Pédagogie interactive de l'apprentissage automatique : développement d'une plateforme Web pour la modélisation prédictive Clinique

Mémoire présenté

dans le cadre du programme de maîtrise en informatique

en vue de l’obtention du grade de maître ès sciences (M. Sc.)

PAR

© Fatemeh Bashardoustjoubjarkouli

Août 2025

Composition du jury :

Mehdi Adda, Directeur de recherche, Université du Québec à Rimouski

Mohamed Tarik Moutacalli, président du jury, Université du Québec à Rimouski

Hamid Mcheick, Membre interne Université du Québec à Chicoutimi

Maryam Yeganeh, examinateur externe, Senior Full Stack developer at DECIDE4ACTION

Dépôt initial le 18-12-2024 Dépôt final le 24-08-2025

À la mémoire de ma chère mère qui a toujours été une source d'inspiration pour moi. Que son âme repose en paix.

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui ont contribué à la réalisation de ce projet. Tout d'abord, je remercie mon professeur Mehdi Adda, mon directeur de recherche, pour son encadrement précieux, ses conseils avisés et son soutien indéfectible tout au long de ce parcours académique. Je tiens également à exprimer ma sincère reconnaissance aux membres du jury pour le temps qu'ils ont consacré à évaluer mon travail, ainsi que pour leurs recommandations et leurs commentaires constructifs qui enrichissent cette recherche.

Je souhaite également remercier ma famille, en particulier mon mari Mohsen, pour sa patience et son encouragement, ainsi que mes enfants Radmehr et Anahita, dont l'amour et la compréhension m'ont donné la force nécessaire pour mener à bien cette recherche. Leur soutien m'a permis de surmonter les défis et de rester concentré sur mes objectifs.

Merci à tous ceux qui ont contribué, directement ou indirectement, à la réalisation de ce travail.

TABLE OF CONTENTS

RÉSUMÉ

Cette Mémoire présente le développement d'une plateforme web visant à améliorer l'accessibilité des outils d'apprentissage automatique (ML) pour les professionnels de la santé. Le but de la recherche est de combler le fossé entre la complexité technique et l'utilisation pratique du ML pour la modélisation prédictive clinique, en particulier pour ceux ayant une expertise limitée en informatique. La plateforme, construite avec Flask, intègre divers modèles de ML, y compris K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes, Decision Tree, ainsi que des algorithmes d'apprentissage non supervisé comme K-Means et DBSCAN. Ces algorithmes sont appliqués pour analyser des ensembles de données cliniques, classifier les maladies et prédire les résultats des patients. L'hypothèse principale de ce mémoire est que le développement d'une plateforme web conviviale peut faciliter l'accès aux outils d'apprentissage automatique pour les utilisateurs disposant de connaissances de base en ML, mais sans compétences en programmation. Cette plateforme vise à leur offrir un environnement structuré pour charger des jeux de données cliniques, sélectionner des algorithmes, et visualiser les résultats, dans le but de soutenir la prise de décision clinique. La méthodologie comprend le développement d'une interface utilisateur modulaire et réactive, permettant aux utilisateurs de télécharger des ensembles de données, de prétraiter les données, et de choisir les algorithmes appropriés pour leur analyse. L'étude conclut que la plateforme améliore considérablement l'accessibilité des outils de ML pour l'analyse des données cliniques. Cependant, la portée actuelle de la plateforme la limite à des ensembles de données spécifiques,et en mettant à jour le système avec d'autres algorithmes d'apprentissage automatique pour maintenir sa pertinence et soutenir les avancées en cours .

Mots clés: Modélisation prédictive clinique, Apprentissage automatique, Flask, KNN, SVM, Naive Bayes, Decision Tree, DBSCAN, K-Means.

ABSTRACT

This study presents the development of a web-based platform aimed at improving the accessibility of machine learning (ML) platform for healthcare professionals. The goal of the research is to bridge the gap between technical complexity and practical usage of ML for clinical predictive modeling, particularly for those with limited expertise in computer science. The platform, built using Flask, integrates various ML models, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes, Decision Trees, and unsupervised learning algorithms like K-Means and DBSCAN. These algorithms are applied to analyze clinical datasets, classifying diseases and predicting patient outcomes. The main hypothesis of this thesis is that the development of a user-friendly web platform can facilitate access to machine learning tools for users who have basic knowledge of ML but no programming skills**.** This platform aims to provide them with a structured environment to upload clinical datasets, select algorithms, and visualize results, with the goal of supporting clinical decision-making. The methodology includes the development of a modular and responsive user interface, allowing users to upload datasets, preprocess data, and choose appropriate algorithms for their analysis. The study concludes that the platform significantly improves the accessibility of ML platform for clinical data analysis. However, the platform's current scope limits it to specific datasets, and future work should focus on expanding its capabilities to include broader types of clinical data and by updating the system with other machine learning algorithms to maintain its relevance and support ongoing advancements.

Keywords: Clinical predictive modeling, Machine learning, Flask, KNN, SVM, Naive Bayes, Decision Trees, DBSCAN, K-Means.

General Introduction

1.Introduction

In the recent past, there has been a substantial increase in the volume of complex documents and texts in the clinical field, demanding a more profound comprehension of machine learning methods to be able to accurately classify texts in many applications [1].Machine learning (ML) has transformed the diagnosis and treatment of diseases, and it is now a crucial component of current healthcare. Particularly in fields like diabetes, cardiovascular illness, and cancer, machine learning techniques have shown impressive promise in the diagnosis of diseases [2]. Machine learning (ML) improves healthcare by helping with decision-making, automating billing, and facilitating tasks for clinicians. It is very good at diagnosis and detection; it promises tailored treatment; it helps with drug discovery; and it can identify diseases from pictures. When these algorithms are widely used, human oversight is necessary [3]. The use of AI in healthcare increased rapidly between 2021 and 2024, according to measurements released[4] , and it is expected to reach $188 billion by 2030. The most popular AI applications in healthcare, according to the survey, is natural language processing (NLP) which helps doctors and clinicians make decisions by predictive analytics combines statistical models and machine learning to estimate/predict outcome using historical data [5].

Machine learning models can be made available to users by deploying them on platforms like websites or mobile apps using web technologies. This study discusses web technologies [6], [7], [8], [9], [10] and machine learning for the prediction of various diseases. Therefore, having web applications as a platform might make things easier and more accessible for people who are unfamiliar with computer science. The World Wide Web has created opportunities for the development of both static and dynamic websites. The widespread use of programming and markup languages, such as JavaScript, and HTML5, has made it simpler than ever to create user-friendly web apps[6] . Therefore, we design and develop platforms that ease access to machine learning models for clinical data analysis. We used Flask [11] in our application development stack. Flask is a Python micro framework that offers basic web functionality and allows for the addition of more plugins, making it simple yet extensible for development. We developed a web application designed to work with a unique clinical dataset . This Flask platform, enhanced with Bootstrap Front-End for a responsive design, integrates several machine learning algorithms. These include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, and Decision Trees, which are employed to classify and predict outcomes based on our clinical symptom datasets. Additionally, we incorporate clustering techniques like DBSCAN and K-Means to uncover hidden patterns within the data.

2. PROBLEMATIC

While numerous studies have explored the application of machine learning models to clinical data, a gap remains in providing accessible and user-friendly platforms that allow users to effectively use these models without coding experts but possess a basic understanding of machine learning knowledge.

Classifying clinical text in electronic medical records poses inherent challenges. Almazaydeh et al. [10] presented a machine learning method for categorizing medical transcriptions, incorporating phases such as text preparation, word representation, feature selection, and classification. Yumeng Guo et al [11] proposed an algorithm for multi-label text classification of clinical records, focusing on robust feature classification while mitigating the adverse effects of training datasets through a forward search strategy.

Although several studies have investigated the existence of web-based platforms that can use machine learning models for prediction and data analysis on a user interface platform for easier accessibly. In [12] researchers use the python flask web development framework to incorporate the models like Decision-Tree (DT), Naïve-Bayes (NB), K-Nearest Neighbor (KNN), Random-Forest (RF), Gradient-Boosting (GB), Logistic-Regression (LR), and Support Vector Machine (SVM) that may be taught using a variety of datasets. This study's findings provide preliminary evidence that using a suitable preprocessing pipeline on clinical data and using ML-based classification might improve the accuracy and efficiency of diabetes prediction. Th researchers in the articles [8] investigated to develop a front-end web application for centralized tracking and management of incidentalomas by the hospital's quality department. They used neural networks for document classification in this research.

In this thesis, we introduce an easy-to-use web-based platform created especially for clinical predictive modelling that makes code-free machine-learning tools easily accessible to healthcare professionals with a basic knowledge of ML . By merging web technologies and automated machine learning pipelines, we aim to bridge the usability gap between complicated machine learning codes and users by ensuring that users can conveniently upload the specific dataset, select models, and interpret results without having to deal into complex code.

3. OBJECTIVE

This project aims to bridge the gap in clinical data analysis by providing healthcare professionals with easier and more accessible platform for predictive modeling.

- The main objective is to develop a user-friendly platform that simplifies access to our specific clinical data analysis, enabling model selection, and result visualization.

- To provide an ecosystem that enables users with no coding experience and limited machine learning knowledge to access a predictive modeling platform, supporting clinical decision-making through integrated web technologies and machine learning models.

4. METHODOLOGY

4.1 GENERAL RESEARCH METHODOLOGY

This thesis adopts a mixed-method approach, combining theoretical research, data-driven analysis, and platform development to create a user-friendly web platform clinical predictive modeling. The research is structured into three main phases:

4.1.1 : Literature Review

A comprehensive review of existing machine learning platforms was conducted to identify their potential in healthcare. Studies covering ML platforms, clinical decision support systems, and healthcare predictive modeling were analyzed. This review helped determine the essential features required for an accessible ML platform tailored to healthcare professionals' needs.

4.1.2: Dataset Selection and Preprocessing

The dataset used in this study consists of patient symptoms, diagnosis labels, and textual medical reports. Preprocessing techniques such as convert to lowercase, stop word, lemmatization was applied to prepare the data for machine learning models.

4.1.3: Machine Learning Model Selection and Evaluation

At this point, some statistical machine-learning models were selected for text analysis and disease classification based on the characteristics of our dataset and the project objectives. The models incorporated into the Flask platform, which are based on user selection, include K-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machine (SVM), and Decision Trees. These models are trained on pre-processed data, which is divided into training and testing sets, with the testing set representing 10% of the data. Additionally, users have the option to manually choose hyper-parameters and the text input vectorization to personalize the training procedure. We assess our machine learning models using various metrics. These metrics encompass accuracy, precision, recall, and analysis of the confusion matrix. Additionally, k-fold cross-validation was applied to evaluate the robustness of the models. Learning curve graphs facilitate visualization of the models' performance across various sizes of training sets.

4.2 Platform Development Methodology

The methodology adopted for the platform in this thesis is appropriate to develop a platform using Flask capable of analyzing our clinical dataset specifically focusing on text descriptions of disease signs and classification of diseases. we can see the main components of the methodology in the future steps, the first step according to our needs is platform analysis.

4.2.1 Platform Analysis

According to [13] System analysis is important for the design and implementation of software initiatives. In this section, we discuss functional and non-functional requirements.

