Université du Québec
à Rimouski

RÉSEAUX DE NEURONES POUR LA DÉTECTION DES SIBILANTS DANS LES SONS ACOUSTIQUES RESPIRATOIRES DANS LE BUT D'UNE IMPLANTATION EFFICACE TEMPS-RÉEL SUR CIRCUIT FPGA

MÉMOIRE PRÉSENTÉ

dans le cadre du programme de maîtrise en ingénierie

en vue de l'obtention du grade de maître ès sciences appliquées (M. Sc. A.)

PAR

©ABDELKRIM SEMMAD

Novembre 2021

Composition du jury :

Jean-François Méthot (Ph.D.), président du jury, Université du Québec à Rimouski

Mohammed Bahoura (Ph.D.), directeur de recherche, Université du Québec à Rimouski

Jean Rouat (Ph.D.), examinateur externe, Université de Sherbrooke

Dépôt initial le 24 août 2021 Dépôt final le 06 novembre 2021

REMERCIEMENTS

Je tiens d'abord à remercier mon directeur de recherche le Professeur Mohammed Bahoura de l'Université du Québec à Rimouski pour sa disponibilité, son soutien indéfectible et ses précieux conseils. Le Professeur Mohammed Bahoura m'a transmis le goût de la recherche, sous sa direction, j'ai eu l'opportunité à participer à une conférence internationale.

Je tiens aussi à remercier le Professeur Jean-François Méthot de l'Université du Québec à Rimouski pour avoir eu l'amabilité d'accepter d'être le président du jury qui va examiner ce mémoire.

J'exprime également mes sincères remerciements au Professeur Jean Rouat de l'Université de Sherbrooke pour bien vouloir examiner ce travail.

Cette recherche a été rendue possible grâce au soutien financier du Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG).

RÉSUMÉ

Les maladies pulmonaires obstructives chroniques (MPOC) sont considérées comme l'une des causes principales de mortalité selon l'organisation mondiale de la santé (OMS). Un diagnostic précoce et un traitement médical adéquat sont recommandés pour diminuer les risques liés à de telles affections. En effet, un bon diagnostic médical repose sur une identification rigoureuse des sons respiratoires produits par un patient. Face aux limites inhérentes aux techniques de diagnostic conventionnelles (auscultation par stéthoscope, spirométrie, etc.), des techniques basées sur le traitement du signal et l'apprentissage machine ont été adoptées pour l'identification automatique des sons respiratoires adventices. L'objet principal de ce travail de recherche est la détection automatique et en temps réel, des sibilants qui sont un symptôme lié le plus souvent à la maladie d'asthme. Pour ce faire, le travail a été divisé en deux parties.

En première partie, un ensemble de méthodes ont été proposées pour la caractérisation des sons respiratoires à l'aide des coefficients cepstraux et chacune de ces méthodes a été combinée avec des techniques d'apprentissage automatique basées sur des réseaux de neurones. Le but étant d'atteindre les meilleurs taux de détection des sibilants. Étant donné la similitude qui existe de point de vue acoustique entre les sons respiratoires et la parole, des techniques telles que les coefficients cepstraux à l'échelle de Mel (MFCC), les coefficients cepstraux à filtres gammatone (GFCC) et les coefficients cepstraux à l'échelle de Bark (BFCC), d'habitude utilisées pour la caractérisation de la parole, ont été employées ici pour l'extraction des caractéristiques à partir des sons respiratoires. Quant à la classification, notre choix s'est porté sur les réseaux de neurones, plus particulièrement le perceptron multicouche (MLP) et le réseau de neurones avec mémoire à court et long terme bidiriectionnel (BiLSTM).

Dans la seconde partie, un modèle de détection basé sur une combinaison MFCC-MLP a été sélectionné afin d'être implanté. Une implantation matérielle sur une puce FPGA a été réalisée pour permettre la détection en temps réel sur un dispositif portable. Pour les applications à faible débit, une architecture complètement sérialisée du classifieur MLP, basée sur une seule unité de traitement a été proposée afin de réduire les ressources matérielles et la puissance totale consommée. L'architecture proposée a été concrétisée à l'aide de Xilinx System Generator (XSG) qui est un outil de programmation de haut niveau pour les circuits FPGA de Xilinx.

Les résultats de classification obtenus ont été évalués en termes de la spécificité et de la sensitivité ainsi que des critères d'exactitude et de performance. Les meilleurs scores de classification ont été obtenus avec la combinaison MFCC-BiLSTM (96% d'exactitude). Quant à la nouvelle architecture proposée pour l'implantation du MLP, elle a été validée étant donné qu'elle produit les mêmes résultats de classification que l'architecture parallèle considérée comme référence.

Mots clés : Sons respiratoires, MFCC, GFCC, BFCC, MLP, LSTM, Sibilants, Détection des sibilants, FPGA.

ABSTRACT

According to the world health organization (WHO), chronic obstructive pulmonary diseases (COPD) are considered as one of the leading causes of mortality. An early diagnosis and an adequate medical treatment are recommended to reduce the risks related to such conditions. Indeed, an accurate medical diagnosis is based on a rigorous identification of the respiratory sounds produced by a patient. Because of the inherent limitations of conventional diagnostic techniques (auscultation by stethoscope, spirometry, etc.), techniques based on signal processing and machine learning have been adopted to automatically identify adventitious respiratory sounds. The primary purpose of this research is the automatic real-time detection, from respiratory sounds, of wheezing, that is a symptom related to asthma disease. This work has been achieved in two steps.

In the first step, a set of methods has been proposed to characterize respiratory sounds using cepstral coefficients, and each of these methods was combined with machine learning techniques based on neural networks. The goal was to achieve the best detection rates of wheezes. Given the similarity, on the acoustic point of view, between breath sounds and speech, techniques such as Mel-scale cepstral coefficients (MFCC), gammatone frequency cepstral coefficients (GFCC), and Bark frequency cepstral coefficients (BFCC), usually used for speech characterization tasks, have been employed here for feature extraction from respiratory sounds. For classification, we chose neural networks, more specifically the multilayer perceptron (MLP) and the bidirectional long short-term memory (BiLSTM).

In the second step, a detection model based on an MFCC-MLP combination was selected to be implemented. A hardware implementation on an FPGA chip has been realized to enable real-time detection on a mobile device. For low data rate applications. A fully serial

architecture of the MLP classifier, based on a single processing unit, has been proposed to reduce the resource utilization and the total power consumption. The proposed architecture has been implemented using Xilinx System Generator (XSG), a high-level programming tool for FPGA chips from Xilinx.

The classification results obtained were evaluated in terms of specificity, sensitivity and the resulting accuracy and performance criteria. The best classification scores were obtained with the MFCC-BiLSTM combination (96% accuracy). The new architecture was validated as it produces the same classification results as the parallel architecture.

Keywords : Respiratory sounds, MFCC, GFCC, BFCC, MLP, LSTM, Wheezing, Wheezing detection, FPGA.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

ANN *Artificial Neural Network*.

Réseau de neurones artificiels.

BiLSTM *Bidirectional Long Short-Term Memory*.

Réseau de neurones avec mémoire à court et long terme bidirectionnel.

BP *Backpropagation*.

Rétropropagation.

ERB *Equivalent Rectangular Bandwidth*.

Largeur de bande rectangulaire équivalente.

FFT *Fast Fourier Transform*.

Transformée de Fourier rapide.

FPGA *Field Programmable Gate Array*.

Réseau de portes logiques programmables.

GFCC *Gammatone Frequency Cepstral Coefficients*.

Coefficients cepstraux à filtres gammatone.

LSTM *Long Short-Term Memory*.

Réseau de neurones avec mémoire à court et long terme.

MAC *Multiplication And accumulation*.

Multiplication et accumulation.

MFCC *Mel-Frequency Cepstral Coefficients*.

Coefficients cepstraux à l'échelle de Mel.

MLP *Multi Layer Perceptron*.

Perceptron multi-couches.

PE *Processing Element*.

Élément de traitement.

BFCC *Bark Frequency Cepstral Coefficients*.

Coefficient cepstraux à l'échelle de Bark.

RNN *Recurrent Neural Network*.

Réseau de neurones récurrent.

STFT *Short Time Fourier Transform*.

Transformée de Fourier à court terme.

SVM *Support Vector Machine*.

Machine à vecteurs de support.

VHDL *Very high speed integrated circuit Hardware Description Language*.

Langage de description de matériel pour circuits à très haute vitesse d'intégration.

WPT *Wavelet Packet Transform*.

Transformée par paquets d'ondelettes.

XSG *Xilinx System Generator*.

Générateur Système de XILINX.

CHAPITRE 0

INTRODUCTION GÉNÉRALE

0.1 État de l'art

Les sibilants sont l'un des trois principaux sons respiratoires pathologiques (avec les crépitants et les ronchi). Les sibilants (*Wheezes* en anglais) sont un symptôme prépondérant dans certaines maladies respiratoires, à savoir l'asthme ou la bronchopneumopathie chronique obstructive (BPCO) (Pasterkamp, 2017; Jácome and Marques, 2015). Selon l'enquête de 2013/2014 sur la santé dans les collectivités canadiennes, la maladie d'asthme est très répandue, son taux de prévalence a atteint au Québec 8.7% de la population de plus de 12 ans. Étant donné le caractère chronique de la maladie d'asthme et les complications qu'elle engendre, beaucoup d'efforts doivent être déployés pour assurer un diagnostic précoce de cette maladie. Cependant, le diagnostic par le médecin spécialiste à l'aide d'un stéthoscope présente un inconvénient majeur lié à la nature subjective de l'examen médical. Également, la spirométrie, une autre technique de diagnostique utilisée par les pneumologues et qui consiste à souffler dans un tube relié à un appareil d'analyse, s'est révélée ne pas être en mesure de fournir un diagnostic exact de l'état du patient. En effet, l'examen de spirométrie surestime la présence de la maladie d'asthme (Schneider et al., 2009). Aussi, selon l'organisation mondiale de la santé (OMS), la maladie d'asthme est sous-diagnostiquée dans les pays sous-développés. Les auteurs dans (van Schayck et al., 2000) ont étudié un échantillon aléatoire de 1155 sujets adultes et ont constaté que 86 (7,4%) présentaient des symptômes compatibles avec de l'asthme. Sur les 86 sujets asthmatiques, 63 (73%) n'ont jamais été diagnostiqué comme souffrant de l'asthme. La cause étant, selon l'étude menée, la minimisation par le

patient des symptômes de la maladie d'asthme, ce phénomène serait lié principalement au niveau d'éducation.

La problématique de recherche est donc définie par le manque d'un dispositif d'aide à la décision basé sur des techniques de classification et de reconnaissance. Un tel système serait capable de fournir un avis objectif complémentaire à celui du médecin et des techniques déjà utilisée comme la spirométrie. Dans la même perspective, une solution portable et à coût modéré, augmenterait l'accessibilité du grand public à un dispositif de classification automatique des sons respiratoires. Pour la détection des sibilants dans les sons respiratoires, nos efforts ont été concentrés sur deux volets. Le premier consiste à tester un ensemble de méthodes de caractérisation des sons respiratoires et à les combiner avec des méthodes de reconnaissance et de classification. Alors que le second concerne l'implantation matérielle du classifieur afin d'accomplir la détection automatique des sibilants en temps réel et à moindre coût. Les travaux précédents dans ce domaine proposent des méthodes qui combinent des techniques de traitement du signal pour l'extraction des caractéristiques ainsi que des algorithmes d'apprentissage machine pour la reconnaissance et la classification des sons respiratoires. Parmi les méthodes de caractérisation classiques, on trouve celles basées sur la Transformée de Fourier (TF) (Oud et al., 2000; Waitman et al., 2000; Rietveld et al., 1999; Riella et al., 2009). Également, la transformée en ondelettes a été proposée pour une analyse temps-échelle multirésolution des sons respiratoires (Kandaswamy et al., 2004; Kahya et al., 2006; Sankur et al., 1996). Les coefficient cepstraux à échelle de fréquence de type Mel (ou MFCC : Mel Frequency Cepstral Coefficients), basés sur une analyse par banc de filtres, sont largement utilisés pour la caractérisation des sons pulmonaires (Sengupta et al., 2016; Bahoura and Pelletier, 2004; Bahoura, 2009; Acharya and Basu, 2020; Lin and Lin, 2016). Un autre algorithme proposé pour la caractérisation des sons respiratoires se base sur la décomposition empirique de modes (ou EMD : Empirical Mode Decomposition). Cet

technique permet une analyse plus fine des sons respiratoires (Lozano et al., 2016; Mondal et al., 2018). Concernant la classification, les réseaux de neurones artificiels (ou ANN : Artificial Neural Networks), en général, et le perceptron multicouches (ou MLP : multilayer perceptron) en particulier sont largement utilisés étant donné leur grande capacité à reconnaitre des formes complexes dans les sons respiratoires (Sezgin et al., 2001; Forkheim et al., 1995; Rietveld et al., 1999). D'autre réseaux de neurones de type récurrent, comme le réseau LSTM (Long Short-Term memory), ont été proposés pour la classification des sons respiratoires (Perna and Tagarelli, 2019; Fraiwan et al., 2021; Messner et al., 2020; Acharya and Basu, 2020). Parmi les techniques d'apprentissage supervisé qui ne sont pas basées sur des réseaux de neurones, nous trouvons celles des machines à vecteurs de support (ou SVM : Support Vector Machine) qui ont été proposées pour la classification des sons respiratoires (Palaniappan and Sundaraj, 2013; Mazić et al., 2015).

Afin de rendre possible le traitement en temps-réel des sons respiratoires une implantation matérielle sur un circuit programable comme les puces FPGA (Field Programmable Gate Array) ou les cartes DSP (Digital Signal Processor) est nécessaire. Parmi les modèles de classification étudiés, un modèle basé sur la combinaison d'une méthode d'extraction de caractéristiques et d'une technique de classification est choisi pour faire l'objet d'une implantation matérielle. Après une revue de littérature exhaustive, seulement 4 implantations sur une puce FPGA de systèmes de détection automatique des sibilants dans les sons respiratoires avaient été dénombrées (Bahoura, 2018; Boujelben and Bahoura, 2018; Lin and Yen, 2014; Gwo-Ching, 2009). Notre choix pour le circuit d'implantation s'est porté vers les puces FPGA étant donné leur flexibilité de configuration et la grande vitesse de traitement, tout ceci combiné à une faible consommation d'énergie.

0.2 Nomenclature et classification des sons respiratoires

Les maladies pulmonaires obstructives chroniques (MPOC) affectent le système respiratoire humain et dont les effets physiologiques donnent lieu à une nouvelle gamme de sons dits sons adventices. La catégorie de son adventice qu'il sera question d'étudier dans le cadre de ce travail de recherche est celle des sibilants (sifflements). Dans cette section, les principaux sons respiratoires sur lesquels on va revenir tout au long de ce mémoire seront définis.

0.2.1 Sons respiratoires normaux

Les sons normaux sont le caractère principal de la respiration chez une personne saine. Ces sons sont produits à la suite du passage de l'air dans l'arbre trachéo-bronchique (Emmanuel et al., 2007). Sachant que la perception des sons normaux dépend de l'endroit où on place le stéthoscope, il existe donc une panoplie de sons respiratoires normaux : son vésiculaire, son trachéal, son trachéo-bronchique, etc. Le principal son pulmonaire normal est appelé murmure vésiculaire et qui s'entend comme un bruit doux durant la phase d'inspiration et au début de l'expiration (Bahoura, 2009).

0.2.2 Sons respiratoires pathologiques

La présence des sons sibilants durant la respiration est un signe d'une pathologie respiratoire. Ces sons sont perçus par l'oreille humaine comme des sifflements (*Wheezing* en anglais) et dont la cause physiologique est le rétrécissement des bronches causé par les muqueuses. Les sibilants sont identifiés comme des sons musicaux ayant une fréquence

inférieure à 2000 Hz et une durée relativement longue (environ 250 ms) (Fiz et al., 2002). Cependant, les fréquences dont les amplitudes sont les plus significatives se situent dans l'intervalle 100-1000 Hz (Taplidou et al., 2004) ou 100-1200 Hz selon (Sánchez Morillo et al., 2013; Pasterkamp et al., 1985) lorsqu'il s'agit de sons enregistrés au niveau de la trachée. Les fréquences fondamentales des sibilants apparaissent comme des lignes quasi-horizontales dans le spectrogramme. Le sifflement peut être monophonique (une seule note de musique) ou polyphonique (plusieurs notes simultanées) (Piirilä et al., 1991).

0.3 Problématique de recherche

La problématique de recherche consiste à concevoir un système de détection automatique des sibilants dans les sons respiratoires. Le manque d'un tel dispositif, capable de faire le dépistage des maladies respiratoires en se basant sur la détection des sons pathologiques, a ouvert la voie afin de mener des travaux de recherche dans ce domaine. Le but de ces travaux est l'expérimentation des méthodes de reconnaissance visant la détection et la classification des sons respiratoires. L'auscultation du patient par stéthoscope dépend fortement de la perception auditive du médecin et de sa capacité de mémorisation. Ces éléments sont des signes que l'avis du spécialiste porte une marque de subjectivité et pourrait donc être erroné. Également, la nature qualitative des sons respiratoires et la terminologie utilisée pour les décrire jouent un rôle considérable dans le diagnostic (Jácome and Marques, 2015). Dès lors, l'intérêt pour un système de détection automatique des sons sibilants se trouve être pleinement justifié.

Un tel système serait capable de déterminer objectivement, c'est-à-dire avec exactitude et conforme à la réalité, la nature d'un son respiratoire et servir comme un dispositif d'aide à la décision pour le médecin responsable d'établir un diagnostic pertinent de l'état du patient.

Un tel système pourrait faire l'objet d'une implantation sur une puce électronique afin de réduire son coût et le rendre abordable. Étant donné que les différentes équipes de recherche utilisent des bases de données différentes, cela rend plus difficile l'établissement d'une base de comparaison ne tenant compte que des méthodes adoptées et des résultats obtenus. Cependant, une base de données a été partagée récemment dans le cadre de la conférence ICBHI 2017 et est de plus en plus exploitée par les chercheurs dans ce domaine. Dans ce mémoire, les bases de données RALE et ASTRA seront utilisées puisque l'intérêt est porté plus particulièrement sur les sibilants. De plus, cela a permis de faire une comparaison avec les travaux antérieurs où ces bases de données ont été utilisées car l'utilisation d'une autre base de données rendrait plus difficile la comparaison avec ces travaux antérieurs, sachant que c'est l'un des objectifs principaux de ce mémoire.

0.4 Objectifs

Afin de résoudre cette problématique, les objectifs qui vont encadrer le travail de recherche ont été fixés :

— Tester un ensemble de méthodes de caractérisation basés sur des techniques d'analyse cepstrale des sons respiratoires et les combiner avec des méthodes de reconnaissance et de classification. Le but étant d'améliorer les performances de classification obtenues avec les méthodes existantes.

— Faire une implantation matérielle sur une puce programmable afin de réaliser un système de traitement temps-réel à faible coût.

0.5 Hypothèses

La première démarche entreprise dans le but de résoudre la problématique était de formuler des hypothèses sur lesquelles doit se baser la méthodologie. Ces hypothèses sont considérées comme des réponses à des problèmes sous-jacents à la problématique générale mais qui doivent néanmoins être vérifiées par la suite. La première hypothèse est basée sur l'analogie qui existe, de point vue acoustique, entre le signal de parole et celui de la respiration. Par conséquent, les méthodes utilisées dans le domaine plus développé de la reconnaissance automatique de la parole seraient adaptées à la classification des sons respiratoires. Cette hypothèse a orientée nos efforts vers les techniques d'analyse cepstrale pour la caractérisation des sons respiratoires étant donné les performances de ces méthodes dans la caractérisation de la parole. En plus, ces méthodes d'analyse cepstrale permettent de réduire l'information spectrale en un nombre limité de coefficients.

La deuxième hypothèse énonce que, étant donné que les sibilants durent une certaine période pendant le cycle de respiration, un réseau de neurones récurrent, le réseau LSTM (Long Short-Term Memory) en l'occurence, serait capable de mémoriser l'évènement (sibilant) en repérant les dépendances temporelles et par conséquent, de donner de meilleurs résultats qu'un réseau non récurrent tel que le réseau MLP. L'information temporelle est traduite par le traitement séquentiel des segments successifs du même son, étant donné que des segments appartenant à la même classe présentent des propriétés acoustiques similaires.

Concernant le volet implantation matérielle, nous partons de l'hypothèse que le calcul des sorties des différents neurones du réseau MLP peut se faire en séquentiel avec une seule unité de calcul dans un intervalle de temps assurant la continuité du traitement. Cette sérialisation va permettre d'implanter des réseaux de neurones de grande taille sur des puces FPGA à faible consomation d'énergie.

0.6 Méthodologie

Après avoir effectué une recherche bibliographique approfondie et analysé les techniques décrites dans les publications les plus pertinentes, une méthode a été élaborée pour atteindre les deux principaux objectifs :

— Amélioration des taux de détection des sibilants.

— Implantation matérielle d'un système de classification à faible coût.

La méthode retenue pour la réalisation du projet de recherche s'articule sur trois points principaux :

0.6.1 Méthodes de caractérisation des sons respiratoires

La méthode principale retenue pour l'extraction des paramètres a été celle des coefficients MFCC (Mel-Frequency Cepstral Coefficients). Ce choix est justifié par le fait que la réponse des filtres sur l'échelle de Mel imite la réponse du système auditif humain. En plus, l'extraction des caractéristiques permet de réduire toute l'information spectrale en un nombre limité de coefficients. D'autres méthodes d'analyse cepstrale ont été testées à savoir les méthodes GFCC (Gammatone-Frequency Cepstral Coefficients) (Dua et al., 2018) et les méthodes BFCC (Bark-Frequency Cepstral Coefficients) (Dash et al., 2021), qui sont jugées plus robustes que les méthodes MFCC dans un milieu bruité. L'algorithme correspondant aux méthodes GFCC est semblable à celui des méthodes MFCC. Cependant, le banc de filtres triangulaires à échelle de Mel est remplacé par une batterie de filtres gammatone avec une largeur de bande rectangulaire équivalente (où ERB : Equivalent Rectangular Bandwidth). Quant aux méthodes BFCC, elles utlisent les mêmes filtres triangulaires que pour les techniques MFCC, excepté que ces filtres sont répartis selon l'échelle de Bark.

0.6.2 Méthodes de Classification/Modélisation

Deux modèles de classification basés sur les réseaux de neurones ont été retenus pour la suite de ce travail : le réseau Multilayer Perceptron (MLP) (Haykin, 1999) et le réseau Bidirectionnal Long-Short Term Memory (BiLSTM) (Hochreiter and Schmidhuber, 1997). Le réseau MLP est un réseau de neurones non récurrent capable d'apprendre à reconnaitre des formes par régressions non linéaires complexes en ajustant ses poids synaptiques à l'aide d'un algorithme d'apprentissage tel que l'algorithme de rétropropagation (ou *Backpropagation*). Cette capacité à apprendre par des exemples concrets rend le réseau MLP parfaitement adapté aux tâches de classification des formes et donc à la classification des sons respiratoires (Kandaswamy et al., 2004). Le réseau MLP est considéré comme une technique classique de classification dans ce domaine et a donc été choisi pour servir comme base de comparaison alors que le choix du réseau BiLSTM est fait pour explorer des nouvelles techniques encore peu utilisées avec les sons respiratoires. Le réseau BiLSTM est une variante du réseau récurrent LSTM qui mémorise dans les deux directions. Le réseau BiLSTM peut extraire de l'information contextuelle en exploitant sa capacité à mémoriser un évènement lorsqu'il se produit (épisode de sibilants) pour améliorer la reconnaissance.

