provides more flexibility in terms of how and when representations learned from Multimodal data are fused. Table 2 discusses the different features of each approach.

Table 2. Data fusion approach specifications.

Attribute	Early	Intermediate/Joint	Late/Decision
Ability to handle missing data	no	no	yes
Scalable	no	yes	yes
Multiple models needed	no	yes	yes
Improved accuracy	yes	yes	yes
Voting/weighting issues	no	yes	yes
Interaction effects across sources	yes	yes	no
Interpretable	yes	no	no
Implemented in health	yes	yes	yes

2.3. Multimodal ML: Technical Perspectives

The goal of Multimodal Machine Learning, also known as Multimodal Deep Learning, is to develop algorithms and models that can interpret and learn from data across multiple modalities, such as text, audio, images, and video. Multimodal ML is a thriving research area with the potential to transform a wide range of applications, from speech recognition and language translation to autonomous cars and medical imaging, among many other areas. Multimodal ML, from a technical perspective, encompasses the various approaches, algorithms, and architectures used in creating and evaluating these models. Data preprocessing, feature extraction, model architecture, training methods, evaluation criteria, generalization, interpretability, and scalability are the most common possible viewpoints. Understanding the technical aspects of Multimodal ML is essential for developing efficient models that can leverage complementary instances across many modalities and make more accurate and robust predictions in the real world. Therefore, the technical perspectives of Multimodal ML are described below.

2.3.1. Data Preparation

Because Multimodal data are often complex and heterogeneous, they must be thoroughly processed before they can be used to train the model. The first step is to recognize the many modalities in action, then learn how to preprocess them, and finally, merge them into a single representation that can be fed into the model [4,5,10].

2.3.2. Model Architecture

Multimodal data can be represented in a variety of ways, including concatenation, fusion, and attentional mechanisms. Choosing the right architecture that can handle the multiple modalities and learn a combined representation is crucial depending on the data and the task to be solved [46,47].

2.3.3. Training Strategies

Pretraining individual modalities, joint training of all modalities, and training individual models and combining them at the time of inference are all viable options for training Multimodal ML models. Selecting the right training methods is a crucial step in achieving the desired goal [4,5,10].

2.3.4. Evaluation Metrics

Following the performance metrics used to evaluate classical ML algorithms, accuracy, precision, recall, sensitivity, specificity, F1 score, and area under the curve (AUC) are just some of the measures that can be used to evaluate Multimodal ML algorithms. It is controversial whether these measures are useful or not when applied to Multimodal data. As a result, the use of evaluation criteria that consider the success of each modality and the overall performance of the model is essential [21–23].

2.3.5. Generalization

Multimodal models are often trained on a specific collection of data and may not generalize well to new data. To assess how well the model can be generalized, it should be tested and validated with data that are very different from the training data [21–23].

2.3.6. Interpretability

Because of their complexity and the relationships between multiple modalities, Multimodal ML algorithms can be difficult to understand and even more difficult to explain and interpret. To decipher the decision process of the model, some tools such as attentional mechanisms and visualization can be used [21–23,48].

2.3.7. Scalability

In Multimodal Machine Learning, scalability is critical because it enables models to deal with real-world situations where datasets are large and complex, and the amount of data is constantly growing. To ensure that the models can cope with the increase in data volume and complexity in the future, it is necessary to develop models that are scalable to enable effective training and deployment, reduce computational costs, and scale the models [25–27,48].

2.4. Multimodal ML and Other Technologies: Borderlines

Multimodal Machine Learning is a new and rapidly growing discipline that focuses on building models that can learn from a variety of data sources. To distinguish Multimodal ML from other areas of Machine Learning, its characteristic aspects should be highlighted, such as the use of many modalities and the need for effective integration of these modalities. Establishing precise terminology and creating an understandable description of the field will help to differentiate it from other techniques. However, because it is a relatively new field, there may be an overlap with other areas of Machine Learning, and it will be critical to accurately define the boundaries of Multimodal ML as the topic evolves.

2.4.1. Multimodal ML vs. Multimodal Datasets

Multimodal datasets are datasets acquired with different sensors, instruments, technologies, or devices to observe a common phenomenon, where the acquired data are considered complementary [49]. Consequently, multimodal datasets define the data itself, regardless of the identity of the algorithms used to analyze the data, whether they have a multimodal or unimodal architecture. However, merging multimodal datasets and unifying their representation into a single vector and then analyzing them with an ML model is considered an early fusion approach that is a type of Multimodal ML.

2.4.2. Multimodal ML vs. Multilabel Models

Multilabel Machine Learning algorithms are used to analyze datasets with more than one target variable. For example, the output of multilabel classification models consists of multiple classification labels. Moreover, when performing predictions using multilabel ML algorithms, a given input may belong to more than one label. For example, predicting the category of a movie may result in horror, action, science fiction, drama, or some or all of these categories simultaneously. In other words, multilabel classification associates data with a set of labels. Classification involves learning from a set of examples associated with

a single label called "l" from a set of disjoint labels called "L", where |L| > 1. When |L| = 2, the learning problem is called a binary classification problem, and when |L| > 2, it is called a multiclass classification problem [50,51]. Thus, Multimodal ML and multilabel learning differ in the data structure itself, where the former uses data from multiple or different sources to obtain a single result, while the latter uses data from only one source to obtain a single classification result with more than two possible outcomes.

2.4.3. Multimodal ML vs. Ensemble Learning

The goal of ensemble Machine Learning is to improve performance and accuracy by combining numerous models into a single prediction. When making predictions, ensemble learning uses multiple interconnected models rather than a single model. Ensemble learning combines the predictions of many models with the goal that the combined predictions are more accurate and robust than any single model. There are several types of ensemble learning techniques, including [52,53]:

- Bagging (Bootstrap Aggregating): is the process of training several models using random subsets of the training data to minimize overfitting;
- Boosting is a technique in which models are trained progressively, and the weights of misclassified data points are raised to enhance performance;
- Stacking is the process of training many models and combining their predictions with another model to obtain the final forecast.

Ensemble Learning has proven useful in a variety of applications, including classification, regression, and anomaly detection. Following this, although Ensemble Learning uses multiple ML models to solve one task, the main difference between these two technologies is that Multimodal ML is able to analyze more than one dataset with more than one model to solve a task, while Ensemble Learning uses multiple models for the same dataset to solve a task. Therefore, unlike Multimodal ML, Ensemble Learning does not perform data fusion to solve the task. Table 3 below summarizes the comparison between Multimodal ML and other technologies.

Table 3. Multimodal ML vs. other technologies.

Technology \Specs	Definition	Main Goal	Perform Better than ML	Merge Datasources	Merge Models
Multimodal Datasets	Datasets that include multiple modalities of data	Enable Multimodal Machine Learning	Not Applicable	Yes	Not Applicable
Multilabel Learning	A supervised learning technique in which an instance can be assigned to multiple labels	Accurately label instances with multiple labels	Not Applicable	No	No
Ensemble Learning	Combines multiple models to improve the accuracy of the prediction	Improve prediction Accuracy	Yes	No	Yes
Multimodal ML	Combines multiple types of models/data to improve performance and feasibility	Improve Performance	Yes	Yes	Yes

2.5. Multimodal ML Available Frameworks

Multimodal Machine Learning frameworks provide a systematic approach for developing models that can learn and integrate information from multiple modalities such as text, audio, images, and other data types. As more and more data are created across multiple modalities, multimodal frameworks for Machine Learning are becoming increasingly important. These frameworks enable the integration of diverse information, allowing for a

more comprehensive understanding of complicated events. They're used in everything from speech recognition and natural language processing to image and video analysis. Some of the existing and commonly used Multimodal ML frameworks are:

- MMF (a framework for multimodal AI models) [54]: is a PyTorch-based modular framework. MMF comes with cutting-edge vision and language pretrained models, a slew of ready-to-use standard datasets, common layers and model components, and training and inference utilities. MMF is also utilized by various Facebook product teams for multimodal understanding of use cases, allowing them to swiftly put research into production;
- TinyM^{2}Net (a flexible system, algorithm co-designed multimodal learning framework for tiny devices) [55]: a unique multimodal learning framework that can handle multimodal inputs of images and audio and can be re-configured for individual application needs. TinyM2Net also enables the system and algorithm to incorporate fresh sensor data that are tailored to a variety of real-world settings. The suggested framework is built on a convolutional neural network, which has previously been recognized as one of the most promising methodologies for audio and visual data classification;
- A Unified Deep Learning Framework for Multimodal Multi-Dimensional Data [56]: is a framework capable of bridging the gap between data sufficiency/readiness and model availability/capability. For successful deployments, the framework is verified on multimodal, multi-dimensional data sets. The suggested architecture, which serves as a foundation, may be developed to solve a broad range of data science challenges utilizing Deep Learning;
- HAIM (unified Holistic AI in Medicine) [57]: It is a framework for developing and testing AI systems that make use of multimodal inputs. It employs generalizable data preprocessing and Machine Learning modeling steps that are easily adaptable for study and application in healthcare settings.
- ML4VocalDelivery [58]: a novel Multimodal Machine Learning technique that uses pairwise comparisons and a multimodal orthogonal fusing algorithm to create large-scale objective assessment findings of teacher voice delivery in terms of fluency and passion;
- Specific Knowledge Oriented Graph (SKOG) [59]: a technique for addressing multimodal analytics within a single data processing approach in order to obtain a streamlined architecture that can fully use the potential of Big Data infrastructures' parallel processing.

2.6. Training and Evaluation of Multimodal ML Algorithms

Multimodal Machine Learning is a technique that combines different modalities in an attempt to solve a complex task. Given that Multimodal ML is based on the concept of data fusion [46], the training process of a multimodal model may differ depending on the type of fusion (early, intermediate, or late fusion). Although it is a Machine Learning concept, it follows the familiar ML workflow, which would be: data preprocessing, model selection, model training, evaluation, fine-tuning, and deployment, but different steps may occur depending on the fusion stage.

First, in the case of early fusion, after preprocessing, the different datasets can be combined and merged into one modality. Once the data are ready and fused, it can be fed into the model to be trained, and then the other steps can be performed. In the second case, called intermediate fusion, the data passed to the same model are merged after preprocessing, then a single model is trained on the fused dataset, and later, the result of the refined model is fused with other models if they exist. Finally, in the late fusion approach, each dataset is passed to a different model after preprocessing, then the models are trained, evaluated, and fine-tuned, and later, the results are merged into a single result. The three approaches are shown in Figure 4 below.

On the other hand, the evaluation of the Multimodal ML model is also influenced by the chosen approach of data fusion. Since data fusion applies a single model to fused data sources, only a single evaluation is required. In the other two approaches,

intermediate and late fusion, each individual model must be evaluated, and later, the final model that merges the different models must be evaluated. The performance measures used to evaluate the Multimodal ML correspond to parameters commonly used in the classical ML domain, such as accuracy, precision, recall, sensitivity, specificity, F1 Score, Area Under Curve (AUC) and others [44]. The evaluation step is also shown in Figure 4 below.

