Analysis and Diagnosis of Newborn Cry Signals based on Signal Processing, Statistical Physics and Deep Learning

by

Salim LAHMIRI

MANUSCRIPT-BASED THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, 11 APRIL 2025

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Prof. Chakib Tadj, Thesis supervisor
Department of Electrical Engineering, École de Technologie Supérieure

Prof. Christian Gargour, Thesis co-supervisor
Department of Electrical Engineering, École de Technologie Supérieure

Prof. Chamseddine Talhi, President of the board of examiners
Department of Electrical Engineering, École de Technologie Supérieure

Prof. Saad Maarouf, Thesis supervisor
Department of Electrical Engineering, École de Technologie Supérieure

Prof. Stephane Gagnon, External examiner
Department of Technology Management, Université du Québec en Outaouais

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON 17 JUNE 2024

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

PREFACE

Signal processing and machine learning are increasingly used in the design of computer-aided diagnosis (CAD) systems. The intent of this PhD dissertation, in one hand, is to improve the understanding the physiology of newborn cry signals by use of signal processing and statistical physics and, in the other hand, to implement deep learning models for improvement of CAD systems in accurate distinction between healthy and unhealthy subjects.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisors Professors Chakib Tadj and Christian Gargour for the continuous support of my Ph.D study and research, for their patience, motivation, continuous support, and enthusiasm.

Besides my advisor, I would like to thank the rest of my thesis committee: Professeur Saad Maarouf and Professor Chamseddine Talhi, for insightful comments.

I would like to thank Professor Stephane Gagnon for serving as external examiner.

I would like to thank Banafshe Salehian for sharing the dataset used in all my research works.

I would like to thank Zahra Khalilzad for editing the manuscript.

I would like to thank Versha Aamir for her grateful help on editing this work. Thank you Versha.

In loving Memory of my Father.

Analyse et diagnostic des cris des bébés par techniques de traitement de signal, physique statistique et apprentissage profond

Salim LAHMIRI

RÉSUMÉ

Les pleurs du nouveau-né sont généralement dus à diverses conditions liées à la physiologie, à la pathologie ou à l'émotion. À cet égard, différents modèles de signaux de cri du nouveau-né sont associés à son état de santé. En conséquence, divers systèmes de diagnostic assisté par ordinateur ont été proposés pour distinguer automatiquement les signaux de cris d'un nouveau-né sain et malsain. Ces systèmes automatiques utilisent des techniques spécifiques de traitement du signal combinées à l'apprentissage automatique pour l'analyse et la classification des signaux de cris du nouveau-né avec une précision acceptable.

L'objectif principal de notre étude est de concevoir de nouveaux systèmes de diagnostic assisté par ordinateur basés sur une combinaison de traitement du signal et d'apprentissage profond pour améliorer la précision de la distinction entre les signaux de cris d'un nouveau-né sain et malsain. De plus, nous étudions la complexité de ces signaux sur la base des fractales, de l'entropie et des multi-fractales afin de mieux comprendre les différences de dynamique non linéaire des signaux de cri d'un sujet à l'autre.

Pour la classification automatique des signaux de cris du nouveau-né, nous avons entraîné divers systèmes d'apprentissage profond (notamment les réseaux neuronaux à rétroaction profonde, à convolution, et à mémoire à long terme et à court terme) avec des informations basées sur l'analyse du cepstre. Nous avons également utilisé la méthode d'optimisation bayésienne pour optimiser les hyperparamètres des machines à vecteurs de support avec fonction à base radiale et l'algorithme du k plus proche voisins (kNN), tous deux entraînés avec différentes caractéristiques audio-acoustiques séparément ou combinées, sélectionnées à l'aide d'un filtre statistique. Pour l'analyse de la complexité, nous avons utilisé la dimension de corrélation, l'entropie approximative et les leaders d'ondelettes.