0.6.3 Implémentation matérielle d'un classifieur de sons respiratoires

Un classifieur basé sur la combinaison MFCC-MLP a été sélectionné pour l'implantation matérielle sur une puce FPGA. L'architecture selon laquelle le réseau MLP a été implanté est complètement sérialisée. Une telle architecture permet d'optimiser les ressources mais demande un grand effort de conception. Elle nécessite un seul neurone seulement, utilisé successivement pour effectuer les calculs qui d'ordinaire sont effectués simultanément par les neurones implantés en parallèle. Notre choix s'est porté sur les puces FPGA (Field Program-

mable Gate Array) pour faire l'implantation étant donné les nombreux avantages que cela représente. En effet les puces FPGA représentent une alternative peu coûteuse et à haute efficacité énergétique, ajouté à cela la flexibilité de configuration en raison de leur architecture interne basée sur des cellules dont la logique est configurable par l'utilisateur. Le perceptron multicouche (MLP) implanté selon une architecture parallèle dans (Bahoura, 2018) a servi de référence pour tester la validité de l'architecture proposée.

0.7 Contributions

Ce mémoire présente le travail de recherche réalisé dans le cadre de la maîtrise en génie. Le travail réalisé a abouti à la rédaction de quatre publications scientifiques dont deux acceptées, un article de journal soumis et un autre qui sera soumis prochainement :

— La première contribution intitulée Long Short-Term Memory based Recurrent Neural Network for Wheezing Detection in Pulmonary Sounds". C'est un article qui a été accepté et présenté à la conférence "64th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS) 2021". Il a été publié dans IEEE Xplore :

- A. Semmad and M. Bahoura, "Long Short Term Memory Based Recurrent Neural Network for Wheezing Detection in Pulmonary Sounds," 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), 2021, pp. 412-415, doi : 10.1109/MWSCAS47672.2021.9531784.

— La deuxième contribution s'intitule "Serial Hardware Architecture Of Multilayer Neural Network for Automatic Wheezing Detection". C'est un article qui a été accepté et présenté à la conférence "64th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS) 2021". Il a été publié dans IEEE Xplore :

- A. Semmad and M. Bahoura, "Serial Hardware Architecture of Multilayer Neural Net-

work for Automatic Wheezing Detection," 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), 2021, pp. 28-31, doi : 10.1109/MWSCAS47672.2021.9531850.

— La troisième contribution s'intitule "Comparative study of Respiratory Sounds Classification Methods based on Cepstral Analysis and Neural Networks". Cet article est une extension de la première contribution, Il sera soumis prochainement à un journal de référence dans le domaine. Il fait l'objet du deuxième chapitre.

— La quatrième contribution s'intitule "Scalable Serial Hardware Architecture of Multilayer Perceptron Network for Automatic Wheezing Detection". Cet article est une version étendue de la deuxième contribution et a été soumis au journal IEEE Access. Cet article représente le troisième chapitre du mémoire.

0.8 Organisation du mémoire

Ce mémoire a été rédigé sous forme d'articles, l'introduction générale représente le premier chapitre, la troisième contribution fait l'objet du deuxième chapitre et la quatrième contribution représente le troisième chapitre du mémoire. Quant au dernier chapitre, il consiste en une conclusion générale ainsi que les perspectives de recherche.

ARTICLE 1

COMPARATIVE STUDY OF RESPIRATORY SOUNDS CLASSIFICATION METHODS BASED ON CEPSTRAL ANALYSIS AND NEURAL NETWORKS

Résumé en français du premier article

Pour mieux répondre à la problématique de la détection automatique des sibilants, plusieurs méthodes basées sur l'apprentissage machine ont été proposées et testées dans cette étude. Ces méthodes ont pu être comparées par la suite en se basant sur les performances de classification obtenues. L'approche principale pour l'élaboration d'un modèle de classification est de combiner une technique d'extraction des caractéristiques des signaux avec une technique, basée sur des réseaux de neurones, permettant leur reconnaissance et classification.

Les méthodes de caractérisation retenues dans cette étude sont des techniques d'analyse cepstrale empruntées au domaine de la reconnaissance de la parole. En effet, les techniques d'extraction basées sur les techniques MFCC (Mel-Frequency Cepstral Coefficient) sont largement utilisées dans le domaine de la classification des sons respiratoires contrairement à celles basées sur les techniques GFCC (Gammatone-Frequency Cepstral Coefficients) et les techniques BFCC (Bark-Frequency Cepstral Coefficients) qui, elles, n'ont jamais été appliquées aux sons respiratoires au mieux de notre connaissance. Pour la classification, notre choix s'est porté sur deux types de réseaux de neurones, le réseau MLP (Multilayer Perceptron) qui est un réseau non récurrent de référence dans ce domaine, ainsi qu'une variante du réseau de neurones récurrent LSTM (Long Short-Term Memory), appelée réseau BiLSTM

(Bidirectional LSTM). Les meilleures performances de classification ont été obtenues avec le modèle MFCC-BiLSTM dont l'exactitude a atteint un taux de 96%.

Cet article fut corédigé par le candidat et son directeur de recherche, le Professeur Mohammed Bahoura. Le directeur de recherche a proposé l'idée qui a donné naissance à cet article et a encadré le candidat dans ces travaux allant de la recherche bibliographique à la rédaction en passant par l'élaboration du modèle et les tests effectués. Une version abrégée de cet article intitulée "Long Short-Term Memory based Recurrent Neural Network for Wheezing Detection in Pulmonary Sounds" a été acceptée et présentée à la conférence MWSCAS 2021 et a été publié dans IEEE Xplore. La version étendue, qui fait l'objet du présent chapitre, sera soumise prochainement à un journal de référence dans le domaine.

Mots clés: Son respiratoire sibilant, MFCC, GFCC, BFCC, Réseaux de neurones, LSTM, BiLSTM, MLP.

Abstract

Several approaches based on machine learning have been proposed and tested in this study to address the problem of automatic wheezing detection. These techniques were compared based on their achieved classification performance. The main approach for developing signal classification model is to combine a feature extraction technique with a classification method allowing its recognition and classification. The characterization techniques used in this study are based on the cepstral analysis, which is extensively used in the automatic speech recognition domain. In fact, feature extraction methods based on the MFCC (Mel-Frequency Cepstral Coefficients) are already used in the field of respiratory sound classification, unlike the feature extraction methods based on the GFCC (Gammatone-Frequency Cepstral Coefficients) and the BFCC (Bark-Frequency Cepstral Coefficients), which have never been applied to respiratory sounds to the best of our knowledge. As for the classification, we chose two types of neural networks, the Multilayer Perceptron (MLP), which is a feedforward neural network, and a variant of the LSTM (Long Short-Term Memory) recurrent neural network, called the BiLSTM (Bidirectional LSTM). The best classification performance was obtained with the MFCC-BiLSTM model that reached an accuracy rate of 96%.

Key words: Wheezing respiratory sound, MFCC, GFCC, BFCC, Neural networks, LSTM, BiLSTM, MLP.

1.1 Introduction

Chronic obstructive pulmonary diseases (COPD) affect the human respiratory system, and their physiological effects produce a new range of so-called abnormal sounds (Wheezes, Ronchi, Crackles). Wheezing is considered a preeminent symptom in the auscultation of

patients with asthma (Pasterkamp, 2017). It results from the narrowing of the bronchi caused by the mucus secretion (Gong, 1990). Wheezes are identified as respiratory musical sounds having a frequency below 2000 Hz and a relatively long duration (about 250 ms) (Lin and Lin, 2016). However, the frequencies with the most significant amplitudes are comprised within the range of 100 Hz to 1000 Hz (Taplidou et al., 2004), or from 100 Hz to 1200 Hz for respiratory sounds collected at the trachea (Sánchez Morillo et al., 2013; Pasterkamp et al., 1985). As shown in Fig.1.1, fundamental wheezing frequencies appear like quasi-horizontal lines in the spectrogram (Bahoura, 2009). Wheezing can be monophonic (single musical note) or polyphonic (multiple simultaneous notes) (Fiz et al., 2002).

The need of using machine learning to detect wheezing in respiratory sounds is justified by the fact that conventional auscultation by stethoscope is a subjective approach based on the

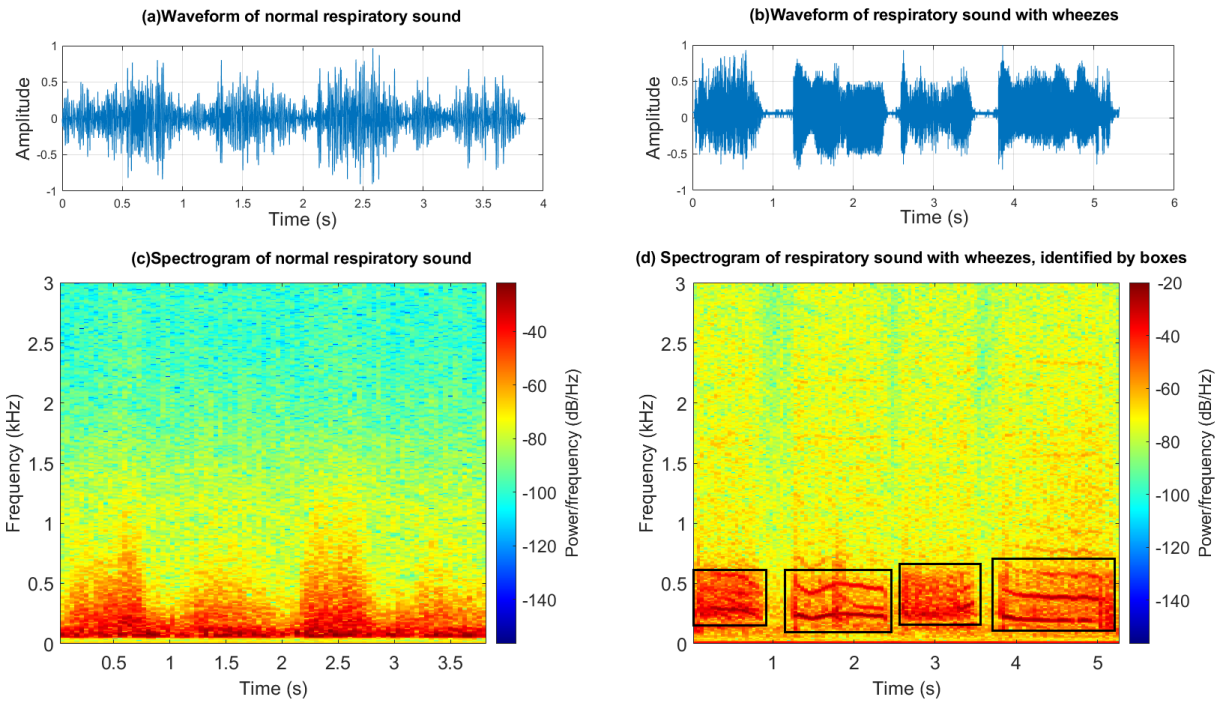


Figure 1.1 – Waveforms and spectrograms of normal pulmonary sound recorded from a healthy patient and wheezing pulmonary sound recorded from patient with asthma.

physician's experience, ability, and auditory perception (Jácome and Marques, 2015). This approach can lead to interpretation errors due to inter-observer variability. Also, auscultation of patients who cannot cooperate (very young children, mentally handicapped subjects) cannot be adequately performed (Rietveld et al., 1999). Therefore, computerized analysis using signal processing techniques and automatic classification with machine learning were used to provide an objective assessment of respiratory sounds.

Literature review of the last decades shows that frequency domain analysis is well adapted to extract wheezing features from recorded respiratory sounds, where the frequency content of the wheezing is very distinctive from that of a normal respiratory sound. The respiratory sounds spectral characterization is mainly based on the Fourier Transform (FT) (Fiz et al., 2002; Rietveld et al., 1999; Riella et al., 2009). Wavelet transform (WT) was also proposed for multiresolution analysis of breath sounds (Kahya et al., 2006; Kandaswamy et al., 2004). The Mel-Frequency Cepstral Coefficients (MFCC) is a powerful approach based on cepstral analysis and Mel-scale filter bank decomposition. MFCC-based feature extraction technique provides a good characterization of the breath sounds and helps to cut down the computational cost by reducing the spectral information to only a few coefficients. MFCCs were used to characterize respiratory sounds resulting in good classification performance (Sengupta et al., 2016; Bahoura and Pelletier, 2004; Bahoura, 2018; Basu and Rana, 2020; Lin and Lin, 2016). There are other types of filter banks that are used to produce other kinds of cepstral coefficients like GFCC (Gammatone-Frequency Cepstral Coefficients) and BFCC (Bark-Frequency Cepstral Coefficients) that have never been used to characterize the respiratory sounds.

The feature extraction techniques are combined with classifiers based on machine learning to build a classification model. The Support Vector Machine (SVM), originally designed for binary classification operations, was adopted to solve respiratory sounds classification

problems (Palaniappan and Sundaraj, 2013; Mazić et al., 2015). k–Nearest Neighbours (k-NN) algorithm was used for the same purpose (Alsmadi and Kahya, 2008; Chen et al., 2015; Yilmaz and Kahya, 2006). Artificial Neural Networks (ANN) are widely used for prediction and pattern classification/recognition tasks. Many application fields, including respiratory sounds classification, rely on the ability of ANNs to learn complex patterns (Sezgin et al., 2001; Forkheim et al., 1995). The multilayer perceptron (MLP), which is a popular ANN, was combined with feature extraction techniques based on Fourier transform (Rietveld et al., 1999; Gouda et al., 2019), wavelet transform (Kandaswamy et al., 2004; Meng et al., 2020; Gouda et al., 2019), or cepstral analysis (Bahoura, 2009; Sengupta et al., 2016; Gouda et al., 2019) for respiratory sounds classification. The MLP belongs to the class of feedforward neural networks, where the data flow between input and output nodes is direct, while feedback links are allowed in the Recurrent Neural Networks (RNNs) topology (Nedjah et al., 2009). The long short-term memory (LSTM) neural network belongs to the RNN class and has been adapted to many application domains, where it is necessary to perform classification or recognition operations on time series such as seizure anticipation in epileptic subjects (Hu et al., 2020). LSTM is extensively used in the context of automatic speech recognition (ASR) as an acoustic model since it efficiently uses the information carried by the dependencies between successive segments of the speech signal (Trianto et al., 2018; He and Droppo, 2016; Graves et al., 2013; Chen and Huo, 2016). Also, since the memory of the LSTM provides a temporal context, it can effectively be combined with a Convolutional Neural Network (CNN) to form a hybrid model that considers the time-space context for the classification of respiratory sounds (Fraiwan et al., 2021; Acharya and Basu, 2020; Gupta et al., 2021).The adopted methodology in this paper consists in using simple neural networks for the classification of respiratory sounds. We have therefore discarded the CNNs because they are not part of the objectives set for this research project.

In addition to the MFCC method which has been already used for respiratory sounds characterization, two novel methods that have never been used before for this purpose, namely GFCC (Gammatone-Frequency Cepstral Coefficients) and BFCC (Bark-Frequency Cepstral Coefficients), are proposed and tested in this article. However, these methods are gaining in popularity in the context of automatic speech recognition (Athulya and Sathidevi, 2018; Chandra Yadav and Pradhan, 2021; Dua et al., 2018), mobile communication (Daalache et al., 2017) or emotion recognition (Chenchah and Lachiri, 2017).

Figure 1.2 illustrates the block diagram of the proposed respiratory sounds classification system into wheezing or normal classes. The input data consists of respiratory sound recordings on which we apply a feature extraction operation in order to reduce the dimensionality and keep the relevant information only. We selected three types of feature extraction techniques based on cepstral analysis and filter bank decomposition (MFCC, GFCC, and BFCC) that will be discussed, in details, in the next section.

For the classification, two neural networks were chosen to complete this task, the MLP feed-forword network and the BiLSTM (Bidirectional LSTM) recurrent network that combines two LSTM layers to process information in opposite directions. After the classifier's modeling through the training step, using the learning dataset, the trained models are used to infer the classes of the test vectors. After testing, the performance of each classifier is compared in terms of specific criteria, which are: sensitivity, specificity, performance and accuracy.

1.2 Feature extraction based on cepstral analysis

Features are extracted from the raw data (respiratory sounds) at the first stage of the classification process. This operation achieves dimensionality reduction while keeping the relevant information only. In general, feature extraction helps to improve the algorithm's speed and reduces the storage space if applicable. Moreover, it improves the signal-to-noise ratio in the projection space and, therefore, the classification performance (Guyon and Elisseeff, 2006). Our approach suggests the use of cepstral coefficients, based on auditory filter banks, as features to be extracted from respiratory sounds. The filter banks are used to decompose the signal being analyzed and help to achieve frequency distribution of multi-component non-stationary signals like the respiratory sounds. A filter bank decomposes the signal to spectral components belonging to different subbands according to each filter's pass-

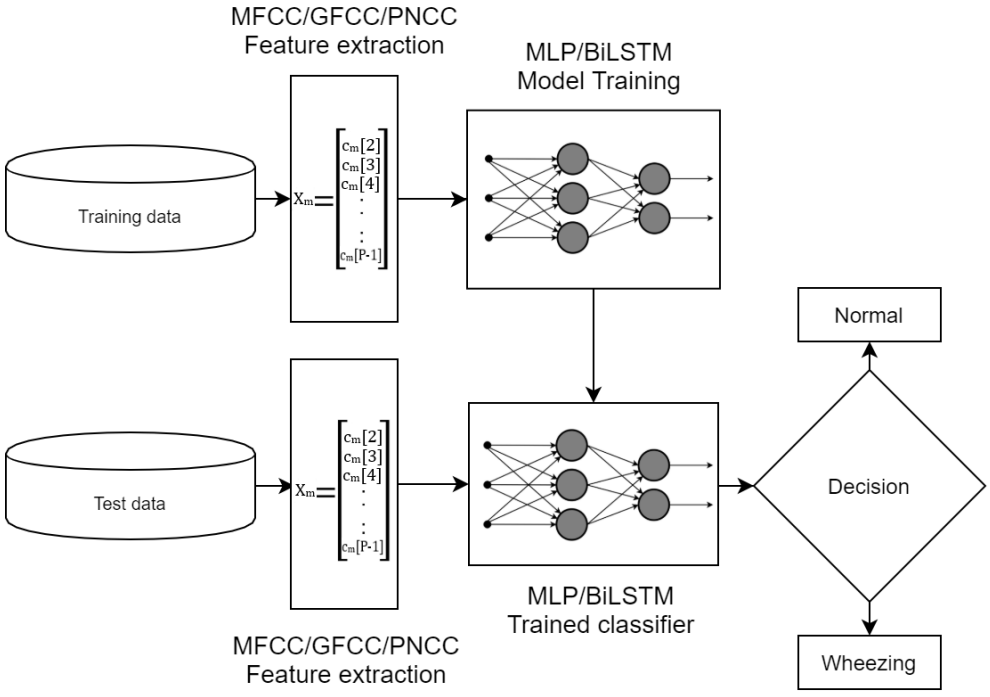


Figure 1.2 – Block diagram of the proposed classification system

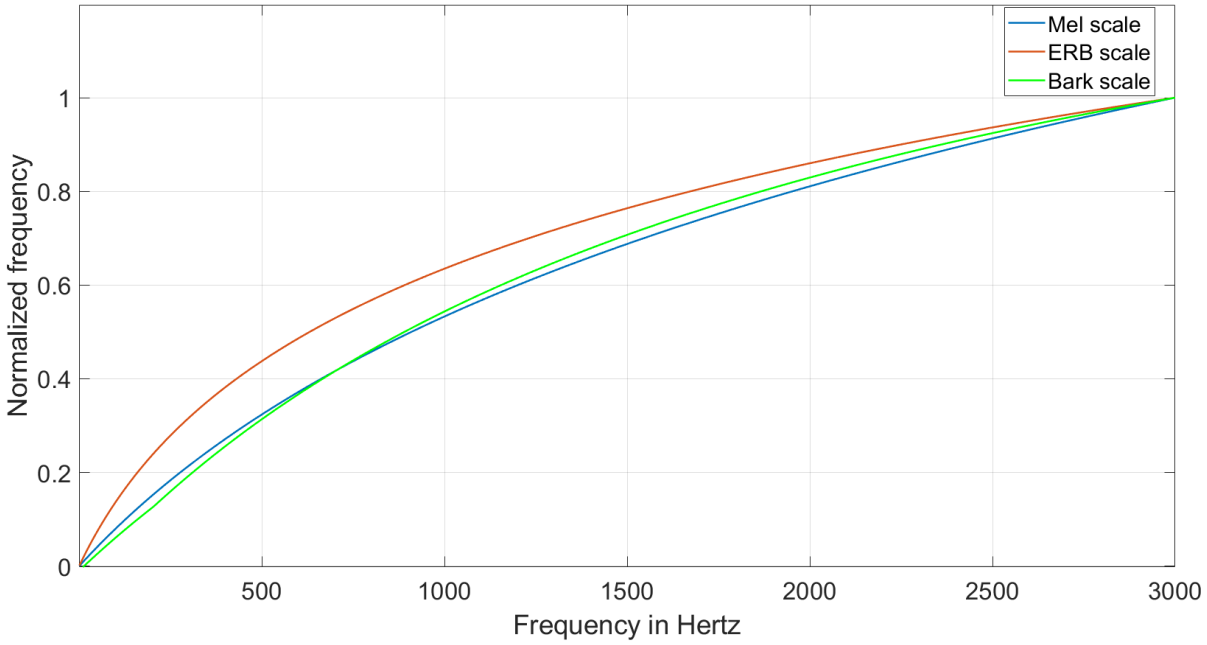
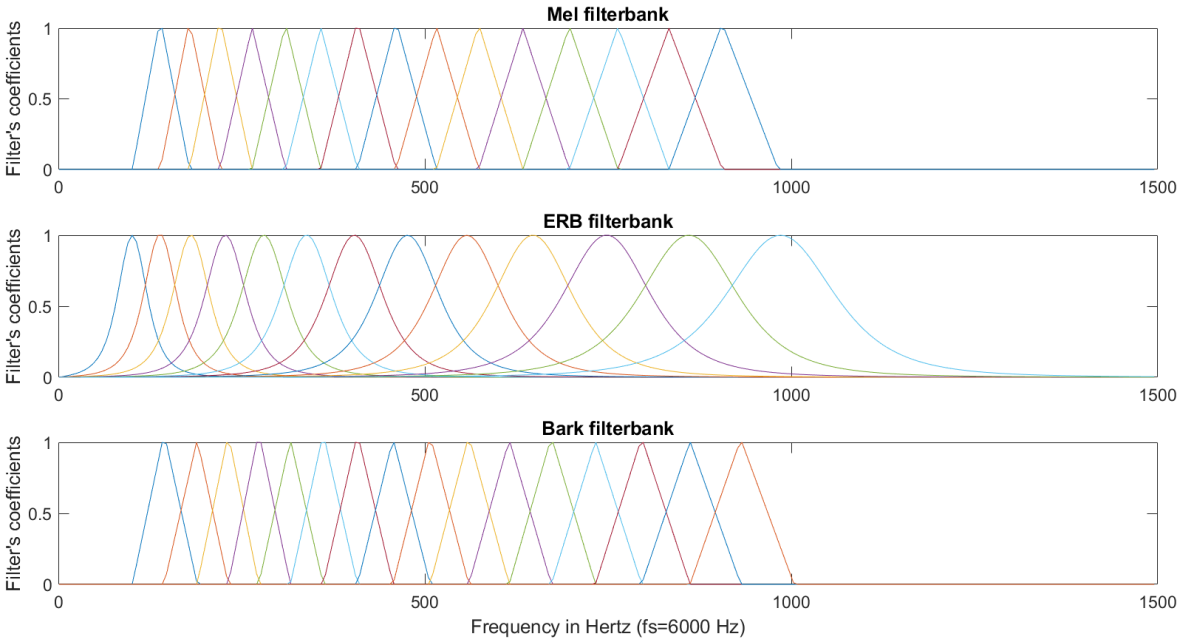


Figure 1.3 – Mapping from linear to Mel, ERB and Bark-scale.

band (Zhi-bin, 2011). Auditory filters mimic the human auditory system and are considered weighting functions with high frequency selectivity around the central frequency (Moore, 2013). Fig.1.3 shows the mapping from linear frequency to the non-linear frequency scales used in the auditory fiterbanks.