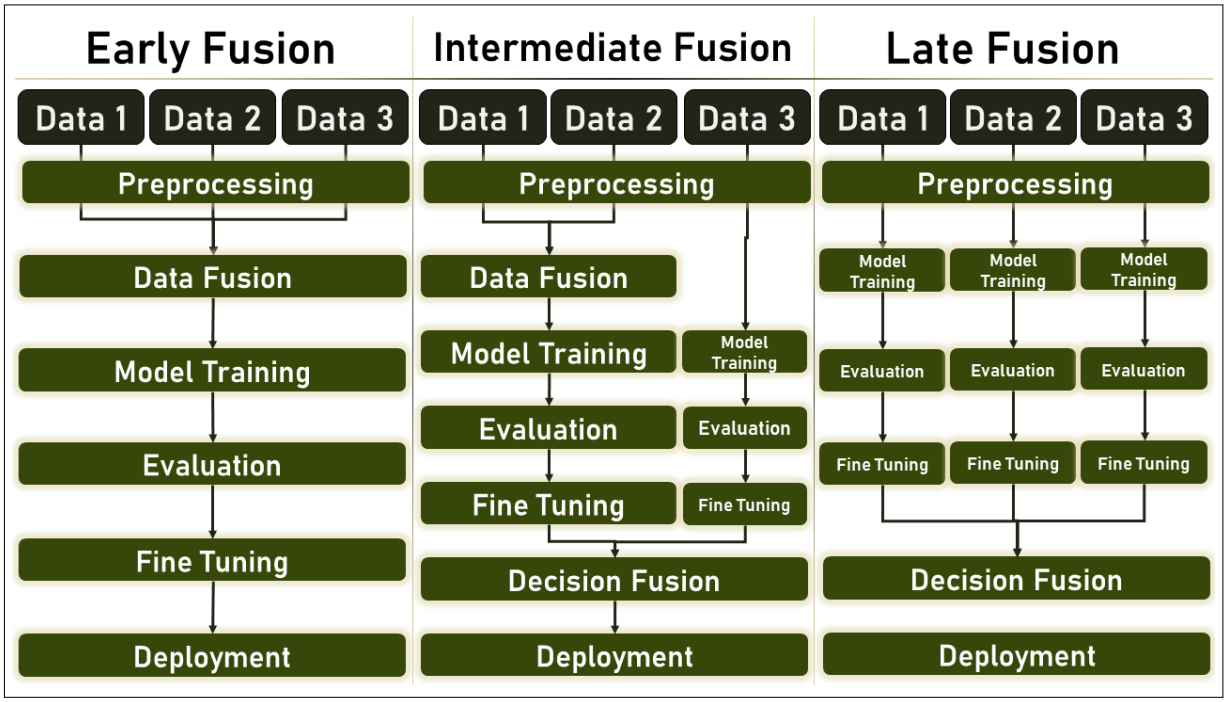

Figure 4. ML workflows based on Multimodal ML approaches.

3. Results: Multimodal ML in Action

Multimodal Machine Learning is a rapidly growing research area that involves the use of many modalities to evaluate and interpret complicated data, such as images, audio, and text [5,47]. Numerous real-world applications, including self-driving vehicles, voice recognition software, and medical imaging, require the ability to integrate and analyze data from multiple sources. Multimodal ML is based on the notion that multiple modalities provide complementary information and that merging these modalities can lead to more accurate and robust models. Multimodal ML has been a hot topic in the scientific community in recent years, and researchers have been striving to develop new algorithms and strategies to improve its performance [5,60–62].

3.1. Multimodal ML: Fields of Implementation

The ability to analyze diverse and complementary data increases the success of Machine Learning algorithms in solving more complex problems. In this context, Multimodal ML has proven its success in a variety of domains. Some of the most promising application areas include [5,60–64]:

- Healthcare: in medical imaging, Multimodal ML can be used to integrate information from different imaging modalities such as MRI, CT, and PET scans to improve diagnosis and treatment planning. It can also be used to classify and predict disease based on a mix of clinical, genetic, and imaging data;
- Autonomous Vehicles: by combining data from numerous sensors, the Multimodal ML can help self-driving vehicles better understand their surroundings. This has the potential to improve object recognition, navigation and safety;

- Natural Language Processing: by blending audio and text data, Multimodal ML can improve speech recognition and natural language comprehension. This can help voice assistants, chatbots and other applications improve their accuracy;
- Robotics: by combining inputs from sensors such as cameras, microphones, and touch sensors, Multimodal ML can be used to improve robot perception and interaction. This has the potential to improve navigation, object recognition, and human–robot interaction;
- Education: this technology is used in education to analyze student data from numerous sources, such as exams, quizzes, and essays, to make individualized learning suggestions and improve student performance;
- Agriculture: this technology can revolutionize agriculture by enabling the optimization of farming processes. It can be used for crop yield prediction, pest and disease detection, precision agriculture, and crop optimization by combining data from multiple sources, such as satellite imagery, weather data, and soil moisture sensors;
- Internet of Things (IoTs): this technology can be used in the context of the Internet of Things to make better use of data provided by networked devices. Multimodal ML can enable more accurate and robust models for predicting, monitoring, and managing IoT systems by incorporating data from many sources, such as sensors, cameras, and audio recordings, leading to advances in areas such as energy management, transportation, and smart cities.

3.2. Multimodal ML in Healthcare

Multimodal ML is still in its infancy but has been studied and applied in many areas of life, including healthcare. Multimodal ML is an effective method for assessing health data from multiple sources and improving predictive ability due to the inherent heterogeneity of such information [5,62,64]. To date, there are 128 applications of Multimodal ML in healthcare, with neurology and cancer being the most prevalent, as reported in [5]. Multimodal machine learning has shown promising results in various medical areas, as illustrated in Figure 5. While the areas depicted in the figure are the most commonly studied to date, it is worth noting that the potential applications of multimodal machine learning extend beyond these domains:

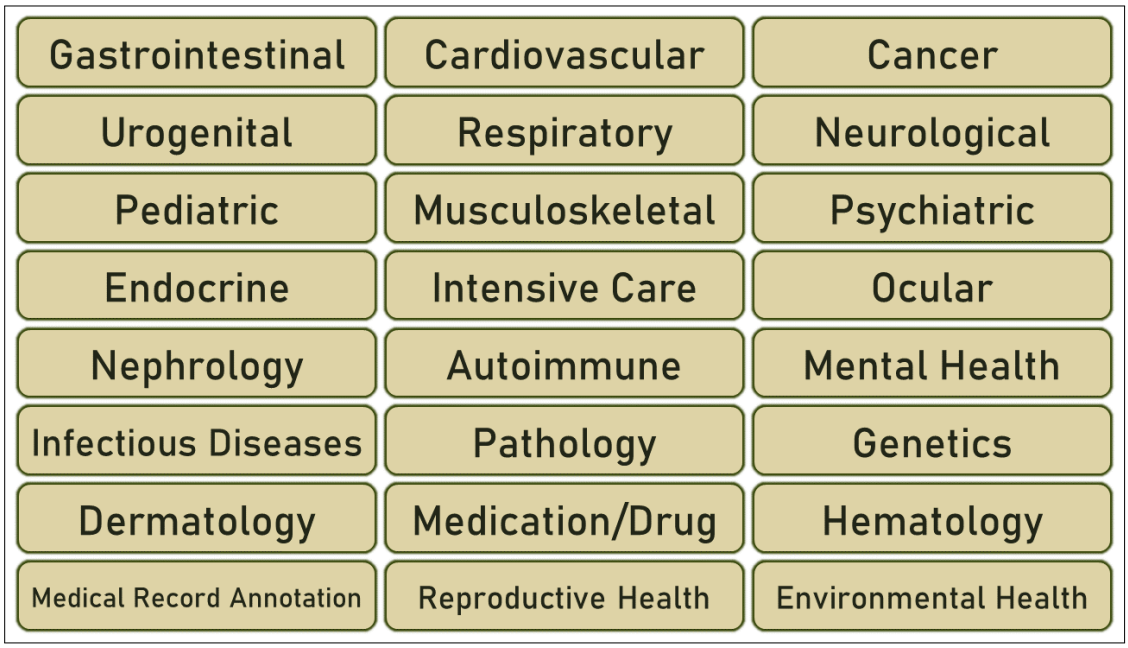

Figure 5. Healthcare sectors where Multimodal ML has been implemented so far.

3.3. Multimodal ML and Cardiovascular Diseases: State-of-the-Art

Cardiovascular Disease, the most deadly disease, is a topic of interest for Multimodal ML implementations. For example, in [65], the authors developed a multimodal data fusion ML model to predict hypertension. Using a Convolutional Neural Network (CNN)-based model, they analyzed different Electronic Health Records (EHRs) that were merged with the multimodal data fusion approach. Their model proved its efficiency with an accuracy that reached 94%. In a similar approach, the authors in [66] created a multimodal data fusion model to predict 30-day hospital readmission of patients with heart failure. For this purpose, they developed a Deep Unified Network (DUNs) trained with EHRs from the Enterprise Data Warehouse (EDW) and the Research Patient Data Repository (RPDR). Their model achieved an accuracy of 76.4%. In addition, the study [67] also implemented a data fusion model to cluster patients with hypertension. The authors proposed a novel Hybrid Non-Negative Matrix Factorization (HNMF) method-based model trained with phenotype and genotype information from the HyperGen dataset [68]. The accuracy of their proposed model reached up to 96%. In addition, the authors also developed a data fusion model in [69]. Their goal was to classify different CVDs, so they developed and trained a Text–Image Embedding network (TieNet) model with Chest X-Ray and free-text radiology clinical reports extracted from ChestX-Ray14 [70] and OpenI [71] Chest X-Rays datasets. The proposed model had an Area Under Curve (AUC) of 0.9, as they mentioned. In the same context, the solution proposed in [72] is a data fusion model developed to classify patients at potential cardiovascular risk. The model was based on Recurrent Neural Networks and trained on EHR data, achieving 96% accuracy.

Other implementations proposed model fusion or hybrid multimodal ML architectures to solve their problems. For example, in [73], the authors proposed a hybrid fusion multimodal ML to predict various cardiac diseases such as atelectasis, pleural effusion, cardiomegaly and edema. They created several ML models to analyze radiographs and associated reports obtained from MIMIC-CXR [74] and OpenI [71] Chest X-Ray datasets. Their proposed solution proved to be better than old implementations in terms of accuracy. Similarly, in [75], a multimodal unsupervised learning approach was proposed for Cardiometabolic Syndrome Detection. The authors applied multimodal hybrid fusion by combining unsupervised ML models to analyze fused data from metabolome, microbiome, genetics, and advanced imaging. Furthermore, in [76], the authors proposed a multimodal fusion-based ML model for stroke prediction. They fused both 3D Convolutional Neural Network and Multilayer Perceptron models to analyze neuroimaging information and clinical metadata extracted from the Hotter [77] dataset, which proved to be efficient and powerful with an AUC of 0.90. In addition, the solution proposed in [78] was used to predict Pulmonary Embolism (PE) by fusing multiple ML models trained with Computed Tomography Pulmonary Angiography scans and EHRs. Their model recorded an AUC of 0.947. Furthermore, in [79], the authors developed a Recurrent Neural Network model with Bidirectional Long-Term Memory (BiLSTM) to predict cardiovascular risk. Their model was trained with EHR data extracted from the Second Manifestations of ARTerial Disease (SMART) Study [80] and recorded an AUC of 0.847.

Similarly, in [81], the authors developed a data fusion model to predict Acute Ischemic Stroke. They used a series of cardiac CT images with EHR recordings to train a Gradient Boosting classifier that achieved an AUC of 0.856. Similarly, the study [82] proposed a Deep Convolutional Neural Network (DCNN) data fusion model to analyze Electrocardiograph (ECG) and Chest X-Ray images to efficiently predict Accessory Pathways (APs) syndrome. Finally, in [83], the authors proposed a novel tensor-based dimensionality reduction method using Naive Bayes, SVM, Random Forest, Adaboost, and LUCCK models. The created models were trained with fused data composed of Salient Physiological Signals and EHR data. Their solution was able to predict Hemodynamic Decompensation with an AUC value of 0.89. Table 4 below summarizes and presents the Multimodal ML implementations in CVDs.

Table 4. Multimodal ML implementations in Cardiovascular Disease diagnosis and prediction.