Dans la tâche de classification automatique des signaux de cris du nouveau-né en utilisant les coefficients du cepstre, nous avons constaté que (*a*) le réseau neuronal à action directe profonde (DFNN) atteignait un taux de classification correcte très proche de la perfection lorsqu'il était appliqué aux signaux de cris d'expiration du nourrisson et donnait une performance parfaite lorsqu'il était appliqué aux signaux de cris d'inspiration du nourrisson, (*b*) DFNN a surpassé les systèmes SVM linéaires et Naïve Bayes lorsqu'ils ont été testés à la fois sur les ensembles d'expiration et d'inspiration, (*c*) DFNN a surpassé les travaux très récents publiés dans la littérature, (*d*) les réseaux neuronaux à convolution (CNN) ont surpassé DFNN et les systèmes à long et courte mémoire (LSTM), et (*e*) les systèmes d'apprentissage profond entrainés avec des descripteurs du cepstre ont obtenu le meilleure taux de classification correcte par rapport aux études similaires dans la littérature.

Dans la tâche de la classification des signaux de cris de nouveau-nés sains et malsains sur la base des caractéristiques acoustiques, nous avons constaté que (*a*) le SVM entraîné avec des caractéristiques de modulation d'amplitude d'inspiration auditive (AAM) atteignait le taux de classification correcte le plus élevé, suivi par l'algorithme kNN entraîné avec une combinaison de fréquences des coefficients cepstrales (MFCC), AAM et prosodie, et (*b*) SVM a surpassé en terme de performance la plupart des travaux existants validés sur la même base de données tout en étant considérablement rapide à exécuter.

Dans la tâche de caractérisation des signaux de cris de nouveau-nés sains et malsains à l'aide de mesures de complexité, nous avons constaté que (*a*) il existe des différences significatives dans l'entropie approximative et la dimension de corrélation entre les deux catégories de patients, (*b*) les signaux de cris de nourrissons en bonne santé présentent un niveau d'entropie approximatif plus élevé que ceux des nourrissons pathologiques, (*c*) les signaux de cris des nourrissons en bonne santé présentent un niveau de dimension de corrélation plus élevé que ceux des nourrissons pathologiques, et (*d*) les cris des nouveau-nés en bonne santé présentent un degré de multi-fractal plus élevé que ceux en mauvaise santé.

En résumé, les systèmes d'apprentissage profond formés avec des descripteurs de cepstre sont prometteurs pour l'analyse et le diagnostic des signaux de cris du nourrisson en milieu clinique. De même, le SVM non linéaire optimisé à l'aide de l'optimisation bayésienne et entraîné par des caractéristiques sélectionnées basées sur le chi-carré de MFCC, AAM, prosodie ou combinaison de ces caractéristiques sélectionnées, peut être prometteur pour le diagnostic des nouveau-nés en milieu clinique. Cependant, l’apprentissage profond permet d’atteindre les plus hautes performances.

Enfin, l’entropie approximative et la dimension de corrélation basées sur le cepstre peuvent être considérées comme des biomarqueurs et pourraient potentiellement aider à comprendre la physiologie des cris du nouveau-né et être utilisées à des fins de diagnostic.

Mots-clés: cri du nouveau-né, traitement de signal, cepstrum, acoustique, apprentissage machine, apprentissage profond, sélection de caractéristiques, classification

Analysis and diagnosis of newborn cry signals based on signal processing, statistical physics and deep learning

Salim LAHMIRI

ABSTRACT

Newborn cry is generally due to various conditions related to physiology, pathology, or emotion. In this regard, different patterns in newborn cry signal are associated with health condition. As a result, various computer aided diagnosis systems have been proposed to automatically distinguish between healthy and unhealthy newborn cry signals. Such CAD systems are used to employ specific signal processing techniques combined with machine learning for the analysis and classification of newborn cry signals with acceptable accuracy.

The main purpose of our research study is to design new computer aided diagnosis systems based on combination of signal processing and deep learning to improve the accuracy to distinguish between healthy and unhealthy newborn cry signals. In addition, we investigate complexity in such signals based on fractals, entropy, and multifractals to better understand the differences in nonlinear dynamics of cry signals across subjects.