In this article, the cepstral coefficients chosen to characterize breath sounds are the MFCC, GFCC, and BFCC. The filter banks were designed to cover the range from 100 Hz to 1000 Hz as shown in Fig.1.4. These values were not chosen arbitrarily but resulted from the fact that the frequency band below 100 Hz is likely to contain energy from heart sounds as well as muscular noise. Whereas, the frequencies of wheezes with the highest amplitudes are located below 1000 Hz in the respiratory sounds collected from the trachea, which is the case for the sounds at our disposal. After setting the filtering frequency boundaries corresponding to the range from 100 Hz to 1000 Hz, we obtained 15 filters for the MFCC, 13 filters for the GFCC, and 16 filters for the BFCC.



Central Frequency (Hz)	**F_{c1}**	**F_{c2}**	**F_{c3}**	**F_{c4}**	**F_{c5}**	**F_{c6}**	**F_{c7}**	**F_{c8}**	**F_{c9}**	**F_{c10}**	**F_{c11}**	**F_{c12}**	**F_{c13}**	**F_{c14}**	**F_{c15}**	**F_{c16}**
Mel filterbank	138	177,8	219,5	263,2	308,9	356,9	407,1	459,7	514,7	572,4	632,9	696,2	762,5	832,0	904,8	-
ERB filterbank	100	137,8	179,9	226,9	279,3	337,7	402,8	475,4	556,3	646,5	747,1	859,3	984,3	-	-	-
Bark filterbank	143,3	188,5	230,8	272,4	315,5	360,4	407,0	455,6	506,2	559,0	614,1	671,6	731,8	794,8	860,8	930,0
Filter's bandwidth (Hz)	**BW_1**	**BW_2**	**BW_3**	**BW_4**	**BW_5**	**BW_6**	**BW_7**	**BW_8**	**BW_9**	**BW_{10}**	**BW_{11}**	**BW_{12}**	**BW_{13}**	**BW_{14}**	**BW_{15}**	**BW_{16}**
Mel filterbank	77,8	81,5	85,4	89,4	93,7	98,1	102,8	107,7	112,8	118,1	123,8	129,6	135,8	142,2	149,0	_
ERB filterbank	36,2	40,3	45,0	50,1	55,9	62,3	69,5	77,5	86,4	96,3	107,3	119,7	133,4	-	-	-
Bark filterbank	88,5	87,5	83,9	84,7	88,0	91,5	95,2	99,2	103,4	107,8	112,6	117,7	123,2	129,0	135,3	142,0

Figure 1.4 – The Mel, ERB and Bark-scale filter banks with filters central frequencies and bandwidths.

As shown in Fig.1.5, the algorithms of these three kinds of cepstral coefficients are the same except for the filter bank step, which will be discussed in detail for each feature extraction technique.

The cepstral coefficients are computed through several processing steps. First, the input signal $s[n]$ is divided into shorter frames that are multiplied by a weighting window to reduce the spectral artifacts caused by the amplitude discontinuities at the frame boundaries. The used smoothing window is a Hamming window of N=1024 samples length. The next

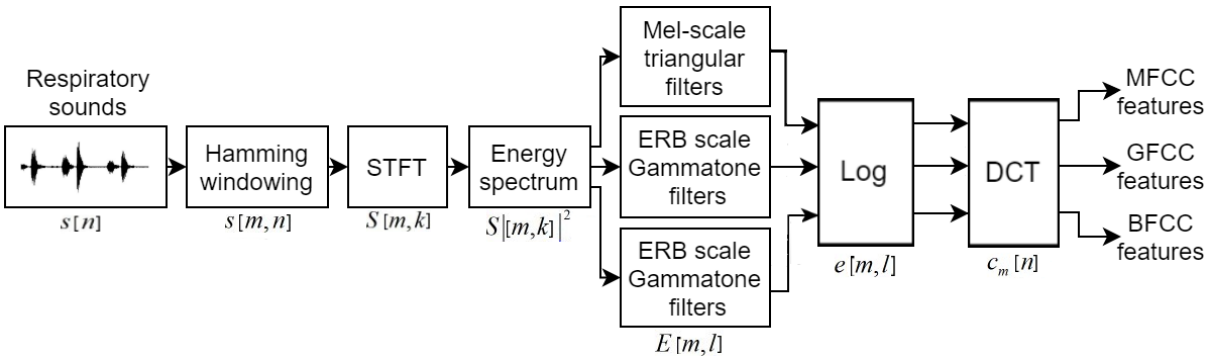


Figure 1.5 – MFCC, GFCC, and BFCC features computation algorithms

equation describes the windowing operation (Bahoura and Ezzaidi, 2013):

$$s[m,n] = s[n]w[n - mL] \tag{1.1}$$

where m is the frame index, L is the shift step (L=1024), and $s[n]$ the analyzed respiratory signal while $w[n]$ is the hamming window $w[n] = 0.54 - 0.46\cos(\frac{2\pi n}{N-1})$. After that, the STFT (Short-Time Fourier Transform) of each windowed frame is computed :

$$S[m,k] = \sum_{n=0}^{N-1} s[m,n]e^{-j2\pi nk/N} \tag{1.2}$$

where $S[m,k]$ is a complex function representing the analyzed frame in the frequency domain. k is the frequency index, and N represents also the number of the discrete frequencies. Then, the energy spectrum is computed based on the STFT outcome.

$$|S[m,k]|^2 = S_r^2[m,k] + S_i^2[m,k] \tag{1.3}$$

where $S_r[m,k]$ and $S_i[m,k]$ are the real and the imaginary parts of the spectrum $S[m,k]$, respectively. Then, the energy spectrum $|S[m,k]|^2$ is filtered by a specific filter bank (Mel-

scale triangular filters, ERB-scale Gammatone filters, or Bark-scale triangular filters) that are defined in the next subsections. The filtering operation consists of a sum operation of the energy spectrum samples weighted by the filter's coefficients.

$$E[m,l] = \sum_{k=0}^{N-1} |S[m,k]|^2 H_l[k] \tag{1.4}$$

where $H_l[k]$ is the frequency response (transfer function) of the lth filter, where $1 \leq l \leq M$, and M is the number of filters (frequency bands).

The logarithmic function is applied for magnitude compression.

$$e[m,l] = \log(E[m,l]) \tag{1.5}$$

The cepstral coefficients associated to the used filter bank are obtained by back transformation in the time domain using DCT (Discrete Cosine Transform).

$$c_m[n] = \sum_{l=1}^{M} e[m,l] \cos\left(n(l-0.5)\frac{\pi}{M}\right) \tag{1.6}$$

The extracted feature vector x_m of the frame m is given by

$$\mathrm{x}_m = [c_m[2], c_m[3], c_m[4] \ldots, c_m[16]] \tag{1.7}$$

The feature vectors, of 15 coefficients each as described in (1.7), are fed to the neural networks for classification. The size of the feature vector was chosen empirically after comparing the results obtained with feature vectors of various lengths. Each feature vector corresponds to a single frame of 170.7 ms of the signal, which approximates the wheezing minimum duration of 250 ms.

1.2.1 Mel-scale triangular filters and MFCC

It is well known that the human ear decomposes frequencies on a nonlinear scale over the audio spectrum. Therefore, MFCCs help to capture relevant information for auditory perception in a similar way a physician can distinguish normal and abnormal sounds by using his auditory perception. MFCCs are computed based on a Mel-scale filter bank that is used to filter the spectral energy in (1.4). Fig.1.3 shows the mapping from linear frequencies to Mel-scale frequencies. This relation is described as follows (O'Shaughnessy, 1999):

$$\text{Mel}(f) = 2595 \log(1 + (\frac{f}{700})) \tag{1.8}$$

The next equation shows how the triangular filters are computed (Dash et al., 2021):

$$H_l[k] = \begin{cases} 0, \ if \ k < k_c[l-1] \\ \frac{k - k_c[l-1]}{k_c[l] - k_c[l-1]}, \ if \ k_c[l-1] \leq k \leq k_c[l] \\ \frac{k_c[l+1] - k}{k_c[l+1] - k_c[l]}, \ if \ k_c[l] \leq k \leq k_c[l+1] \\ 0, \ if \ k > k_c[l+1] \end{cases} \tag{1.9}$$

where $k_c[l]$ is the frequency bin corresponding to the central frequency of the lth filter. The complete algorithm for the computation of the MFCC coefficients is described in Fig.1.5.

1.2.2 ERB-spaced gammatone filters and GFCC

In GFCC, the Mel-scale triangular filters are replaced by Equivalent Rectangular Bandwidth (ERB) gammatone filters. The impulse response of a gammatone filter centered at f_c is

given by (Qi et al., 2013):

$$g(t) = at^{n-1}e^{-2\pi bt}\cos(2\pi f_c t + \varphi) \tag{1.10}$$

with f_c the filter's central frequency, φ is the phase (generally, $\varphi = 0$). The constant a is used to control the gain, n is the order of the filter and b is a factor that determines the filter's bandwidth and depends on f_c.

ERB is a psychoacoustic measure of the filters bandwidth. It is defined by analogy with the human auditory filters located at each point of the inner ear's cochlea. Fig.1.3 shows the mapping from linear frequencies to ERB-scale frequencies. The linear frequencies are converted to the ERB-scale according to the following equation (Moore and Glasberg, 1983):

$$\mathrm{ERB}(f) = 6.23f^2 + 93.39f + 28.52 \tag{1.11}$$

The complete algorithm for the computation of the GFCC coefficients is described in Fig.1.5.

1.2.3 Bark-scale triangular filters and BFCC

BFCCs computation is based on the Bark nonlinear frequency scale. This scale has resulted from the fact that each region of the basilar membrane is excited according to a specific stimulus. After measurements, the different widths of the critical frequency bands are determined from these stimuli where each critical band corresponds to 1 Bark. Each critical band corresponds to one filter of the filter bank and is characterized by its central frequency and bandwidth. Fig.1.3 shows the mapping from linear frequencies to Bark-scale frequencies. The following equation explains the relationship between these two frequency

scales (Traunmüller, 1990):

$$\text{Bark}(f) = 13\arctan(0.00076f) + 3.5\arctan(\frac{f}{7500})^2 \tag{1.12}$$

BFCCs computation is based on a Bark scale triangular filter bank. The triangular filters are computed similarly as in (2.7). Even though the filters shape is the same as for the MFCCs, these filters don't have the same characteristics (central frequency and bandwidth) as shown in Fig.1.4. As for the cepstral coefficients mentioned previously, the Bark scale triangular filter bank is used to filter the spectral energy in (1.4). The complete algorithm for the computation of the MFCC coefficients is described in Fig.1.5.

1.3 Sound classifiers based on neural networks

As indicated previously, two neural networks, the MLP and the BiLSTM, are used to perform respiratory sound classification into normal and wheezing classes. Each of these two neural networks belongs to a distinct class: feedforward or recurrent, respectively. The MLP, an established technique, will serve as a benchmark to assess the BiLSTM's classification performance. Such comparison will determine the impact of the inter-frame temporal dependencies on the wheezing detection.

1.3.1 Multilayer Perceptron (MLP) network

The MLP (Multilayer Perceptron) is a popular neural network that is widely used for modeling and classification tasks. As shown in Fig 1.6, it has a layered topology which consists of an input layer, one or more hidden layers, and an output layer. Just like the brain, the

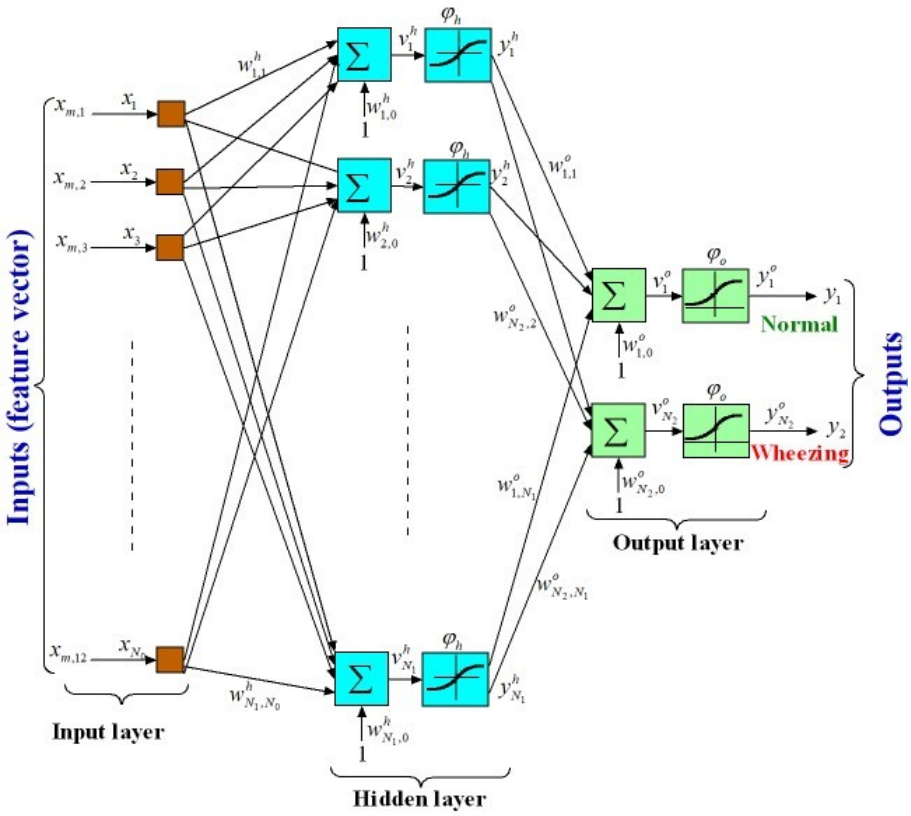


Figure 1.6 – Topology of a typical MLP neural network (Bahoura, 2018)

information is processed by artificial neurons that compose the MLP's hidden and output layers. The MLP can learn to recognize patterns by complex nonlinear regressions by adjusting its synaptic weights using a learning algorithm such as Back-Propagation (BP) which is based on error-correction learning (Haykin, 1999). A training phase needs to be performed by the MLP to learn to correctly classify respiratory sounds into normal and wheezing classes.

BP algorithm requires two passes of computations, a forward pass and a backward pass. In the forward pass the MLP's synaptic weights are fixed and the input signal propagates from the input layer to the output layer and the computations are performed layer-by-layer. The next equations show how an output of a neuron located in the hidden layer of the MLP is

computed (Haykin, 1999):

$$v_j^h = \sum_{i=1}^{N_0} w_{j,i}^h x_i + w_{j,0}^h \qquad j = 1, \ldots, N_1 \tag{1.13}$$

$$y_j^h = \varphi_h \left[v_j^h \right] \tag{1.14}$$

$w_{j,i}^h$ is the corresponding synaptic weight and $w_{j,0}^h$ represents the bias. For the output layer's neurons:

$$v_k^o = \sum_{j=1}^{N_1} w_{k,j}^o x_j^h + w_{k,0}^o \qquad k = 1, \ldots, N_2 \tag{1.15}$$

$$y_k^o = \varphi_o \left[v_k^o \right] \tag{1.16}$$

Here, the hidden neurons use hyperbolic tangent activation $\varphi_h[v] = (1 - e^{-2v})/(1 + e^{-2v})$, whereas output neurons use logistic-sigmoid function $\varphi_o[v] = 1/(1 + e^{-v})$. The use of the hyperbolic tangent is appropriate for the hidden layer because it's a bipolar function, i.e., it produces output values in the range of $[-1, 1]$ while the logistic sigmoid is more appropriate for the output layer given that, in our case, the MLP network performs a classification task. The synaptic weights are computed during the training phase using a set of learning data (inputs for which outputs are known). The learning operation is performed by the BP algorithm (Bahoura, 2009). In this case, we tested the MLP network with 10, 15, 20, and 25 neurons in the hidden layer.

1.3.2 Bidirectional Long Short-Term Memory (BiLSTM) network

In any feed-forward network, past outputs are not involved in making predictions about the present. In the pattern recognition context, we may have as a task to model time series, e.g. speech recognition. The MLP (Multilayer Perceptron) archetype inspired RNN (Recurrent

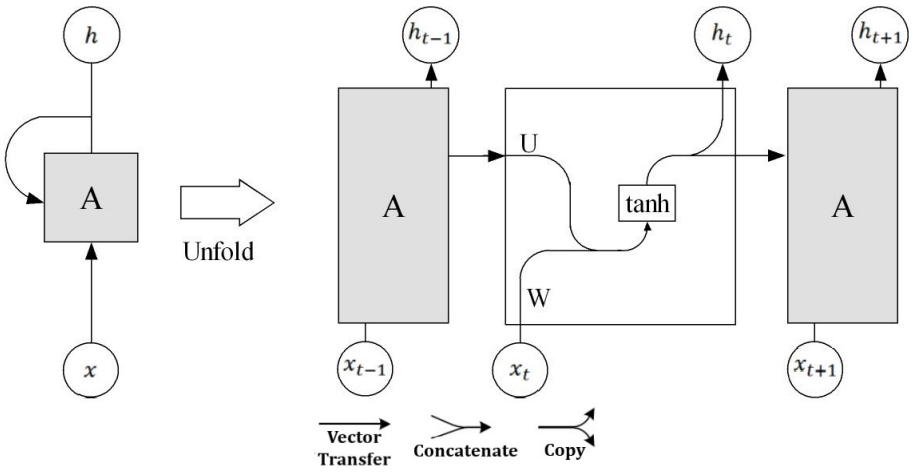


Figure 1.7 – Basic diagram of RNN cell unfolded through time.

Neural Networks), except that RNN involves time-lagged recurrent connections (loop-back) that help the model save relevant information about the past to spot similarities timewise (Pascanu et al., 2013). Fig. 1.7 shows an example of a basic RNN structure.

The hidden state vector h_t, at the time t, is given by:

$$h_t = \tanh\left[Wx_t + Uh_{t-1} + b\right] \tag{1.17}$$

where x_t is the input vector at time t, h_{t-1} is the hidden vector at time $(t-1)$, W is the input connection weights matrix, and U is the recurrent connection weights matrix.

Even though, at first glance, recurrent networks may seem like a powerful, straightforward alternative to the feed-forward networks, the reality is quite different. The major problems encountered when training a recurrent network are the vanishing or exploding gradient due to its inefficiency to handle long-term dependencies (Bengio et al., 1994). Long-term dependencies refer to predicting the current output depending on an input that existed in the remote past. The RNN can be challenging to train since the BP algorithm does not perform

well at discovering dependencies that span long temporal gaps (Mozer, 1991). This inability of the BP algorithm is the vanishing gradient phenomenon (Graves et al., 2009). Moreover, the exploding gradient may be as compromising to the learning as the vanishing gradient. The exploding gradient designates the situation where the norm of the long-term dependencies-related gradient increases exponentially, faster than the short-term ones (Pascanu et al., 2013).

To overcome the limitations of RNNs to deal with long-term dependencies, the LSTM (Long Short-Term Memory) recurrent neural network has been introduced as a gradient-based method assumed to maintain a constant error flow by preventing vanishing or exploding gradient. The novelty in LSTM, as presented by (Hochreiter and Schmidhuber, 1997), is a new element called CEC (Constant Error Carrousel), which consists of a linear and self-connected unit included, as its name suggests, in order to keep the error flow steady. However, the CEC is connected to other units to set up an architecture that meets the constant error flow needs. An input gate is added to protect the content stored in memory from perturbations caused by irrelevant inputs. Likewise, an output gate is implemented to protect other memory units from being spoiled by potentially irrelevant outputs of the current memory. The combination of the three elements (CEC, input gate, and output gate) is called a memory cell. The cell state should be capable of being reset by itself as the starting point of the new sequence is not explicitly indicated. The cell should also prevent LSTM from storing memory for arbitrary timespans and considering new events when they occur. A new type of gates labeled as forget gates has been introduced to circumvent this problem (Gers et al., 2000). The forget gates substitute the CEC's constant weight of 1 by a multiplicative activation unit. Fig. 1.8 shows an example of an LSTM cell with its respective gates. The typical layout of a network based on LSTM cells is similar to any conventional neural network layered topology, consisting of an input layer, a hidden layer, and an output layer. The hidden layer is fully self-connected and contains LSTM memory cells.

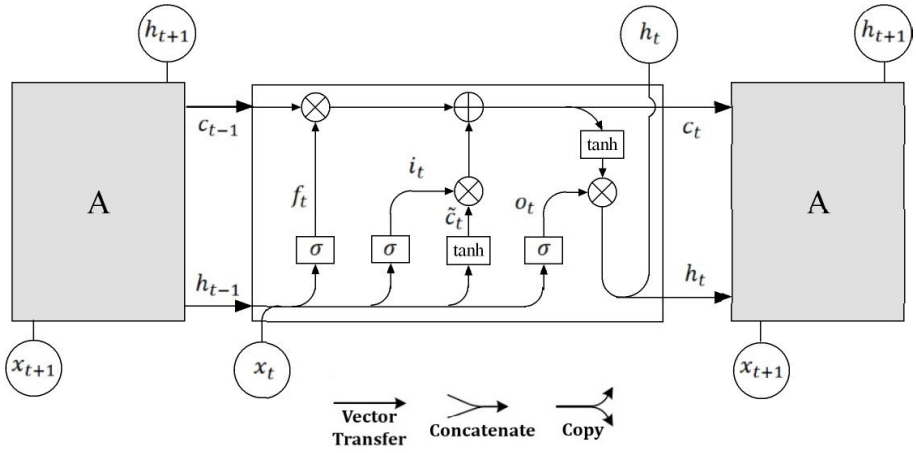


Figure 1.8 – Basic diagram of an LSTM cell unfolded through time.