Ref	Year	Type	Parameter Studied	Predicted Outcome	Model	Architecture	Datasets Used	Performance
[65]	2017	Classification	EHR Data	Hypertension	Convolutional Neural Network	Data Fusion	Private Data	Accuracy: 94.8%
[66]	2018	Classification	EHR Data	Thirty-day readmission risk for heart failure patients	Deep Unified Networks (DUNs)	Data Fusion	Enterprise Data Warehouse (EDW) Research Patient Data Repository (RPDR)	Accuracy: 76.4%
[67]	2018	Clustering	Phenotype and Genotype Information	Hypertension	Hybrid Non-Negative Matrix Factorization (HNMF) model	Data Fusion	HyperGEN dataset [68]	Accuracy: 96%
[69]	2018	Classification	Chest X-Ray Clinical Free-Text Radiological Report Scan	Several CVDs	Text-Image Embedding network (TieNet)	Data Fusion	ChestX-Ray14 dataset [70] OpenI Chest X-Ray dataset [71]	AUC: 0.9
[72]	2019	Classification	EHR Data	Cardiovacsular Risk Prediction	Recurrent Convolutional Neural Network	Data Fusion	obtained from a grade-A hospital of second class in Wuhan	Accuracy: 96%
[73]	2020	Classification	MIMIC-CXR Radiographs and Associated Reports	Atelectasis, Pleural Effusion, Cardiomegaly, Edema	four pre-trained Vision+Language models: LXMERT / VisualBERT / UNIER / PixelBERT	Hybrid Fusion	MIMIC-CXR Chest X-Ray Dataset [74] OpenI Chest X-Ray Dataset [71]	Enhanced accuracy of classification
[75]	2020	Clustering	Metabolome Microbiome Genetics Advanced Imaging	Cardiometabolic Syndrome	Combianation of unsupervised ML Models	Hybrid Fusion	Private Data	-
[76]	2020	Classification	Neuroimaging Information Clinical Metadata	Stroke	3D Convolutional Neural Network Multilayer Perceptron	Model Fusion	Hotter Dataset [77]	AUC: 0.90
[78]	2020	Classification	Computed Tomography Pulmonary Angiography Scans EHR	Pulmonary Embolism (PE)	Combination of ML Models	Hybrid Fusion	Data obtained from Stanford University Medical Center (SUMC)	AUC: 0.947
[79]	2020	Classification	EHR Data	Cardiovascular Risk	Bidirectional Long Short-Term Memory (BiLSTM) Recurrent Neural Network	Hybrid Fusion	Second Manifestations of ARTerial Disease (SMART) Study [80]	AUC: 0.847
[81]	2020	Classification	Different Cardiac CT Images and EHR Data	Acute Ischemic Stroke	Gradient Boosting Classifiers	Data Fusion	obtained from Department of Neuroradiology at Heidelberg University Hospital (Heidelberg, Germany)	AUC: 0.856
[82]	2021	Classification	Electrocardiograph (ECG) Chest X-Ray	Cardiac Accessory Pathways (APs) Syndrome	Deep Convolutional Neural Network (DCNN)	Data Fusion	Private Data	-
[83]	2021	Classification	Salient Physiological Signals EHR Data	Hemodynamic Decompensation	Used a novel tensor-based dimensionality reduction with the below models: Naive Bayes SVM Random Forest Adaboost LUCCK	Data Fusion	Collected retrospectively from Michigan Medicine data systems	AUC: 0.89

3.4. Multimodal ML and CVDs: Discussion

Multimodal ML is a method for training different modalities using heterogeneous data that may not fit the same structure, format, or type that can be used for traditional ML algorithms. In the field of disease diagnosis, Multimodal ML could be used to train models on a huge distributed dataset of patient data from different hospitals or clinics. This method allows information and knowledge to be fused to solve complex problems. Using a larger, more diverse dataset also allows for more accurate and robust models. However, the implementation of Multimodal Machine Learning for disease prediction, especially Cardiovascular Disease, can be discussed from different angles, which are detailed in this section.

3.4.1. Models Performance: Competition between Multimodal and Classical ML

Data collection is the starting point for the operation of the established pipeline in the classical ML. It is generally accepted that more data can be used to increase the accuracy of an already trained Machine Learning model. It is generally accepted that due to the ability of Multimodal ML to analyze heterogeneous data, the accuracy of the models far exceeds that of typical ML models where more data are analyzed simultaneously.

In this context, the results presented in Table 4 reflect the high feasibility and accuracy that Multimodal ML cope with the diagnosis and prediction of Cardiovascular Disease. For example, the studies [65,67,72] achieved high accuracy records, with the first recording 94.8% and the other two, 96%. These results are highly comparable to the state of the art of conventional ML models used for the detection and prediction of CVDs and cerebrovascular events, with the highest recorded accuracy reaching 91.80%, as shown in [84]. In addition, the studies [69,76,78] recorded high values for Area Under Curve (AUC), with the first and second reaching a value of 0.9 and the third up to 0.95 for this parameter. These values demonstrate the high feasibility of these studies, which are consistent with and even exceed conventional ML algorithms. Moreover, the authors mention in [73] that their results show improved classification accuracy compared to conventional ML algorithms.

On the other hand, the results in [66] failed to outperform or even match conventional ML algorithms, where the recorded accuracy was 76.4%, which is lower than the values obtained by the latest ML algorithms in predicting ML models [84]. In addition, the studies [79,81,83] obtained different AUC values of 0.85, 0.86, and 0.89, respectively. These values are high and feasible, but they are close to but do not exceed the highest results obtained with classical ML models. Finally, the studies [75,82] did not mention the results obtained, which makes it impossible to compare their results with the classical ML models in the field of CVD diagnosis.

Overall, of the thirteen studies presented in Table 4, seven exceeded the results of the classical ML in terms of accuracy, three matched those results, and only one was obviously lower than them, and the other two are not comparable because they did not report their results. In this context, these figures help to confirm the hypothesis that the ability to analyze heterogeneous data increases the performance and accuracy of the models, which is a major strength in the field of multimodal ML since more than three-quarters of the Multimodal ML algorithms either match or exceed the results of the classic ML in the diagnosis of Cardiovascular Disease.

3.4.2. Real World vs. Research Implementations

The concept of Multimodal ML can be traced back to the early 2000s in the technology field, where authors in [85] suggested using this concept because the combination of communication modalities and acquisition devices can produce a wide range of unimodal and multimodal interface techniques. However, advances in computer technologies, data transmission, communication techniques, and other aspects have helped to increase the efficiency of Multimodal ML technology.

As a result, studies [65,75,85] have used their own data. Although these datasets are not publicly available, the authors assured that the data are real datasets collected from various health centers in compliance with medical standards and norms. This confirms that these studies can be classified as real-world studies. The same is true for [66,72,78,81,83], where each study used a dataset collected in different medical facilities in compliance with standard medical norms, making these studies real-world implementations.

On the other hand, the studies [67,69,73,76,79] used publicly available datasets, which are listed in Table 4. Although these datasets were collected under real-world conditions and obtained from patients, the study itself cannot be described as a real-world implementation. Real-world use of multimodal ML models in healthcare can provide a number of significant benefits, including:

- Improved Diagnostic Accuracy: Multimodal ML models can evaluate multiple sources of patient data, such as medical imaging, electronic health records, and genetic information, to make more accurate and thorough diagnoses. This can help physicians identify diseases and conditions at an early stage when they are more curable;
- Personalized Treatment: multimodal ML models can be trained on large data sets to identify trends and predict outcomes for individual patients. This can help physicians tailor treatments and therapies to the unique needs of each patient, leading to better outcomes and fewer side effects;
- Efficient Resource Allocation: Multimodal ML models can help physicians allocate resources more efficiently by identifying patients who are at higher risk for poor outcomes or need more intensive care. This has the potential to reduce healthcare costs while improving overall system efficiency;
- Improved patient experience: Multimodal ML models can help clinicians identify patients who need more individualized care or are at risk for problems or adverse events. This can help improve patient satisfaction and overall quality of care.

Overall, real-world adoption of Multimodal ML models in healthcare has the potential to enhance patient outcomes, lower costs, and improve healthcare delivery efficiency. However, it is critical that these models be created and used in an ethical manner, with proper protections for patient privacy and data security. That being said, the progress of Multimodal ML implementations and their real-world execution are promising where most of the carried applications are applied outside of labs, with real data, which enhances the trust in this technology and assists its adoption in the production stages.

3.4.3. Use of Smart Wearables and IoTs

Continuous monitoring of patients' heart rate, blood pressure and other biometric data through smart wearables and Internet of Things devices could revolutionize medical treatment. This has the potential to enable earlier detection of medical problems, more accurate diagnosis, and more personalized treatment approaches. Wearable technologies that can monitor and interact with the user's health could enable individuals to participate more fully in their treatment. In addition, Internet of Things (IoT) devices can enable physicians to monitor patients remotely and deliver treatments more effectively, reducing demand on healthcare systems and improving access to care for people in underserved or extremely remote and isolated areas. Smart wearables and Internet of Things (IoT) devices could increase hospital efficiency, save costs, and improve patient outcomes [86,87].

Consequently, only studies [67,75] considered the use of smart wearables or IoTs devices in their implementations. The other studies used data collected with other devices. Therefore, there is a lot of catching up to do in the implementation of multimodal ML in wearables and IoTs for CVD detection and prediction. Considering the fact that these technologies can revolutionize healthcare, as mentioned earlier, there is a great need to increase the use of wearables and IoTs in this field. In Table 5 below, the comparison between the performance of Multimodal ML and classical ML, the validation in practice,

and the use of smart wearables and IoTs for the state of the art in predicting CVDs with Multimodal ML is summarized.

Table 5. Key findings in state-of-the-art of Multimodal ML in CVDs diagnosis.

Ref#	Multimodal ML Beats ML (Performance)	Real-World Implementation	Smart Wearables/IoTs Included
[65]	Yes	Yes	No
[66]	No	Yes	No
[67]	Yes	Public Dataset(s)	Yes
[69]	Yes	Public Dataset(s)	No
[72]	Yes	Yes	No
[73]	Yes	Public Dataset(s)	No
[75]	Not Available	Yes	Yes
[76]	Yes	Public Dataset(s)	No
[78]	Results Match	Yes	No
[79]	Results Match	Public Dataset(s)	No
[81]	Results Match	Yes	No
[82]	Not Available	Yes	No
[83]	Results Match	Yes	No

3.4.4. Limitations in the Use of Multimodal ML for Disease Prediction

From this perspective, the use of Multimodal Machine Learning for the diagnosis and prognosis of CVDs is still in its infancy. Apart from the fact that not all implementations of Multimodal Machine Learning are superior to traditional ML models, vivid real-world examples can be observed when discussing this topic. Moreover, it has been rare to see FL researchers using smart wearables or IoTs in their experiments. This highlights the need to further investigate the use of such technologies due to their high degree of practicality and applicability in the field. Other limitations and difficulties encountered in the field of multimodal ML and its applications in disease prediction are discussed in Section 4.1, which can also be seen below.

3.5. Multimodal ML in CVDs: A Technical Overview

In Multimodal Machine Learning technology, the main goal is to analyze different data with different structures, such as merging EHR data with medical images to predict the occurrence of Cardiovascular Disease. In this context, each Multimodal ML implementation follows its own workflow and goes through its own steps to achieve its goal. In the aforementioned implementations of Cardiovascular Disease detection using Multimodal ML, different workflows, model structures, and hyperparameters were used for different implementations. All the related data provided by the authors are listed in Table 6 below.

Table 6. Technical details for Multimodal models used in the prediction of CVDs.