- Functional requirements

The primary goal of our platform is to make clinical data analysis easier for healthcare professionals with less technical knowledge in the computer science field. The platform must manage user authentication, allowing users to log in and access various features based on their roles. The platform must use MySQL to authenticate users and administrators. After logging in, the administrator can choose between the supervised and unsupervised learning sections. In each component, the administrator should be able to upload a dataset and select their preferred preparation methods. They can also select their preferred machine learning models and customize the hyper-parameters of each method. The application will then show the results of the selected models.

- Non-Functional requirements

Non-functional requirements identify the essential features and properties that a system must have, regardless of its core capabilities. They cover topics such as performance, dependability, usability, and security. These requirements ensure the system's quality and efficacy [14]. For this platform, the requirements include a responsive and accessible system appropriate for healthcare professionals, capable of analyzing specific datasets. Additionally, a multi-layered design for a user-friendly interface, development using Flask Python for the back end, Bootstrap for the front-end, and ensuring optimal security through a MySQL-based authentication system are critical.

4.2.2 Quick design

A use case diagram shows structure of the system's required features. The use cases are gathered after an evaluation of the system's functional needs [15] The use case diagram in Fig. 2 displays the quick design of key functionalities of our system, with a focus on interactions between the user (admin) and the system components. The workflow begins when the administrator accesses the system through the login page. This step involves user authentication through a MySQL Workbench database. Upon successful authentication, the administrator is redirected to the main dashboard, where they can choose between supervised or unsupervised learning pathways.

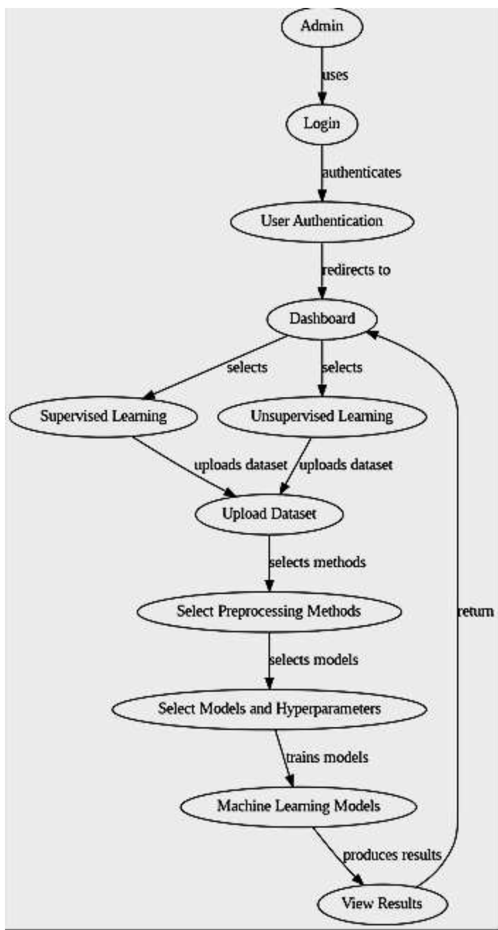


Figure 2: use case diagram for quick design

In both learning directions for supervised and unsupervised learning, the administrator can upload datasets, select pre-processing methods, and choose machine learning models. The system then trains the selected models, producing results that are then visualized for the user. This comprehensive workflow allows the administrator to efficiently control machine learning processes while also tracking the results via a clear online interface.

This structured workflow enables the administrator to manage the full cycle of machine learning from data preparation to result interpretation through a single web-based platform. The interface ensures ease of use, particularly for non-expert users in coding, while maintaining flexibility in model configuration and performance monitoring.

4.2.3 Platform Architecture

The platform architecture presents a three-tier architecture consisting of a comprehensive web application structured into a front-end, an API layer, and a database. The front-end consists of various HTML templates that communicate with the API REST layer to execute actions like GET, POST, PUT, and DELETE. The API, developed using Python, is responsible for managing routes for functionalities such as user login, data retrieval, and machine learning model operations. Dynamic

HTML rendering is achieved using Jinja templates, enhancing user interaction and experience.

The API layer processes data through dedicated modules for pre-processing, model training, and model testing. The MySQL database plays a role in storing user information, including usernames, passwords, and roles (user or admin). This database structure is responsible for handling user data while supporting the diverse functionalities required for clinical data analysis. The integration of the front-end, API, and database components allows seamless data flow and processing, providing a user-friendly environment for healthcare professionals with limited technical expertise. This architecture ensures that users can upload datasets, select pre-processing options, train machine learning models, and view results through an intuitive web interface.

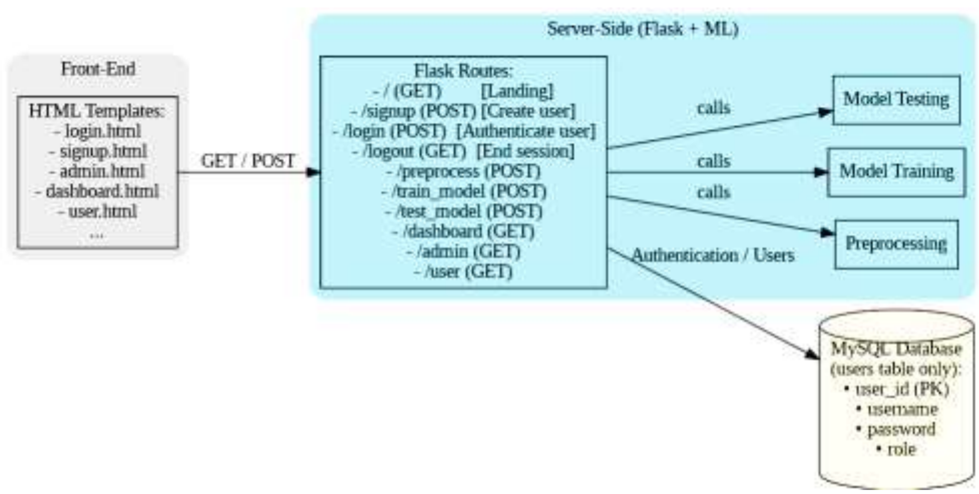


Figure 3: Platform Front-End Back-End Architecture

4.2.4 Development

Flask is a Python web framework that provides tools and pre-written code that simplifies website construction, minimizing the need to start from scratch. We picked Flask because it is a micro-framework with few tools and libraries, making it portable and resource-efficient [16]. In the development phase Fig. 3, the initial emphasis was on setting up the program structure, which specifically focused on creating a user interface framework and how to display the results to the user. The primary file structures of the developed application consist of four distinct program modules:

- 'model.pkl': stores the machine learning model for our prediction.
- 'app.py': Manages user authentication and session handling, coordinates data preprocessing, machine learning model training, and result visualization within the web application.
- 'templates/index.html': The HTML form allows users to upload datasets, select preprocessing models, choose machine learning algorithms, and view the predicted results.
- 'static/css/style.css': Contains the necessary styling for the HTML form to ensure a user-friendly interface.

applications by offering a compilation of libraries and modules for building back-end applications, while the front-end is constructed using HTML, CSS, and JavaScript. The framework employs techniques like "get" and "post" to manage data requests and responses, transform string inputs into integers, and transmit them as parameters to models for prediction. The Flask application is configured to handle HTTP requests, render templates, and manage routes efficiently.

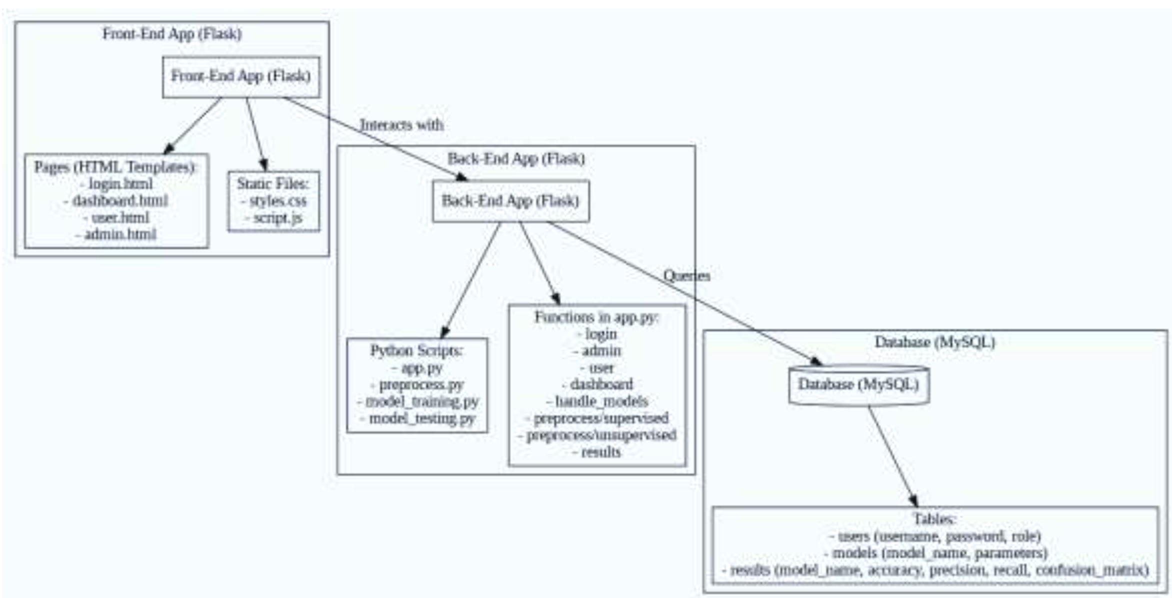


Figure 4: Platform Development diagram

4.2.5 Deployment

This web application is currently designed to operate on a local host, providing users with a platform for clinical data analysis and model visualization. While the current deployment is suitable for testing and small-scale use, future work may involve extending the application to a cloud or hosted environment.

5. ORGANIZATION

This thesis is based on two articles. It is broken into four main parts. The Introduction establishes the problem statement, study aims, and methodology. It also emphasises the importance of developing a user-friendly platform for analysing clinical data. Chapter 1 contains the first article, which acts as a literature review of machine learning platforms. This article critically evaluates machine learning platforms, focussing on their use in healthcare. It conducts a thorough review of existing solutions, assessing their strengths, weaknesses, and applicability to clinical needs. The second article is presented in Chapter 2, which explores “Interactive Machine Learning Pedagogy: Developing a Web-Based Platform for Clinical Predictive Modeling.” This article was published by Elsevier in 2024 [10] The thesis concludes with a general conclusion, which summarizes the development and contributions of a web-based platform for clinical predictive modeling. It highlights the platform's strengths, limitations, and future improvements necessary for broader applicability in healthcare settings.

CHAPTER 1

LITERATURE REVIEW

RÉSUMÉ EN FRANÇAIS DU PREMIER ARTICLE

Cette revue de la littérature étudie le développement et les applications des plateformes d'apprentissage automatique, en mettant l'accent sur leur utilisation dans le domaine de la santé. Les maladies chroniques telles que le diabète, les maladies cardiaques et les accidents vasculaires cérébraux se généralisent, l'utilisation de techniques d'apprentissage automatique présente un potentiel transformateur pour l'identification précoce et les soins personnalisés aux patients. La revue résume les recherches récentes sur divers algorithmes d'apprentissage automatique utilisés pour la prédiction des maladies, en mettant l'accent sur plateformes web qui améliorent l'accessibilité et l'engagement des utilisateurs. Cet article souligne la nécessité de développer des plateformes d'apprentissage automatique pour améliorer les résultats des soins de santé et rationaliser les processus de prise de décision clinique.