The equations below describe the forward propagation and are presented separately for each element (Zhang et al., 2020; Apaydin et al., 2020; Ameur et al., 2020):

$$f_t = \sigma[W_f x_t + U_f h_{t-1} + b_f] \tag{1.18}$$

where f_t is the forget gate, σ represents the sigmoid activation function, x_t is the current input vector, h_{t-1} is the hidden output vector, b_f is the bias of the forget gate, while W_f and U_f are their input and recurrent weights, respectively.

$$i_t = \sigma[W_i x_t + U_i h_{t-1} + b_i] \tag{1.19}$$

where i_t is the input gate, b_i is its bias, while W_i and U_i are their input and recurrent weights, respectively.

$$o_t = \sigma[W_o x_t + U_o h_{t-1} + b_o] \tag{1.20}$$

where o_t is the output gate, b_o is its bias, while W_o and U_o are their input and recurrent

weights, respectively.

$$\tilde{c}_t = \tanh[W_c x_t + U_c h_{t-1} + b_c] \tag{1.21}$$

where $\tilde{c}_t$ is the candidate state, b_c is its bias, while W_c and U_c are their input and recurrent weights, respectively. *tanh* is the hyperbolic tangent function.

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \tag{1.22}$$

where c_t and c_{t-1} are the current and the previous memory states.

$$h_t = o_t \tanh [c_t] \tag{1.23}$$

where h_t is the current hidden layer output of the cell.

LSTM has gained much popularity in the last few years. It has been adapted to many application fields where it is necessary to perform classification or recognition operations involving sequential data (e.g., respiratory sounds) since it effectively uses the information conveyed by the dependencies between successive respiratory sound segments (Perna and Tagarelli, 2019; Sreeram et al., 2020). BiLSTM neural network is a variant of LSTM that combine two LSTM networks, processing information in opposite directions, forward (from the past) and backward (from the future). Fig. 1.9 represents a neural network containing a BiLSTM layer. BiLSTM is also extensively used in this field as it is likely to give better results because the contextual information is captured in both directions. Since BiLSTM processes data in both directions, this will cause latency issues for real-time applications. However, in this case, there's no real-time constraint because data are processed off-line. In this study, we tested a BiLSTM network with 10, 15, 20, and 25 hidden units.

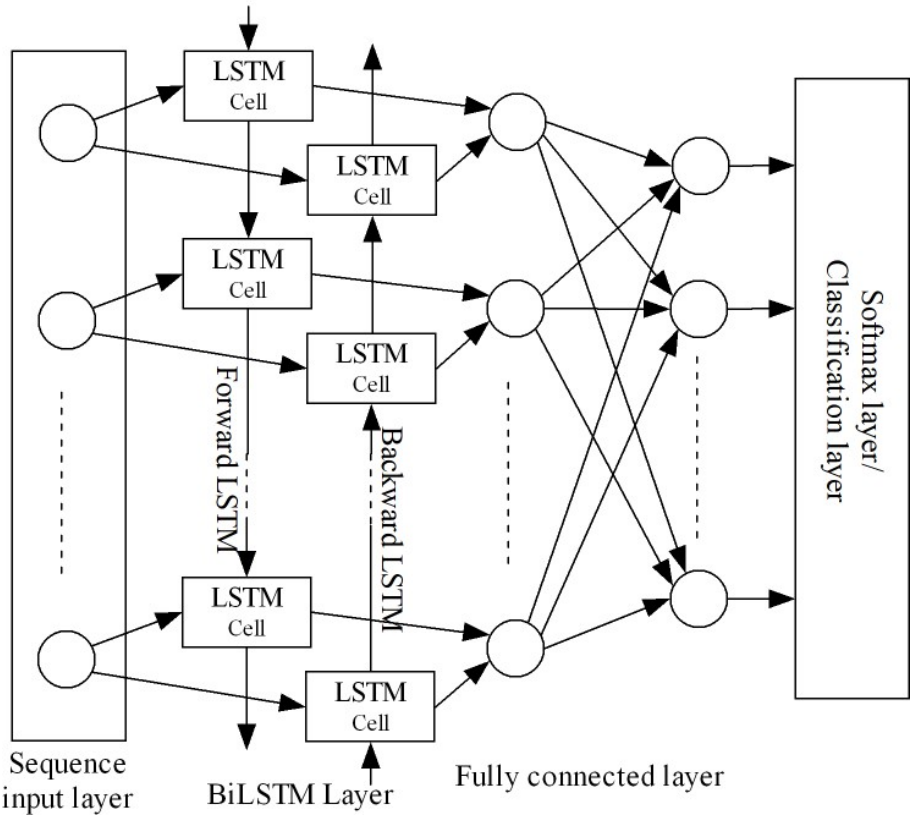


Figure 1.9 – Structure of the proposed BiLSTM-based network.

1.3.3 Class recognition

The assignment of a respiratory sound segment to a specific class (normal or wheezing) depends on the output values of the two neurons of the output layer of the MLP or the BiLSTM neural networks. We define a score function as the difference between the two neural network outputs, this function is given by:

$$\mathrm{Sc}[m] = y_2^o - y_1^o \tag{1.24}$$

The outcome of this function is compared to a threshold θ that is used to optimize the

classification. The classification decision is made with respect to this comparison as shown in the following equation:

$$\mathrm{s}[m,n] = \begin{cases} \text{wheezing} & if\ \mathrm{Sc}[m] \geq \theta \\ \text{normal} & otherwise \end{cases} \tag{1.25}$$

In our case, $\theta = 0$ (default value).

1.4 Results and discussion

1.4.1 Database and protocol

The database used in this research contains respiratory sound recordings collected from RALE database-CD, ASTRA database-CD, and other websites (Bahoura, 2009). The recordings were made using a microphone placed at the trachea. It consists of 12 pulmonary sounds recorded from healthy subjects and considered normal respiratory sounds, and 12 sounds recorded from people with asthma and identified as wheezes. The sounds were downsampled at 6 kHz at the preprocessing stage, and each sample is 16 bits wide. Wheeze recordings were manually cleaned up to keep wheeze segments only. Each recording was segmented into frames of 1024 samples, resulting in a database consisting of 915 normal sound segments and 497 wheezing sound segments of 170.7 ms each. The number of segments in each recording and its length in seconds are presented in Table 1.1. The segments length of 170.7 ms is appropriate for wheezing detection considering that it approximates the wheezing minimal duration of 250 ms.

The tests were carried out on MATLAB R2021a using the Leave-One-Out Cross-

Validation (LOOCV) procedure over the respiratory sound recordings set to ensure a reliable and unbiased estimate of the performance. Since the connection weights of the neural networks are initialized randomly, each learning/testing combination of the LOOCV procedure is repeated ten times for statistical reliability.

Table 1.1 – The database used in this research

Normal respiratory sounds			Wheezing respiratory sounds		
Filename	Duration (s)	Number of segments	Filename	Duration (s)	Number of segments
N01	15.8	92	W01	6.7	39
N02	17.1	100	W02	4.6	27
N03	32.4	189	W03	9.6	56
N04	10.1	59	W04	2.8	16
N05	16.8	98	W05	2.7	15
N06	17.8	104	W06	17.5	102
N07	7.7	44	W07	4.2	24
N08	8.2	48	W08	12.4	72
N09	6.8	40	W09	6.2	36
N010	7.2	42	W10	8.1	47
N011	9.2	53	W11	4.1	24
N012	7.9	46	W12	6.7	39
Total Nomal	157.0	915	Total Wheezes	85.6	497

1.4.2 Evaluation criteria

The assessment criteria according to which we evaluate our model are: sensitivity, specificity, performance and accuracy defined as follows :

$$\text{Sensitivity (SE)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1.26}$$

$$\text{Specificity (SP)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{1.27}$$

$$\text{Performance (PER)} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \tag{1.28}$$

$$\text{Accuracy (ACC)} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \tag{1.29}$$

TP, FN, TN, FP are the number of true positives, false negatives, true negatives, and false positives, respectively.

1.4.3 Classification results

The classification results are discussed in terms of the performance criteria defined in section 1.4.2 and whose average values are shown in Tables 1.2-1.4. Each table details the classification performance achieved with each feature extraction method in order to assess their impact on classification. The performance of the MLP and BiLSTM networks is compared according to each feature extraction method and for a different number of hidden units. Overall, we can see that BiLSTM outperforms the MLP in terms of the achieved sensitivity, specificity, performance, and accuracy irrespective of the feature extraction method used. As shown in Fig.1.10, the highest sensitivity value was scored with the MFCC-BiLSTM classification model (SE=0.9252) using 25 hidden BiLSTM units. In fact, the BiLSTM-based classifiers achieve high values of sensitivity compared to the MLP-based classifiers. The gap in terms of sensitivity can reach up to 20%. This gap is explained by the fact that the BiLSTM takes advantage of the temporal dependencies, unlike the MLP, which results in a better detection of wheezing. However, the MLP best sensitivity score was achieved with GFCC features using 25 hidden neurons (SE=0.778).

Fig.1.10 shows that the BiLSTM achieved a perfect score in terms of specificity (SP=1) with both MFCC and BFCC features, significantly surpassing the MLP-based classifiers

Table 1.2 – Average classification results obtained with both BiLSTM and MLP using MFCC features.

Neural network	MLP				BiLSTM			
Hidden units	10	15	20	25	10	15	20	25
Accuracy	0,837	0,836	0,839	0,848	0,961	0,951	0,960	0,963
Sensitivity	0,701	0,707	0,704	0,721	0,924	0,904	0,919	0,925
Specificity	0,911	0,906	0,912	0,917	0,998	0,998	1,000	1,000
Performance	0,799	0,800	0,801	0,813	0,960	0,950	0,959	0,962

Table 1.3 – Average classification results obtained with both BiLSTM and MLP using GFCC features.

Neural network	MLP				BiLSTM			
Hidden units	10	15	20	25	10	15	20	25
Accuracy	0,844	0,858	0,859	0,862	0,922	0,942	0,943	0,947
Sensitivity	0,744	0,776	0,768	0,778	0,870	0,900	0,898	0,913
Specificity	0,898	0,902	0,909	0,907	0,974	0,985	0,988	0,981
Performance	0,817	0,837	0,836	0,840	0,920	0,941	0,942	0,946

Table 1.4 – Average classification results obtained with both BiLSTM and MLP using BFCC features.

Neural network	MLP				BiLSTM			
Hidden units	10	15	20	25	10	15	20	25
Accuracy	0,834	0,842	0,852	0,858	0,949	0,945	0,951	0,945
Sensitivity	0,721	0,736	0,737	0,747	0,905	0,891	0,903	0,890
Specificity	0,896	0,899	0,915	0,918	0,993	0,999	0,999	1,000
Performance	0,804	0,813	0,821	0,828	0,948	0,944	0,950	0,943

scores (up to 10% higher). A perfect specificity score means that all the normal respiratory sounds segments were classified as such. The MLP best score in terms of specificity was achieved with BFCC features (SP=0.918). Also, as we can see on Fig.1.10, the classification performance and accuracy achieved with the BiLSTM based classifiers are significantly higher than the performance and accuracy of the MLP-based ones. The highest performance and accuracy (ACC=0.963 and PER=0.962) were achieved using MFCC features fed to a BiLSTM neural network with 25 hidden units. The MLP's highest performance and accuracy were achieved using GFCC features and 25 hidden neurons (ACC=0.862 and PER=0.840).

With respect to the classifier used, the three characterization methods produce the same classification results to a certain extent, yet with a slight advantage of the MFCC features. In light of the conclusive results achieved, it is worth mentioning here that the GFCC and BFCC characterization techniques have never been applied to breath sounds before to the best of our knowledge.

1.5 Conclusion

In this article, three methods of cepstral coefficients extraction were used to characterize respiratory sounds for wheezing detection. The cepstral features used are the MFCC, GFCC, and BFCC. The MFCCs are well known in this application area, while the two other methods are applied for the first time to characterize respiratory sounds. These features were fed to neural network-based classifiers, namely the MLP and the BiLSTM. The classification results achieved with each classification model were assessed according to specific performance criteria and then compared with each other. The highest wheezing detection accuracy of 96.3% was achieved with an MFCC-BiLSTM combination, and a perfect specificity (100%) was also reached with the same model. Overall, the BiLSTM neural network performs better

than MLP in detecting wheezes in respiratory sounds, irrespective of the feature extraction method used. The BiLSTM, as a recurrent neural network, captures temporal characteristics of respiratory sounds resulting in higher wheezing detection accuracy. Based on the achieved

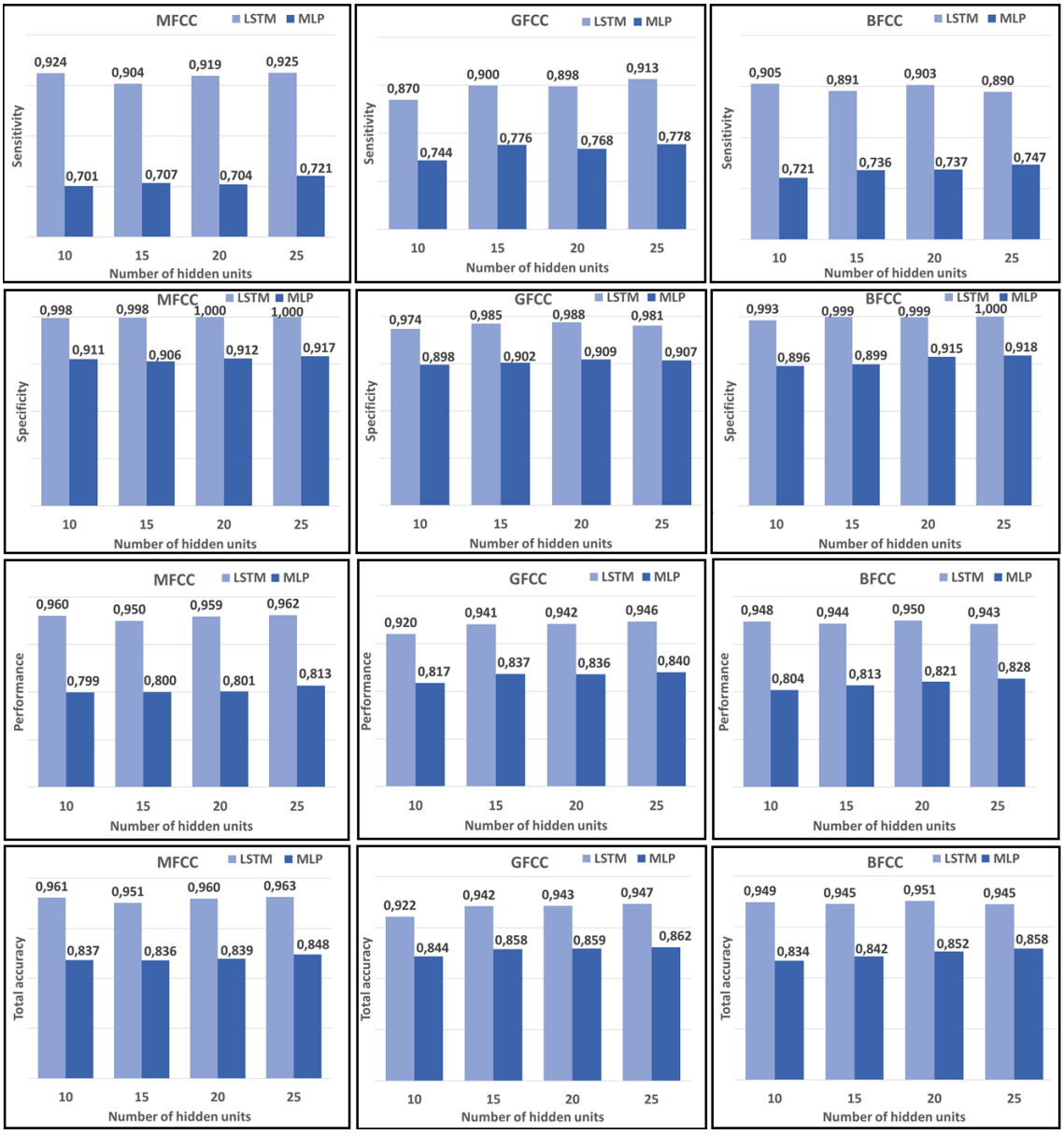


Figure 1.10 – Sensitivity, specificity, performance, and total accuracy achieved by the classification models.

results, the cepstral coefficients provide a good characterization of the respiratory sounds with 15 coefficient only.

In future works, we will extend this study to cover other classes of adventitious respiratory sounds and we will also work towards a hardware implementation for real-time processing.

Acknowledgement

This research is financially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Références

Acharya, J., Basu, A., 2020. Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. IEEE Transactions on Biomedical Circuits and Systems 14, 535–544.

Alsmadi, S., Kahya, Y.P., 2008. Design of a dsp-based instrument for real-time classification of pulmonary sounds. Comput Biol Med 38, 53–61.

Ameur, S., Ben Khalifa, A., Bouhlel, M., 2020. A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with leap motion. Entertainment Computing 35, 100373.

Apaydin, H., Feizi, H., Sattari, M.T., Colak, M.S., Shamshirband, S., Chau, K.W., 2020. Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. Water 12, 1500.

Athulya, M.S., Sathidevi, P.S., 2018. Speaker verification from codec distorted speech for forensic investigation through serial combination of classifiers. Digital Investigation 25, 70–77.

Bahoura, M., 2009. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Computers in Biology and Medicine 39, 824–843.

Bahoura, M., 2018. FPGA implementation of an automatic wheezing detection system. Biomedical Signal Processing and Control 46, 76–85.

Bahoura, M., Ezzaidi, H., 2013. Hardware implementation of MFCC feature extraction for respiratory sounds analysis. 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA) , 226–229.

Bahoura, M., Pelletier, C., 2004. Respiratory sounds classification using cepstral analysis and gaussian mixture models. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 1, 9–12.

Basu, V., Rana, S., 2020. Respiratory diseases recognition through respiratory sound with the help of deep neural network, in: 2020 4th International Conference on Computational Intelligence and Networks (CINE), pp. 1–6.

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks 5, 157–166.

Chandra Yadav, I., Pradhan, G., 2021. Pitch and noise normalized acoustic feature for children's ASR. Digital Signal Processing 109, 1128–1132.

Chen, C.H., Huang, W.T., Tan, T.H., Chang, C.C., Chang, Y.J., 2015. Using k-nearest neighbor classification to diagnose abnormal lung sounds. Sensors 15, 13132–13158.

Chen, K., Huo, Q., 2016. Training deep bidirectional LSTM acoustic model for LVCSR by a context-sensitive-chunk BPTT approach. IEEE/ACM Transactions on Audio, Speech, and Language Processing 24, 1185–1193.

Chenchah, F., Lachiri, Z., 2017. A bio-inspired emotion recognition system under real-life conditions. Applied Acoustics 115, 6–14.

Daalache, M.R.L., Addou, D., Boudraa, M., 2017. An efficient distributed speech processing in noisy mobile communications, in: 2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), pp. 1–4.

Dash, T.K., Mishra, S., Panda, G., Satapathy, S.C., 2021. Detection of COVID-19 from speech signal using bio-inspired based cepstral features. Pattern Recognition 117, 107999.

Dua, M., Aggarwal, R.K., Biswas, M., 2018. Performance evaluation of hindi speech recognition system using optimized filterbanks. Engineering Science and Technology, an International Journal 21, 389–398.

Fiz, J.A., Jané, R., Homs, A., Izquierdo, J., García, M.A., Morera, J., 2002. Detection of wheezing during maximal forced exhalation in patients with obstructed airways. Chest 122, 186–191.

Forkheim, K.E., Scuse, D., Pasterkamp, H., 1995. A comparison of neural network models for wheeze detection, in: IEEE WESCANEX 95. Communications, Power, and Computing. Conference Proceedings, pp. 214–219.

Fraiwan, M., Fraiwan, L., Alkhodari, M., Hassanin, O., 2021. Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. Journal of Ambient Intelligence and Humanized Computing 4, 1–13.

Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: continual prediction with lstm. Neural Comput 12, 2451–71.

Gong, H.J., 1990. Wheezing and asthma, in: Walker, H.K., Hall, W.D., Hurst, J.W. (Eds.), Clinical Methods: The History, Physical, and Laboratory Examinations. © 1990, Butterworth Publishers, a division of Reed Publishing., Boston, pp. 100–239.

Gouda, A., El Shehaby, S., Diaa, N., Abougabal, M., 2019. Classification techniques for diagnosing respiratory sounds in infants and children, in: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0354–0360.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J., 2009. A novel connectionist system for unconstrained handwriting recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 855–868.

Graves, A., Mohamed, A., Hinton, G., 2013. Speech recognition with deep recurrent neural networks, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6645–6649.

Gupta, S., Agrawal, M., Deepak, D., 2021. Gammatonegram based triple classification of lung sounds using deep convolutional neural network with transfer learning. Biomedical Signal Processing and Control 70, 102947.

Guyon, I., Elisseeff, A., 2006. An introduction to feature extraction, in: Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A. (Eds.), Feature Extraction. Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, Berlin, Heidelberg. volume 207, pp. 1–25.

Haykin, S., 1999. Multilayer perceptrons, in: Neural Networks: A Comprehensive Foundation. Prentice Hall, pp. 122–221.

He, T., Droppo, J., 2016. Exploiting LSTM structure in deep neural networks for speech recognition, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5445–5449.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780.

Hu, X., Yuan, S., Xu, F., Leng, Y., Yuan, K., Yuan, Q., 2020. Scalp EEG classification using deep Bi-LSTM network for seizure detection. Computers in Biology and Medicine 124, 103919.

Jácome, C., Marques, A., 2015. Computerized respiratory sounds in patients with copd: a systematic review. COPD 12, 104–12.

Kahya, Y.P., Yeginer, M., Bilgic, B., 2006. Classifying respiratory sounds with different feature sets, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2856–2859.

Kandaswamy, A., Kumar, C.S., Ramanathan, R.P., Jayaraman, S., Malmurugan, N., 2004. Neural classification of lung sounds using wavelet coefficients. Comput Biol Med 34, 523–37.

Lin, B.S., Lin, B.S., 2016. Automatic wheezing detection using speech recognition technique. Journal of Medical and Biological Engineering 36, 545–554.

Mazić, I., Bonković, M., Džaja, B., 2015. Two-level coarse-to-fine classification algorithm for asthma wheezing recognition in childrenś respiratory sounds. Biomedical Signal Processing and Control 21, 105–118.

Meng, F., Shi, Y., Wang, N., Cai, M., Luo, Z., 2020. Detection of respiratory sounds based on wavelet coefficients and machine learning. IEEE Access 8, 155710–155720.