Ref#	Model	Workflow Description	Training Parameters
[65]	CNN-Based Multimodal Disease Risk Prediction (CNN-MDRP) Algorithm	1. Data Representation: text is represented in the form of vector 2. Convolution Layer: perform convolution operation on vectors of 5 words 3. Pool Layer: use the max pooling (1-max pooling) operation on the input of the convolution layer 4. Full Connection Layer: pooling layer is connected with a fully connected neural network 5. Classifier: the full connection layer links to a softmax classifier	Iterations: 200 Sliding Window: 7 Running Time: 1637.2 s
[66]	Deep Unified Networks (DUNs)	1. All inner layers of DUNs can learn the prediction task from the training data to avoid over-fitting 2. The DUNs architecture has horizontally shallow and vertically deep layers to prevent gradient vanishing and explosion 3. There are only two horizontal layers from the data unit nodes to the output node, regardless of how many layers deep the architecture is vertically 4. Only the harmonizing and decision units have learning parameters	Number of epochs: 100 Number of inner layers: 5 Number of inner neurons: 759 Number of maxout: – Activation function: Sigmoid Dropout rate of: input layer: 0.397/inner layers 0.433

Table 6. *Cont.*

Ref#	Model	Workflow Description	Training Parameters
[67]	Hybrid Non-Negative Matrix Factorization (HNMF) model	1. Impute missing values in the phenotypic variables 2. For genetic variants, first annotate the variants and then keep those that are likely gene disruptive (LGD) 3. The preprocessed phenotypic measurements and genetic variants are then used as input to the HNMF model 4. The patient factor matrix is then used as the feature matrix to perform regression analysis to predict main cardiac mechanistic outcomes	Up to 50 iterations
[69]	Text–Image Embedding Network (TieNet)	1. Data Preprocessing and word embedding 2. Training TieNet model 3. Joint Learning for results fusion 4. Evaluation	Dropout: 0.5 L2 Regularization: 0.0001 for. Adam optimizer with a mini-batch size of 32 Learning Rate of: 0.001 Hidden Layer with 350 units
[72]	Recurrent Convolutional Neural Network	1. Structured Data: extract relevant data, supplement missing data, make correlation analysis to look for the relation among data and apply dimension reduction to obtain corresponding structured features 2. Unstructured Textual Data: first, use numerical values to present unstructured textual data based on work embedding. Then, the features of textual data are extracted based on RCNN 3. Use Deep Belief Network (DBN) to fuse features and predict disease risks	up to 200 iterations
[73]	VisualBERT, UNITER, LXMERT, and PixelBER	1. The feature map ($7 \times 7 \times 1024$) of CheXNet is first flattened by spatial dimensions (49×1024) then down-sampled to 36 1024-long visual features 2. Models are then trained with the data 3. Results are fused	Epochs: PixelBERT: 18 / other 3 models 6 SGD optimizer weight decay 5×10^{-4} learning rate 0.01 Each model can be fit into 1 Tesla K40 GPU when using a batch size of 16
[75]	Collection of unsupervised ML models	1. Data collection and data features 2. Data preprocessing 3. Network analysis 4. Key biomarker selection and Markov network construction 5. Stratifying individuals with similar biomarker signatures 6. Validation cohort	-
[76]	3D Convolutional Neural Network Multilayer Perceptron	All models were trained on a binary classification task using binary cross-entropy loss	Loss function: Binary cross-entropy loss Adam optimizer Initial weights were sampled from a Glorot uniform distribution Output layer activation function: Softmax function Early stopping used to prevent over-fitting
[78]	Different ML models	Seven different workflows based on the difference between models	Batch Size: 256 Epochs: 200
[79]	Bidirectional Long Short-Term Memory (BiLSTM) Recurrent Neural Network	1. Embedding Layer: To extract the semantic information of radiology reports 2. Bidirectional-LSTM Layer: to achieve another representation of radiology reports 3. Dropout 4. Concatenation Layer 5. Dense Layers	Embedding dimension (d): 500 #neurons in LSTM layer: 100 CNN filter size: 5 filters in CNN: 128 neurons in dense layers: 64 Dropout rate: 0.2 Recurrent dropout rate: 0.2 Batch size: 64 epochs: 20 Optimization method ADAM
[81]	Gradient Boosting Classifiers	Integrative assessment of clinical, multimodal imaging, and angiographic characteristics with Machine Learning Allowed to accurately predict the clinical outcome following endovascular treatment for acute ischemic stroke	-
[82]	Deep Convolutional Neural Networks (DCNN)	First Model to analyze ECG 1. Convolutional Neural Network (CNN) 2. A one-dimensional CNN model was used to input the ECG data 3. The network model contained 16 convolution layers Followed by a fully connected layer 4. Then a Softmax layer, which calculated the probability of each of the four as the output in the last layer Second Model to analyze X-Ray images 1. A two-dimensional CNN model Then apply fusion to merge results	First Model Parameters: Adamax optimizer with the default parameters $\beta 1 = 0.9$, $\beta 2 = 0.999$, and a mini-batch size of 32

Table 6. *Cont.*

Ref#	Model	Workflow Description	Training Parameters
[83]	Random Forest Naive Bayes Support Vector Machine Adaboost Learning Using Concave and Convex Kernels (LUCCK)	1. Apply feature extraction on fused data composed of Salient Physiological Signals and EHR data 2. Apply Tensor reduction functionality 3. Train the Machine Learning model	Naive Bayes: (NB) no hyperparameter tuning was trained Support Vector Machines: used linear, radial basis function (RBF), and 3rd-order polynomial kernels Random Forest: number of trees: 50, 75, and 100/minimum leaf size: 1, 5, 10, 15, and 20/node splitting criterion: cross entropy and Gini impurity/number of predictors to sample: [10, 20, . . . , 100]/maximum number of decision splits for the decision trees: 0.25, 0.50, 0.75, or 1.0 Adaboost: learning rate: 1

4. Discussion: Challenges and Future Perspectives

Recently, Multimodal Machine Learning (ML) has emerged as an effective method for studying and analyzing complex data from multiple sources and modalities. However, dealing with diverse data presents researchers with unique challenges that must be overcome for efficient analysis and interpretation to increase the feasibility and usability of multimodal ML [10,48,49,62]. Unifying and standardizing multiple data sources and establishing links between them are significant obstacles. In addition, data must be normalized and preprocessed to ensure reliability and accuracy. However, future research could take several approaches to mitigate these challenges and overcome future obstacles. This section addresses these issues and identifies future perspectives needed to overcome them and improve multimodal FL.

4.1. Challenges

Multimodal Machine Learning still struggles with various challenges arising from the use of heterogeneous data with different structures and formats. Moreover, the fusion process, whether applied to the data itself or to different trained models to recognize a single result, is a challenging process that requires further research. Therefore, the most common challenges can be summarized in the following points [10,48,49,62].

4.1.1. Data Availability and Quality

To efficiently train multimodal ML models, large amounts of high-quality data are needed. However, collecting and processing large amounts of high-quality data in healthcare can be challenging, especially for rare or complex diseases. Data scarcity or poor data quality can lead to biased or unreliable models, compromising the accuracy of predictions and treatment decisions. To develop more robust and effective multimodal ML models for healthcare, researchers must seek to identify and address data quality and quantity issues.

4.1.2. Data Representation

Multimodal ML promotes the use of data from multiple sources for presentation. As a result, there is a high likelihood of dealing with heterogeneous data, which presents a number of problems. For example, it may be difficult to merge heterogeneous data that do not overlap in common characteristics or overlap only in a very limited area. In addition, data from different sources may need to be processed to different extents, especially with respect to noise reduction and missing data management. This hurdle is clearly reflected in the fact that until recently, most multimodal representations were simply the concatenation of unimodal ones [88]. Smoothness, temporal and spatial coherence, sparsity, and natural grouping have been cited by authors in [89] as qualities for excellent data representation.

4.1.3. Data Integration and Interoperability

Multimodal Machine Learning models are used to integrate and analyze data from multiple sources, such as electronic health records, medical imaging, and genetic data.

However, data from different sources may use different formats, standards, or terminologies, posing significant challenges for data integration and interoperability. Medical images, for example, may use different file formats or imaging techniques, making it difficult to compare and analyze data from different studies or sources.

4.1.4. Fusion

It is not easy to learn the ability to merge information from two modalities and determine the optimal fusion strategy. This is due to the different predictive capacities and noise structures of the different information coming from different senses. In addition, the ability to deal with missing data at different levels has a significant impact on the ability to perform fusion tasks.

4.1.5. Translation

The challenge in translation is not only the heterogeneity of data but also the relationships between modalities. The translation or mapping of data is subjective; for example, two models may describe the same image in more than one correct way, and a perfect or uniform translation or mapping may not exist. Several studies argue that while translations can be quite broad and modality-specific, they still have a number of unifying features. Accordingly, there are two forms of translation, namely the "Example-Based" and the "Generative" models. The former relies on a dictionary to translate data across modalities, while the latter relies on the creation of a model that manages translation according to uniform or at least explicit standards.

4.1.6. Alignment

Finding connections and correspondences between subelements from two or more different modalities is called multimodal alignment. This also involves distinguishing between these linear connections rather than just recognizing them. In this context, there are few data sets with obvious and identifiable correlations. Therefore, it is challenging to perform similarity measurements across modalities. Moreover, there may be numerous alignments without being able to select the optimal one, and not all components in one modality may match in another.

4.1.7. Explainability and Interpretability

Multimodal Machine Learning models (ML) have shown great promise in healthcare by enabling more accurate and tailored diagnosis and treatment recommendations. However, these models can be very complicated and difficult to understand, making it difficult for physicians to understand how the models arrived at a particular decision or recommendation. The lack of interpretability and openness of these models can affect their clinical acceptance and confidence.

4.1.8. Co-Learning

Merging different modalities, such as images, text, and sensor data, can increase model performance and enable more comprehensive analysis of complicated data in Multimodal Machine Learning. However, there are significant hurdles to this fusion, including the difficulty of transferring knowledge, representation, and predictive models across modalities. Each modality has its own characteristics and advantages, and it can be difficult to successfully integrate these aspects into a coherent representation. In addition, different modalities may require different strategies for feature engineering, preprocessing, and modeling.

4.1.9. Increased Computation Cost

When multiple modalities and features are introduced into a Multimodal Machine Learning model, the complexity of the model may increase, and the performance of the model may degrade due to the increased difficulty in computing the desired outcome.

Complex models have higher processing requirements, which can increase inference times and memory consumption. The complexity of a model makes it more difficult to optimize, which can lead to an increased risk of over- or under-fitting the data.

4.1.10. Regulatory and Ethical Considerations

Apart from the technical hurdles in developing and implementing multimodal ML models in healthcare, there are also legal and ethical factors to consider. Depending on their intended use, these models may be subject to regulatory restrictions, such as the European Union's General Data Protection Regulation (GDPR) [90], China's Cyber Security Law of the People's Republic of China [91], the General Principles of the Civil Law of the People's Republic of China [92], the PDPA in Singapore [93], and hundreds of principles that apply around the world. In addition, researchers and clinicians must ensure that these models are created and used in an ethical manner and that patient privacy and data security are adequately protected. For example, patient data must be de-identified and protected from illegal access or disclosure. In addition, maintaining the fairness and openness of these models is critical to minimize bias and discrimination. Responsible development and adoption of multimodal ML models therefore require careful evaluation of these legal and ethical factors to ensure that they deliver safe, effective, and fair outcomes for patients.

4.1.11. Implementation and Adoption

To fully deliver on their promise to improve healthcare, Multimodal Machine Learning models (ML) must be integrated into current healthcare processes and systems. However, several barriers stand in the way of this integration, such as technological, organizational, and cultural. In addition to the technical challenges mentioned above, resistance to change, lack of stakeholder participation, and concerns about accountability and obligations are all examples of organizational and cultural hurdles that may arise.