Mots-clés : apprentissage automatique ; soins de santé ; plateformes basées sur le Web, prise de décision clinique ; analyse prédictive ; qualité des données

Literature Review on Machine Learning Platforms: Applications and Innovations

Fatemeh Bashardoustjobjarkouli, Mehdi Adda[1]

[1]Université du Québec à Rimouski, Département de mathématiques, informatique et génie, Lévis, Canada ,
Fatemeh.Bashardoustjoubjarkouli@uqar.ca,Mehdi_Adda@uqar.ca

May 23, 2025

Abstract

This literature review investigates the development and applications of machine learning platforms, with a focus on their use in the healthcare. As chronic diseases like diabetes, heart disease, and strokes become more common, using machine learning techniques has transformative potential for early identification and personalised patient care. The review summaries recent research on various machine learning platforms used for disease prediction, that improve accessibility and user engagement. This paper emphasises the need of developing machine learning platforms for improving healthcare outcomes and streamlining clinical decision-making processes.

Keywords: Machine Learning; Health Care; Web-Based Platforms, Clinical Decision-Making; Predictive Analytics; Data Quality

1 Introduction

People's lifestyles and the environment are progressively shifting as society changes. As a result, there are more hidden risks associated with certain diseases. In general, major diseases such as diabetes, heart disease, and stroke have a significant influence. Significant progress has been made in the use of machine learning and deep learning approaches for disease prediction in the healthcare industry in recent years [1].Machine learning has demonstrated advances in the medical field, particularly in disease prediction, medical imaging, and clinical decision support systems [2]. There are many Researchers have previously investigated a variety of machine learning methods for predicting diseases. The table I that follows contains a number of studies that have been published in the medical and health domains throughout a range of years, together with working algorithms.

Machine learning, which was once limited to experts in research labs, has now become widely accessible. Today, both machine learning specialists and other users can easily begin utilizing it by taking advantage of popular front-end technologies [13]. Although extensive studies have been done in this domain, as shown by the examples included in this table, there are also some limitations.These studies did not propose any platform or online tools that users could use, so they did not suggest any workable systems that could be readily applied in day-to-day situations. In our suggested work, we have presented an intuitive web-based platform. The use of web technologies and machine learning for disease prediction is a developing topic that could completely transform healthcare.Therefore, we have decided to investigate several articles that propose platforms utilizing machine learning.

In the following sections, I will discuss the literature review focusing on machine learning platforms in general across various fields, with a particular emphasis on healthcare applications. I will also address some of the risks and limitations identified in these studies. The final chapter will provide a conclusion summarizing the key findings of this research.

Table I: Summary of Machine Learning Applications in Healthcare

Author(s)	Article Title	Applications in Healthcare	ML Techniques
Sarvamangala and Kulkarni (2021)	A comprehensive survey on medical image analysis using deep learning techniques [3]	Medical image understanding	CNN
Shahid et al. (2019)	Machine learning in health care: A review of the literature [4]	Cancer prediction, clinical diagnosis, speech recognition, length of stay prediction	ANN
Son et al. (2010)	Predicting medication adherence in patients with heart failure: A machine learning approach [5]	Predicting medication adherence	SVM
Soleimanian Gharehchopogh et al. (2012)	Application of decision tree algorithm for the prediction of the need for cesarean section [6]	Decision-making in obstetrics	Decision Tree
Enriko et al. (2018)	Heart disease diagnosis system with k-nearest neighbors method using real clinical medical records [7]	Diagnosing heart disease	K-Nearest Neighbor (KNN)
Bhargava et al. (2017)	An approach for classification using simple cart algorithm in WEKA [8]	Real-world male heart disease	CART
Dhomse and Mahale (2016)	Study of ML algorithms for special disease prediction using PCA [9]	Heart disease, diabetic patients	DT, SVM, Naive Bayes
Ramzan (2016)	Comparing and evaluating the performance of WEKA classifiers on critical diseases [10]	Disease classification	Random Forests, J48, Naive Bayes
Polaraju and Prasad (2017)	Prediction of heart disease using multiple linear regression model [11]	Heart disease	Multiple Linear Regression
Ambekar and Phalnikar (2018)	Disease risk prediction by using convolutional neural network [12]	Illness risk	CNN

2 Overview of Machine Learning Platforms

This section provides an overview of some web-based platforms that may be used machine learning approaches accessible through user interface modules and visualisations.

In their article [14] the authors explore the significant role of machine learning in the banking sector, particularly in predicting loan approvals. The study discusses the development of a loan prediction system utilizing various machine learning algorithms, including Logistic Regression, Decision Trees, K Nearest Neighbor, and Random Forest. The authors emphasize the importance of data analysis in identifying patterns within a dataset, which consists of 641 samples with 13 attributes relevant to loan applications. They detail the processes of data cleaning, model training, and testing, highlighting the effectiveness of these algorithms in achieving high accuracy rates in loan prediction tasks. The implementation of the model in Python, along with the use of Flask for creating a web application, facilitates user-friendly interaction with the system. The deployment of the loan prediction model on a web application using Heroku is a critical aspect of this research, as it enables real-time loan assessments for users. The authors present their results through confusion matrices and ROC curves, demonstrating that Logistic Regression achieved the highest accuracy (84.55%) among the tested algorithms. They conclude that the developed web application can significantly assist both bank employees and customers in the loan application process, providing a reliable tool for evaluating creditworthiness. The article suggests that further enhancements could be made by incorporating more diverse datasets and refining the model to improve prediction accuracy for different banking environments.

Assessment of slope stability is vital in geotechnical engineering, particularly for infrastructure projects such as earth dams and landfills. [15] address this challenge by developing a GUI-based platform that utilizes machine learning algorithms to predict slope stability under seismic conditions. Their study evaluates four algorithms: Logistic Regression (LR), Quadratic Discriminant Analysis (QDA), Light Gradient Boosting Machine (LGBM), and Linear Discriminant Analysis (LDA), using a dataset of 700 slope samples. Through tenfold cross-validation, the LGBM model is identified as the most accurate, demonstrating the effectiveness of machine learning in enhancing predictive capabilities. Additionally, the SHapley Additive exPlanations (SHAP) method is employed to interpret model predictions, revealing key factors such as peak ground acceleration and friction angle that influence slope stability. This innovative platform not only improves prediction accuracy but also makes advanced analytical tools accessible to geotechnical engineers, facilitating better decision-making in infrastructure design and maintenance.

In the article [16] they explain about Exploratory Data Analysis.(EDA) is an essential technique employed by machine learning engineers to derive insights from their datasets. EDA enables data scientists to uncover patterns, correlations, and anomalies that may not be immediately apparent in raw data. By utilizing various visualizations, EDA facilitates a deeper understanding of the data's structure, including its distribution, the presence of outliers, and any missing values. This foundational analysis is critical for making informed decisions when building machine learning models, as it helps to prevent inaccurate assumptions and enhances the overall accuracy of the models. The increasing popularity of EDA in machine learning stems from its ability to identify key relationships between variables, which are vital for developing precise models. EDA not only highlights the most significant features but also provides visual representations of trends and patterns, enabling engineers to focus on the most relevant aspects of the data. Additionally, EDA plays a crucial role in evaluating model performance by visualizing results, which allows for the identification of areas needing improvement.Table II presents a list of the machine learning tools and programming languages.These tools are also used for data visualization.

Table II: Some of the machine learning tools and supporting programming languages available.

Tool	Company	Open Source	Reference
Matlab	MathWorks	No	[17]
Microsoft Azure Auto ML	Microsoft, Inc.	No	[18]
Python	Python Software Foundation	Yes	[19]
R	Microsoft	Yes	[20]
Amazon AWS	Amazon, Inc.	No	[21]
IBM SPSS	IBM, Inc.	No	[22]
Weka tool	University of Waikato	Yes	[23]
DataRobot	DataRobot, Inc.	No	[24]
Google Cloud AutoML	Google LLC	No	[25]
Amazon SageMaker	Amazon, Inc.	No	[26]
KNIME	Privately Held	Yes	[27]
Alteryx	Alteryx, Inc.	No	[28]

In the field of education the article [29] presents how AI can transform educational environments into adaptive, personalized learning experiences. It emphasizes the use of machine learning and natural language processing to develop self-updating platforms tailored for continuous education and professional development. Key features include personalized learning pathways that analyze user data to provide tailored recommendations, enhancing engagement and retention. The article critiques traditional e-learning models, which rely on static content and manual updates, advocating for dynamic systems that continuously refresh their content based on the latest information and user feedback. The authors also call for national-level initiatives to support the integration of AI into educational frameworks, emphasizing the importance of technology in improving learning outcomes. By addressing the limitations of existing systems, this research highlights the transformative potential of AI in creating effective and responsive educational platforms.

Also in the same filed we have another paper which [30] Junping Zhou addresses the inefficiencies of traditional English listening classes in colleges, highlighting their inability to cater to individual student needs and learning preferences. The study reveals that despite the increasing integration of AI technologies in education, many existing AI-aided listening platforms fail to engage students effectively, resulting in poor listening comprehension and low performance in standardized assessments like the CET-4. A survey conducted among 100 college students indicates that while these platforms offer extensive resources, they often lack personalized exercise guidance and effective evaluation systems, making it challenging for students to track their progress and receive tailored feedback. To overcome these challenges, Zhou proposes a novel AI-based self-learning platform structured around a three-layer design model—service, technical, and data layers. This architecture facilitates the integration of various resources and support mechanisms tailored to individual learning needs. The platform employs advanced technologies such as machine learning and deep learning to provide intelligent feedback and adapt learning paths based on continuous performance analysis. By incorporating features like real-time assistance and personalized recommendations, the platform aims to create an engaging self-learning environment that enhances students' English listening proficiency. Ultimately, the article underscores the potential of AI to transform language education by addressing the limitations of traditional teaching methods.

3 Machine Learning Platforms in Healthcare

Specialized ML platforms, particularly those designed for the healthcare sector, address unique challenges inherent to medical applications. In this section, we will explore a selection of recent articles focused on the development of platforms tailored specifically for the healthcare and medical sectors. This article [31] presents a web application for diabetes prediction using machine learning algorithms. The authors compared two machine learning techniques: K-Nearest Neighbor (KNN) and Random Forest (RF). After preprocessing the data and selecting influential features, they found that the Random Forest model with Bagging Meta-Estimator offers an accuracy of 83%, higher than KNN (75%). The web application allows users to input their medical information (number of pregnancies, glucose level, blood pressure, insulin level, body mass index, diabetes pedigree function, age) to predict their risk of developing diabetes-related diseases.The model used is the Random Forest with Bagging Meta-Estimator, providing a user-friendly interface to help patients understand and manage their diabetes risk.the proposed architecture in this study is like the figure 1.

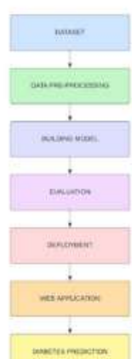


Figure 1: Proposed Architecture

In conclusion, this approach proposes a new method for predicting diabetes using machine learning algorithms and an interactive web application to improve patient health management. The following table is a MODEL COMPARISON table and it shows the precision score, recall score, and f1-Score score for each model that used in this study 2.