For automatic classification of newborn cry signals, we trained various deep learning systems (including deep feedforward, convolution neural networks, and long short-term memory neural networks) with cepstrum-based information. We also used Bayesian optimization method to optimize the hyper-parameters of the support vector machines (SVM) with radial basis function and k-nearest neighbors (kNN), both trained with different audio acoustic features separately or combined which were selected by using a statistical filter. For complexity analysis, we used correlation dimension, approximate entropy, and wavelet leaders.

In the task of automatic classification of newborn cry signals based on cepstrum analysis we found that (*a*) deep feedforward neural network (DFNN) achieved very close to perfect accuracy when applied to expiration infant cry signals and yielded to perfect accuracy when applied to inspiration infant cry signals, (*b*) DFNN outperformed the linear SVM and the Naïve Bayes systems when tested both on the expiration and inspiration sets, (*c*) DFNN outperformed very recent works found in the literature, (*d*) convolution neural networks (CNN) outperformed DFNN and long short-term memory (LSTM) system, and (*e*) deep learning systems trained with cepstrum descriptors obtained the highest accuracy compared to similar studies in the literature.

In the task of classification of healthy versus unhealthy newborn cry signals based on acoustic features we found that (*a*) the SVM trained with auditory-inspired amplitude modulation (AAM) features achieved the highest accuracy followed by kNN algorithm trained with combination of Mel frequency cepstral coefficients (MFCC), AAM, and prosody, and (*b*) SVM outperformed most existing works validated on the same database while being considerably fast to perform.

In the task of characterization healthy and unhealthy newborn cry signals by using complexity measures, we found that (*a*) there are significant differences in approximate entropy and correlation dimension across two categories of subjects, (*b*) healthy infant cry signals show higher approximate entropy level than those of pathological infants, (*c*) healthy infant cry signals show higher correlation dimension level than those of pathological infants, and (*d*) healthy signals exhibit a higher degree of multifractality than unhealthy ones.

In summary, deep learning systems trained with cepstrum descriptors are promising for analysis and diagnosis of infant cry signals in clinical milieu. Likewise, the nonlinear SVM optimized by using Bayesian optimization and trained by Chi-square based selected features from MFCC, AAM, prosody or combination of those selected features, can be promising for diagnosis of newborns based on their cry signals in clinical milieu. However, deep learning allows achieving the highest performance.

Finally, the cepstrum-based approximate entropy and correlation dimension can be considered as biomarkers and could potentially help understanding the physiology of newborn cries and be used for diagnosis purpose.

Keywords: newborn cry signal, signal processing, cepstrum, acoustics, machine learning, deep learning, features selection, classification

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABREVIATIONS

AAM	Auditory-inspired amplitude modulation
AE	Approximate entropy
BO	Bayesian optimization
CAD	Computer-aided diagnosis
CD	Correlation Dimension
CNN	Convolutional neural networks
DFFNN	Deep feedforward neural networks
EXP	Expiration
INS	Inspiration
kNN	k-nearest neighbors
LSTM	Long short-term memory
MFCC	Mel frequency cepstral coefficients
RBF	Radial basis function
SVM	Support vector machine
WHO	World Health Organization

INTRODUCTION

0.1 Research context

The automated analysis of the cry of newborns and infants is an important topic in biological processes and signals since it is interesting in basic science and relevant to clinical and medical practice (Valdes et al., 2022). Nowadays, studying cry is recommended to be a part of clinical pediatrics to recognize healthy and unhealthy newborns, to monitor sick newborns, and perform the prognosis of newborns recovering from brain damage, and for fast diagnosis of specific abnormalities (Valdes et al., 2022). In this regard, the development of computer-aided diagnosis (CAD) systems based on biomedical signal processing and artificial intelligence systems can provide visual and statistical valuable information to accurately perform an automatic diagnosis task.

The literature on the development of CAD systems for newborn condition detection is receiving growing attention. Hence, the discovery of new patterns in newborn cry records is essential to understand the physiology of pathologies and for the development of accurate CAD systems.