These challenges give rise to the study questions in the list below (the abbreviation RQ in the list below refers to the term "research question"):

- **RQ1:** Multimodal ML needs sufficient data to be trained. Are the needed data sets available? And is their quality acceptable?
- **RQ2:** Multimodal ML deals with heterogeneous data that has different formats and structures. What approaches can be taken to represent the data used in this technology?
- **RQ3:** How can the heterogeneous data used in Multimodal ML be integrated and shared?
- **RQ4:** What are the best approaches for fusion, and how to choose between the different options available?
- **RQ5:** Given that different models can lead to the same result in different ways, how does one choose the optimal path?
- **RQ6:** How to align and link two different modalities, especially in the middle and late fusion cases?
- **RQ7:** The Multimodal ML is known for its black box identity. Is there a way to explain the methods by which a model arrives at its result?
- **RQ8:** In Multimodal ML, different models can be integrated to solve a complex task. What techniques can be applied to ensure efficient knowledge transfer between these models?
- **RQ9:** Heterogeneity and diversity in both models and data add to computational costs. How can this problem be dealt with to improve the usability and feasibility of the models?
- **RQ10:** How to ensure data exchange between multimodal ML facilities to comply with existing regulations and laws?
- **RQ11:** How can trust in multimodal ML be strengthened to promote its adoption in different areas of life?

4.2. Future Perspectives

The challenges faced in Multimodal Machine Learning can be solved through different approaches and perspectives. These solutions have either already been considered but should be more widely used in the field of Cardiovascular Disease prediction to improve and increase their usability and feasibility. In this context, the following solutions can serve as future recommendations.

4.2.1. Use Convenient Tools to Collect More Data

Modern technology has changed the method of data collection and analysis. The use of smart wearables and Internet-of-Things (IoT) devices has enabled the real-time collection of vast amounts of data [33,39,86,87]. These data can provide useful insights in a variety of areas, particularly in healthcare. In addition to these new data sources, current data sources should be used to create more complete databases. Researchers can gain access to larger and more diverse data sets by collaborating with other institutions, which can help them identify patterns and correlations that would not be obvious with smaller data sets. Collaboration between different institutions could be achieved using a variety of techniques such as Federated Machine Learning technology, which can help train Machine Learning models by sharing parameters rather than the data itself [9].

4.2.2. Automate and Boost Data Preprocessing

Creating larger and more comprehensive datasets could help improve the quality of Machine Learning models but is not yet sufficient. To gain valuable insights, data must be processed and analyzed using advanced techniques. These techniques include artifact automation and noise removal, as performed in [94,95]. In addition, it may be necessary to use techniques such as data augmentation [96] or data normalization [97] and data resampling [98] to ensure that the data are balanced and ready for model training and to improve the quality of the overall process.

4.2.3. Employment of Advanced Data Integration Tools

To address the problems posed by the diversity of data formats and structures, improved methods for data harmonization [99], standardization [100], and normalization [97] need to be developed, as well as the use of AI and ML algorithms to automate these processes. Multimodal ML has the potential to revolutionize healthcare by enabling thorough and tailored analysis of patient data from numerous sources if these barriers are overcome.

4.2.4. Embedding Modern Techniques to Enhance Explainability

To address the problems associated with the black-box nature of multimodal ML models, more explainable and interpretable models are needed that give healthcare professionals insight into how the models arrive at their judgments. Approaches such as feature relevance ranking [101], model visualization [102], decision rules [103], probabilistic [104] and neuro-fuzzy approaches [105], and many others can improve the interpretability of multimodal ML models so that interested parties can make more informed and confident treatment decisions. In the list below, a brief definition for each of these tools is presented:

- Feature relevance ranking: include methods such as permutation significance and partial dependency plots to give insights into the importance and correlations of input variables, allowing for a better understanding of the model's decision-making process and boosting transparency and interpretability in healthcare applications;
- Model visualization: such as decision trees and heatmaps that provide a graphical representation of the model's decision-making process, allowing for better understanding of the factors that influence the model's predictions and increasing the transparency and interpretability of the technology;

- Decision rules: by providing clear and understandable rationales for the model's predictions, decision rules that specify explicit decision criteria based on the input data improve the interpretability and transparency of machine learning models in healthcare.
- Probabilistic approach: employ probabilistic reasoning to represent and manage the uncertainty inherent in medical data allowing for transparent decision-making that can be easily understood by healthcare practitioners;
- Neuro-fuzzy techniques: combine the benefits of neural networks and fuzzy logic to generate more interpretable models that can deal with imprecise and uncertain inputs.

4.2.5. Implementing Necessary Methods to Guarantee Knowledge Transfer

The diversity of datasets and models in the field of multimodal ML can lead to knowledge transfer problems. Therefore, researchers need to develop novel strategies for multimodal feature selection [106], fusion [46], and modeling that can capture complementary information from many modalities while minimizing redundancy or overfitting. Overcoming these obstacles will allow for more robust and accurate multimodal ML models that will lead to improved diagnosis, treatment, and patient outcomes in healthcare settings.

4.2.6. Reducing Computation Cost

Reducing computational costs in multimodal ML is a critical issue. Therefore, researchers need to explore methods for model compression [107] and optimization [108] that can reduce the computational complexity of the model without compromising its performance. As an added bonus, Multimodal Machine Learning can benefit from efficient hardware and software implementations, such as specialized hardware accelerators and distributed computing frameworks, that can reduce computational load. The use of such techniques can help build multimodal ML models that are more robust, efficient, and scalable, and therefore applicable to a wider variety of health problems, leading to faster and more accurate solutions.

4.2.7. Increase Trust and Feasibility to Raise the Technology Adoption

Researchers, clinicians, information technology experts, and healthcare administrators must work together to increase confidence in multimodal ML technology. In addition, cultural and organizational barriers can be reduced by promoting trust and transparency through open dialog and training. The best way to improve patient outcomes and revolutionize healthcare delivery is to properly integrate multimodal ML models into current healthcare delivery processes and systems.

The results of the mapping of challenges and solutions can be summarized in the following topics (the symbol TR in the list below refers to the term "Trending Research Topic"):

- **TR1:** Data collection tools such as smart wearables and IoTs are very helpful in augmenting the data collected for multimodal ML algorithms;
- **TR2:** Data harmonization, standardization, and normalization are highly feasible for integrating heterogeneous data in the multimodal ML domain;
- **TR3:** Multimodal feature selection and modeling are techniques that can help ensure knowledge transfer between different modalities in a multimodal ML system;
- **TR4:** For better explainability and interpretability of a multimodal ML model, decision rules, feature relevance ranking, and model visualization are practical and feasible methods;
- **TR5:** Model compression and optimization are great tools for reducing computational costs in multimodal ML;
- **TR6:** Current and trending ML topics, such as Federated Machine Learning, can help overcome privacy and confidentiality issues in the Multimodal ML domain;
- **TR7:** Increasing feasibility, improving performance, and implementation in real-world scenarios are all factors that can help expand the adoption of multimodal ML technology in healthcare and, in particular, in Cardiovascular Disease detection.

Finally, the challenges that hinder the progress of Multimodal Machine Learning techniques, along with the solutions and future perspectives that could be pursued, are presented in Figure 6 below.

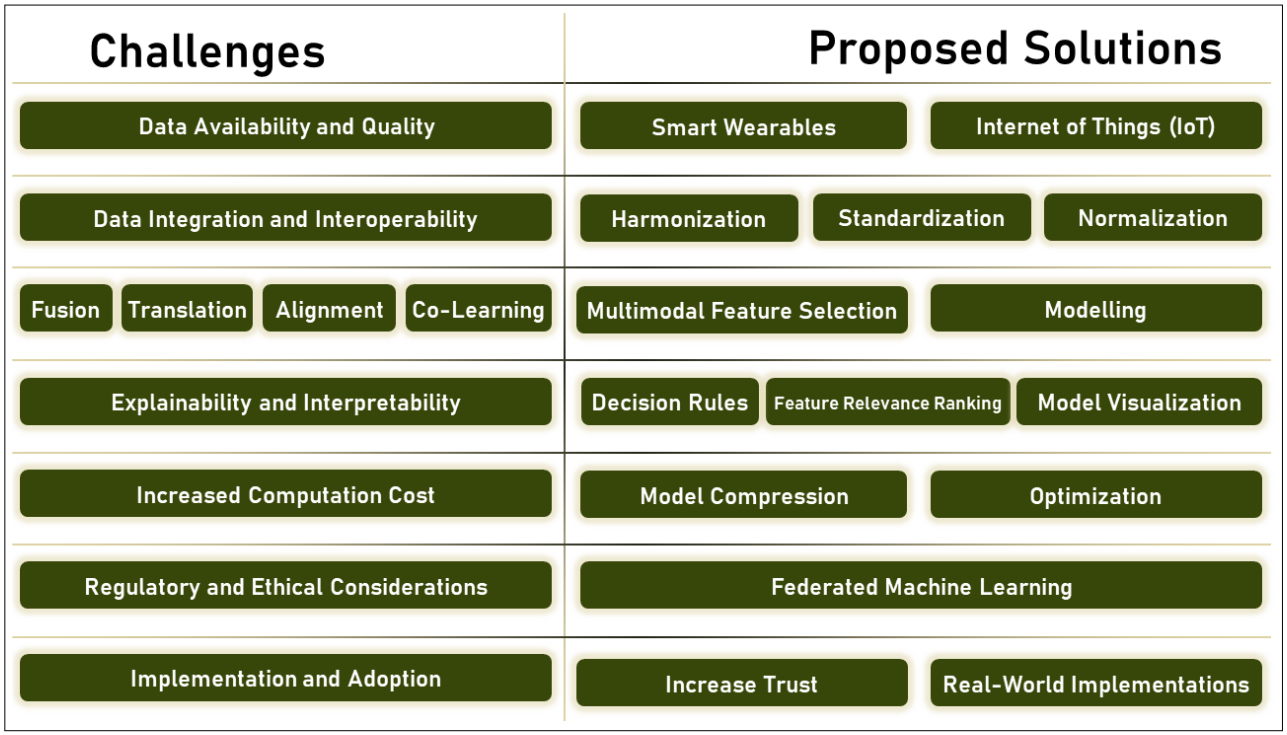

Figure 6. Multimodal machine learning challenges–solutions mapping.

5. Conclusions

In summary, Multimodal ML is a new technique that enables the simultaneous use of multiple models and data types in the creation of complex ML and DL models. Multimodal ML has the potential to significantly improve the accuracy and effectiveness of AI applications, especially in healthcare, where it has already become an important part of everyday patient care by addressing the problem of data heterogeneity. In particular, the technical features of Multimodal ML, such as data fusion and workflows, were covered, and the differences with other technologies, such as Ensemble Learning, were highlighted. In addition, an overview of the application of Multimodal ML in the diagnosis and prediction of Cardiovascular Disease was provided, highlighting the encouraging results to date and the room for growth in this area. Privacy, bias, and interpretability of results are just some of the remaining difficulties that need to be addressed, as with any rapidly evolving technology. However, it is likely that these obstacles can be addressed through further research and development and that multimodal ML will continue to play an important role in the development of AI applications in a variety of sectors, particularly healthcare.