Model	Precision Score	Recall Score	f1-Score
K-Nearest Neighbor	0.7636	0.7826	0.7730
Random Forest Classifier	0.7250	0.6505	0.6888
Random Forest Classifier with Bagging Meta-estimator	0.7692	0.6382	0.6976

Figure 2: MODELS COMPARISON

The paper [1] presents an integrated approach to predict multiple diseases using machine learning (ML) and deep learning techniques, alongside web technology for user accessibility. The study focuses on developing a user-friendly web-based platform that allows patients to predict diseases such as diabetes, heart disease, lung cancer, Parkinson's disease, and brain stroke by entering their symptoms. The authors employ various machine learning algorithms, including Decision Trees, Random Forest, Support Vector Machine (SVM), and k-nearest neighbors (KNN), as well as deep learning models like Long Short-Term Memory (LSTM). The platform integrates these models to achieve high accuracy rates, providing patients with early predictions that could potentially improve outcomes by encouraging timely medical consultations. The Random Forest model, for instance, achieved up to 99% accuracy for predicting brain stroke, demonstrating the effectiveness of ensemble methods in disease prediction. By leveraging a web-based architecture developed with Django, the platform offers a practical, scalable, and easily accessible interface for disease prediction. This approach highlights the importance of integrating machine learning capabilities into web platforms to enable early, data-driven health assessments, thereby democratizing access to predictive healthcare tools. The implementation of multiple disease detection in a single platform also eliminates the need for users to access separate applications, providing a seamless experience that supports comprehensive health monitoring.

The article [32] introduces FetalAI, a web-based application designed for predicting newborn birthweights using deep learning techniques. The primary objective of this tool is to forecast the risks associated with excessive fetal growth conditions, such as large-for-gestational-age (LGA) and macrosomia. The development of FetalAI encompasses a comprehensive process that includes data sourcing, preprocessing, back-end model design utilizing mixed-effects models and recurrent neural networks (RNNs), and user interaction through a graphical user interface (GUI) built on Streamlit. FetalAI relies on maternal baseline characteristics—such as height, weight, and health history—as well as prenatal ultrasound measurements. This combination of data is crucial for generating accurate predictions. The application is capable of providing real-time predictions, enabling clinicians to receive results within seconds of input. Additionally, the interface allows for intuitive data input and result visualization, making it user-friendly for both medical professionals and researchers. The article emphasizes the potential of FetalAI to enhance prenatal care by facilitating timely medical decision-making based on accurate risk assessments of fetal conditions .

The shinyDeepDR [33] platform represents 3 a significant advancement in precision oncology by providing a user-friendly web application for predicting the response of cancer cells to a diverse panel of 265 anti-cancer drugs. Developed using the DeepDR deep learning model, this platform enables researchers to input mutation and gene expression data from cancer samples—be it cell lines or tumors—without the need for extensive programming skills or high-performance computing resources. By bridging the gap between complex computational models and end-users, shinyDeepDR aims to make sophisticated predictive tools more accessible in the field of cancer research. Key functionalities of shinyDeepDR include its "Find Drug" module, which predicts the sensitivity of samples to various anti-cancer compounds, and the "Find Sample" module, which identifies cell lines or tumors with genomic profiles similar to the input sample. The platform provides an interactive interface for result interpretation, facilitating a deeper understanding of potential drug efficacy and informing therapeutic decisions. Notably, the tool showcases its utility by identifying effective treatments for previously deemed "undruggable" mutations in liver cancer. The authors emphasize that while shinyDeepDR enhances accessibility to drug response predictions, the accuracy of its predictions heavily relies on the quality of the underlying genomic data.

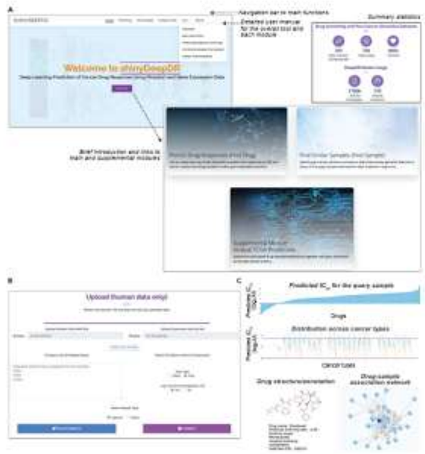


Figure 3: The screenshots of the pages of the application proposed in [33]

In summary, shinyDeepDR not only aids in the identification of effective anti-cancer treatments but also represents a significant step towards democratizing access to advanced computational tools in the oncology research community.

Recent trends underscore the growing importance of online medical services, particularly as the prevalence of infectious diseases accelerates [34]. Leveraging data mining has been shown to enhance the efficiency of disease diagnosis by cutting down on travel expenses and wait times, thus making healthcare more accessible. Over the last decade, machine learning (ML) approaches have proliferated in the medical domain, with many researchers investigating variants of these algorithms for precise and rapid disease detection.

One strategy employs chest X-ray data in combination with multiple ML algorithms to support near-instant diagnosis. This approach recognizes the lack of a standard system capable of diagnosing more than one illness through a single application. By integrating frameworks such as TensorFlow and deploying models via a Flask-based web interface, patients can upload an X-ray image and receive real-time results. Large datasets are used to train each model, ensuring robust and accurate predictions. Ultimately, the goal of such solutions is to identify widespread diseases early on, provide timely alerts to patients, and thereby help reduce mortality rates through prompt medical intervention.

Recent advancements in machine learning for healthcare have predominantly focused on analyzing single diseases, such as diabetes or cancer. This article [35] proposes a novel system that utilizes a Flask API to predict multiple diseases, including diabetes, diabetic retinopathy, heart disease, and breast cancer, with plans to expand to other conditions like skin disorders and fever analysis. The system employs machine learning algorithms along with TensorFlow, leveraging Python's pickling and unpickling methods to save and load model states efficiently. A significant advantage of this approach is its comprehensive parameter analysis; unlike many existing models that rely on a limited set of factors, this system incorporates additional variables such as serum creatinine, potassium levels, Glasgow Coma Scale scores, and lipid profiles. This broader scope enhances prediction accuracy and provides a more thorough patient assessment. When users access the Flask API, they submit relevant parameters along with the disease name, prompting the system to invoke the appropriate predictive model and return the patient's status. By facilitating early detection and monitoring of multiple diseases, this system aims to improve patient outcomes and reduce mortality rates.

The writers of the paper [36] G. Shobana and S. Nikkath Bushra, discuss how the incidence of cardiovascular diseases is rising and how crucial it is to identify them early on utilising cutting-edge analytical methods. The study highlights how machine learning can be used to pinpoint important variables that lead to heart ailments, including heart failure. The authors implement seven machine learning algorithms, including classic models like Logistic Regression and Naïve Bayes, as well as ensemble techniques like Random Forest and Boosting algorithms, using two well-known machine learning platforms: Scikit-Learn and Orange. Their investigation shows that, when tested on a Heart Failure dataset taken from the UCI repository, the Boosting algorithms achieved the maximum prediction accuracy of 89%.

One prominent example is the MODELHealth platform, which is designed to integrate disparate health data from various sources, including electronic health records (EHRs), imaging data, and patient-generated health information. This integration is critical for enabling the development of accurate predictive models that can inform clinical decision-making. By consolidating data from different health units, MODELHealth aims to create a comprehensive repository of patient information that can be analyzed to uncover insights into health trends, treatment outcomes, and disease progression [37].A key feature of healthcare-specific ML platforms like MODELHealth is their emphasis on data anonymization and security. Given the sensitive nature of health data, these platforms implement robust anonymization techniques to protect patient privacy. This is crucial not only for ethical reasons but also for ensuring compliance with healthcare regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States. By addressing data privacy concerns, healthcare ML platforms facilitate the secure sharing of medical data for research and clinical purposes, thereby promoting collaboration between healthcare providers and researchers [37]. Furthermore, healthcare-specific ML platforms often incorporate algorithms tailored for medical applications. These algorithms are optimized for the unique characteristics of health data, such as dealing with missing values, understanding the implications of class imbalances in disease prevalence, and adapting to the multifactorial nature of health conditions. For example, platforms may employ ensemble learning techniques or advanced neural networks that can analyze complex datasets to identify patterns indicative of diseases, predict patient outcomes, and recommend personalized treatment plans.

Chongtham et al [38] developed a web-based application for early-stage diabetes risk prediction using both supervised and unsupervised machine learning algorithms. Their study utilized the "Early Stage Diabetes Risk Prediction Dataset" from the UCI Machine Learning Repository, which includes 17 features related to diabetes risk factors. The authors implemented several supervised machine learning algorithms, including Naïve Bayes, Random Forest, Support Vector Machine, Decision Tree, and K-nearest Neighbor, achieving accuracy rates between 86% and 99%.

In addition to supervised learning, the study employed Kohonen's Self-Organizing Maps (KSOM) for unsupervised learning, clustering the data into groups to provide probabilities for positive and negative diabetes risk. This approach demonstrated the potential of machine learning to predict diabetes risk based on patient questionnaires, reducing the need for traditional testing kits.

The article [39] explores the innovative use of smartphones as wearable and wireless gyroscope platforms for the quantification and classification of hemiplegic gait using machine learning. By leveraging the gyroscope sensors integrated into smartphones, the research aims to differentiate between the affected and unaffected legs of individuals with hemiplegia. The authors argue that this approach not only harnesses the widespread accessibility of smartphones but also enables remote data collection and processing through wireless internet connectivity. The study involves mounting a smartphone on both the affected and unaffected legs of a participant while walking on a treadmill at a constant speed. The collected gyroscopic data is then transmitted via email for post-processing, where it is converted into a feature set suitable for machine learning classification. Utilizing a multilayer perceptron neural network, the research achieves significant classification accuracy, demonstrating the potential of this method for enhancing rehabilitation practices. The findings suggest that such a system can facilitate remote monitoring and therapy, ultimately improving the therapeutic strategies for patients with mobility impairments.

The integration of artificial intelligence (AI) in healthcare is transforming traditional approaches to disease prediction and management. In the study by [40], a web-based application is developed that utilizes machine learning techniques to predict the risk of diabetes at an early stage. This innovative platform leverages a dataset that includes patient responses gathered through questionnaires, enabling the system to classify patients without the need for traditional testing methods. The authors employ a combination of supervised and unsupervised machine learning algorithms, including Decision Trees, Random Forests, and Support Vector Machines (SVM), to analyze the dataset. The model's performance is evaluated based on accuracy, with results indicating that the Random Forest algorithm achieves the highest accuracy rate, reaching up to 99%. Additionally, the study utilizes Kohonen's Self-Organizing Maps (KSOM) for clustering patients into distinct risk categories, enhancing the system's ability to provide tailored predictions based on individual characteristics. This approach not only demonstrates the efficacy of machine learning in healthcare but also highlights the potential for developing scalable, user-friendly applications that can assist in early disease detection. By focusing on personalized risk assessments, the system contributes to smarter healthcare solutions that empower patients and healthcare providers alike.

.