0.2 Statement of research problem

According to the World Health Organization (WHO, 2020), 2.4 million children died in the first month of life in 2019, equating to approximately 6 700 newborn deaths every day, making up 47% of all child deaths under the age of 5-years, up from 40% in 1990. For instance, sub-Saharan Africa had the highest neonatal mortality rate in 2019 at 27 deaths per 1,000 live births, followed by Central and Southern Asia with 24 deaths per 1,000 live births. Indeed, according to the World Health Organization (WHO, 2020), a child born in sub-Saharan Africa or in Southern Asia is 10 times more likely to die in the first month than a child born in a high-income country. Therefore, understanding the physiology of newborn cry is helpful to perform an early and accurate diagnosis. This can determine appropriate medical treatment, thereby increasing life expectancy.

There is recently a growing interest in the application of signal processing and artificial intelligence in biomedical engineering ranging from pattern and biomarker discovery to automatic classification of pathologies. In this regard, extensive research has been devoted to the development of suitable machine learning and data mining systems for accurate classification of biomedical and physiological data. This thesis seeks to extract new patterns from newborn cry records for better understanding and discrimination between pathologies.

0.3 Research objectives

The main purpose of our research study is to develop CAD systems used to distinguish between healthy and unhealthy newborn cry signals in an accurate manner. In addition, we aim to understand the dynamics in healthy and unhealthy newborn cry signals for better understanding of their respective physiology. Specifically, we intend to develop CAD systems based on deep learning to achieve high accuracy classification of healthy and unhealthy newborns. Additionally, we aim to examine the statistical physics of newborn cry signals for better understanding of the nonlinear dynamics in normal and abnormal signals.

The main objectives of our research study follow:

1. Design of deep learning feedforward neural networks for automatic classification of healthy and unhealthy newborn cry signals.
2. Design of advanced deep learning systems for automatic classification of healthy and unhealthy newborn cry signals.
3. Automatic selection of acoustic features and optimization of standard machine learning models for automatic classification of healthy and unhealthy newborn cry signals.
4. Comparison of the results of the models in 1-3.
5. Characterization of newborn cry signal using statistical physics measures.
6. Characterization of newborn cry signal using multiscale statistical physics measures.

Highlighting the discriminative power of statistical physics measures in 5-6 for better understanding of the nonlinear dynamics in newborn cry signals.

0.4 Datasets and pre-processing

In this study, we consider two different datasets. In the first one, recorded newborn signals will be processed by cepstrum analysis to extract coefficients used to describe the original signals in cepstrum domain. These coefficients will be used to train deep learning models. In addition, statistical physics measures will be extracted in the cepstrum domain. Hence, we get labeled cepstrum and their associated statistical physics measures. Besides, the second dataset contains labeled newborn cry signals and their corresponding acoustic features.

The first dataset is used for:

1. Evaluating the deep learning feedforward neural networks for the automatic classification of healthy and unhealthy newborn cry signals.
2. Assessing the performance of advanced deep learning systems for the automatic classification of healthy and unhealthy newborn cry signals.
3. Characterizing newborn cry signal using statistical physics measures.
4. Characterizing newborn cry signal using multiscale statistical physics measures.

The second dataset is used for:

5. Performing automatic selection of acoustic features
6. Evaluating the performance of optimized machine learning models for automatic classification of healthy and unhealthy newborn cry signals.

Given the differing nature of the two datasets, distinct experiments and tasks are being considered. Below, a description of each dataset is provided.

0.4.1 The first dataset and cepstrum analysis

The first dataset is composed of two subsets: expiration (EXP) set and inspiration (INS) set. The EXP set has 2638 cry signals and INS set has 1860 cry signals. Specifically, there are 1319 healthy signals and 1319 unhealthy signals in the EXP set. The group of unhealthy subjects suffers from various pathologies such as diseases affecting the central nervous system, and respiratory system. Other pathologies include blood disorder, chromosomal abnormality, and congenital cardiac anomaly. Besides, there are 930 healthy signals and 930 unhealthy signals in the INS set. To record cry signals, a two-channel sound recorder with a sampling frequency

of 44.1 kHz and a resolution of 16 bits was placed at 10cm to 30cm from the infant. The time duration of each recorded signal is within 2-3 minutes. Each original recorded cry signal has been pre-processed to remove background noise and artifacts. It is also segmented to keep only respiration and expiration episodes. The segmentation task is manually performed by using the Wave Surfer tool.