Author Contributions: Conceptualization: M.M. and M.A.; formal analysis: M.M.; investigation: M.M.; methodology: M.M. and M.A.; supervision: M.A., A.B., H.I. and A.R.; visualization: M.M.; writing—original draft: M.M.; writing—review and editing: M.A., A.B., H.I. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number 06351.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Acknowledgments: We acknowledge the support of Centre d'Entrepreneuriat et de Valorisation des Innovations (CEVI).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Moor, J. The Dartmouth College artificial intelligence conference: The next fifty years. *AI Mag.* **2006**, *27*, 87–87.
2. Simone, N.; Ballatore, A. Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence* **2020**, *26*, 3–18.
3. John, M. *What Is Artificial Intelligence*; Stanford University: Stanford, CA, USA, 2007
4. Ramach, ram, D.; Taylor, G.W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process. Mag.* **2017**, *34*, 96–108.
5. Kline, A.; Wang, H.; Li, Y.; Dennis, S.; Hutch, M.; Xu, Z.; Luo, Y. Multimodal Machine Learning in Precision Health. *arXiv* **2022**, arXiv:2204.04777.
6. Ngiam, J.; Khosla, A.; Kim, M.; Nam, J.; Lee, H.; Ng, A.Y. Multimodal deep learning. In Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011; pp. 689–696.
7. Giuseppe, B. *Machine Learning Algorithms*; Packt Publishing Ltd.: Birmingham, UK, 2017.
8. Yann, L.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
9. Mohammad, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Reviewing Federated Machine Learning and Its Use in Diseases Prediction. *Sensors* **2023**, *23*, 2112.
10. Tadas, B.; Ahuja, C.; Morency, L. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 423–443.
11. Pallathadka, H.; Mustafa, M.; Sanchez, D.T.; Sajja, G.S.; Gour, S.; Naved, M. Impact of machine learning on management, healthcare and agriculture. *Mater. Today Proc.* 2021, *in press*. http://doi.org/10.1016/j.matpr.2021.07.042
12. Ghazal, T.M.; Hasan, M.K.; Alshurideh, M.T.; Alzoubi, H.M.; Ahmad, M.; Akbar, S.S.; Al Kurdi, B.; Akour, I.A. IoT for smart cities: Machine learning approaches in smart healthcare—A review. *Future Internet* **2021**, *13*, 218.
13. Erickson, B.J.; Korfiatis, P.; Akkus, Z.; Kline, T.L. Machine learning for medical imaging. *Radiographics* **2017**, *37*, 505.
14. Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 1–21.
15. Sharma, N.; Sharma, R.; Jindal, N. Machine learning and deep learning applications-a vision. *Glob. Transitions Proc.* **2021**, *2*, 24–28.
16. Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D. A review of machine learning and IoT in smart transportation. *Future Internet* **2019**, *11*, 94.
17. Xin, Y.; Kong, L.; Liu, Z.; Chen, Y.; Li, Y.; Zhu, H.; Gao, M.; Hou, H.; Wang, C. Machine learning and deep learning methods for cybersecurity. *IEEE Access* **2018**, *6*, 35365–35381.
18. Nagarhalli, T.P.; Vaze, V.; Rana, N.K. Impact of machine learning in natural language processing: A review. In Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, Tirunelveli, India, 4–6 February 2021; pp. 1529–1534.
19. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674.
20. Larrañaga, P.; Atienza, D.; Diaz-Rozo, J.; Ogbechie, A.; Puerto-Santana, C.; Bielza, C. *Industrial Applications of Machine Learning*; CRC Press: Boca Raton, FL, USA, 2018.
21. L'heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A. Machine learning with big data: Challenges and approaches. *IEEE Access* **2017**, *5*, 7776–7797.
22. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. *Neurocomputing* **2017**, *237*, 350–361.
23. Injadat, M.; Moubayed, A.; Nassif, A.B.; Shami, A. Machine learning towards intelligent systems: Applications, challenges, and opportunities. *Artif. Intell. Rev.* **2021**, *54*, 3299–3348.
24. Leskovec, J.; Rajaraman, A.; Ullman, J.D. *Mining of Massive Data Sets*; Cambridge University Press: Cambridge, UK, 2020.
25. Paleyes, A.; Urma, R.G.; Lawrence, N.D. Challenges in deploying machine learning: A survey of case studies. *ACM Comput. Surv.* **2020**, *55*, 1–29.
26. Char, D.S.; Shah, N.H.; Magnus, D. Implementing machine learning in health care—Addressing ethical challenges. *N. Engl. J. Med.* **2018**, *378*, 981.
27. Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.D. Machine learning in manufacturing: Advantages, challenges, and applications. *Prod. Manuf. Res.* **2016**, *4*, 23–45.
28. Rosario, M.; Mukhopadhyay, S.C.; Liu, Z.; Slomovitz, D.; Samantaray, S.R. Advances on sensing technologies for smart cities and power grids: A review. *IEEE Sens. J.* **2017**, *17*, 7596–7610.
29. Total Data Volume Worldwide 2010–2025 | Statista. Petroc Taylor. 8 September 2022. Statista. Available online: https://www.statista.com/statistics/871513/worldwide-data-created/ (accessed on 15 February 2023).
30. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144.
31. Lidong, W. Heterogeneous data and big data analytics. *Autom. Control. Inf. Sci.* **2017**, *3*, 8–15.

32. Geert, L.; Ciompi, F.; Wolterink, J.M.; de Vos, B.D.; Leiner, T.; Teuwen, J.; Išgum, I. State-of-the-art deep learning in cardiovascular image analysis. *JACC Cardiovasc. Imaging* **2019**, *12*, 1549–1565.
33. Mohammad, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Cardiovascular Diseases: A Systematic Literature Review. *Sensors* **2023**, *23*, 828.
34. Aamir, J.; Zghyer, F.; Kim, C.; Spaulding, E.M.; Isakadze, N.; Ding, J.; Kargillis, D.; Gao, Y.; Rahman, F.; Brown, D.E.; et al. Medicine 2032: The future of cardiovascular disease prevention with machine learning and digital health technology. *Am. J. Prev. Cardiol.* **2022**, *12*, 100379.
35. Ramesh, A.N.; Kambhampati, C.;Monson, J.R.L.; Drew, P.J. Artificial intelligence in medicine. *Ann. R. Coll. Surg. Engl.* **2004**, *86*, 334.
36. Maddox, T.M.; Rumsfeld, J.S.; Payne, P.R. Questions for artificial intelligence in health care. *JAMA* **2019**, *321*, 31–32.
37. Amine, M.M.; Adda, M.; Bouzouane, A.; Ibrahim, H. Machine learning and smart devices for diabetes management: Systematic review. *Sensors* **2022**, *22*, 1843.
38. Shweta, C.; Biswas, N.; Jones, L.D.; Kesari, S.; Ashili, S. Smart Consumer Wearables as Digital Diagnostic Tools: A Review. *Diagnostics* **2022**, *12*, 2110.
39. Mohammad, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Smart Wearables for the Detection of Occupational Physical Fatigue: A Literature Review. *Sensors* **2022**, *22*, 7472.
40. Yukang, X. A review on intelligent wearables: Uses and risks. *Hum. Behav. Emerg. Technol.* **2019**, *1*, 287–294.
41. Marie, C.; Estève, D.; Fourniols, J.; Escriba, C.; Campo, E. Smart wearable systems: Current status and future challenges. *Artif. Intell. Med.* **2012**, *56*, 137–156.
42. Chinthaka, J.S.M.D.A.; Ganegoda, G.U. Involvement of machine learning tools in healthcare decision making. *J. Healthc. Eng.* **2021**, *2021*, 6679512.
43. Sameer, Q. Artificial intelligence and machine learning in precision and genomic medicine. *Med. Oncol.* **2022**, *39*, 120.
44. Arjun, P. *Machine Learning and AI for Healthcare*; Apress: Coventry, UK, 2019.
45. Nitish, S.; Salakhutdinov, R.R. Multimodal learning with deep boltzmann machines. *Adv. Neural Inf. Process. Syst.* **2014**, *15*, 2949–2980.
46. White, F.E. *Data Fusion Lexicon*; Joint Directors of Labs: Washington, DC, USA, 1991.
47. Baronio, M.A.; Cazella, S.C. Multimodal Deep Learning for Computer-Aided Detection and Diagnosis of Cancer: Theory and Applications. *Enhanc. Telemed. Health Adv. Iot Enabled Soft Comput. Framew.* **2021**, 267–287.
48. Baltrušaitis, T.; Ahuja, C.; Morency, L.P. Challenges and applications in multimodal machine learning. In *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition—Volume 2*; Association for Computing Machinery and Morgan & Claypool: San Rafael, CA, USA, 2018; pp. 17–48. https://doi.org/10.1145/3107990.
49. Anil, R.; Walambe, R.; Ramanna, S.; Kotecha, K. Multimodal co-learning: challenges, applications with datasets, recent advances and future directions. *Inf. Fusion* **2022**, *81*, 203–239.
50. Grigorios, T.; Katakis, I. Multi-label classification: An overview. *Int. J. Data Warehous. Min.* **2007**, *3*, 1–13.
51. Min-Ling, Z.; Zhou, Z. A review on multi-label learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 1819–1837.
52. Xibin, D.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2020**, *14*, 241–258.
53. Omer, S.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249.
54. Announcing MMF: A Framework for Multimodal AI Models. Announcing MMF: A Framework for Multimodal AI Models. Available online: https://ai.facebook.com/blog/announcing-mmf-a-framework-for-multimodal-ai-models/ (accessed on 18 February 2023).
55. Hasib-Al, R.; Ovi, P.R.; Gangopadhyay, A.; Mohsenin, T. TinyM2Net: A Flexible System Algorithm Co-designed Multimodal Learning Framework for Tiny Devices. *arXiv* **2022**, arXiv:2202.04303.
56. Pengcheng, X.; Shu, C.; Goubran, R. A Unified Deep Learning Framework for Multi-Modal Multi-Dimensional Data. In Proceedings of the 2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Istanbul, Turkey, 26–28 June 2019; pp. 1–6.
57. Ma, S.L.R.Y.; Zeng, C.; Boussioux, L.; Carballo, K.V.; Na, L.; Wiberg, H.M.; Li, M.L.; Fuentes, I.; Bertsimas, D. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit. Med.* **2022**, *5*, 149.
58. Hang, L.; Kang, Y.; Hao, Y.; Ding, W.; Wu, Z.; Liu, Z. A Multimodal Machine Learning Framework for Teacher Vocal Delivery Evaluation. In *Proceedings of the Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, 14–18 June 2021*; Springer International Publishing: Cham, Switzerland, 2021; Part II, pp. 251–255.
59. Valerio, B.; Ceravolo, P.; Maghool, S.; Siccardi, S. Toward a general framework for multimodal big data analysis. *Big Data* **2022**, *10*, 408–424.
60. YJing, A.; Liang, N.; Pitts, B.J.; Prakah-Asante, K.O.; Curry, R.; Blommer, M.; Swaminathan, R.; Yu, D. Multimodal Sensing and Computational Intelligence for Situation Awareness Classification in Autonomous Driving. *IEEE Trans. -Hum.-Mach. Syst.* **2023**, *53*, 270–281.
61. Azin, A.; Saha, R.; Jakubovitz, D.; Peyre, J. AutoFraudNet: A Multimodal Network to Detect Fraud in the Auto Insurance Industry. *arXiv* **2023**, arXiv:2301.07526.
62. Arnab, B.; Ahmed, M.U.; Begum, S. A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions. *IEEE Access* **2023**, 11, 14804–14831.