Table III: Summary of ML Models and Platform Technologies in Healthcare Applications

Ref. No.	Machine Learning Models	Platform Type / Technology
31	KNN, Random Forest (with Bagging Meta-Estimator)	Web application
1	Decision Tree, Random Forest, SVM, KNN, LSTM	Web application (Django)
32	Mixed-effects models, RNN	Web application (Streamlit)
33	DeepDR (Deep Learning)	Web application (R Shiny)
34	Various ML and DL models	Web application (Flask, TensorFlow)
35	Decision Tree, Random Forest, SVM, etc.	Flask API (Web service)
36	Logistic Regression, Naïve Bayes, Random Forest, Boosting	Desktop platforms (Scikit-Learn, Orange)
37	Ensemble techniques, Neural Networks	Software platform (MODELHealth, multi-source integration)
38	Naïve Bayes, Random Forest, SVM, Decision Tree, KNN, KSOM	Web-based (Early Stage Diabetes Risk Platform)
39	MLP Neural Network	Mobile smartphone-based (Wearable gyroscope)
40	Decision Tree, Random Forest, SVM, KSOM	Web-based application (Questionnaire-driven)

4 Limitation and Challenges

This section explores limitations and challenges associated with implementing machine learning in healthcare. Issues such as the lack of standardized data formats, concerns over data quality, and reluctance to share sensitive information significantly hinder effective data analysis and decision-making. Additionally, the high costs and time requirements for establishing centralized data repositories further complicate efforts to leverage healthcare data for improved patient outcomes. Understanding these challenges is essential for developing more effective machine learning solutions in the healthcare sector.

For instance the article [41] mentioned that healthcare data is generated and stored by various organizations, but it faces challenges in making proper decisions. The lack of a standard format for data storage in healthcare organizations can make epidemic situations worse. If an epidemic spreads across a country, the health ministry requires all organizations to share their data with a centralized data warehouse for analysis. However, this may take longer than usual, potentially leading to out-of-control situations. Quality of healthcare data is crucial for extracting meaningful information for improving patient services. However, maintaining quality depends on factors like removing noise and missing data. Data sharing is another challenge, as neither patients nor healthcare organizations are interested in sharing their private data. This could lead to worsening epidemic situations, difficulty in providing better treatments, and difficulty in detecting fraud and abuse in healthcare insurance companies. Building a data warehouse for sharing data among healthcare organizations is a costly and time-consuming process. Additionally, we can see in the article [42] in health data analysis, missing data can be a serious challenge, thus it's critical to use approaches that deal with. The necessity of robust techniques, like MImp, in healthcare research is highlighted by the identification of missing data as a possible source of uncertainty. The use of MImp becomes crucial when thorough data gathering is not done during the study's design phase [43].

Healthcare solutions based on machine learning offer novel and forward-thinking prospects, but they also bring with them special risks, difficulties, and an appropriate measure of fear. The chance of prediction mistake and its impact, the privacy and security vulnerabilities of the systems, and even the lack of data availability to provide repeatable results are some of the primary risk issues that are covered here. The loss of the personal factor in healthcare, ethical issues, and the interpretability and are a few of the difficulties [44].

The following table IVpresents the limitations identified in the articles we reviewed in our literature review.

References	**Limitations**
[32]	Limited interpretability and scope; primarily focused on birthweight forecasting.
[33]	Dependent on data quality; focused on gene expression/mutation data; not for clinical use.
Youn	Small sample size; employed a cross-sectional design; reliance on patient-reported information.
[45]	Dependence on quality of data; selection of models depends on various parameters.

Table IV: Summary of References and Limitations

5 Conclusion

This literature review presents advancements and applications of machine learning platforms in healthcare. The paper has demonstrated how various machine learning algorithms have been effectively utilized for disease prediction, medical imaging, and clinical decision support systems. The review also presented several web-based platforms that integrate these algorithms, providing accessible tools for healthcare professionals and patients.

Despite these advancements, the review identified several challenges that need to be addressed to fully realize the potential of machine learning in healthcare. These include issues related to data quality, standardization, and the ethical implications of using sensitive health data. Future research should focus on developing robust frameworks to overcome these challenges, ensuring that machine learning applications are both effective and secure.

In summary, the integration of machine learning platforms in healthcare has the potential to improve patient outcomes and streamline clinical decision-making processes. By addressing the identified limitations, these technologies can be further refined to provide even greater benefits to the healthcare industry.

References

[1] M. Rahman, S. Islam, S. B. Sarowar, and M. T. Zaman, "Multiple disease prediction using machine learning and deep learning with the implementation of web technology," in *2023 IEEE International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)*, 2023, pp. 1–7. DOI: 10.1109/AIBThings58340.2023.10292488.

[2] M. Shehab, L. Abualigah, Q. Shambour, *et al.*, "Machine learning in medical applications: A review of state-of-the-art methods," *Computers in Biology and Medicine*, vol. 145, p. 105 458, 2022, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2022.105458. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482522002505.

[3] D. R. Sarvamangala and R. V. Kulkarni, "A comprehensive survey on medical image analysis using deep learning techniques," *Journal of Medical Systems*, vol. 45, no. 5, pp. 1–14, 2021. DOI: 10.1007/s10916-021-01782-7.

[4] N. Shahid, T. Rappon, and W. Berta, "Machine learning in health care: A review of the literature," *Health Informatics Journal*, vol. 25, no. 1, pp. 1–10, 2019. DOI: 10.1177/1460458218755485.

[5] Y. J. Son, M. Choi, and J. Kim, "Predicting medication adherence in patients with heart failure: A machine learning approach," *Computers in Biology and Medicine*, vol. 40, no. 9, pp. 856–860, 2010. DOI: 10.1016/j.compbiomed.2010.05.007.

[6] F. Soleimanian Gharehchopogh, P. Mohammadi, and P. Hakimi, "Application of decision tree algorithm for the prediction of the need for cesarean section: A study of obstetric operations," *Health Information Science and Systems*, vol. 1, no. 1, pp. 1–8, 2012. DOI: 10.1186/2047-2501-1-3.

[7] K. Enriko, M. Suryanegara, and D. Gunawan, "Heart disease diagnosis system with k-nearest neighbors method using real clinical medical records," in *Proceedings of the 4th International Conference on Frontiers of Educational Technologies*, New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 127–131. DOI: 10.1145/3233347.3233386.

[8] N. Bhargava, S. Dayma, A. Kumar, and P. Singh, "An approach for classification using simple cart algorithm in weka," in *2017 11th International Conference on Intelligent Systems and Control (ISCO)*, IEEE, 2017. DOI: 10.1109/ISCO.2017.7955667.

[9] K. B. Dhomse and K. M. Mahale, "Study of machine learning algorithms for special disease prediction using principal component analysis," in *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication*, IEEE, 2016. DOI: 10.1109/ICGTSP.2016.7915638.

[10] M. Ramzan, "Comparing and evaluating the performance of weka classifiers on critical diseases," in *2016 1st India International Conference on Information Processing (IICIP)*, IEEE, 2016. DOI: 10.1109/IICIP.2016.7950792.

[11] K. Polaraju and D. Prasad, *Prediction of heart disease using multiple linear regression model*, Unpublished work, 2017.

[12] S. Ambekar and R. Phalnikar, "Disease risk prediction by using convolutional neural network," in *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation (ICCUBEA)*, Jul. 2018. DOI: 10.1109/ICCUBEA.2018.8697423.

[13] H.-A. Goh, C.-K. Ho, and F. S. Abas, "Front-end deep learning web apps development and deployment: A review," *Applied Intelligence*, vol. 53, no. 12, pp. 15 923–15 945, 2023, ISSN: 1573-7497. DOI: 10.1007/s10489-022-04278-6. [Online]. Available: https://doi.org/10.1007/s10489-022-04278-6.

[14] C. N. Sujatha, A. Gudipalli, B. Pushyami, N. Karthik, and B. N. Sanjana, "Loan prediction using machine learning and its deployment on web application," in *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Hyderabad, India: IEEE, 2021. DOI: 10.1109/i-PACT51564.2021.978-1-6654-2691-6.

[15] M. S. Barkhordari, M. M. Barkhordari, D. J. Armaghani, E. T. Mohamad, and B. Gordan, "Gui-based platform for slope stability prediction under seismic conditions using machine learning algorithms," *Architecture, Structures and Construction*, vol. 4, no. 2, pp. 145–156, Dec. 2024, ISSN: 2730-9894. DOI: 10.1007/s44150-024-00112-4. [Online]. Available: https://doi.org/10.1007/s44150-024-00112-4.

[16] S. Tufail, H. Riggs, M. Tariq, and A. Sarwat, "Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms," *Electronics*, vol. 12, no. 8, p. 1789, 2023. DOI: 10.3390/electronics12081789.

[17] Mathworks.com, *Machine learning with matlab*, Available online: https://www.mathworks.com (accessed on 22 November 2022), 2022.

[18] Microsoft, *Azure machine learning documentation*, Available online: https://docs.microsoft.com/en-us/azure/machinelearning/ (accessed on 6 February 2023), 2022.

[19] Python, *Python software foundation*, Available online: https://www.python.org/psf/ (accessed on 6 February 2023), 2022.

[20] r-project.org, *R: What is r?* Available online: https://www.r-project.org/about.html (accessed on 6 February 2023), 2022.

[21] Amazon, *Cloud computing services—amazon web services (aws)*, Available online: https://aws.amazon.com/ (accessed on 6 February 2023), 2022.

[22] IBM, *Spss software — ibm*, Available online: https://www.ibm.com/analytics/spss-statistics-software (accessed on 6 February 2023), 2022.

[23] cs.Waikato.ac.nz, *Weka 3—data mining with open source machine learning software in java*, Available online: https://www.cs.waikato.ac.nz/ml/weka/ (accessed on 6 February 2023), 2022.

[24] DataRobot.com, *Datarobot ai cloud—the next generation of ai*, Available online: https://www.datarobot.com/ (accessed on 5 February 2023), 2022.

[25] Google, *Cloud automl custom machine learning models*, Available online: https://cloud.google.com/automl (accessed on 4 February 2023), 2022.

[26] Amazon, *Machine learning—amazon web services*, Available online: https://aws.amazon.com/sagemaker/ (accessed on 5 February 2023), 2022.

[27] KNIME.com, *Open for innovation*, Available online: https://www.knime.com (accessed on 4 February 2023), 2022.

[28] Alteryx.com, *Self-service analytics, data science & process automation — alteryx*, Available online: https://www.alteryx.com (accessed on 4 February 2023), 2022.

[29] S. S. Gulyamov, S. Mamanazarov, and A. A. Rodionov, "Creating self-updating digital platforms using artificial intelligence technologies for continuous education and professional development," *Journal of Educational Technology*, vol. 15, no. 1, pp. 1–14, 2024. DOI: 10.1007/s00779-024-01635-2.

[30] J. Zhou, "Design of ai-based self-learning platform for college english listening," in *2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, IEEE, Xi'an, China, 2020. DOI: 10.1109/MLBDBI51377.2020.00114.

[31] P. T. S. Gununathan, R. S, R. S, and N. S, "Web application-based diabetes prediction using machine learning," in *Proceedings of the 7th International Conference on Computing Methodologies and Communication (ICCMC-2023)*, IEEE Xplore Part Number: CFP23K25-ART, ISBN: 978-1-6654-6408-6, IEEE, 2023.