Moore, B., 2013. Frequency selectivity, masking and the critical band, in: An Introduction to the Psychology of Hearing: Sixth Edition. Brill, Leiden, The Netherlands, pp. 19–234.

Moore, B.C., Glasberg, B.R., 1983. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am 74, 750–3.

Mozer, M.C., 1991. Induction of multiscale temporal structure, in: Proceedings of the 4th International Conference on Neural Information Processing Systems, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. p. 275–282.

Nedjah, N., da Silva, R., Mourelle, L., da Silva, M., 2009. Dynamic MAC-based architecture of artificial neural networks suitable for hardware implementation on FPGAs. Neurocomputing 72, 2171–2179.

O'Shaughnessy, D., 1999. Hearing, in: Speech Communications: Human and Machine, 2nd Edition. Wiley-IEEE Press, Piscataway, NJ, pp. 109–138.

Palaniappan, R., Sundaraj, K., 2013. Respiratory sound classification using cepstral features and support vector machine, in: 2013 IEEE Recent Advances in Intelligent Computational Systems RAICS, pp. 132–136.

Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks, in: Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, JMLR.org. p. III–1310–III–1318.

Pasterkamp, H., 2017. The highs and lows of wheezing: A review of the most popular adventitious lung sound. Pediatric Pulmonology 53, 243–254.

Pasterkamp, H., Tal, A., Leahy, F., Fenton, R., Chernick, V., 1985. The effect of anticholinergic treatment on postexertional wheezing in asthma studied by phonopneumography and spirometry. Am Rev Respir Dis 132, 16–21.

Perna, D., Tagarelli, A., 2019. Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks, in: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), pp. 50–55.

Qi, J., Wang, D., Jiang, Y., Liu, R., 2013. Auditory features based on gammatone filters for robust speech recognition, in: 2013 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 305–308.

Riella, R., Nohama, P., Maia, J., 2009. Method for automatic detection of wheezing in lung sounds. Brazilian Journal of Medical and Biological Research 42, 674–684.

Rietveld, S., Oud, M., Dooijes, E.H., 1999. Classification of asthmatic breath sounds: Preliminary results of the classifying capacity of human examiners versus artificial neural networks. Computers and Biomedical Research 32, 440–448.

Sengupta, N., Sahidullah, M., Saha, G., 2016. Lung sound classification using cepstral-based statistical features. Comput Biol Med 75, 118–29.

Sezgin, M.C., Dokur, Z., Olmez, T., Korurek, M., 2001. Classification of respiratory sounds by using an artificial neural network, in: 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 697–699.

Sánchez Morillo, D., Astorga Moreno, S., Fernández Granero, M., León Jiménez, A., 2013. Computerized analysis of respiratory sounds during copd exacerbations. Comput Biol Med 43, 914–21.

Sreeram, A.S.K., Ravishankar, U., Sripada, N.R., Mamidgi, B., 2020. Investigating the potential of mfcc features in classifying respiratory diseases, in: 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), pp. 1–7.

Taplidou, S.A., Hadjileontiadis, L.J., Kitsas, I.K., Panoulas, K.I., Penzel, T., Gross, V., Panas, S.M., 2004. On applying continuous wavelet transform in wheeze analysis. Conf Proc IEEE Eng Med Biol Soc 2004, 3832–5.

Traunmüller, H., 1990. Analytical expressions for the tonotopic sensory scale. The Journal of the Acoustical Society of America 88, 97–100.

Trianto, R., Tai, T., Wang, J., 2018. Fast-LSTM acoustic model for distant speech recognition, in: 2018 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–4.

Yilmaz, C.A., Kahya, Y.P., 2006. Multi-channel classification of respiratory sounds, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2864–2867.

Zhang, X., Zhao, M., Dong, R., 2020. Time-series prediction of environmental noise for urban iot based on long short-term memory recurrent neural network. Applied Sciences 10, 1144.

Zhi-bin, G., 2011. Time frequency analysis of multi-component non-stationary signal with filter bank decomposition, in: 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), pp. 2035–2038.

ARTICLE 2

SCALABLE SERIAL HARDWARE ARCHITECTURE OF MULTILAYER PERCEPTRON NETWORK FOR AUTOMATIC WHEEZING DETECTION

Résumé en français du deuxième article

Dans cet article, on propose une nouvelle architecture série pour l'implantation du perceptron mlticouches (MLP) dans le cadre d'une application pour la détection en temps réel des sibilants dans les sons respiratoires. Étant donné la capacité du réseau MLP à détecter des formes complexes présentes dans ces sons, ce réseau de neurones s'est imposé comme outil de classification pour de telles applications. La particularité de l'architecture proposée réside dans le fait que celle-ci utilise une seule unité de calcul et cela indépendamment de la taille du réseau (le nombre de neurones dans chaque couche). Également, l'architecture proposée est parfaitement modulable à savoir qu'elle peut s'adapter à n'importe quelle topologie du MLP à implanter sans avoir à modifier le schéma ou le câblage. L'architecture proposée a été implantée sur une puce FPGA à coût modéré et à faible consommation d'énergie à l'aide d'un langage de programmation à haut niveau d'abstraction. Les résultats de classification obtenus à l'aide de cette implantation sont évalués en termes de la sensibilité, la spécificité, la performance et l'exactitude de classification. Pour valider notre implantation, ces résultats ont été comparés à ceux donnés par l'architecture parallèle. Effectivement le réseau MLP sérialisé donne les mêmes résultats que le réseau MLP parallèle avec, néanmoins, l'avantage d'utiliser moins de ressources.

Cet article fut corédigé par le candidat et son directeur de recherche, le Professeur Mo-

hammed Bahoura. Le directeur de recherche a proposé l'idée qui a donné naissance à cet article et a encadré le candidat dans ces travaux allant de la recherche bibliographique à la rédaction en passant par l'élaboration du modèle et les tests effectués. Cet article a été soumis à un journal (IEEE Access) et dont une version abrégée intitulée "Serial Hardware Architecture Of Multilayer Neural Network for Automatique Wheezing Detection" a été acceptée et présentée dans une conférence internationale (MWSCAS 2021) et a été publié dans IEEE Xplore.

Mots clés: réseau de neurones artificiels, Perceptron multicouches, MLP, Analyse cepstrale, MFCC, Sons respiratoires, Détection des sibilants, FPGA.

Abstract

This paper proposes a serial hardware architecture of a multilayer perceptron (MLP) neural network for real-time wheezing detection in respiratory sounds. As an established classification tool, the MLP has proven its ability to identify complex patterns within respiratory sounds. The proposed fully serial architecture uses a single calculation unit, independently of the number of neurons in the MLP network. It is also a fully scalable architecture that permits to implement MLP networks, of any size, easily and efficiently without modifying the design or the wiring. The proposed serial architecture has been implemented on a low-cost and power-efficient field programmable gate array (FPGA) chip using a high-level programming tool. The respiratory sounds classification rates are evaluated in terms of sensitivity, specificity, performance, and accuracy. The proposed serial architecture reaches the same classification performance as the parallel one, but it presents the main advantage of using much fewer hardware resources.

Key words: Artificial neural network, Multilayer perceptron, MLP, Cepstral analysis, MFCC, Respiratory sounds, Wheezing detection, FPGA.

2.1 Introduction

The multilayer perceptron (MLP) is a popular feed-forward artificial neural network (ANN), a class of bio-inspired learning machines that mimic the brain in the way it processes information. Due to its relative simplicity, MLP is the most widely used ANN for prediction, classification, and pattern recognition tasks. It is used as a model in various application fields, such as image processing and automatic speech recognition (ASR). Generally, the MLP-based models are first tested in software to validate their behavior, then adapted to

be implemented on-chip. The main advantage of such an implementation is that it enables real-time processing. The flexibility and high performance combined with the low-power consumption of the field programmable gate arrays (FPGAs) have made them the ideal candidate to perform hardware implementation of the MLP (Omondi et al., 2006; Zhu and Sutton, 2003; Caelli et al., 1999; Zhu et al., 1999).

Just like the brain, the MLP elementary units are neurons. Artificial neurons process information using arithmetic operations that involve many computationally costly multiplication operations. Therefore, large-scale neural networks cannot be implemented on low-cost FPGA chips because the demand in terms of resources exceeds the chip's capacity. The typical hardware implementation of a basic parallel neuron is presented in detail in 2.3. However, artificial neural networks can be implemented with different levels of parallelism following a trade-off between speed, data precision, and resources utilization. When increasing the degree of parallelism, we improve the implemented neural network's speed at the expense of the necessary chip resources, and conversely, the serialization reduces the used resources but will, in contrast, affect the processing speed (Savich et al., 2012). The literature in this area is abundant with examples of MLP implementations on FPGA chips. In general, the FPGA chips are presented as a low-cost, energy-efficient solution for hardware implementation of MLP. The trade-off between resources utilization and performance (calculation speed and accuracy) optimization, leads to a wide range of implementation architectures depending on different degrees of parallelism. The fully parallel MLP achieves the highest processing rates; however, it is shown that it is the most resource-demanding (Bahoura, 2016; Özdemir and Danişman, 2015, 2011; Gaikwad et al., 2019; Baptista et al., 2017; Baptista and Morgado-Dias, 2017; Mellit et al., 2010; Bahoura, 2018). Moreover, their accuracy remains dependant on the available resources on the used chips. The nonlinear activation function (sigmoid or hyperbolic tangent), in each neuron, can not directly be implemented in hardware sice both division and exponential operations require a lot of time and a large amount

of material resources (Gomperts et al., 2011). The remained practical hardware approach is to approximate the function (Tommiska, 2003; Armato et al., 2011). An important parameter to consider when choosing the hardware architecture is the processing time and data precision required by the given application to be implemented. The processing time is the time needed to compute an output from the moment the corresponding input is available. Fully parallel architectures are also less flexible and impose many constraints when the needs of the situation require a modification of the network topology. On the other hand, the fully serial architecture requires a substantial design effort to transform the native parallel computing operations of the MLP network into iterative operations using a single calculation unit, with the subsequent data synchronization challenge, in order to reduce resources. At the best of our knowledge, only one fully serial architecture has been proposed in the past (Ortigosa et al., 2006). However, this architecture is not scalable and requires a lot of effort to adapt the MLP to the desired topology. Furthermore, to reach an appropriate trade-off between performance (calculation speed and accuracy) optimization and resources consumption, many partially parallel architectures were proposed. These architectures implement material neurons within the same sequentially reusable layer (Nedjah et al., 2009, 2012) or only one physical neuron for each layer (Mellal et al., 2017; Oniga, 2005). On the other hand, the fully serial architecture uses a single calculation unit to drastically reduce the needed resources. This resource saving can be used to improve the calculation accuracy (data precision) of the network.

The main contribution presented in this article consists on designing and implementing a scalable fully serialized MLP network on a low-cost FPGA chip for real-time wheezing detection in respiratory sounds. Wheezes are polyphonic or monophonic adventitious respiratory sounds that appear at frequencies below 2000 Hz and last for a duration of about 250 ms (Lin and Lin, 2016; Fiz et al., 2002). They are a prevalent symptom observed in patients suffering from asthma (Pasterkamp, 2017). The physiological cause that produces

such sounds is the narrowing of the bronchi caused by mucus secretions (Gong, 1990). Four FPGA-based implementations of wheezing detection applications were found in the literature (Lin and Yen, 2014; Boujelben and Bahoura, 2018; Bahoura, 2018; Gwo-Ching, 2009). The new proposed architecture is assumed to reduce resource utilization by using a single neuron based on a multiply-and-accumulate (MAC) unit. The fully serialized MLP architecture thus implemented provides, on top of being less resources consuming, the same classification performance as the existing fully parallel architecture (Bahoura, 2018). This single neuron is reused to perform the computations of all the neurons in both the hidden and output layers. The reuse of the same physical neuron is made possible because, in MLP, all the neurons have the same hardware design and perform the same operations, except the activation function that may differ from one layer to another. However, the hardware implementation of fully serial MLP is an onerous and time-consuming process because of the need of controlling the temporary storage of the hidden neuron outputs and synchronization data and their corresponding synaptic weights at the input of the shared calculation unit. Usually, FPGA implemented ANNs are based on efficient yet complex processing elements (PEs), compromising flexibility and adaptability (Fe et al., 2015). In fact, a big implementation inconvenience identified in the literature is the complexity of the neuron's inner architecture (Nedjah et al., 2009, 2012). However, the PE of the proposed architecture has the particularity of being efficient and providing the necessary simplicity and flexibility at the same time. The architecture that we put forward in this article is effectively minimalist since it involves fewer elements than the previously mentioned architectures. Furthermore, the proposed architecture presents another significant advantage. In fact, the proposed architecture is scalable (can be easily expanded), so it can be adapted to any MLP topology (size of the input, hidden, and output layers). On the contrary, the parallel architecture does not allow such topology modifications without adjusting the design, including the wiring. The high flexibility of the implemented MLP architecture, achieved via serialization, supports the modularity and thus

the extendibility. A set of predefined parameters allows us to adjust the size of the network, according to the application needs. The design of the proposed serial architecture is presented in detail in subsection 2.2.3.

In short, compared to the parallel architectures, the proposed fully-serial architecture presents four advantages: 1) drastic reduction of the required resources, 2) the saved resources can be used to improve accuracy (data precision), 3) lower power consumption, and 4) higher flexibility (scalability). To validate the main advantages (reduction of the required resources, reduction of power consumption, accuracy, and scalability) of the proposed serial architecture, four systems, based on the Mel frequency cepstral coefficients (MFCC) and MLP networks, have been developed and implemented to classify respiratory sounds into normal and wheezing classes. The MFCC block handles the feature extraction operation while the MLP performs the classification. Classifiers based on parallel architectures, which require more hardware resources, have been implemented only on the Virtex-6 XC6VLX240T FPGA chip using Xilinx System Generator (XSG), which is an FPGA high level abstraction programming language operating in MATLAB/SIMULINK environment (Xilinx, 2012). However, those based on serial architectures, which require much less resources, have also been implemented on the low-cost Artix-7 XC7A100T FPGA circuit. The MLP network's synaptic weights were fixed after training in MATLAB, then transferred to XSG for hardware implementation on an FPGA chip.

2.2 Methods

The block diagram of a typical automatic classifier of respiratory sounds into normal and wheezing classes is presented in Fig. 2.1. Feature extraction and classification are the two principal computation stages performed by the classifier. The feature extraction block reduces

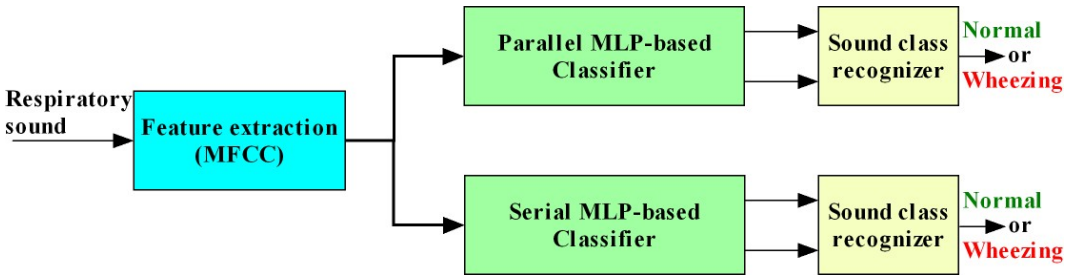


Figure 2.1 – Block diagram of two respiratory sounds classifiers based on parallel and serial architectures of MLP network.

the dimensionality of the input signal while capturing the relevant information in a set of low dimensionality feature vectors. The MFCC-based feature extraction technique was used for this purpose as it ensures good characterization of respiratory sounds. On the other hand, the classification block assigns the signal segment to one of the two classes (normal or wheezing) based on the outcome of the MFCC feature extraction block. MLP, which is considered as a reference approach in respiratory sounds classification field, was used to perform this task (Bahoura, 2018, 2009; Palaniappan et al., 2013; Forkheim et al., 1995; Rietveld et al., 1999; Kandaswamy et al., 2004; Sezgin et al., 2001; Waitman et al., 2000; Hashemi et al., 2012).

2.2.1 MFCC-based feature extraction

Based on the existence of some form of similarity in production and transmission between voice sounds and respiratory sounds, several methods of speech characterization have been adapted to respiratory sounds (Xu et al., 1998; Bahoura and Pelletier, 2004; Lin and Lin, 2016). The MFCC extraction technique is one of the most widely used for respiratory sounds characterization (Bahoura and Pelletier, 2004; Bahoura, 2009; Hashemi et al., 2012; Sengupta et al., 2016; Lin and Lin, 2016). Based on a filter bank, this technique ensures a good analysis of the respiratory sounds while reducing the spectral information to only a

few coefficients. On the other hand, it is known that the human ear decomposes frequencies on a nonlinear scale (such as the Mel scale) over the whole audio spectrum. Therefore, the decomposition of respiratory signals using such a filter bank improves the recognition performance by providing an adequate frequency resolution. As shown in Fig.2.2, computation of the MFCC features is performed in several steps. First, in order to split the signal into smaller frames, we apply a window function that consists of multiplying the signal by a Hamming window of length N=1024 samples. This frame corresponds to a segment of 170.7 ms of the signal and is appropriate for wheezing detection considering that it approximates the wheezing minimal duration of 250 ms. The following equation represents the windowing operation:

$$s[m,n] = s[n]w[n - mL] \tag{2.1}$$

where $w[n] = 0.54 - 0.46\cos(\frac{2\pi n}{N-1})$ is the Hamming window function, L is the time shift (L=1024), m the index of the current frame (window) and $s[n]$ the input signal. After this, we calculate the short-time Fourier transform (STFT) of each windowed segment:

$$S[m,k] = \sum_{n=0}^{N-1} s[m,n]e^{-j2\pi nk/N} \tag{2.2}$$

where $0 \leq k \leq N-1$ is the index of the frequency components. N represents also the number of the discrete frequencies. The energy spectrum is then computed based on the STFT outcome.

$$|S[m,k]|^2 = S_r^2[m,k] + S_i^2[m,k] \tag{2.3}$$

with $S_r[m,k]$ and $S_i[m,k]$ represent the real and the imaginary part of $S[m,k]$, respectively. The next step is to filter the energy spectrum with a Mel-scale filter bank. The filtering operation consists of summing the energy spectrum weighted by the filter's frequency response

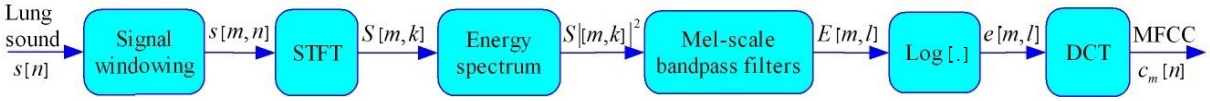


Figure 2.2 – Block diagram of MFCC-based feature extraction method.

$H_l[k]$.

$$E[m,l] = \sum_{k=0}^{N-1} |S[m,k]|^2 H_l[k] \tag{2.4}$$

where $1 \leq l \leq M$ is the filter index. Then, we apply a compressive logarithmic function to the filtered signal.

$$e[m,l] = \log(E[m,l]) \tag{2.5}$$

Finally, the discrete cosine transform (DCT) is applied to obtain the cepstral coefficients as follows :

$$c_m[n] = \sum_{l=1}^{M} e[m,l] \cos\left(n(l-0.5)\frac{\pi}{M}\right) \tag{2.6}$$

where $0 \leq n \leq P-1$ is the cepstral coefficient index, and $P \leq M$ is the number of coefficients computed.

The results of the DCT are then concatenated to give the feature vector x_m representing the acoustic respiratory frame $s[m,n]$. For the experimental test of this article, two feature vectors of, 12 MFCCs and 15 MFCCs, were defined according to the following equation:

$$\mathrm{x}_m = \begin{cases} [c_m[2], c_m[3] \ldots, c_m[13]], 12MFCCs \\ [c_m[2], c_m[3] \ldots, c_m[16]], 15MFCCs \end{cases} \tag{2.7}$$

where the first two coefficients c_0 and c_1 have been discarded because they do not represent the inter-class variability. These feature vectors of 12 and 15 MFCCs are provided to the MLP network as an input to perform classification of the original frame $s[m,n]$.

2.2.2 Natively Parallel Topology of MLP

MLP neural network is a basic approach used in most of modeling and classification tasks. It has a layered topology which consists of an input layer of N_0 variables, one (or more) hidden layers of N_1 neurons, and an output layer of N_2 neurons (Fig.2.3). The size of the input layer N_0 is determined by the number of coefficients in the feature vector, whereas the number of hidden neurons depends on the task complexity. The MLP network can have more than one hidden layer. However, a single hidden layer that contains an adequate number of neurons can effectively approximate any function (Hornik et al., 1989). The output layer, in turn, contains as many neurons (N_2) as there are classes since it is used to perform a classification task. In an MLP network, each neuron of a given layer is connected to the next layer's neurons via weighted connections. This interconnectivity provides the native parallelism of the MLP topology, as shown in Fig.2.3. Given the parallelism of the MLP topology, the calculations can obviously be performed simultaneously. The following equations show how an outcome of an MLP that has N_0 inputs, one hidden layer of N_1 neurons and N_2 output neurons is computed (Bahoura, 2018). For the hidden layer's neurons:

$$v_j^h = \sum_{i=1}^{N_0} w_{j,i}^h x_i + w_{j,N_0+1}^h \qquad j = 1, \ldots, N_1 \tag{2.8}$$

$$y_j^h = \varphi_h \left[v_j^h \right] \tag{2.9}$$

with $w_{j,i}^h$ the synaptic weight, whereas w_{j,N_0+1}^h represents the bias. For the output layer's neurons:

$$v_k^o = \sum_{j=1}^{N_1} w_{k,j}^o x_j^h + w_{k,N_1+1}^o \qquad k = 1, \ldots, N_2 \tag{2.10}$$

$$y_k^o = \varphi_o \left[v_k^o \right] \tag{2.11}$$

with $w^o_{k,j}$ the synaptic weight, whereas w^o_{k,N_1+1} represents the bias. The activation function used with hidden neurons is an hyperbolic tangent $\varphi_h[v] = (1 - e^{-2v})/(1 + e^{-2v})$, while the logistic-sigmoid function $\varphi_o(v) = 1/(1 + e^{-v})$ is used with output neurons. The use of the hyperbolic tangent is appropriate for the hidden layer because it's a bipolar function, i.e., it produces output values in the range of $[-1, 1]$, while the logistic sigmoid is more convenient for the output layer given that, in our case, the MLP network performs a classification task. For this article four topologies were tested in order to assess the scalability, these topologies are: MLP-12-12-2, MLP-12-16-2, MLP-15-12-2, and MLP-15-16-2.