63. Lemay, P.C.S.D.G.; Owen, C.L.; Woodward-Greene, M.J.; Sun, J. Multimodal AI to Improve Agriculture. *IT Prof.* **2021**, *23*, 53–57.
64. Yuchen, Z.; Barnaghi, P.; Haddadi, H. Multimodal federated learning on iot data. In Proceedings of the 2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI), Milano, Italy, 4–6 May 2022; pp. 43–54.
65. Min, C.; Hao, Y.; Hwang, K.; Wang, L.; Wang, L. Disease prediction by machine learning over big data from healthcare communities. *IEEE Access* **2017**, *5*, 8869–8879.
66. Bersche, G.S.; Shibahara, T.; Agboola, S.; Otaki, H.; Sato, J.; Nakae, T.; Hisamitsu, T.; Kojima, G.; Felsted, J.; Kakarmath, S.; et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *Bmc Med. Inform. Decis. Mak.* **2018**, *18*, 1–17.
67. Yuan, L.; Mao, C.; Yang, Y.; Wang, F.; Ahmad, F.S.; Arnett, D.; Irvin, M.R.; Shah, S.J. Integrating hypertension phenotype and genotype with hybrid non-negative matrix factorization. *Bioinformatics* **2019**, *35*, 1395–1403.
68. Rao, W.R.R.D.C.; Ellison, R.C.; Arnett, D.K.; Heiss, G.; Oberman, A.; Eckfeldt, J.H.; Leppert, M.F.; Province, M.A.; Mockrin, S.C.; Hunt, S.C.; et al. NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. *Ann. Epidemiol.* **2000**, *10*, 389–400.
69. Xiaosong, W.; Peng, Y.; Lu, L.; Lu, Z.; Summers, R.M. Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 9049–9058.
70. Xiaosong, W.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2097–2106.
71. Dina, De.; Kohli, M.D.; Rosenman, M.B.; Shooshan, S.E.; Rodriguez, L.; Antani, S.; Thoma, G.R.; McDonald, C.J. Preparing a collection of radiology examinations for distribution and retrieval. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 304–310.
72. Yixue, H.; Usama, M.; Yang, J.; Hossain, M.S.; Ghoneim, A. Recurrent convolutional neural network based multimodal disease risk prediction. *Future Gener. Comput. Syst.* **2019**, *92*, 76–83.
73. Yikuan, L.; Wang, H.; Luo, Y. A comparison of pre-trained vision-and-language models for multimodal representation learning across medical images and reports. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, South Korea, 16–19 December 2020; pp. 1999–2004.
74. Pollard, J.A.E.W.T.J.; Greenbaum, N.R.; Lungren, M.P.; Deng, Ch.; Peng, Y.; Lu, Z.; Mark, R.G.; Berkowitz, S.J.; Horng, S. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv* **2019**, arXiv:1901.07042.
75. Ilan, S.; Cirulli, E.T.; Huang, L.; Napier, L.A.; Heister, R.R.; Hicks, M.; Cohen, I.V.; Yu, H.C.; Swisher, C.L.; Schenker-Ahmed, N.M.; et al. An unsupervised learning approach to identify novel signatures of health and disease from multimodal data. *Genome Med.* **2020**, *12*, 1–14.
76. Esra, Z.; Madai, V.I.; Khalil, A.A.; Galinovic, I.; Fiebach, J.B.; Kelleher, J.D.; Frey, D.; Livne, M. Multimodal Fusion Strategies for Outcome Prediction in Stroke. In Proceedings of the 13th International Conference on Health Informatics, Valletta, Malta, 24–26 February 2020; pp. 421–428.
77. Benjamin, H.; Pittl, S.; Ebinger, M.; Oepen, G.; Jegzentis, K.; Kudo, K.; Rozanski, M.; Schmidt, W.U.; Brunecker, P.; Xu, C.; et al. Prospective study on the mismatch concept in acute stroke patients within the first 24 h after symptom onset-1000Plus study. *BMC Neurol.* **2009**, *9*, 60.
78. Shih-Cheng, H.; Pareek, A.; Zamanian, R.; Banerjee, I.; Lungren, M.P. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Sci. Rep.* **2020**, *10*, 1–9.
79. Ayoub, B.; Groenhof, T.K.J.; Veldhuis, W.B.; de Jong, P.A.; Asselbergs, F.W.; Oberski, D.L. Multimodal learning for cardiovascular risk prediction using EHR data. *arXiv* **2020**, arXiv:2008.11979.
80. Gerarda, S.P.C.; Algra, A.; Laak, M.F.V.D.; Grobbee, D.E.; Graaf, Y.V.D. Second manifestations of ARTerial disease (SMART) study: rationale and design. *Eur. J. Epidemiol.* **1999**, *15*, 773–781.
81. Gianluca, B.; Neuberger, U.; Mahmutoglu, M.A.; Foltyn, M.; Herweh, C.; Nagel, S.; Schönenberger, S.; Heiland, S.; Ulfert, C.; Ringleb, P.A.; et al. Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke* **2020**, *51*, 3541–3551.
82. Makoto, N.; Kiuchi, K.; Nishimura, K.; Kusano, K.; Yoshida, A.; Adachi, K.; Hirayama, Y.; Miyazaki, Y.; Fujiwara, R.; Sommer, P.; El Hamriti, M. et al. Accessory pathway analysis using a multimodal deep learning model. *Sci. Rep.* **2021**, *11*, 8045.
83. Larry, H.; Kim, R.; Tokcan, N.; Derksen, H.; Biesterveld, B.E.; Croteau, A.; Williams, A.M.; Mathis, M.; Najarian, K.; Gryak, J. Multimodal tensor-based method for integrative and continuous patient monitoring during postoperative cardiac care. *Artif. Intell. Med.* **2021**, *113*, 102032.
84. Mohammad, M.; Adda, M.; Bouzouane, A.; Ibrahim, H.; Raad, A. Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability. *Procedia Comput. Sci.* **2022**, *203*, 231–238.
85. Matthew, T. Gesture recognition. In *Handbook of Virtual Environments*; CRC Press: Boca Raton, FL, USA, 2002; pp. 263–278.
86. Armando, P.; Mital, M.; Pisano, P.; Giudice, M.D. E-health and wellbeing monitoring using smart healthcare devices: An empirical investigation. *Technol. Forecast. Soc. Chang.* **2020**, *153*, 119226.
87. Nasiri, A.Z.; Rahmani, A.M.; Hosseinzadeh, M. The role of the Internet of Things in healthcare: Future trends and challenges. *Comput. Methods Programs Biomed.* **2021**, *199*, 105903.
88. K, D.S.; Kory, J. A review and meta-analysis of multimodal affect detection systems. *Acm Comput. Surv.* **2015**, *47*, 1–36.

89. Yoshua, B.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828.
90. Albrecht, J.P. How the GDPR will change the world. *Eur. Data Prot. L. Rev.* **2016**, *2*, 287.
91. Parasol, M. The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams. *Comput. Law Secur. Rev.* **2018**, *34*, 67–98.
92. Gray, W.; Zheng, H.R. General Principles of Civil Law of the People's Republic of China. *Am. J. Comp. Law* **1986**, *34*, 715–743.
93. Chik, W.B. The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform. *Comput. Law Secur. Rev.* **2013**, *29*, 554–575.
94. Islam, M.K.; Rastegarnia, A.; Sanei, S. Signal Artifacts and Techniques for Artifacts and Noise Removal. In Signal Processing Techniques for Computational Health Informatics; Springer: Cham, Switzerland, 2021; pp. 23–79.
95. Daly, I.; Billinger, M.; Scherer, R.; Müller-Putz, G. On the automated removal of artifacts related to head movement from the EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013**, *21*, 427–434.
96. Dyk, V.; A, D.; Meng, Xi. The art of data augmentation. *J. Comput. Graph. Stat.* **2001**, *10*, 1–50.
97. Dalwinder, S.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2020**, *97*, 105524.
98. Jameel, M.A.; Hassan, M.M.; Kadir, D.H. Improving classification performance for a novel imbalanced medical dataset using SMOTE method. *Int. J.* **2020**, *9*, 3161–3172.
99. Ganesh, K.; Basri, S.; Imam, A.A.; Khowaja, S.A.; Capretz, L.F.; Balogun, A.O. Data harmonization for heterogeneous datasets: a systematic literature review. *Appl. Sci.* **2021**, *11*, 8275.
100. Michal S.G.; Rubinfeld, D.L. Data standardization. *NYUL Rev.* **2019**, *94*, 737.
101. Maksymilian, W.; Chen, K. Feature importance ranking for deep learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 5105–5114.
102. Angelos, C.; Martins, R.M.; Jusufi, I.; Kerren, A. A survey of surveys on the use of visualization for interpreting machine learning models. *Inf. Vis.* **2020**, *19*, 207–233.
103. Alberto, Bl.; Domingo-Ferrer, J. Machine learning explainability through comprehensible decision trees. In *Machine Learning and Knowledge Extraction: Third IFIP TC 5, TC 12, WG 8.4, WG 8.9, WG 12.9 International Cross-Domain Conference, CD-MAKE 2019, Canterbury, UK, 26–29 August 2019*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 15–26.
104. Stephan, W. *Towards Explainable Artificial Intelligence: Interpreting Neural Network Classifiers with Probabilistic Prime Implicants*; Technische Universitaet: Berlin, Germany, 2022.
105. Edwin, L. Evolving fuzzy and neuro-fuzzy systems: Fundamentals, stability, explainability, useability, and applications. In *Handbook on Computer Learning and Intelligence: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation*; World Scientific: Singapore, 2022; pp. 133–234.
106. Shima, K.; Eftekhari, M. Feature selection using multimodal optimization techniques. *Neurocomputing* **2016**, *171*, 586–597.
107. Tejalal, C.; Mishra, V.; Goswami, A.; Sarangapani, J. A comprehensive survey on model compression and acceleration. *Artif. Intell. Rev.* **2020**, *53*, 5113–5155.
108. Shiliang, S.; Cao, Z.; Zhu, H.; Zhao, J. A survey of optimization methods from a machine learning perspective. *IEEE Trans. Cybern. 2019*, **50**, 3668–3681.

References

[1] Moor, James. "The Dartmouth College Artificial Intelligence conference: The next fifty years." Ai Magazine 27, no. 4 (2006): 87-87.

[2] Natale, Simone, and Andrea Ballatore. "Imagining the thinking machine: Technological myths and the rise of Artificial Intelligence." Convergence 26, no. 1 (2020): 3-18.

[3] McCarthy, John. "What is Artificial Intelligence." (2007): 2020.

[4] Ramachandram, Dhanesh, and Graham W. Taylor. "Deep multimodal learning: A survey on recent advances and trends." IEEE signal processing magazine 34, no. 6 (2017): 96-108.

[5] Ramachandram, Dhanesh, and Graham W. Taylor. "Deep multimodal learning: A survey on recent advances and trends." IEEE signal processing magazine 34, no. 6 (2017): 96-108.

[6] Ngiam, Jiquan, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng. "Multimodal Deep Learning." In Proceedings of the 28th international conference on Machine Learning (ICML-11), pp. 689-696. 2011.

[7] Giuseppe Bonaccorso, Machine Learning Algorithms. Packt Publishing Ltd. 2017. Birmingham, United Kingdom.

[8] Moshawrab, Mohammad, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, and Ali Raad. "Reviewing Federated Machine Learning and Its Use in Diseases Prediction." Sensors 23, no. 4 (2023): 2112.

[9] Sarker, Iqbal H. "Machine Learning: Algorithms, real-world applications and research directions." SN computer science 2, no. 3 (2021): 160.

[10] Sharma, Neha, Reecha Sharma, and Neeru Jindal. "Machine Learning and Deep Learning applications-a vision." Global Transitions Proceedings 2, no. 1 (2021): 24-28.

[11] Pallathadka, Harikumar, Malik Mustafa, Domenic T. Sanchez, Guna Sekhar Sajja, Sanjeev Gour, and Mohd Naved. "Impact of Machine Learning on management, healthcare and agriculture." Materials Today: Proceedings 80 (2023): 2803-2806.

[12] Ghazal, Taher M., Mohammad Kamrul Hasan, Muhammad Turki Alshurideh, Haitham M. Alzoubi, Munir Ahmad, Syed Shehryar Akbar, Barween Al Kurdi, and Iman A. Akour. "IoT for smart cities: Machine Learning approaches in smart healthcare—A review." Future Internet 13, no. 8 (2021): 218.

[13] Erickson, Bradley J., Panagiotis Korfiatis, Zeynettin Akkus, and Timothy L. Kline. "Machine Learning for medical imaging." Radiographics 37, no. 2 (2017): 505-515.

[14] Zantalis, Fotios, Grigorios Koulouras, Sotiris Karabetsos, and Dionisis Kandris. "A review of Machine Learning and IoT in smart transportation." Future Internet 11, no. 4 (2019): 94.

[15] Xin, Yang, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixia Hou, and Chunhua Wang. "Machine Learning and Deep Learning methods for cybersecurity." Ieee access 6 (2018): 35365-35381.