[32] Y. Yao, R. Liu, and B. Zhang, "Fetalai: A deep learning web-based application for predicting birthweight from prenatal ultrasound measurements," *Informatics in Medicine Unlocked*, vol. 49, p. 101 555, 2024, ISSN: 2352-9148. DOI: https://doi.org/10.1016/j.imu.2024.101555. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914824001114.

[33] L.-J. Wang, M. Ning, T. Nayak, *et al.*, "Shinydeepdr: A user-friendly r shiny app for predicting anti-cancer drug response using deep learning," *Patterns*, vol. 5, no. 2, p. 100 894, 2024, ISSN: 2666-3899. DOI: https://doi.org/10.1016/j.patter.2023.100894. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666389923002969.

[34] N. D. Navghare, P. Rale, S. Kolhe, V. Bhadale, and R. Wasnik, "Web application for multiple disease prediction using machine learning and deep learning," in *AIP Conference Proceedings*, vol. 2981, American Institute of Physics (AIP), 2023, p. 020 010. DOI: 10.1063/5.0182673.

[35] A. Yaganteeswarudu, "Multi disease prediction model by using machine learning and flask api," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, pp. 1242–1246. DOI: 10.1109/ICCES48766.2020.9137896.

[36] G. Shobana and S. N. Bushra, "Prediction of cardiovascular disease using multiple machine learning platforms," in *Proceedings of the International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Chennai, India: IEEE, 2021. DOI: 10.1109/ICSES51954.2021.9233056.

[37] A. Anastasiou, E. Kostalas, S. Pitoglou, G. Matsopoulos, T. Androutsou, and D. Koutsouris, "Modelhealth: An innovative software platform for machine learning in healthcare leveraging indoor localization services," in *20th IEEE International Conference on Mobile Data Management (MDM)*, IEEE, Athens, Greece, 2019. DOI: 10.1109/MDM.2019.000-5.

[38] P. Chongtham, K. V. Singh, and K. R. Singh, "Artificial intelligence enabled web-based prediction of diabetes using machine learning approach," in *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, IEEE, 2021, pp. 60–64. DOI: 10.1109/CENTCON52345.2021.9688236.

[39] R. LeMoyne and T. Mastroianni, "Implementation of a smartphone as a wearable and wireless gyroscope platform for machine learning classification of hemiplegic gait through a multilayer perceptron neural network," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Flagstaff, AZ, USA: IEEE, 2018, pp. 946–951. DOI: 10.1109/ICMLA.2018.00153.

[40] N. Chongtham, A. Kumar, S. Khokhar, and P. Kumar, "Artificial intelligence enabled web-based prediction of diabetes using machine learning approach," *Journal of Healthcare Engineering*, vol. 2023, pp. 1–10, 2023. DOI: 10.1155/2023/1234567.

[41] P. Ahmad, S. Qamar, and S. Q. A. Rizvi, "Techniques of data mining in healthcare: A review," *International Journal of Computer Applications*, vol. 120, no. 15, pp. 38–48, 2015, ISSN: 0975-8887. [Online]. Available: http://www.ijcaonline.org/research/volume120/number15/21122-5154.

[42] M. Marino, J. Lucas, E. Latour, and J. D. Heintzman, "Missing data in primary care research: Importance, implications and approaches," *Family practice*, vol. 38, no. 2, pp. 199–202, 2021.

[43] J. Di, C. Demanuele, A. Kettermann, *et al.*, "Considerations to address missing data when deriving clinical trial endpoints from digital health technologies," *Contemporary Clinical Trials*, vol. 113, p. 106 661, 2022.

[44] H. Habehh and S. Gohel, "Send orders for reprints to reprints@benthamscience.net machine learning in healthcare," *Current Genomics*, vol. 22, p. 291, 2021. DOI: 10.2174/1389202922666210705124359.

[45] S. Sobale, H. More, S. Mulay, and M. Kshirsagar, "Supervised machine learning model builder for regression and classification analysis featured as a web application," in *2023 International Conference on Advancement in Computation Computer Technologies (InCACCT)*, 2023, pp. 16–21. DOI: 10.1109/InCACCT57535.2023.10141809.

CHAPTER 2 INTERACTIVE MACHINE LEARNING PEDAGOGY: DEVELOPING A WEB-BASED EDUCATIONAL PLATFORM FOR CLINICAL PREDICTIVE MODELING

RÉSUMÉ EN FRANÇAIS DU DEUXIÈME ARTICLE

Cet article [10] présente une plateforme interactive basée sur le web pour la modélisation prédictive clinique, conçue pour les professionnels de la santé. Développée avec Flask, la plateforme intègre des algorithmes d'apprentissage automatique, y compris les K-Nearest Neighbours (KNN), les Machines à Vecteurs de Support (SVM), le Naive Bayes (NB) et les Arbres de Décision (DT). Elle offre des outils conviviaux pour la préparation des données, l'entraînement des modèles et la visualisation des résultats, simplifiant des processus complexes grâce à l'automatisation. L'étude vise à améliorer l'accès aux ressources d'apprentissage automatique (ML) dans le domaine de la santé, facilitant ainsi l'analyse de jeux de données cliniques spécifiques. Cette plateforme démontre le potentiel d'avancer la prise de décision basée sur les données dans des environnements cliniques.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2018) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

The 14th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2024)
October 28-30, 2024, Leuven, Belgium

Interactive Machine Learning Pedagogy: Developing a Web-Based Educational Platform for Clinical Predictive Modeling

Fatemeh Bashardoustjoubjarkouli*, Mehdi Adda

Department of Mathematics, Computer Science and Engineering University of Quebec at Rimouski, 1595 Bd Alphonse-Desjardins, Levis, Quebec G6V 0A6, , Canada

Abstract

This paper presents an interactive web-based platform for clinical predictive modeling, designed for healthcare professionals. Developed using Flask, the platform integrates machine learning algorithms including K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Naive Bayes (NB), and Decision Trees (DT). It features user-friendly tools for data preparation, model training, and result visualization, simplifying complex processes through automation. The study aims to improve accessibility to Machine Learning (ML) resources in healthcare, facilitating the analysis of specific clinical data sets. This platform demonstrates the potential to advance data-driven decision-making in clinical settings.

Keywords: Machine Learning, Classification, Clinical Data Analysis, Web Application, Flask, K-Nearest Neighbors, Support Vector Machine, Naive Bayes, Decision Trees, Clustering.

1. Introduction

In recent years, there has been a notable increase in the complexity of documents and texts, necessitating a deeper understanding of machine learning techniques for accurate text categorization across various contexts [1]. The field of Machine Learning (ML) has matured significantly, with applications now spanning various business domains including web search, online advertising, product recommendations, and object recognition [2].

Machine learning, which originated from computational learning theory and pattern recognition, has become an effective approach to data analytics to predict results and uncover hidden patterns. It enables researchers, engineers,

* Corresponding author. Tel.: +1-418-570-0949.
E-mail address: Fatemeh.bashardoustjoubjarkouli@uqar.ca

data scientists, and analysts to generate valid results by creating models and algorithms that learn from historical data and trends [3].

Despite the growing need for machine learning techniques in many sectors, specialists in these fields may lack the necessary computer science skills to effectively utilize them. To address this challenge, we propose the development of a web-based platform that makes machine learning more accessible to non-technical users, particularly in the healthcare domain. Our study focuses on designing and developing a platform that enhances the training and evaluation of machine learning models for clinical data analysis. We utilize Flask [4], a Python micro-framework, for our platform development. It integrates several Machine Learning algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes (NB), and Decision Trees (DT). In addition, we incorporate clustering techniques like DBSCAN and K-Means to uncover hidden patterns within the data.

The primary goal of this paper is to create a Graphical User Interface (GUI) platform for accessible data analysis, targeted at non-technical healthcare professionals. By incorporating machine learning models within the platform and providing easy access to model findings, our aim is to simplify the understanding and application of these techniques.

The remainder of this paper is organized as follows. In Section 2, we review relevant previous work. Section 3 describes the proposed method and system design. Experimental procedure and discussion of result analysis are presented in Section 4. Finally Section 5 provides concluding remarks.

2. Literature Review

This section presents an overview of various recent studies on the deployment of applications or tools that utilize predictive models based on Machine Learning algorithms. The machine learning model builder is a web-based tool that generates and processes various machine learning models, primarily using regression or classification, as proposed by Sobale and Shital [5]. It allows for dynamic file acceptance and visualizations and can handle complex datasets. The tool includes extensions for processing rich datasets and explores regression, classification, and ensemble learning approaches. It uses the Python-based framework Django and machine learning foundations for architecture creation. The user-centered approach helps consumers understand machine learning's capabilities as a web application. Akkem [6] proposed a framework that employs Flask API and TensorFlow for the prediction of diseases such as diabetes, heart disease, and breast cancer. This framework integrates Python pickling for model retention and addresses a range of factors responsible for the onset of diseases. Swati [7] investigated the utilization of interactive machine learning (iML) tools to democratize machine learning for individuals lacking expertise in the field. Their creation of a system employing Flask API for comprehensive disease analysis underscores the obstacles in obtaining training data and the promise of transfer learning in interactive user settings. Ahmed [8] et al. organized a web-based platform deploying an Artificial Neural Network (ANN) to anticipate diabetes with a remarkable accuracy rate of 82.3%. They also utilized K-Nearest Neighbors (KNN) and Naive Bayes for supplementary classification assignments, underscoring the significance of efficient detection techniques for diabetes. Sujatha [9] et al. directed their attention toward loan prediction within the banking sector utilizing logistic regression. Their system, crafted in Python and launched through Visual Studio Code, attained a high level of precision, accentuating the significance of pattern scrutiny in loan prediction and the capacity for financial institutions to embrace web applications for this objective. Dey [10] et al. established a web-based tool for predicting diabetes employing various machine learning methodologies, encompassing SVM, KNN, NB, LR, and ANN. Their tool, leveraging the Puma Indian dataset, achieved an accuracy level of 82.35%, exhibiting the efficacy of machine learning in enhancing predictive precision. In the article [11], we can find that one of the trickiest issues in computer science's machine learning sector is now disease prediction. The majority of machine learning algorithms now in use are obsessed with the prediction of a specific disease. As an illustration, consider machine learning models for forecasting fever, brain cancer, malaria, and other diseases. This study suggests a method for predicting several different diseases. As of right now, the suggested method can forecast pancreatic cancer, Alzheimer's disease, malaria, and tuberculosis. The suggested method analyzed and predicted several diseases using CNN, Random Forest, and Logistic Regression. The system is developed with Python ML Libraries like TensorFlow, Scikit Learn, Pandas, and Flask API, a framework for building web applications in Python. The study [12] looks into the use of logistic regression models to predict diabetes mellitus. The researchers used data from the Pima Indians Diabetes Database to compare the performance of logistic regression to other classification algorithms such as K-Nearest Neighbors (KNN) and Support Vector Machine (SVM). They used data pre-processing to improve model performance and evaluated the models based on accuracy, precision, recall, F1-score, and confusion matrix. The findings revealed that logistic regression is a robust and interpretable model for diabetes prediction, with comparable performance to KNN and SVM. The study suggests that logistic regression is an excellent strategy for predicting diabetes and can be used as a dependable tool in clinical settings to help healthcare workers diagnose.