According to (2.8) and (2.10), the fully parallel architecture is very resources demanding, it requires $N_0 \times N_1 + N_1 \times N_2$ multipliers and as the same number of adders. For this reason,

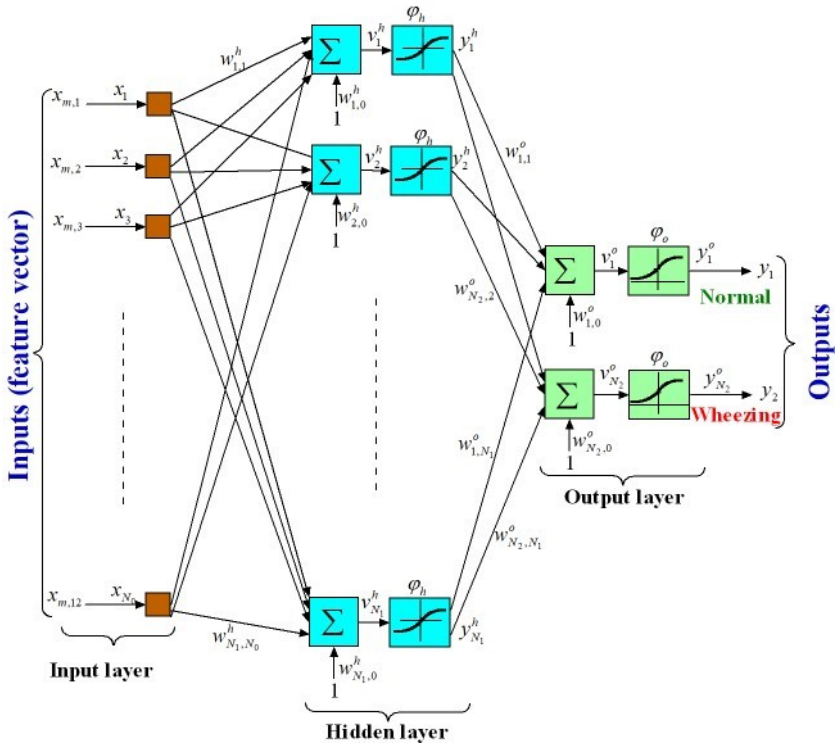


Figure 2.3 – Parallel architecture of a typical MLP neural network of N_0 inputs, one hidden layer of N_1 nodes, and $N_2 = 2$ output nodes.(Bahoura, 2018).

it is impossible to fit parallel architecture of a large size MLP in low-cost FPGA chips. In the next section, we present our innovative serial architecture to resolve this issue.

2.2.3 Proposed Serial Topology of MLP

The novel architecture proposed in this article is fully serial at both synapse and node levels, meaning that it uses only one material neuron fed successively by a single data line to reduce the number of multipliers and adders. The proposed single neuron's core element is a MAC (multiply-and-accumulate) unit whose task is to compute the weighted sum sequentially using a multiplier followed by an accumulator. The accumulator consists of an adder and a register whose output is looped back to the adder to perform the same operations described in (2.8) and (2.10). Actually, the multiplier computes the products, in pairs, of the inputs with their corresponding synaptic weights. At the same time, the accumulator handles the sum operations over a given number of cycles corresponding to the number of computations performed by a single neuron with fully synapse parallelism. The following equation describes the operation performed by the MAC unit:

$$v[n+1] = v[n] + w[n]z[n] \tag{2.12}$$

where $v[n]$ represents the cumulative product, at the nth iteration, of the input $z[n]$ and its corresponding weight $w[n]$.

The MAC unit calculates the weighted sum of both hidden and output layers. Once all the outputs of the hidden layer's neurons are calculated, they are stored in a memory block in order to be used in the next phase as inputs to calculate the outcomes of the output layer's neurons. The block diagram of the proposed fully serial architecture for the multilayer per-

ceptron (MLP) neural network, presented in Fig. 2.4, is partly based on the same idea of (Ortigosa et al., 2006). It consists of the following blocks:

— Serialization block : Since the MFCC-based feature extraction block provides parallel MFCC vectors, a serialization unit is required in the front line of the MLP block to feed the MAC-based neuron sequentially using a single data line. This operation is handled by a multiplexer that includes $N_0 + 1$ data inputs corresponding to the size of the feature vector plus a channel dedicated to the bias ($x_{N_0+1} = 1$), as described in equation (2.8). The controller block drives the address port of the multiplexer through the controller's Serial_Sel bus. This signal is presented in Fig.2.5.

— Data Multiplexer: To make sure to use the appropriate data at every stage of computations, we set up a switching system based on a 2-to-1 multiplexer whose selector input is driven by the controller block. When calculating the weighted sum of the hidden layer neurons ($v(n) \rightarrow v_j^h$), the selector Data_Sel is on 0 allowing the multiplexer to transmit the MFCC-based feature values ($x_i \rightarrow z[n]$) to the calculation unit. On the other hand, if the selector

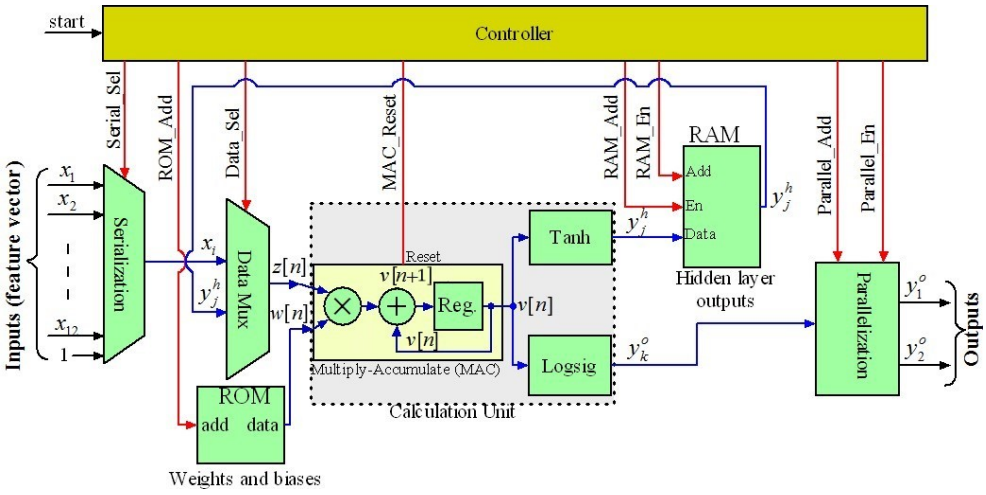


Figure 2.4 – Serial architecture of a typical MLP neural network.

value is 1, the multiplexer transmits the outcomes of the hidden layer stored beforehand in the RAM to the MAC ($y_j^h \rightarrow z[n]$), when calculating the weighted sum of the output layer neurons ($v[n] \rightarrow v_k^o$).

— ROM block: The ROM block is more suitable to store the weights and biases of the MLP as it can be loaded once, at the beginning, and allows only read access to the memory space during the processing. The MLP's weights and biases are fixed after off-line training on MATLAB then implemented in the ROM block. The ROM size is determined by the MLP topology, therefore, it must contain $(N_0 + 1) \times N_1 + (N_1 + 1) \times N_2$ addresses at least. The ROM address port is driven by the ROM_Add bus whose signal is generated using a cyclic counter located in the controller block. This counter is triggered by the *start* signal every time a new feature vector is available. These two signals are shown in Fig.2.5.

— Calculation Unit: It is the central element of the proposed MLP. It consists of a MAC block that performs the neurons computations sequentially, first, for the hidden layer ($v[n] \rightarrow v_j^h$) then for the output layer ($v[n] \rightarrow v_k^o$). The other elements of the calculation unit are the two activation function blocks *Tanh* and *Logsig* that provide the neurons outputs y_j^h and y_k^o, respectively. The MAC_Reset signal generated by the controller block resets the MAC unit at the beginning of each calculation cycle. This signal is presented in Fig.2.5.

— RAM block: The RAM block stores the intermediate outcomes of the hidden layer y_j^h that are then used later as inputs in the output layer computations of the weighted sum ($v[n] \rightarrow v_k^o$). The RAM size is equal to the number of hidden neurons plus one memory slot ($N_1 + 1$). This slot is used to add the bias w_{k,N_1+1}^o to the weighted sum described in equation (2.10). The RAM block is addressed using RAM_Add bus while the writing enable is controlled by RAM_En bit that is set when the hidden output y_j^h is calculated

and ready to be stored. Fig.2.5 shows the aforementioned control signals.

— Parallelization block: This block mainly operates like a 1-to-N_2 demultiplexer whose outputs are stored in N_2 parallel placed registers. It is fed serially by a single data line that carries y_k^o produced by the calculation unit. The address carried by the controller's Parallel_Add bus is decoded, and the input data y_k^o is transmitted to its corresponding block output y_k^o, according to the decoded address. For synchronization purposes, the Parallel_En bit shown in Fig.2.5 enables the acquisition of y_k^o once the calculations are achieved.

— Controller block: The controller block is the key element of the proposed fully serial architecture. It handles control signals of the above described blocks, including multiplexer's selector, block's reset/enable ports, and ROM/RAM address ports. It mainly consists of cyclic counters operating at different clocking as well as delay elements, logic gates and comparators used for synchronization purposes. This block permits the scalability of the proposed serial architecture to any MLP topology since the counting limits of its counters are imposed by the size of the MLP layers and are expressed as a function of N_o, N_1 and N_2. Fig.2.5 presents all control and synchronization signals provided by the controller block for MLP neural network having $N_0 = 15$, $N_1 = 12$, and $N_2 = 2$.

2.2.4 Sound class recognizer

It is the last block of the automatic respiratory sounds classifier. Its task is to assign the processed sound segment to one of the two predefined classes based on the MLP outputs. These outputs are used to compute a score function $Sc[m]$ as follows:

$$\mathrm{Sc}[m] = y_2^o - y_1^o \tag{2.13}$$

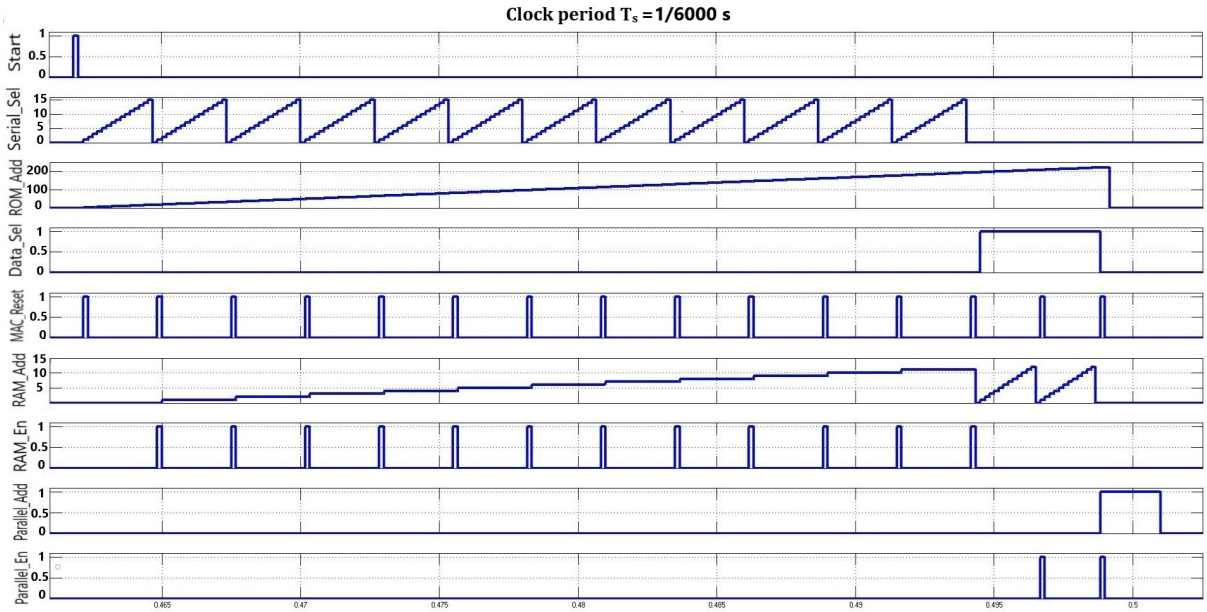


Figure 2.5 – Control and synchronization signals provided by the controller block for MLP neural network having $N_0 = 15$, $N_1 = 12$, and $N_2 = 2$.

where y_1^o and y_2^o are the MLP outputs. This score function is compared to a threshold:

$$s[m,n] = \begin{cases} \text{wheezing} & if\ \text{Sc}[m] \geq \theta \\ \text{normal} & otherwise \end{cases} \tag{2.14}$$

The discriminating threshold θ can be defined using an optimization criterion (Bahoura, 2009). Here, it is chosen to be zero.

2.3 Hardware implementation

The proposed novel serial architecture of MLP-based respiratory sound classifier has been designed using Xilinx system generator (XSG) tool and two development boards (ML-605 and Nexys-4). XSG is a plug-in to MATLAB/SIMULINK that allows to develop ad-

vanced digital signal processing systems for Xilinx FPGA chips. Fig.2.6 presents the top-level testbed diagram of the parallel and serial MLP-based respiratory sounds classifiers, as described in Fig. 2.1, using XSG blockset (counters, RAMs, ROM, multiplexers, accumulator, registers, logic gates, etc). In other words, the proposed fully-serial hardware architecture of MLP-based respiratory sounds classifier is implemented and compared, in terms of the classification performance and resources utilization, to its corresponding fully-parallel architecture using the same input MFCC-based feature vector. In fact, four MLP topologies, identified by MLP-12-12-2, MLP-12-16-2, MLP-15-12-2, and MLP-15-16-2 are implemented both in parallel and in serial in order to be compared, where MLP-N_0-N_1-N_2 refers to an MLP of N_0 inputs, N_1 hidden neurons, and N_2 output neurons. Therefore, Fig. 2.6 corresponds to both parallel and serial implementation of the MLP-12-12-2 topology, using 2's complement signed 24-bit fixed-point format having 16 fractional bits.

The first hardware part (block) is the MFCC-based feature extraction block that provides MFCC features to both parallel and serial topologies. This block has been used and thoroughly discussed in (Bahoura, 2018; Bahoura and Ezzaidi, 2013). The next is the parallel MLP-based classifier. As presented in subsection 2.2.2, the MLP topology is parallel by nature at both synapse and node levels (Fig.2.3). Therefore MLP network requires as many PEs as the number of neurons when implemented in parallel. Moreover, in the parallel implementation, each hidden neuron contains N_0 XSG multipliers and as many XSG adders, whereas the output neurons contain N_1 XSG multipliers and as many XSG adders. Therefore, the four MLP topologies MLP-12-12-2, MLP-12-16-2, MLP-15-12-2, and MLP-15-16-2 require 168, 224, 204, and 272 XSG multiplier blocks and as many of XSG adder blocks, respectively. The synaptic weights, including the bias, are stored separately in XSG constant blocks. Also, full parallelism requires that each neuron of the parallel MLP contains its own activation function block. More details about the fully parallel MLP implementation

can be found in (Bahoura, 2018).

The hardware resources required by the parallel architecture of the MLP network drastically increase with its size, which requires high-density and power-demanding FPGA chips. Our fully-serial architecture has been proposed in order to overcome this issue and makes it possible to implement large size MLP networks on low-density and low-power FPGA chips. As shown in Fig. 2.7, the proposed serial MLP architecture is implemented as follows:

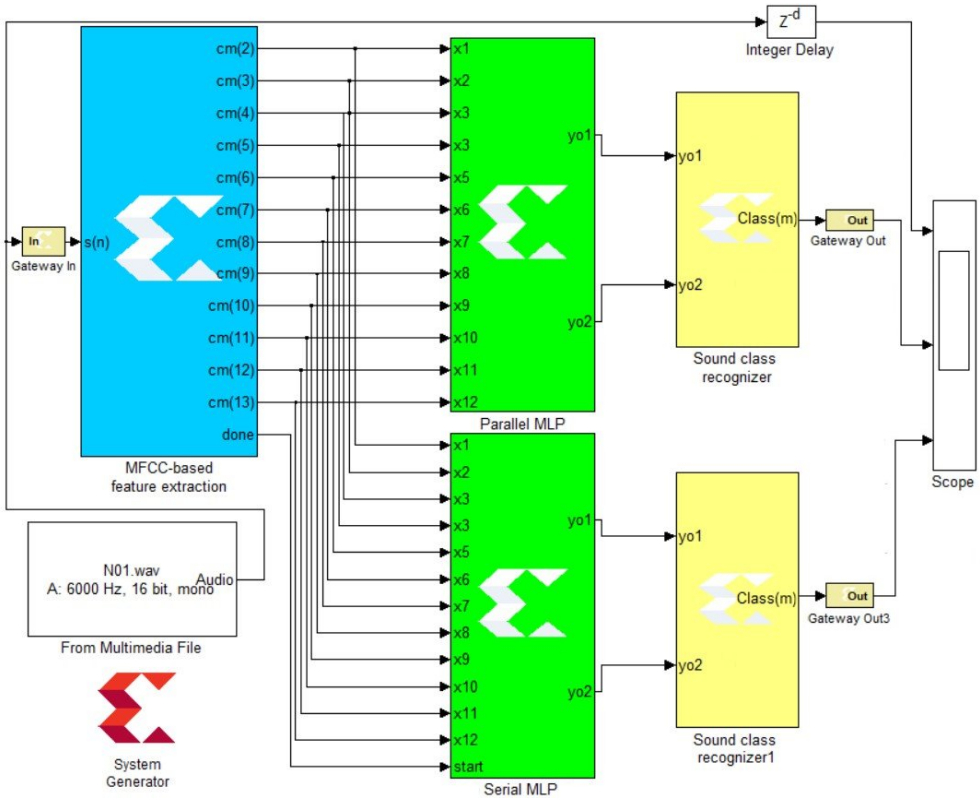


Figure 2.6 – Top-level testbed diagram of parallel and serial MLP-based respiratory sounds classifiers, as described in Fig 2.1, implemented on FPGA circuit using XSG blockset for the MLP-12-12-2 topology ($N_0 = 12, N_1 = 12, N_2 = 2$).

— Serialization and Data Select: The serialization unit was implemented using an ($N_0 + 1$)-to-1 Mux block available in XSG library, while the data select uses an 2-to-1 Mux.

— ROM and RAM blocks: The ROM memory used to store the MLP synaptic weights

and biases was implemented using one XSG RAM block, while the RAM memory that stores the hidden layer's outputs was implemented using a single port XSG RAM with distributed memory. The size of the ROM and RAM blocks are parameterizable with the MLP size (N_0, N_1, and N_2), ensuring the scalability of the proposed architecture.

— Calculation Unit: As mentioned previously, a single MAC-based calculation unit is used to calculate the output of all MLP neurons. This MAC unite is implemented using an XSG Multiplier block followed by an XSG accumulator block. This means that only one XSG multiplier block is used instead of 168, 224, 204, and 272 for the MLP-12-12-, MLP-12-16-2, MLP-15-12-2, and MLP-15-16-2 topologies, respectively. This calculation unit uses one Tangent hyperbolic (*Tanh*) block and one logistic sigmoid (*Logsig*) block instead of N_1 *Tanh* block and N_2 *Logsig* block for the fully-parallel architecture (Bahoura, 2018). The XSG-based implementation of these two activation functions is presented in Fig. 2.7.

— Parallelization block : Unlike the multiplexer, the demultiplexer is not provided by the XSG library. Therefore, it was implemented using XSG logical gates and N_2 XSG registers to make the parallel outputs simultaneously available.

Finally, the sound class recognizer is implemented using one XSG relational, one XSG Subtractor, and XSG constant blocks. Details can be found in (Bahoura, 2018).

From a hardware point of view, scalability is made possible by the fact that the proposed serial architecture does not require any addition/removal of hardware components nor wiring modification, when the size of the network changes. Simply, a small set of parameters performs the modifications (size of ROM and RAM blocks) required by the network's size. Unlike (Ortigosa et al., 2006) that used a hardware description language (VHDL) and Handel-C to implement a serial MLP on FPGA, the proposed architecture presents the advantage of being implemented using a higher abstraction level tool XSG that permits to design and test, easily and quickly, these architectures with all powerful simulation capabilities of MAT-

LAB/SIMULINK environment. Fig.2.7 shows the details of the XSG-based implementation. The proposed fully-serial architecture presents also the advantage of scalability.

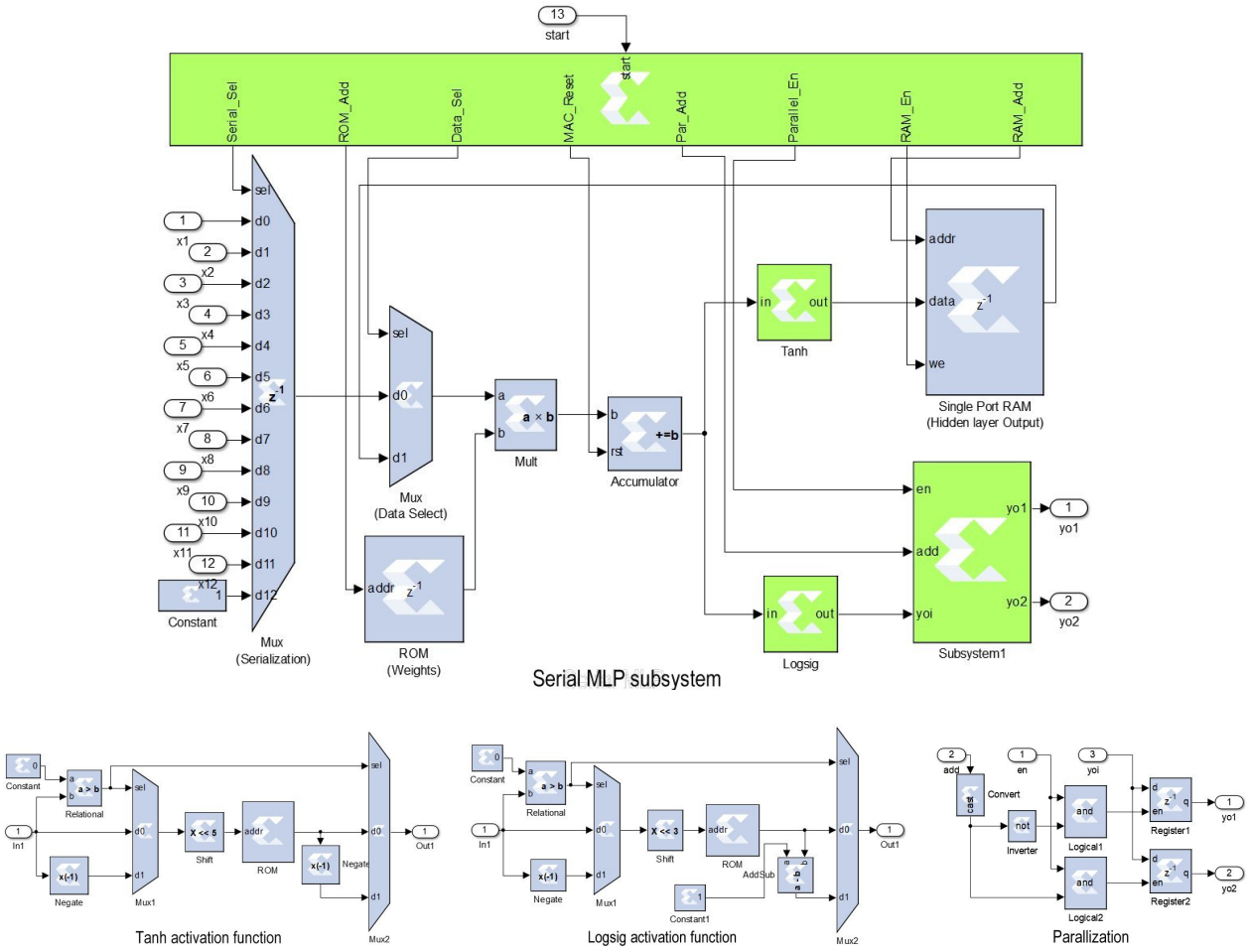


Figure 2.7 – Details of the proposed MLP and its sub-blocks implemented using XSG.