All infant cry signals have been recorded in the neonatology departments of the following hospitals: Sainte-Justine hospital (Montreal, Canada), and Al-Sahel and Al-Raee hospitals, both in Lebanon. The infants who entered the study are preterm and full term and their respective ages range from 1 to 53 days. The sample includes both healthy and unhealthy babies and both males and females. The group of unhealthy babies suffers from various pathologies such as diseases affecting the central nervous system, and respiratory system. Other pathologies include blood disorder, chromosomal abnormality, and congenital cardiac anomaly. For illustration purpose, Fig.1.1 displays examples of healthy and unhealthy newborn cry signals.

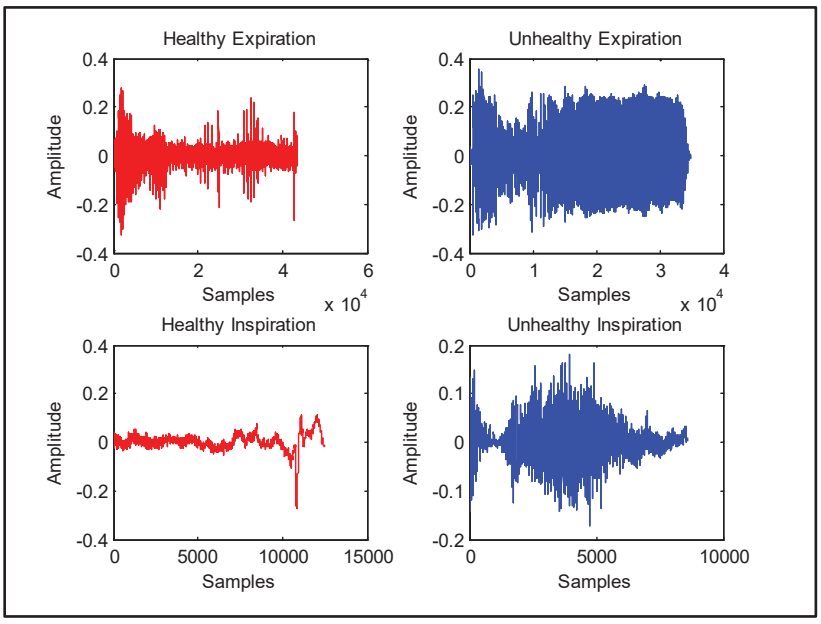

Figure 0.1 Example of healthy and unhealthy newborn cry signals

In this research study, we pre-process the recorded newborn cry signals by using the cepstrum analysis (Childers et al., 1977; Oppenheim, 2010) for two primary reasons. First, the recorded newborn cry signals are not having the same length. However, a similar length across all signals is required for training deep learning neural networks. Thus, the cepstrum analysis is applied to obtain a representation of all recorded signals in the cepstrum domain with the same length. Specifically, the cepstrum analysis provides a vector of coefficients used to describe the original signal. In general, the user determines the number of the resulting coefficients. Hence, we set the same number of representative coefficients for all newborn cry signals when processed by cepstrum analysis to obtain new vectors all having the same length. Second, the cepstrum analysis allows extracting a determined number of coefficients used to provide description of the spectrum envelope and spectral richness and to characterize the harmonic and noise components of the original signal (Travieso et al., 2014).