3. Methodology And Platform Design

The methodology adopted in this paper is appropriate to develop a platform using Flask capable of analyzing our clinical dataset focusing specifically on text descriptions of disease signs and classification of diseases. we can see the main components of the methodology in Fig. 1, the first step according to our needs is the platform analysis.

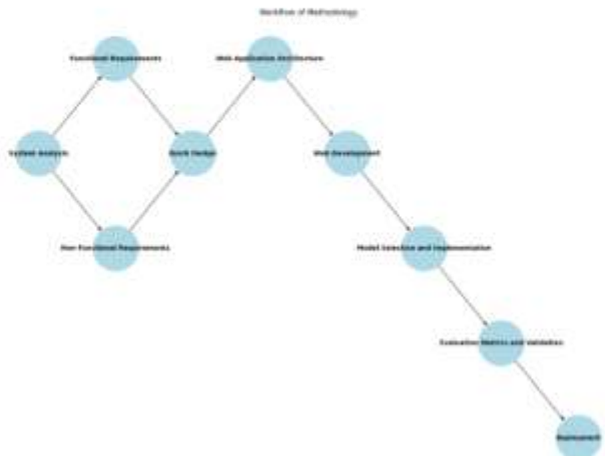


Fig. 1. workflow of methodology

3.1. Platform Analysis

According to [13] System analysis is important for the design and implementation of software initiatives. In this section, we discuss functional and non-functional requirements.

3.1.1. Functional requirements

The platform must manage user authentication, allowing users to log in and access various features based on their roles. The web application must use MySQL to authenticate users and administrators. After logging in, the administrator can choose between the supervised and unsupervised learning sections. In each component, the administrator should be able to upload a dataset and select their preferred preparation methods. They can also select their preferred machine learning models and customize the hyper-parameters of each method. The application will then show the results of the selected models.

3.1.2. Non-Functional requirements

Non-functional requirements identify the essential features and properties that a system must have, regardless of its core capabilities. They cover topics such as performance, dependability, usability, and security. These requirements ensure the quality and efficacy of the system [14]. For this platform, the requirements include a responsive and accessible system appropriate for healthcare professionals, capable of analyzing specific datasets. Additionally, a multi-layered design for a user-friendly interface, development using Flask Python for the back-end, Bootstrap for the front-end, and ensuring optimal security through a MySQL-based authentication system are critical.

3.2. Quick design

A use case diagram shows the structure of the required features of the system. The use cases are collected after an evaluation of the functional needs of the system[15]. The use case diagram in Fig. 2 displays the key functionalities of our system, with a focus on interactions between the user (admin) and the system components. The administrator starts the process by using the login feature, which requires user identification using MySQL. Successful authentication takes the administrator to the dashboard, where they can select between supervised and unsupervised learning.

In both learning directions, the administrator can upload datasets, select pre-processing methods, and choose machine learning models with hyperparameter adjustment. The system then trains the selected models, producing results that are then visualized for the user. This comprehensive workflow allows the administrator to efficiently control machine learning processes while also tracking the results via a clear online interface.

Fig. 2. Use case diagram.

3.3. Platform Architecture

The platform architecture diagram Fig. 3 illustrates a comprehensive web application structured into a front-end, an API layer, and a database. The front-end consists of various HTML templates that communicate with the API REST layer to execute actions like GET, POST, PUT, and DELETE. The API, developed using Python, is responsible for managing routes for functionalities such as user login, data retrieval, and machine learning model operations. Dynamic HTML rendering is achieved using Jinja templates, enhancing user interaction and experience.

The API layer processes data through dedicated modules for pre-processing, model training, and model testing. The MySQL database plays a role in storing user information, including usernames, passwords, and roles (user or admin). This database structure is responsible for handling user data while supporting the diverse functionalities required for clinical data analysis. The integration of the front-end, API, and database components allows seamless data flow and processing, providing a user-friendly environment for healthcare professionals with limited technical expertise. This architecture ensures that users can upload datasets, select pre-processing options, train machine learning models, and view results through an intuitive web interface.

This back-end manages tasks like user authentication, data pre-processing, model training, and result visualization. Secure user authentication is ensured by connecting to a MySQL database using the PyMySQL library. In the back-end, pre-processing executes based on the user selection of available Methods. In addition, each machine learning

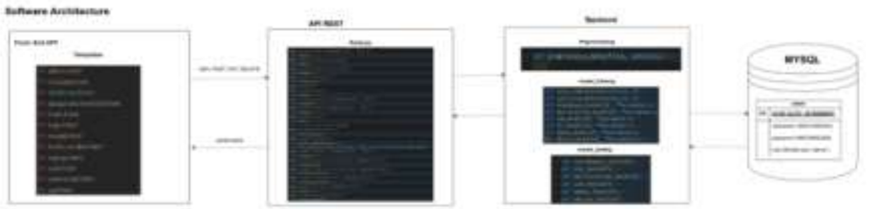


Fig. 3. The platform architecture diagram

algorithm is managed by dedicated modules for both training and testing. Once the models are trained, the results are displayed on the front-end through dynamic visualizations.We used Pandas and NumPy for efficient data handling, while Scikit-learn powered the platform's machine learning algorithms like SVM, K-Nearest Neighbors, and Decision Trees, as well as evaluation metrics. Text preprocessing was done using NLTK for tasks like stopword removal and lemmatization, and Matplotlib provided data visualizations. To handle imbalanced datasets, SMOTE from Imbalanced-learn was used, while Joblib saved trained models for future use.

3.4. Development

Flask is a Python web framework that provides tools and pre-written code that simplifies website construction, minimizing the need to start from scratch. We chose Flask because it is a microframework with few tools and libraries, making it portable and resource-efficient [16]. In the development phase in Fig. 4, the initial emphasis was on setting up the program structure, which specifically focused on creating a user interface framework and how to display the results to the user. The primary file structures of the developed application consist of four distinct program modules:

- **'model.pkl':** stores the machine learning model for our prediction.
- **'app.py':**Manages user authentication and session handling, coordinates data preprocessing, machine learning model training, and result visualization within the web application.
- **'templates/index.html':**The HTML form allows users to upload datasets, select preprocessing models, choose machine learning algorithms, and view the predicted results.
- **'static/css/style.css':** Contains the necessary styling for the HTML form to ensure a user-friendly interface.

applications by offering a compilation of libraries and modules for building back-end applications, while the front-end is constructed using HTML, CSS, and JavaScript. The framework employs techniques like "get" and "post" to manage data requests and responses, transform string inputs into integers, and transmit them as parameters to models for prediction. The Flask application is configured to handle HTTP requests, render templates, and manage routes efficiently.

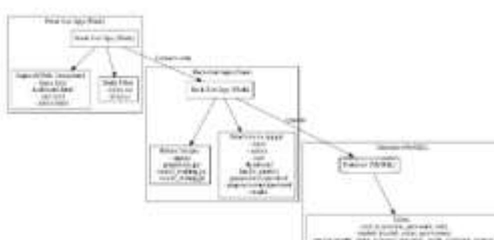


Fig. 4. Web development Diagram

3.5. Model Selection and Implementation

At this point, some statistical machine-learning models were selected for text analysis and disease classification based on the characteristics of our dataset and the project objectives. At first, we have some information about our dataset. This study's dataset consists of textual records related to health issues, including categories like "COVID-19" and "Fatigue, exhaustion, or overload at work." Each entry is labeled for classification purposes. The platform is designed specifically to process and analyze this dataset, enabling the prediction and categorization of new data entries based on these predefined labels.The models incorporated into the Flask platform, which are based on user selection, include K-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machine (SVM), and Decision Trees. These models are trained on pre-processed data, which is divided into training and testing sets, with the testing set representing 10% of the data. Additionally, users have the option to manually choose hyper-parameters and the text input vectorization to personalize the training procedure.

3.6. Evaluation Metrics and Validation

We assess our machine learning models using various metrics. These metrics encompass accuracy, precision, recall, and analysis of the confusion matrix. Additionally, k-fold cross-validation was applied to evaluate the robustness and generalized of the models. Learning curve graphs facilitate visualization of the models' performance across various sizes of training sets.

4. Results

In this section, we present the results obtained from developing our web application platform for clinical predictive modeling. While the primary focus of our research was on creating this web platform, we also report the accuracy of the model obtained during the training and validation phases using various algorithms.

4.1. User Interface workflow

When designing the User Interface, the platform is separated into distinct layers for data input, model selection, and result visualization, streamlining the user experience. To further enhance usability, we integrated pop-up features that provide detailed explanations for each parameter, ensuring that users can understand the function and impact of each setting. These enhancements help non-technical users to easily navigate through the platform, confidently configure model parameters, and interpret results.

Fig. 5 represents our developed platform workflow that has been developed based on Flask framework, integrating various machine learning algorithms for clinical data analysis The web application for clinical predictive modeling begins with a login interface Fig. 5 (a). Users, including administrators, can log in by entering their username and password and selecting their role. Upon successful authentication, the user is redirected to the dashboard in Fig. 5 (b). The dashboard presents options for supervised and unsupervised learning algorithms. Users can choose from various algorithms such as Naive Bayes, K-Nearest Neighbors, SVM, Decision Tree for supervised learning, and K-means and DBSCAN for unsupervised learning. The dashboard provides a brief description of each algorithm to assist users in their selection. Next, users proceed to the data pre-processing section Fig. 5 (c). Here, they can upload their dataset and select pre-processing operations such as converting text to lowercase, removing stop words, and lemmatization. After pre-processing the data, users can choose their desired classification model and configure hyper-parameters accordingly. Once the pre-processing and model selection are complete, users can initiate the model training process. The results page Fig. 5 (d) displays the performance metrics of the selected models, including accuracy, precision, recall, and a confusion matrix. Additionally, learning curves are presented to visualize the model's training and validation performance.

4.2. Machine Learning Algorithms:

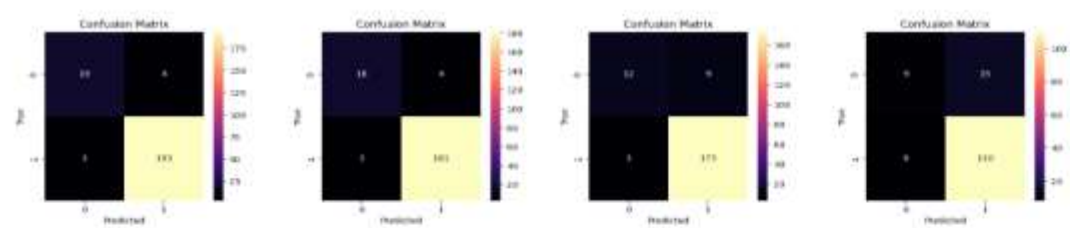
Our primary objective was to develop a user-friendly platform tailored for analyzing our clinical data. The platform's core design aims to deliver adaptability in the training of machine learning models through a range of hyper-

(a) Login Page (b) Admin Page (c) Multi Models (d) Result of Models

Fig. 5. User Interface workflow

parameter configurations. This adaptability empowers users to have varied outcomes depending on their chosen hyper-parameters. Every machine learning algorithm included in our platform is furnished with a specific set of hyper-parameters that are modifiable to improve performance. Modifying a single parameter within algorithms like SVM, K-Nearest Neighbors, or Decision Tree can impact the model's performance metrics. However, in this section, we examine the results obtained using the default hyper-parameters typically employed in data science. By presenting these baseline results, users can easily compare the outcomes and understand how changes in hyper-parameters might affect the performance. The outcomes presented herein include accuracy, precision, recall, and the confusion matrix for each algorithm, offering a comprehensive overview of the model's baseline performance.