2.4 Results and discussion

2.4.1 Database and protocol

This work was conducted using respiratory sound recordings collected from RALE database-CD, ASTRA database-CD, and other websites (Bahoura, 2009). Our database contains 12 respiratory sounds recorded from 12 healthy subjects and thereby placed in the category of normal sounds, and 12 sounds recorded from people with asthma and identified as wheezing class. All these sounds were down-sampled at 6 kHz, and wheeze recordings have been manually screened to have sounds composed of wheeze segments exclusively. The short-term analysis and windowing is done on non-overlapping frames of N=1024 samples leading to a database of 915 normal sound segments and 497 wheezing sound segments.

Knowing that MLP-based classifier needs to be trained first, this step was achieved off-line in MATLAB environment, and the trained model was implemented on FPGA. The tests were carried out using the leave-one-out cross-validation (LOOCV) procedure over the sound recording set, where 11 sounds of each sound class are used for training, and one sound is left for testing. This procedure is used to ensure a reliable and unbiased estimate of the performance. Also, each iteration of the LOOCV procedure was repeated ten times to provide statistically credible results by calculating the average of the outcomes of the ten repetitions.

Four classifiers built on the four MLP topologies (MLP-12-12-2, MLP-12-16-2, MLP-15-12-2, and MLP-15-16-2) are implemented in software (MATLAB) and in hardware (XSG) for both parallel and serial architectures. For each LOOCV iteration, the classification models are trained in MATLAB and their synaptic weights and biases are transferred to XSG-based architectures in order to be implemented and tested on FPGA.

2.4.2 Performance criteria

The classification performance obtained with these MLP-based classifiers is discussed in terms of four parameters: Sensitivity, specificity, performance, and accuracy. Each one of these parameters is defined as follows:

$$\text{Sensitivity (SE)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.15}$$

$$\text{Specificity (SP)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2.16}$$

$$\text{Performance (PER)} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \tag{2.17}$$

$$\text{Accuracy (ACC)} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{TP} + \text{FN}} \tag{2.18}$$

where TP, FN, TN, FP are the number of true positives, false negatives, true negatives, and false positives, respectively.

2.4.3 Classification results

The average values of TP, FN, TN, FP constitute the confusion matrix, as shown in Tables 2.1 and 2.2, whereas Tables 2.3 and 2.4 show the average values of the classification performance.

Table 2.1 – Average confusion matrix obtained with MATLAB software and both parallel and serial XSG architectures implemented for the MLP-12-12-2 and MLP-12-16-2 topologies, using 2's complement signed 24-bit fixed-point format with 16 fraction bits.

	Assigned class											
	MLP-12-12-2						MLP-12-16-2					
	XSG-Parallel		XSG-Serial		MATLAB		XSG-Parallel		XSG-Serial		MATLAB	
True class	Wheeze	Normal	Wheeze	Normal	Wheeze	Normal	Wheeze	Normal	Wheeze	Normal	Wheeze	Normal
Wheeze	405.1 (TP)	91.9 (FN)	405 (TP)	92 (FN)	406.1 (TP)	90.9 (FN)	396(TP)	101 (FN)	396 (TP)	101 (FN)	396.3 (TP)	100.7 (FN)
Normal	105.1 (FP)	809.9 (TN)	105.1 (FP)	809.9(TN)	105.8 (FP)	809.2(TN)	99.2 (FP)	815.8 (TN)	99.2 (FP)	815.8 (TN)	97.8 (FP)	817.2(TN)

Table 2.2 – Same as Table 2.1 but for the MLP-15-12-2 and MLP-15-16-2 topologies.

	Assigned class											
	MLP-15-12-2						MLP-15-16-2					
	XSG-Parallel		XSG-Serial		MATLAB		XSG-Parallel		XSG-Serial		MATLAB	
True class	Wheeze	Normal	Wheeze	Normal	Wheeze	Normal	Wheeze	Normal	Wheeze	Normal	Wheeze	Normal
Wheeze	404.9 (TP)	92.1 (FN)	404.6 (TP)	92.4 (FN)	406.7 (TP)	90.3 (FN)	406.9(TP)	90.1 (FN)	407(TP)	90 (FN)	406.9 (TP)	91.1 (FN)
Normal	77.8 (FP)	837.2 (TN)	77.3 (FP)	837.7 (TN)	77.6 (FP)	837.4 (TN)	64.4 (FP)	850.6 (TN)	64.2 (FP)	850.8 (TN)	64.9 (FP)	850.1 (TN)

Table 2.3 – Average classification results obtained with MATLAB software and both parallel and serial XSG architectures for the MLP-12-12-2 and MLP-12-16-2 topologies, using 2's complement signed 24-bit fixed-point format with 16 fraction bits.

Architecture	MLP12-12-2								MLP-12-16-2							
	TP	TN	FP	FN	SE	SP	PER	ACC	TP	TN	FP	FN	SE	SP	PER	ACC
XSG-Parallel	405.1	809.9	105.1	91.9	0.8152	0.8851	0.8494	0,8605	396.0	815.8	99.2	101.0	0.7967	0.8915	0.8428	0.8582
XSG-Serial	405.0	809.9	105.1	92.0	0.8152	0.8852	0.8493	0.8605	396.0	815.8	99.2	101.0	0.7968	0.8916	0.8429	0.8582
MATLAB	406.1	809.2	105.8	90.9	0.8171	0.8843	0.8500	0,8606	396.3	817.2	97.8	100.7	0.7973	0.8929	0.8438	0.8593

Table 2.4 – Same as Table 2.3 but for the MLP-15-12-2 and MLP-15-16-2 topologies.

Architecture	MLP-15-12-2								MLP-15-16-2							
	TP	TN	FP	FN	SE	SP	PER	ACC	TP	TN	FP	FN	SE	SP	PER	ACC
XSG-Parallel	404.9	837.2	77.8	92.1	0.8146	0.9149	0.8633	0.8796	406.9	850.6	64.4	90.1	0.8187	0.9296	0.8724	0.8905
XSG-Serial	404.6	837.7	77.3	92.4	0.8140	0.9155	0.8633	0.8798	407.0	850.8	64.2	90.0	0.8189	0.9298	0.8726	0.8907
MATLAB	406.7	837.4	77.6	90.3	0.8183	0.9147	0.8652	0.8808	406.9	850.1	64.9	91.1	0.8170	0.9290	0.8712	0.8896

For the four tested topologies, it can be shown that the serial architecture gives similar classification performance than that of the parallel one. The best classification performance (SE=81.9%, SP=83%, PER=87.3% and ACC=89.1%) is obtained with the MLP-15-16-2. Overall, the classification performance metrics are insignificantly different from the performance obtained with MATLAB software (less than 1%). The insignificant discrepancies noted are due to quantization errors, given that MATLAB handles floating-point data while the hardware implementation uses the 2's complement signed 24-bit fixed-point format with 16 fraction bits. However, the classification performance metrics of the proposed MLP serial architecture are practically the same as those obtained with parallel architecture, the maximum gap is 0.02% of the accuracy irrespective of the classifier's configuration. Based on these results, we considered that the tests are conclusive, and the implementation of the proposed serial architecture for wheezing detection was successful. The fact that the proposed novel FPGA-based serial architecture yields the same results as the corresponding parallel architecture, as well as the software-based implementation confirms the validity of our design methodology.

2.4.4 Evaluation of the hardware implementation

The wheezing detector used for hardware prototyping is based on MFCC for feature extraction and MLP for classification tasks. To highlight the gain in resources provided by our serial architecture, four different combinations (MFCC/MLP-12-12-2, MFCC/MLP-12-16-2, MFCC/MLP-15-12-2, and MFCC/MLP-15-16-2) have been implemented on a Virtex-6 XC6VLX240T FPGA chip using Xilinx system generator (XSG) and ML-605 development board. In fact, Virtex-6 XC6VLX240T FPGA circuit has sufficient resources to allow the hardware implementation of the four classifiers for both parallel and serial architectures. Table 2.5 presents the resources utilization (slices, Flip Flops, LUTs, RAMB36E1s,

RAMB18E1s, and DSP48E1s), as provided by the Xilinx ISE 14.7 tool, for the complete classification systems (MFCC/MLP-12-12-2, MFCC/MLP-12-16-2, MFCC/MLP-15-12-2, and MFCC/MLP-15-16-2), whereas, Table 2.6 presents the required resources only for the MLP topologies (MLP-12-12-2, MLP-12-16-2, MLP-15-12-2, and MLP-15-16-2). Maximum operating frequency and power consumption are also estimated by the Xilinx ISE 14.7 tool. As expected, the serial MLP architecture requires much less resources than the parallel one. This gap between the two designs is explained by the fact that the parallel version uses as many processing elements as it contains neurons, unlike the serial architecture, where only one processing element is used. For example, the parallel MLP-12-12-2 requires 12×12+12×2=168 DSP48E1s blocks, but the serial MLP-12-12-2 require only one DSP48E1s block. The parallel MLP-15-16-2 requires 15×16+16×2=272 DSP48E1s blocks, whereas the serial MLP-12-12-2 require only one DSP48E1s block. It can be noted that each DSP48E1 block includes one 25×8 multiplier. Therefore, the number of DSP48E1s used by a XSG multiplier depends on the used data width (Bahoura, 2018).

Overall, the amount of used Slices, LUTs, and Flip-Flops has drastically dropped in the serial architecture. The implemented serial MLP-based respiratory sound classifier (excluding the MFCC block) uses only 2-3% of the slices used in the parallel architecture, 1% of the Flip-Flops, and 2% of the LUTs depending on the MLP's size. The computations were performed using fixed-point data and the fact that the proposed architecture uses fewer processing elements allows us to handle larger fixed-point data representation thus reduces the impact of the quantization error. As for the parallel architecture (Bahoura, 2018), the proposed serial architecture has been pipelined in order to improve its operating frequency. For the Virtex-6 XC6VLX240T FPGA chip, both parallel and serial architecture of the complete MFCC/MLP classifier (Table 2.5) reach approximatively the same maximum operating frequency (about 115 MHz). However, the operating frequency is increased for the MLP

Table 2.5 – Comparison in terms of resource utilization, and total power consumption of the FPGA-based complete MFCC-MLP classifier implemented with the proposed serial MLP architecture and a fully parallel architecture using a Virtex-6 XC6VLX240T FPGA chip. Both architectures use 2's complement signed 24-bit fixed-point format with 16 fraction

	Architecture							
	MFCC/MLP-12-12-2		MFCC/MLP-12-16-2		MFCC/MLP/15-12-2		MFCC/MLP-15-16-2	
	Parallel	Serial	Parallel	Serial	Parallel	Serial	Parallel	Serial
Resource utilization								
Slices (37,680)	6,148	4,560	6,681	4,578	6,632	4,839	7,374	4,607
Flip Flops (301,440)	22,036	15,484	24,312	15,486	23,924	15,932	26,680	15,934
LUTs (150,720)	19,871	15,179	21,525	15,232	21,367	15,669	23,195	15,643
RAMB36E1s (416)	24	2	32	2	24	2	32	2
RAMB18E1s (832)	6	6	6	6	6	6	6	6
DSP48E1s (768)	217	50	273	50	262	59	330	59
Max.Op. freq. (MHz)	119.360	115.234	119.546	119.818	119.589	112.854	115.982	119.332
Total power consum. (W)	3.484	3.459	3.490	3.464	3.485	3.464	3.491	3.465

Table 2.6 – Comparison in terms of resource utilization, and total power consumption of the FPGA-based MLP alone implemented on with the proposed serial MLP architecture and a fully parallel architecture using a Virtex-6 XC6VLX240T FPGA chip. Both architectures use 2's complement signed 24-bit fixed-point format with 16 fraction

	Architecture							
	MLP-12-12-2		MLP-12-16-2		MLP-15-12-2		MLP-15-16-2	
	Parallel	Serial	Parallel	Serial	Parallel	Serial	Parallel	Serial
Resource utilization								
Slices (37,680)	1,579	131	2,101	128	2,013	129	2,515	133
Flip Flops (301,440)	6,766	179	9,042	181	8,206	179	10,962	181
LUTs (150,720)	5,252	337	7,116	341	5,760	335	8,320	340
RAMB36E1s (416)	24	2	32	2	24	2	32	2
RAMB18E1s (832)	2	2	2	2	2	2	2	2
DSP48E1s (768)	168	1	224	1	204	1	272	1
Max.Op. freq. (MHz)	191.351	193.087	185.460	198.255	184.809	191.571	182.183	194.818
Total power consum. (W)	3.449	3.423	3.456	3.429	3.452	3.429	3.460	3.429

Table 2.7 – Resource utilization, and total power consumption of the proposed serial MLP architecture implemented on the low-cost Artix-7 XC7A100T FPGA chip. Both architectures use 2's complement signed 24-bit fixed-point format with 16 fraction

	Architecture							
	MFCC/Serial MLP				Serial MLP			
	12-12-2	12-16-2	15-12-2	15-16-2	12-12-2	12-16-2	15-12-2	15-16-2
Resource utilization								
Slices(15,850)	4,762	4,536	4,783	4,822	126	130	145	142
Flip Flops (126,800)	15,282	15,284	15,730	15,732	179	181	179	181
LUTs (63,400)	15,478	15,473	15,919	15,924	334	337	330	336
RAMB36E1s (135)	2	2	2	2	2	2	2	2
RAMB18E1s (270)	6	6	6	6	2	2	2	2
DSP48E1s (240)	50	50	59	59	1	1	1	1
Max.Op. freq. (MHz)	102.417	104.102	109.601	109.039	141.663	140.469	141.323	139.684
Total power consum. (W)	0.111	0.114	0.114	0.114	0.084	0.085	0.085	0.085

classifier alone (Table 2.6) to about 185 MHz. This can be explained by the fact that complete MFCC-MLP classifier presents a longer critical path than the MLP classifier alone. However, both parallel and serial architectures, of the complete MFCC-MLP classifier or the MLP classifier alone, present the same power consumption of 3.4 W for this Virtex-6 FPGA chip (Tables 2.5 and 2.6).

The selected MLP parallel architectures, except the 12-12-2, cannot be implemented on a low-cost FPGA board such as the Artix-7 XC7A100T because the demand for DSP48E1s exceeds the chip's capabilities. However, the MLP-based systems designed according to the fully-serial architecture proposed in this paper can be implemented on this FPGA chip since they use only one processing element. As shown in Table 2.7 the proposed MLP serial architectures, when implemented on the Artix-7 XC7A100T FPGA chip, use approximatively the same number of the hardware resources as for the Virtex-6 FPGA. However, the maximum operating frequencies are decreased (from about 115 MHz to 105 MHz, and from about

185 MHz to 140 MHz). Differences in the operating frequency can be explained by the speed grades of these FPGA families. The additional advantage on using such FPGA is the low-power consumption. As shown in Table 2.7, the total power consumption of this implementation does not exceed 114 mW for the Artix-7 XC7A100T FPGA, compared to 3.4 W for the Virtex-6 XC6VLX240T FPGA chip. This fully serial architecture is an adequate solution to implement an autonomous real-time classification based on a large-size MLP neural network.

Finally, the hardware implementation is successfully achieved since the proposed serial architecture is less resources demanding than the parallel one and can thus be implemented on low-cost FPGA chips. Also, the low-power consumption and the high processing speed are key features that validate the hardware implementation.

2.5 Conclusion

The proposed fully-serial architecture of the MLP-based respiratory sound classifier helps to reduce the needed resources significantly while providing the same classification performance as those obtained with the parallel architecture. Scalability is also a key feature of the proposed architecture. The MLP serial architecture implemented in this way does not require any addition/removal of hardware components nor wiring changes and can easily be adapted to any topology dictated by the situation's needs. The sequential calculation in the proposed MLP fully-serial architecture could have an impact on the processing speed. However, since the response of the MLP-based classifier depends on the computational speed of the MFCC coefficients provides at its input, the effect of serialization on the overall speed is minor. The automatic classifier discussed in this paper has been implemented on a Artix-7 XC7A100T chip. Our choice was motivated by the low-cost and the flexibility of implemen-

tation of FPGA technology. Moreover, one of the advantageous features of this architecture is the possibility to implement a large-size MLP network on low-power FPGA families. Future work will focus on the refinement of this architecture. We are planning to implement a serial architecture that handles layer-wise scalability to ensure more flexibility in case we have an MLP with more than one hidden layer.

Acknowledgment

This research is financially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Références

Armato, A., Fanucci, L., Scilingo, E., De Rossi, D., 2011. Low-error digital hardware implementation of artificial neuron activation functions and their derivative. Microprocessors and Microsystems 35, 557–567.

Bahoura, M., 2009. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Computers in Biology and Medicine 39, 824–843.

Bahoura, M., 2016. Fpga implementation of blue whale calls classifier using high-level programming tool. Electronics 5, 1–8.

Bahoura, M., 2018. FPGA implementation of an automatic wheezing detection system. Biomedical Signal Processing and Control 46, 76–85.

Bahoura, M., Ezzaidi, H., 2013. Hardware implementation of MFCC feature extraction for respiratory sounds analysis. 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA) , 226–229.

Bahoura, M., Pelletier, C., 2004. Respiratory sounds classification using cepstral analysis and gaussian mixture models. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 1, 9–12.

Baptista, D., Abreu, S., Travieso-Gonzalez, C., Morgado-Dias, F., 2017. Hardware implementation of an artificial neural network model to predict the energy production of a photovoltaic system. Microprocessors and Microsystems 49, 77–86.

Baptista, F.D., Morgado-Dias, F., 2017. Automatic general-purpose neural hardware generator. Neural Computing and Applications 28, 25–36.

Boujelben, O., Bahoura, M., 2018. Efficient FPGA-based architecture of an automatic wheeze detector using a combination of MFCC and SVM algorithms. Journal of Systems Architecture 88, 54–64.

Caelli, T., Ling Guan, Wen, W., 1999. Modularity in neural computing. Proceedings of the IEEE 87, 1497–1518.

Fe, J.D., Aliaga, R.J., Gadea-Gironés, R., 2015. Evolutionary optimization of neural networks with heterogeneous computation: study and implementation. The Journal of Supercomputing 71, 2944–2962.

Fiz, J.A., Jané, R., Homs, A., Izquierdo, J., García, M.A., Morera, J., 2002. Detection of wheezing during maximal forced exhalation in patients with obstructed airways. Chest 122, 186–191.

Forkheim, K.E., Scuse, D., Pasterkamp, H., 1995. A comparison of neural network models for wheeze detection, in: IEEE WESCANEX 95. Communications, Power, and Computing. Conference Proceedings, pp. 214–219.

Gaikwad, N.B., Tiwari, V., Keskar, A., Shivaprakash, N.C., 2019. Efficient FPGA implementation of multilayer perceptron for real-time human activity classification. IEEE Access 7, 26696–26706.

Gomperts, A., Ukil, A., Zurfluh, F., 2011. Development and implementation of parameterized FPGA-based general purpose neural networks for online applications. IEEE Transactions on Industrial Informatics 7, 78–89.

Gong, H.J., 1990. Wheezing and asthma, in: Walker, H.K., Hall, W.D., Hurst, J.W. (Eds.), Clinical Methods: The History, Physical, and Laboratory Examinations. © 1990, Butterworth Publishers, a division of Reed Publishing., Boston, pp. 100–239.

Gwo-Ching, C., 2009. FPGA realization of an automatic wheeze detector based on wavelet filter bank, in: 2009 IEEE 13th International Symposium on Consumer Electronics, pp. 832–833.

Hashemi, A., Arabalibeik, H., Agin, K., 2012. Classification of wheeze sounds using cepstral analysis and neural networks. Studies in health technology and informatics 173, 161–165.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Kandaswamy, A., Kumar, C.S., Ramanathan, R.P., Jayaraman, S., Malmurugan, N., 2004. Neural classification of lung sounds using wavelet coefficients. Comput Biol Med 34, 523–37.

Lin, B.S., Lin, B.S., 2016. Automatic wheezing detection using speech recognition technique. Journal of Medical and Biological Engineering 36, 545–554.

Lin, B.S., Yen, T.S., 2014. An FPGA-based rapid wheezing detection system. International Journal of Environmental Research and Public Health 11, 1573–1593.

Mellal, I., Laghrouche, M., Bui, H.T., 2017. Field Programmable Gate Array (FPGA) respiratory monitoring system using a flow microsensor and an accelerometer. Measurement Science Review 17, 61 – 67.

Mellit, A., Mekki, H., Messai, A., Salhi, H., 2010. FPGA-based implementation of an intelligent simulator for stand-alone photovoltaic system. Expert Systems with Applications 37, 6036–6051.

Nedjah, N., da Silva, R., Mourelle, L., da Silva, M., 2009. Dynamic MAC-based architecture of artificial neural networks suitable for hardware implementation on FPGAs. Neurocomputing 72, 2171–2179.

Nedjah, N., da Silva, R.M., de Macedo Mourelle, L., 2012. Compact yet efficient hardware implementation of artificial neural networks with customized topology. Expert Systems with Applications 39, 9191–9206.

Omondi, A.R., Rajapakse, J.C., Bajger, M., 2006. FPGA neurocomputers, in: Omondi, A.R., Rajapakse, J.C. (Eds.), FPGA Implementations of Neural Networks. Springer US, Boston, MA, pp. 1–36.

Oniga, S., 2005. A new method for FPGA implementation of artificial neural network used in smart devices. In: international computer science conference microCAD 37, 31–36.

Ortigosa, E.M., Cañas, A., Ros, E., Ortigosa, P.M., Mota, S., Díaz, J., 2006. Hardware description of multi-layer perceptrons with different abstraction levels. Microprocessors and Microsystems 30, 435–444.

Özdemir, A.T., Danişman, K., 2011. Fully parallel ANN-based arrhythmia classifier on a single-chip FPGA: FPAAC. Turkish Journal of Electrical Engineering and Computer Science 19, 667–687.

Özdemir, A.T., Danişman, K., 2015. A comparative study of two different FPGA-based arrhythmia classifier architectures. Turkish Journal of Electrical Engineering & Computer Sciences 23, 2089–2106.

Palaniappan, R., Sundaraj, K., Ahamed, N.U., 2013. Machine learning in lung sound analysis: A systematic review. Biocybernetics and Biomedical Engineering 33, 129–135.