[16] Nagarhalli, Tatwadarshi P., Vinod Vaze, and N. K. Rana. "Impact of Machine Learning in natural language processing: A review." In 2021 third international conference on intelligent communication technologies and virtual mobile networks (ICICV), pp. 1529-1534. IEEE, 2021.

[17] Liakos, Konstantinos G., Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. "Machine Learning in agriculture: A review." Sensors 18, no. 8 (2018): 2674.

[18] Larrañaga, Pedro, David Atienza, Javier Diaz-Rozo, Alberto Ogbechie, Carlos Esteban Puerto-Santana, and Concha Bielza. Industrial applications of Machine Learning. CRC press, 2018.

[19] Aggarwal, Karan, Maad M. Mijwil, Abdel-Hameed Al-Mistarehi, Safwan Alomari, Murat Gök, Anas M. Zein Alaabdin, and Safaa H. Abdulrhman. "Has the future

started? The current growth of Artificial Intelligence, Machine Learning, and Deep Learning." Iraqi Journal for Computer Science and Mathematics 3, no. 1 (2022): 115-123.

[20] Carta, Silvio, ed. Machine Learning and the city: applications in architecture and urban design. John Wiley & Sons, 2022.

[21] Hardt, Moritz, and Benjamin Recht. "Patterns, predictions, and actions: A story about Machine Learning." arXiv preprint arXiv:2102.05242 (2021).

[22] Dasgupta, Ariruna, and Asoke Nath. "Classification of Machine Learning algorithms." International Journal of Innovative Research in Advanced Engineering (IJIRAE) 3, no. 3 (2016): 6-11.

[23] Ray, Susmita. "A quick review of Machine Learning algorithms." In 2019 International conference on Machine Learning, big data, cloud and parallel computing (COMITCon), pp. 35-39. IEEE, 2019.

[24] Alzubi, Jafar, Anand Nayyar, and Akshi Kumar. "Machine Learning from theory to algorithms: an overview." In Journal of physics: conference series, vol. 1142, p. 012012. IOP Publishing, 2018.

[25] Verbraeken, Joost, Matthijs Wolting, Jonathan Katzy, Jeroen Kloppenburg, Tim Verbelen, and Jan S. Rellermeyer. "A survey on distributed Machine Learning." Acm computing surveys (csur) 53, no. 2 (2020): 1-33.

[26] Panayiotou, Tania, Giannis Savvas, Ioannis Tomkos, and Georgios Ellinas. "Centralized and distributed Machine Learning-based QoT estimation for sliceable optical networks." In 2019 IEEE Global Communications Conference (GLOBECOM), pp. 1-7. IEEE, 2019.

[27] L'heureux, Alexandra, Katarina Grolinger, Hany F. Elyamany, and Miriam AM Capretz. "Machine Learning with big data: Challenges and approaches." Ieee Access 5 (2017): 7776-7797.

[28] Zhou, Lina, Shimei Pan, Jianwu Wang, and Athanasios V. Vasilakos. "Machine Learning on big data: Opportunities and challenges." Neurocomputing 237 (2017): 350-361.

[29] Paleyes, Andrei, Raoul-Gabriel Urma, and Neil D. Lawrence. "Challenges in deploying Machine Learning: a survey of case studies." ACM Computing Surveys 55, no. 6 (2022): 1-29.

[30] Char, Danton S., Nigam H. Shah, and David Magnus. "Implementing machine learni

[31] Wuest, Thorsten, Daniel Weimer, Christopher Irgens, and Klaus-Dieter Thoben. "Machine Learning in manufacturing: advantages, challenges, and applications." Production & Manufacturing Research 4, no. 1 (2016): 23-45.

[32] Injadat, MohammadNoor, Abdallah Moubayed, Ali Bou Nassif, and Abdallah Shami. "Machine Learning towards intelligent systems: applications, challenges, and opportunities." Artificial Intelligence Review 54 (2021): 3299-3348.

[33] Albrecht, Jan Philipp. "How the GDPR will change the world." Eur. Data Prot. L. Rev. 2 (2016): 287.

[34] Parasol, Max. "The impact of China's 2016 Cyber Security Law on foreign technology firms, and on China's big data and Smart City dreams." Computer law & security review 34, no. 1 (2018): 67-98.

[35] Gray, Whitmore, and Henry Ruiheng Zheng. "General principles of civil law of the People's Republic of China." The American Journal of Comparative Law 34, no. 4 (1986): 715-743.

[36] Chik, Warren B. "The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform." Computer Law & Security Review 29, no. 5 (2013): 554-575.

[37] Lyu, Lingjuan, Han Yu, and Qiang Yang. "Threats to federated learning: A survey." arXiv preprint arXiv:2003.02133 (2020).

[38] Bonawitz, Keith, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. "Practical secure aggregation for privacy-preserving Machine Learning." In proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1175-1191. 2017.

[39] Pillutla, Krishna, Sham M. Kakade, and Zaid Harchaoui. "Robust aggregation for Federated Learning." IEEE Transactions on Signal Processing 70 (2022): 1142-1154.

[40] Varma, Kamala, Yi Zhou, Nathalie Baracaldo, and Ali Anwar. "Legato: A layerwise gradient aggregation algorithm for mitigating byzantine attacks in Federated Learning." In 2021 IEEE 14th international conference on cloud computing (CLOUD), pp. 272-277. IEEE, 2021.

[41] Jeon, Beomyeol, S. M. Ferdous, Muntasir Raihan Rahman, and Anwar Walid. "Privacy-preserving decentralized aggregation for Federated Learning." In IEEE INFOCOM 2021-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), pp. 1-6. IEEE, 2021.

[42] Zhao, Lingchen, Jianlin Jiang, Bo Feng, Qian Wang, Chao Shen, and Qi Li. "Sear: Secure and efficient aggregation for byzantine-robust Federated Learning." IEEE Transactions on Dependable and Secure Computing 19, no. 5 (2021): 3329-3342.

[43] Song, Jingcheng, Weizheng Wang, Thippa Reddy Gadekallu, Jianyu Cao, and Yining Liu. "Eppda: An efficient privacy-preserving data aggregation Federated Learning scheme." IEEE Transactions on Network Science and Engineering (2022).

[44] Elkordy, Ahmed Roushdy, and A. Salman Avestimehr. "Heterosag: Secure aggregation with heterogeneous quantization in Federated Learning." IEEE Transactions on Communications 70, no. 4 (2022): 2372-2386.

[45] Zhang, Zaixi, Xiaoyu Cao, Jinyuan Jia, and Neil Zhenqiang Gong. "FLDetector: Defending Federated Learning against model poisoning attacks via detecting malicious clients." In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2545-2555. 2022.

[46] Cao, Xiaoyu, Zaixi Zhang, Jinyuan Jia, and Neil Zhenqiang Gong. "Flcert: Provably secure Federated Learning against poisoning attacks." IEEE Transactions on Information Forensics and Security 17 (2022): 3691-3705.

[47] Rathee, Mayank, Conghao Shen, Sameer Wagh, and Raluca Ada Popa. "Elsa: Secure aggregation for Federated Learning with malicious actors." In 2023 IEEE Symposium on Security and Privacy (SP), pp. 1961-1979. IEEE, 2023.

[48] So, Jinhyun, Ramy E. Ali, Başak Guler, Jiantao Jiao, and A. Salman Avestimehr. "Securing secure aggregation: Mitigating multi-round privacy leakage in Federated Learning." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, no. 8, pp. 9864-9873. 2023.

[49] Hardy, Stephen, Wilko Henecka, Hamish Ivey-Law, Richard Nock, Giorgio Patrini, Guillaume Smith, and Brian Thorne. "Private Federated Learning on vertically partitioned data via entity resolution and additively homomorphic encryption." arXiv preprint arXiv:1711.10677 (2017).

[50] Ou, Wei, Jianhuan Zeng, Zijun Guo, Wanqin Yan, Dingwan Liu, and Stelios Fuentes. "A homomorphic-encryption-based vertical Federated Learning scheme for rick management." Computer Science and Information Systems 17, no. 3 (2020): 819-834.

[51] Sav, Sinem, Apostolos Pyrgelis, Juan R. Troncoso-Pastoriza, David Froelicher, Jean-Philippe Bossuat, Joao Sa Sousa, and Jean-Pierre Hubaux. "POSEIDON: Privacy-preserving federated neural network learning." arXiv preprint arXiv:2009.00349 (2020).

[52] Liu, Xiaoyuan, Hongwei Li, Guowen Xu, Zongqi Chen, Xiaoming Huang, and Rongxing Lu. "Privacy-enhanced Federated Learning against poisoning adversaries." IEEE Transactions on Information Forensics and Security 16 (2021): 4574-4588.

[53] Tian, Haibo, Fangguo Zhang, Yunfeng Shao, and Bingshuai Li. "Secure linear aggregation using decentralized threshold additive homomorphic encryption for Federated Learning." arXiv preprint arXiv:2111.10753 (2021).

[54] Stripelis, Dimitris, Hamza Saleem, Tanmay Ghai, Nikhil Dhinagar, Umang Gupta, Chrysovalantis Anastasiou, Greg Ver Steeg et al. "Secure neuroimaging analysis using Federated Learning with homomorphic encryption." In 17th International Symposium on Medical Information Processing and Analysis, vol. 12088, pp. 351-359. SPIE, 2021.

[55] Zhang, Li, Jianbo Xu, Pandi Vijayakumar, Pradip Kumar Sharma, and Uttam Ghosh. "Homomorphic encryption-based privacy-preserving Federated Learning in iot-enabled healthcare system." IEEE Transactions on Network Science and Engineering (2022).

[56] Alenezi, Mohammed N., Haneen Alabdulrazzaq, and Nada Q. Mohammad. ”Symmetric encryption algorithms: Review and evaluation study.” International Journal of Communication Networks and Information Security 12, no. 2 (2020): 256-272.

[57] Daemen, Joan, and Vincent Rijmen. ”Reijndael: The advanced encryption standard.” Dr. Dobb's Journal: Software Tools for the Professional Programmer 26, no. 3 (2001): 137-139.

[58] Li, P.; Li, J.; Huang, Z.; Li, T.; Gao, C.Z.; Yiu, S.M.; Chen, K. Multi-key privacyp-reserving deep learning in cloud computing. Future Gener. Comput. Syst. 2017, 74, 76–85.

[59] Moshawrab, Mohammad. “GitHub - MohMsh/PolyFLAG_SVM.” GitHub, January 1, 2024. https://github.com/MohMsh/PolyFLAG_SVM.

[60] Moshawrab, Mohammad. “GitHub - MohMsh/PolyFLAM.” GitHub, January 1, 2024. https://github.com/MohMsh/PolyFLAM.

[61] Moshawrab, Mohammad. “GitHub - MohMsh/PolyFLAP.” GitHub, January 1, 2024. https://github.com/MohMsh/PolyFLAP.

[62] Moshawrab, Mohammad. “GitHub - MohMsh/HP_FLAP.” GitHub, January 1, 2024. https://github.com/MohMsh/HP_FLAP.

[63] McMahan, Brendan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. "Communication-efficient learning of deep networks from decentralized data." In Artificial Intelligence and statistics, pp. 1273-1282. PMLR, 2017.

[64] Jamil, Bushra, Humaira Ijaz, Mohammad Shojafar, Kashif Munir, and Rajkumar Buyya. "Resource allocation and task scheduling in fog computing and internet of everything environments: A taxonomy, review, and future directions." ACM Computing Surveys (CSUR) 54, no. 11s (2022): 1-38.

[65] Feng, Yong, Jinglong Chen, Jingsong Xie, Tianci Zhang, Haixin Lv, and Tongyang Pan. "Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: Algorithms, applications, and prospects." Knowledge-Based Systems 235 (2022): 107646.