A confusion matrix illustrates the efficacy of a classification model by comparing the observed values with the anticipated values produced by the model. Elevated counts of true positives and true negatives, coupled with diminished counts of false positives and false negatives, signify high-quality performance of the model [9]. Using K Nearest Neighbors for testing: The K Nearest Neighbors algorithm is employed to make predictions on the dataset points. It determines the point that closely matches based on the dataset values. The resulting accuracy is 80.29%. The confusion matrix is displayed in Fig. 6 (a). The Naïve Bayes classifier, based on the Bayes theorem, also known as Bayes rule or Bayes law, is at the heart of the Naïve Bayes algorithm, providing a framework for calculating conditional probabilities based on existing information about events [17]. As we can see in the table 1 the accuracy of NB is 80.29% a confusion matrix displayed in Fig. 6 (b). Testing with SVM: Support vector machine is a supervised learning algorithm primarily utilized for classification tasks. SVM features a linear hyperplane with a margin that set apart the dataset into positive and negative samples. SVM chooses the hyper-planes that offer the maximum distance possible [18, 19]. In Fig. 6 (c) we can observe the result of the confusion matrix. As previously indicated, our application is structured to produce varying outcomes contingent upon each specific modification in the hyper-parameters. The results of the studies indicate that the KNN and Naive Bayes algorithms outperformed the other algorithms in terms of accuracy and performance.



(a) KNN (b) Naive Bayes (c) SVM (d) Desision Tree

Fig. 6. Confusion Matrix

Furthermore, our platform integrates clustering techniques to provide further insight into the structure of the data. Using DBSCAN and K-Means algorithms, users can uncover hidden patterns and groupings within their datasets.

Table 1. Performance Comparison

Model	Accuracy	Precision	Recall
NB	80.29	84.68	84.29
KNN	80.11	84.99	85.83
SVM	32.99	30.0	64.17
DT	42.88	41.23	42.88

5. Conclusion

This study presented the development of a Web-based platform for clinical predictive modeling, built on the Flask framework. It supports several machine learning algorithms, including K-Nearest Neighbors, Support Vector Machines, Naive Bayes, and Decision Trees, along with clustering methods such as DBSCAN and K-means. Healthcare professionals can easily upload, analyze data, make predictions, and classify new items without requiring advanced technical expertise. The platform's seamless integration of data pre-processing, model training, and visualization enhances its accessibility and effectiveness for non-experts. Future work should focus on incorporating more advanced models, improving performance, strengthening data security, scaling for larger datasets, and validating with real-world clinical data to further support data-driven decision-making in healthcare.

References

[1] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," 2019.
[2] G. Roy, R. Rastogi, and V. Chaoji, "Machine learning (ml) in the real world," in *VLDB 2016*, 2016.
[3] S. Angra and S. Ahuja, "Machine learning and its applications: A review," in *2017 international conference on big data analytics and computational intelligence (ICBDAC)*, pp. 57–60, IEEE, 2017.
[4] F. A. Aslam, H. N. M. J. M. M. M. M. A. Gulamgaus, and P. S. L. A. Professor, "Efficient way of web development using python and flask."
[5] S. Sobale, H. More, S. Mulay, and M. Kshirsagar, "Supervised machine learning model builder for regression and classification analysis featured as a web application," in *2023 International Conference on Advancement in Computation and Computer Technologies (InCACCT)*, pp. 16–21, 2023.
[6] Y. Akkem, "Multi disease prediction model by using machine learning and flask api," 2020.
[7] M. Swati and M. R. Jeffrey, "Designing interactive transfer learning tools for ml non-experts," 2021.
[8] N. Ahmed, R. Ahammed, M. M. Islam, M. A. Uddin, A. Akhter, M. A. A. Talukder, and B. K. Paul, "Machine learning based diabetes prediction and development of smart web application," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 229–241, 6 2021.
[9] C. N. Sujatha, A. Gudipalli, B. Pushyami, N. Karthik, and B. N. Sanjana, "Loan prediction using machine learning and its deployement on web application," Institute of Electrical and Electronics Engineers Inc., 2021.
[10] S. K. Dey, A. Hossain, and M. M. Rahman, "Implementation of a web application to predict diabetes disease: An approach using machine learning algorithm," Institute of Electrical and Electronics Engineers Inc., 1 2019.
[11] S. Narayanan, M. Balamurugan, M. K, and P. B. Palas, "Leveraging machine learning methods for multiple disease prediction using python ml libraries and flask api," 2022.
[12] H. B. Sriram M, Nithish J, "Prediction of diabetes using logistic regression algorithms," *International Journal for Research in Applied Science and Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*, p. 10, 2022.
[13] E. Foster and B. T. Jr., *Software Engineering: A Methodical Approach, 2nd Edition*. Auerbach Publications, 2nd ed., 2021.
[14] I. Sommerville, *Software Engineering 9th Edition*. 2011.
[15] U. Isikdag and J. Underwood, "Two design patterns for facilitating building information model-based synchronous collaboration," *Automation in Construction*, vol. 19, no. 5, pp. 544–553, 2010.
[16] D. A. Anggoro and N. C. Aziz, "Implementation of k-nearest neighbors algorithm for predicting heart disease using python flask," *Iraqi Journal of Science*, vol. 62, pp. 3196–3219, 9 2021.
[17] M. J. Islam, Q. M. J. Wu, M. Ahmadi, and M. A. Sid-Ahmed, "Investigating the performance of naive- bayes classifiers and k- nearest neighbor classifiers," in *2007 International Conference on Convergence Information Technology (ICCIT 2007)*, pp. 1541–1546, 2007.
[18] V. Chandra and S. Hareendran, *Artificial Intelligence and Machine Learning*. Delhi 110092: PHI Learning Private Limited, 2014.
[19] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Pearson Education, Inc, 2006.

GENERAL CONCLUSION

This work presents the development of a web-based platform designed to facilitate clinical predictive modeling by integrating various machine learning (ML) techniques into an accessible user interface. By providing a web-based platform that applies machine learning algorithms to a specific clinical data. Our proposed solution does not require coding skills. However, a minimal knowledge on ML techniques and their evaluation is required.

The user-friendly platform allows the implementation of ML applications in our specific clinical settings for users. Key features such as dataset uploads, result visualization, and preprocessing options contribute to its accessibility.. It supports several supervised and unsupervised machines learning techniques, including K-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machines (SVM), Decision Trees, K-Means, and DBSCAN.

Moreover, the integration of performance indicators such as confusion matrices, accuracy, precision, and recall enables users to gain a comprehensive understanding of the models' efficacy. The primary goal of this thesis was to design and develop a user-friendly, web-based platform that makes machine learning more accessible to healthcare professionals with limited programming skills. Rather than focusing on the validation or optimization of machine learning models, the emphasis was placed on usability and the integration of multiple algorithms into an intuitive interface. Nevertheless, preliminary testing of the platform revealed that algorithms such as Naïve Bayes (NB) and K-Nearest Neighbors (KNN) performed relatively well on the clinical dataset used, while others were less effective. These initial results highlight the potential of the platform but also suggest opportunities for future improvement. In particular, integrating more advanced model such as neural networks or deep learning technique could [9], [17]enhance our platform efficiency. Future work may also involve expanding the platform's capabilities and testing it with larger and more diverse datasets.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, “Text classification algorithms: A survey,” 2019, *MDPI AG*. doi: 10.3390/info10040150.

[2] C. Laizhong, Y. Shu, C. Fei, and M. Zhong, “A survey on application of machine learning for Internet of Things,” *International Journal of Machine Learning and Cybernetics* , pp. 1399–1417, Jun. 2018.

[3] K. Vishwanatth, S. Satish, P. Enagandula, A. A. Khade, and Z. Mirza, “A Comprehensive Survey of Artificial Intelligence and Machine Learning Application in Healthcare,” 2024*, J. Electrical Systems 20-6s (2024): 1425-1431*

[4] C. Stewart, ‘‘AI in healthcare market size worldwide 2030,.’’, *https://www.statista.com/statistics/1334826/ai-in-healthcare-market-size-worldwide/*

[5] S. Nadakuditi, B. Kumar, and T. Kumar, “AI and Machine Learning in Healthcare-Applications, Challenges and Ethics Sr IT Business System Analyst (Florida Blue),” *Int J Health Sci (Qassim)*, vol. 7, no. 4, pp. 36–43, 2024, [Online]. Available: www.carijournal.or

[6] E. Jahanshiri and A. R. M. Shariff, “Developing web-based data analysis tools for precision farming using R and Shiny,” in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, 2014. doi: 10.1088/1755-1315/20/1/012014.

[7] A. Verma, C. Kapoor, A. Sharma, and B. Mishra, “Web Application Implementation with Machine Learning,” in *Proceedings of 2021 2nd International Conference on Intelligent Engineering and Management, ICIEM 2021*, Institute of Electrical and

Electronics Engineers Inc., Apr. 2021, pp. 423–428. doi: 10.1109/ICIEM51511.2021.9445368.

[8] W. Bala *et al.*, “A Web application for adrenal incidentaloma identification, tracking, and management using machine learning,” *Appl Clin Inform*, vol. 11, no. 4, pp. 606–616, Aug. 2020, doi: 10.1055/s-0040-1715892.

[9] Y. Yu, M. Li, L. [9], [17], Y. Li, and J. Wang, “Clinical big data and deep learning: Applications, challenges, and future outlooks,” Dec. 01, 2019, *Tsinghua University Press*. doi: 10.26599/BDMA.2019.9020007.

[10] F. Bashardoustjoubjarkouli and M. Adda, “Interactive Machine Learning Pedagogy: Developing a Web-Based Educational Platform for Clinical Predictive Modeling,” *Procedia Comput Sci*, vol. 251, pp. 382–389, 2024, doi: 10.1016/j.procs.2024.11.124.

[11] F. Armash Aslam, H. Nabeel Mohammed Jummal Musab Mohd Munir Murade Aaraf Gulamgaus, and P. S. Lokhande Assistant Professor, “Efficient Way Of Web Development Using Python And Flask.” [Online]. Available: www.ijarcs.info

[12] R. K. G, R. Reddyvari Venkateswara, and M. Jayarathna, “Web application based Diabetes prediction using Machine Learning,” *Conference: 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Mar. 2023.

[13] E. Foster and B. Towle Jr, *Software Engineering: A Methodical Approach, 2nd Edition*. 2021.

[14] I. Sommerville, *Software Engineering 9th Edition*. 2011.

[15] U. Isikdag and J. and Underwood, “Two design patterns for facilitating Building Information Model-based synchronous collaboration,” *Autom Constr*, vol. 19, no. 5, pp. 544–553, 2010.

[16] D. Anggoro and N. Aziz, “Implementation of K-Nearest Neighbors algorithm for predicting heart disease using Python Flask,” *Iraqi Journal of Science*, vol. 62, no. 9, pp. 3196–3219, 2021.

[17] D. Chen *et al.*, “Deep learning and alternative learning strategies for retrospective real-world clinical data,” Dec. 01, 2019, *Nature Publishing Group*. doi: 10.1038/s41746-019-0122-0.