Pasterkamp, H., 2017. The highs and lows of wheezing: A review of the most popular adventitious lung sound. Pediatric Pulmonology 53, 243–254.

Rietveld, S., Oud, M., Dooijes, E.H., 1999. Classification of asthmatic breath sounds: Preliminary results of the classifying capacity of human examiners versus artificial neural networks. Computers and Biomedical Research 32, 440–448.

Savich, A., Moussa, M., Areibi, S., 2012. A scalable pipelined architecture for real-time computation of MLP-BP neural networks. Microprocessors and Microsystems 36, 138–150.

Sengupta, N., Sahidullah, M., Saha, G., 2016. Lung sound classification using cepstral-based statistical features. Comput Biol Med 75, 118–29.

Sezgin, M.C., Dokur, Z., Olmez, T., Korurek, M., 2001. Classification of respiratory sounds by using an artificial neural network, in: 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 697–699.

Tommiska, M., 2003. Efficient digital implementation of the sigmoid function for reprogrammable logic. IEEE proceedings-computers and Digital Techniques 150, 403–411.

Waitman, L.R., Clarkson, K.P., Barwise, J.A., King, P.H., 2000. Representation and classification of breath sounds recorded in an intensive care setting using neural networks. Journal of Clinical Monitoring and Computing 16, 95–105.

Xilinx, 2012. System generator for DSP, user guide UG640 (v14.3-14.7). URL: https://www.xilinx.com/support/documentation/sw_manuals/xilinx14_7/sysgen_user.pdf.

Xu, J., Cheng, J., Wu, Y., 1998. A cepstral method for analysis of acoustic transmission characteristics of respiratory system. IEEE Trans. Biomed. Eng. 45, 660–664.

Zhu, J., Milne, G.J., Gunther, B.K., 1999. Towards an FPGA based reconfigurable computing environment for neural network implementations, in: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), pp. 661–666.

Zhu, J., Sutton, P., 2003. FPGA implementations of neural networks – a survey of a decade of progress, in: Field Programmable Logic and Application, Springer Berlin Heidelberg. pp. 1062–1066.

CONCLUSION GÉNÉRALE

Le travail de recherche présenté ici consiste en une étude approfondie des techniques de détection automatique des sibilants dans les sons respiratoires et dont le but ultime est de réaliser une implémentation matérielle afin de permettre le traitement en temps réel. La détection automatique des sibilants a été rendue possible grâce aux techniques avancées issues du domaine du traitement du signal ainsi que des techniques de reconnaissance de formes basée sur l'apprentissage machine. Les modèles de détection et de classification que nous proposons dans ce travail de recherche se basent essentiellement sur des techniques qui ont fait leurs preuves en reconnaissance de la parole et d'identification du locuteur. Des méthodes de caractérisation basées sur l'analyse cepstrale ont été utilisées pour extraire l'information pertinente à partir des sons respiratoires. Ces caractéristiques sont par la suite utilisées pour entraîner un modèle de classification. Un tel modèle, basé sur l'apprentissage machine, sera capable de distinguer avec précision entre les deux classes de sons respiratoires à savoir les sons normaux et les sibilants.

Dans les travaux précédents ayant trait à ce domaine, plusieurs méthodes de caractérisation des sons respiratoires ont été proposées. Pour notre part, nous avons sélectionné des méthodes de caractérisation basées sur des coefficients cepstraux, ces méthodes sont : les coefficients cepstraux à échelle de Mel (MFCC), les coefficients cepstraux à filtres gammatone (GFCC) et les coefficients cepstraux à échelle de Bark (BFCC). Chacune de ces techniques d'extraction a été combinée avec des techniques d'apprentissage machine pour la classification, le but étant d'identifier la combinaison qui donne les meilleurs taux de classification des sons respiratoires. Deux types de réseaux de neurones ont été retenus pour cette étape à savoir le perceptron multicouche (MLP) et le BiLSTM (Bidirectional Long Short-Term Memory) qui est un réseau de neurone qui combine deux couches LSTM. Un paramétrage adéquat du

modèle (Taille du vecteur de caractéristiques, nombre de neurones cachés dans les réseaux de neurones,etc.) a aussi été pris en considération afin d'assurer une meilleure performance. Dans la deuxième partie de ce travail, nous avons réalisé une implémentation matérielle sur un circuit FPGA d'un classifieur de sons respiratoires basé sur la combinaison MFFC/MLP. Le MLP dont il est question ici a été implémenté selon une nouvelle architecture qui permet de réduire l'utilisation des ressources du circuit FPGA. Une telle implémentation permet d'effectuer la détection des sibilants en temps réel tout en ayant pour support un dispositif portable et à bas coût.

La base de données dont nous disposons contient 24 enregistrements de sons respiratoires équitablement répartis sur les deux classe (sons normaux, sons sibilants). Après segmentation, nous avons obtenus une base de 915 segments de sons normaux et 497 segments sibilants. Dans la phase des tests, l'approche Leave-One-Out a été adoptée afin de tester toute la base de donnée. Les résultats obtenus ont été évalués en fonction de la spécificité, la sensitivité, la performance et l'exactitude.

Au terme de la phase de test, la meilleure performance de classification a été obtenue à l'aide de la combinaison MFCC-BiLSTM avec 96.3% des segments correctement classifiés, tandis que la nouvelle architecture pour l'implémentation du MLP a été validée étant donné qu'elle produit les mêmes résultats de classification que la méthode d'implémentation de référence.

Dans les travaux futurs, nous proposons d'investiguer d'autres méthodes d'extraction de caractéristiques et de classification afin d'optimiser les performances. Également, il est envisageable d'étendre le champ d'application à d'autres catégories de sons pathologiques tels que les crépitants et les râles.

RÉFÉRENCES

Acharya, J., Basu, A., 2020. Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. IEEE Transactions on Biomedical Circuits and Systems 14, 535–544.

Alsmadi, S., Kahya, Y.P., 2008. Design of a dsp-based instrument for real-time classification of pulmonary sounds. Comput Biol Med 38, 53–61.

Ameur, S., Ben Khalifa, A., Bouhlel, M., 2020. A novel hybrid bidirectional unidirectional LSTM network for dynamic hand gesture recognition with leap motion. Entertainment Computing 35, 100373.

Apaydin, H., Feizi, H., Sattari, M.T., Colak, M.S., Shamshirband, S., Chau, K.W., 2020. Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. Water 12, 1500.

Armato, A., Fanucci, L., Scilingo, E., De Rossi, D., 2011. Low-error digital hardware implementation of artificial neuron activation functions and their derivative. Microprocessors and Microsystems 35, 557–567.

Athulya, M.S., Sathidevi, P.S., 2018. Speaker verification from codec distorted speech for forensic investigation through serial combination of classifiers. Digital Investigation 25, 70–77.

Bahoura, M., 2009. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Computers in Biology and Medicine 39, 824–843.

Bahoura, M., 2016. Fpga implementation of blue whale calls classifier using high-level programming tool. Electronics 5, 1–8.

Bahoura, M., 2018. FPGA implementation of an automatic wheezing detection system. Biomedical Signal Processing and Control 46, 76–85.

Bahoura, M., Ezzaidi, H., 2013. Hardware implementation of MFCC feature extraction for respiratory sounds analysis. 2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA) , 226–229.

Bahoura, M., Pelletier, C., 2004. Respiratory sounds classification using cepstral analysis and gaussian mixture models. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 1, 9–12.

Baptista, D., Abreu, S., Travieso-Gonzalez, C., Morgado-Dias, F., 2017. Hardware implementation of an artificial neural network model to predict the energy production of a photovoltaic system. Microprocessors and Microsystems 49, 77–86.

Baptista, F.D., Morgado-Dias, F., 2017. Automatic general-purpose neural hardware generator. Neural Computing and Applications 28, 25–36.

Basu, V., Rana, S., 2020. Respiratory diseases recognition through respiratory sound with the help of deep neural network, in: 2020 4th International Conference on Computational Intelligence and Networks (CINE), pp. 1–6.

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks 5, 157–166.

Boujelben, O., Bahoura, M., 2018. Efficient FPGA-based architecture of an automatic wheeze detector using a combination of MFCC and SVM algorithms. Journal of Systems Architecture 88, 54–64.

Caelli, T., Ling Guan, Wen, W., 1999. Modularity in neural computing. Proceedings of the IEEE 87, 1497–1518.

Chandra Yadav, I., Pradhan, G., 2021. Pitch and noise normalized acoustic feature for children's ASR. Digital Signal Processing 109, 1128–1132.

Chen, C.H., Huang, W.T., Tan, T.H., Chang, C.C., Chang, Y.J., 2015. Using k-nearest neighbor classification to diagnose abnormal lung sounds. Sensors 15, 13132–13158.

Chen, K., Huo, Q., 2016. Training deep bidirectional LSTM acoustic model for LVCSR by a context-sensitive-chunk BPTT approach. IEEE/ACM Transactions on Audio, Speech, and Language Processing 24, 1185–1193.

Chenchah, F., Lachiri, Z., 2017. A bio-inspired emotion recognition system under real-life conditions. Applied Acoustics 115, 6–14.

Daalache, M.R.L., Addou, D., Boudraa, M., 2017. An efficient distributed speech processing in noisy mobile communications, in: 2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), pp. 1–4.

Dash, T.K., Mishra, S., Panda, G., Satapathy, S.C., 2021. Detection of COVID-19 from speech signal using bio-inspired based cepstral features. Pattern Recognition 117, 107999.

Dua, M., Aggarwal, R.K., Biswas, M., 2018. Performance evaluation of hindi speech recognition system using optimized filterbanks. Engineering Science and Technology, an International Journal 21, 389–398.

Emmanuel, A., Sandra, R., Raymond, G., Christian, B., Yvan, K., 2007. La sémiologie pulmonaire à l'ère de la médecine factuelle. Médecine thérapeutique 13, 353–356.

Fe, J.D., Aliaga, R.J., Gadea-Gironés, R., 2015. Evolutionary optimization of neural networks with heterogeneous computation: study and implementation. The Journal of Supercomputing 71, 2944–2962.

Fiz, J.A., Jané, R., Homs, A., Izquierdo, J., García, M.A., Morera, J., 2002. Detection of wheezing during maximal forced exhalation in patients with obstructed airways. Chest 122, 186–191.

Forkheim, K.E., Scuse, D., Pasterkamp, H., 1995. A comparison of neural network models for wheeze detection, in: IEEE WESCANEX 95. Communications, Power, and Computing. Conference Proceedings, pp. 214–219.

Fraiwan, M., Fraiwan, L., Alkhodari, M., Hassanin, O., 2021. Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. Journal of Ambient Intelligence and Humanized Computing 4, 1–13.

Gaikwad, N.B., Tiwari, V., Keskar, A., Shivaprakash, N.C., 2019. Efficient FPGA implementation of multilayer perceptron for real-time human activity classification. IEEE Access 7, 26696–26706.

Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: continual prediction with lstm. Neural Comput 12, 2451–71.

Gomperts, A., Ukil, A., Zurfluh, F., 2011. Development and implementation of parameterized FPGA-based general purpose neural networks for online applications. IEEE Transactions on Industrial Informatics 7, 78–89.

Gong, H.J., 1990. Wheezing and asthma, in: Walker, H.K., Hall, W.D., Hurst, J.W. (Eds.), Clinical Methods: The History, Physical, and Laboratory Examinations. © 1990, Butterworth Publishers, a division of Reed Publishing., Boston, pp. 100–239.

Gouda, A., El Shehaby, S., Diaa, N., Abougabal, M., 2019. Classification techniques for diagnosing respiratory sounds in infants and children, in: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0354–0360.

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J., 2009. A novel connectionist system for unconstrained handwriting recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 855–868.

Graves, A., Mohamed, A., Hinton, G., 2013. Speech recognition with deep recurrent neural networks, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6645–6649.

Gupta, S., Agrawal, M., Deepak, D., 2021. Gammatonegram based triple classification of lung sounds using deep convolutional neural network with transfer learning. Biomedical Signal Processing and Control 70, 102947.

Guyon, I., Elisseeff, A., 2006. An introduction to feature extraction, in: Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A. (Eds.), Feature Extraction. Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, Berlin, Heidelberg. volume 207, pp. 1–25.

Gwo-Ching, C., 2009. FPGA realization of an automatic wheeze detector based on wavelet filter bank, in: 2009 IEEE 13th International Symposium on Consumer Electronics, pp. 832–833.

Hashemi, A., Arabalibeik, H., Agin, K., 2012. Classification of wheeze sounds using cepstral analysis and neural networks. Studies in health technology and informatics 173, 161–165.

Haykin, S., 1999. Multilayer perceptrons, in: Neural Networks: A Comprehensive Foundation. Prentice Hall, pp. 122–221.

He, T., Droppo, J., 2016. Exploiting LSTM structure in deep neural networks for speech recognition, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5445–5449.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Hu, X., Yuan, S., Xu, F., Leng, Y., Yuan, K., Yuan, Q., 2020. Scalp EEG classification using deep Bi-LSTM network for seizure detection. Computers in Biology and Medicine 124, 103919.

Jácome, C., Marques, A., 2015. Computerized respiratory sounds in patients with copd: a systematic review. COPD 12, 104–12.

Kahya, Y.P., Yeginer, M., Bilgic, B., 2006. Classifying respiratory sounds with different feature sets, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2856–2859.

Kandaswamy, A., Kumar, C.S., Ramanathan, R.P., Jayaraman, S., Malmurugan, N., 2004. Neural classification of lung sounds using wavelet coefficients. Comput Biol Med 34, 523–37.

Lin, B.S., Lin, B.S., 2016. Automatic wheezing detection using speech recognition technique. Journal of Medical and Biological Engineering 36, 545–554.

Lin, B.S., Yen, T.S., 2014. An FPGA-based rapid wheezing detection system. International Journal of Environmental Research and Public Health 11, 1573–1593.

Lozano, M., Fiz, J.A., Jané, R., 2016. Automatic differentiation of normal and continuous adventitious respiratory sounds using ensemble empirical mode decomposition and instantaneous frequency. IEEE J Biomed Health Inform 20, 486–97.

Mazić, I., Bonković, M., Džaja, B., 2015. Two-level coarse-to-fine classification algorithm for asthma wheezing recognition in childrenś respiratory sounds. Biomedical Signal Processing and Control 21, 105–118.

Mellal, I., Laghrouche, M., Bui, H.T., 2017. Field Programmable Gate Array (FPGA) respiratory monitoring system using a flow microsensor and an accelerometer. Measurement Science Review 17, 61 – 67.

Mellit, A., Mekki, H., Messai, A., Salhi, H., 2010. FPGA-based implementation of an intelligent simulator for stand-alone photovoltaic system. Expert Systems with Applications 37, 6036–6051.

Meng, F., Shi, Y., Wang, N., Cai, M., Luo, Z., 2020. Detection of respiratory sounds based on wavelet coefficients and machine learning. IEEE Access 8, 155710–155720.

Messner, E., Fediuk, M., Swatek, P., Scheidl, S., Smolle-Jüttner, F.M., Olschewski, H., Pernkopf, F., 2020. Multi-channel lung sound classification with convolutional recurrent neural networks. Computers in Biology and Medicine 122, 103831.

Mondal, A., Banerjee, P., Tang, H., 2018. A novel feature extraction technique for pulmonary sound analysis based on EMD. Computer Methods and Programs in Biomedicine 159, 199–209.

Moore, B., 2013. Frequency selectivity, masking and the critical band, in: An Introduction to the Psychology of Hearing: Sixth Edition. Brill, Leiden, The Netherlands, pp. 19–234.

Moore, B.C., Glasberg, B.R., 1983. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am 74, 750–3.

Mozer, M.C., 1991. Induction of multiscale temporal structure, in: Proceedings of the 4th International Conference on Neural Information Processing Systems, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. p. 275–282.

Nedjah, N., da Silva, R., Mourelle, L., da Silva, M., 2009. Dynamic MAC-based architecture of artificial neural networks suitable for hardware implementation on FPGAs. Neurocomputing 72, 2171–2179.

Nedjah, N., da Silva, R.M., de Macedo Mourelle, L., 2012. Compact yet efficient hardware implementation of artificial neural networks with customized topology. Expert Systems with Applications 39, 9191–9206.

Omondi, A.R., Rajapakse, J.C., Bajger, M., 2006. FPGA neurocomputers, in: Omondi, A.R., Rajapakse, J.C. (Eds.), FPGA Implementations of Neural Networks. Springer US, Boston, MA, pp. 1–36.

Oniga, S., 2005. A new method for FPGA implementation of artificial neural network used in smart devices. In: international computer science conference microCAD 37, 31–36.

Ortigosa, E.M., Cañas, A., Ros, E., Ortigosa, P.M., Mota, S., Díaz, J., 2006. Hardware description of multi-layer perceptrons with different abstraction levels. Microprocessors and Microsystems 30, 435–444.

O'Shaughnessy, D., 1999. Hearing, in: Speech Communications: Human and Machine, 2nd Edition. Wiley-IEEE Press, Piscataway, NJ, pp. 109–138.

Oud, M., Dooijes, E.H., Zee, J.S.v.d., 2000. Asthmatic airways obstruction assessment based on detailed analysis of respiratory sound spectra. IEEE Transactions on Biomedical Engineering 47, 1450–1455.

Özdemir, A.T., Danişman, K., 2011. Fully parallel ANN-based arrhythmia classifier on a single-chip FPGA: FPAAC. Turkish Journal of Electrical Engineering and Computer Science 19, 667–687.

Özdemir, A.T., Danişman, K., 2015. A comparative study of two different FPGA-based arrhythmia classifier architectures. Turkish Journal of Electrical Engineering & Computer Sciences 23, 2089–2106.

Palaniappan, R., Sundaraj, K., 2013. Respiratory sound classification using cepstral features and support vector machine, in: 2013 IEEE Recent Advances in Intelligent Computational Systems RAICS, pp. 132–136.

Palaniappan, R., Sundaraj, K., Ahamed, N.U., 2013. Machine learning in lung sound analysis: A systematic review. Biocybernetics and Biomedical Engineering 33, 129–135.

Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks, in: Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, JMLR.org. p. III–1310–III–1318.

Pasterkamp, H., 2017. The highs and lows of wheezing: A review of the most popular adventitious lung sound. Pediatric Pulmonology 53, 243–254.

Pasterkamp, H., Tal, A., Leahy, F., Fenton, R., Chernick, V., 1985. The effect of anticholinergic treatment on postexertional wheezing in asthma studied by phonopneumography and spirometry. Am Rev Respir Dis 132, 16–21.

Perna, D., Tagarelli, A., 2019. Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks, in: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), pp. 50–55.

Piirilä, P., Sovijärvi, A.R., Kaisla, T., Rajala, H.M., Katila, T., 1991. Crackles in patients with fibrosing alveolitis, bronchiectasis, copd, and heart failure. Chest 99, 1076–83.

Qi, J., Wang, D., Jiang, Y., Liu, R., 2013. Auditory features based on gammatone filters for robust speech recognition, in: 2013 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 305–308.

Riella, R., Nohama, P., Maia, J., 2009. Method for automatic detection of wheezing in lung sounds. Brazilian Journal of Medical and Biological Research 42, 674–684.

Rietveld, S., Oud, M., Dooijes, E.H., 1999. Classification of asthmatic breath sounds: Preliminary results of the classifying capacity of human examiners versus artificial neural networks. Computers and Biomedical Research 32, 440–448.

Sankur, B., Güler, E.a., Kahya, Y.P., 1996. Multiresolution biological transient extraction applied to respiratory crackles. Computers in Biology and Medicine 26, 25–39.

Savich, A., Moussa, M., Areibi, S., 2012. A scalable pipelined architecture for real-time computation of MLP-BP neural networks. Microprocessors and Microsystems 36, 138–150.

van Schayck, C.P., van Der Heijden, F.M., van Den Boom, G., Tirimanna, P.R., van Herwaarden, C.L., 2000. Underdiagnosis of asthma: is the doctor or the patient to blame? the dimca project. Thorax 55, 562–5.

Schneider, A., Gindner, L., Tilemann, L., Schermer, T., Dinant, G.J., Meyer, F.J., Szecsenyi, J., 2009. Diagnostic accuracy of spirometry in primary care. BMC Pulm Med 9, 31.

Sengupta, N., Sahidullah, M., Saha, G., 2016. Lung sound classification using cepstral-based statistical features. Comput Biol Med 75, 118–29.

Sezgin, M.C., Dokur, Z., Olmez, T., Korurek, M., 2001. Classification of respiratory sounds by using an artificial neural network, in: 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 697–699.

Sánchez Morillo, D., Astorga Moreno, S., Fernández Granero, M., León Jiménez, A., 2013. Computerized analysis of respiratory sounds during copd exacerbations. Comput Biol Med 43, 914–21.

Sreeram, A.S.K., Ravishankar, U., Sripada, N.R., Mamidgi, B., 2020. Investigating the potential of mfcc features in classifying respiratory diseases, in: 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), pp. 1–7.

Taplidou, S.A., Hadjileontiadis, L.J., Kitsas, I.K., Panoulas, K.I., Penzel, T., Gross, V., Panas, S.M., 2004. On applying continuous wavelet transform in wheeze analysis. Conf Proc IEEE Eng Med Biol Soc 2004, 3832–5.

Tommiska, M., 2003. Efficient digital implementation of the sigmoid function for reprogrammable logic. IEEE proceedings-computers and Digital Techniques 150, 403–411.

Traunmüller, H., 1990. Analytical expressions for the tonotopic sensory scale. The Journal of the Acoustical Society of America 88, 97–100.

Trianto, R., Tai, T., Wang, J., 2018. Fast-LSTM acoustic model for distant speech recognition, in: 2018 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–4.

Waitman, L.R., Clarkson, K.P., Barwise, J.A., King, P.H., 2000. Representation and classification of breath sounds recorded in an intensive care setting using neural networks. Journal of Clinical Monitoring and Computing 16, 95–105.

Xilinx, 2012. System generator for DSP, user guide UG640 (v14.3-14.7). URL: https://www.xilinx.com/support/documentation/sw_manuals/xilinx14_7/sysgen_user.pdf.

Xu, J., Cheng, J., Wu, Y., 1998. A cepstral method for analysis of acoustic transmission characteristics of respiratory system. IEEE Trans. Biomed. Eng. 45, 660–664.

Yilmaz, C.A., Kahya, Y.P., 2006. Multi-channel classification of respiratory sounds, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2864–2867.

Zhang, X., Zhao, M., Dong, R., 2020. Time-series prediction of environmental noise for urban iot based on long short-term memory recurrent neural network. Applied Sciences 10, 1144.

Zhi-bin, G., 2011. Time frequency analysis of multi-component non-stationary signal with filter bank decomposition, in: 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), pp. 2035–2038.

Zhu, J., Milne, G.J., Gunther, B.K., 1999. Towards an FPGA based reconfigurable computing environment for neural network implementations, in: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), pp. 661–666.

Zhu, J., Sutton, P., 2003. FPGA implementations of neural networks – a survey of a decade of progress, in: Field Programmable Logic and Application, Springer Berlin Heidelberg. pp. 1062–1066.