For instance, the complex cepstrum $\hat{s}(n)$ of a signal $s(n)$ is given by the inverse Fourier transform of its log spectrum as follows:

$$\hat{s}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln S(e^{J\omega})\, e^{J\omega n} d\omega \tag{0.1}$$

where, $S(e^{J\omega})$ is the discrete Fourier Transform of s(n) and is given by:

$$S(e^{J\omega}) = \sum_{n=-\infty}^{\infty} s(n)e^{-J\omega n} = |S(e^{J\omega})|e^{J\theta(\omega)} \tag{0.2}$$

where $|S(e^{J\omega})|$ and $\theta(\omega)$ are respectively the amplitude and the phase spectra. Besides, the real cepstrum of a signal takes into consideration only its amplitude spectrum. Specifically, the real cepstrum $\hat{r}(n)$ of a signal $s(n)$ is given by:

$$\hat{r}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|S(e^{J\omega})|\, e^{J\omega n} d\omega \tag{0.3}$$

In this study, the real cepstrum is calculated from healthy and pathological newborn cry signals. The number of extracted coefficients to represent each newborn cry signal in cepstrum domain is set to 1000. Hence, all cepstrum representations have the same length and could be used for training deep learning networks. There is no rule on how to set the number of

predetermined coefficients; however, we set it at 1000 coefficients to obtain large representation while making the learning process of the deep learning systems fast. Examples of cepstrum coefficients are shown in Fig.0.2.

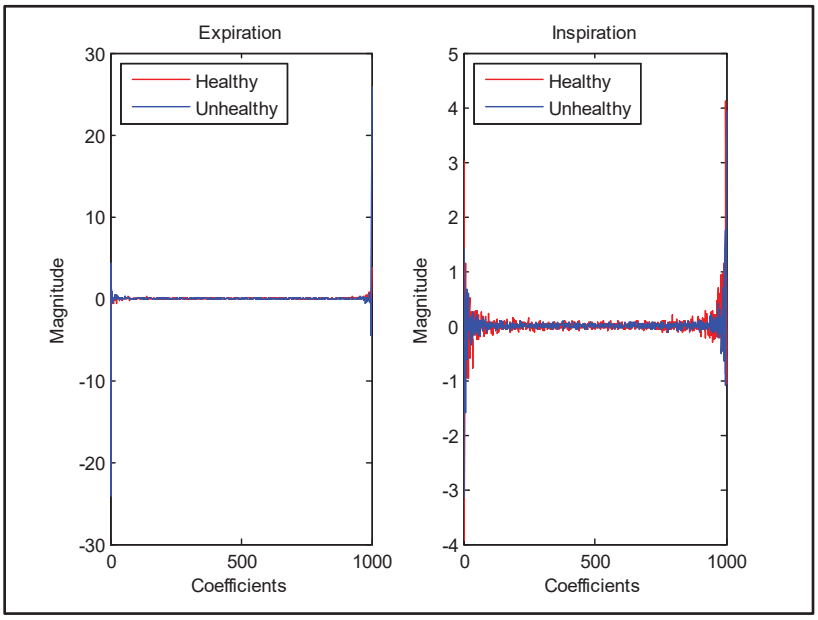

Figure 0.2 Examples of cepstrums from healthy and unhealthy newborn cry signals

0.4.2 The second dataset and acoustic features

The second dataset cry signals recorded from 763 healthy newborns and 320 unhealthy ones under the same conditions as the first dataset. The group of unhealthy subjects suffers from various pathologies such as diseases affecting the central nervous system, and respiratory system. Other pathologies include blood disorder, chromosomal abnormality, and congenital cardiac anomaly. Each labeled cry signal is associated with features set composed of Mel frequency cepstral coefficients (MFCC), auditory-inspired amplitude modulation (AAM) features, and a prosody feature set of tilt, intensity, and rhythm features. This dataset was used in two studies by Matikolaie and Tadj (2022) and by Matikolaie, Kheddache, and Tadj (2022).

0.5 Proposed methodologies

In this research study, we aim to develop various CAD systems for automatic classification of healthy and unhealthy newborn cry signals. Specifically, we design deep learning systems to distinguish between healthy and unhealthy signals. In addition, we aim to assess complexity in newborn cry signals by means of various statistical physics measures and apply formal statistical tests to check if complexity is different across healthy and unhealthy subjects.

For classification task, we develop CAD systems based on deep learning models trained with cepstrum coefficients to be applied to the first dataset (See Fig.0.3). The included deep learning systems are deep feedforward neural networks (DFFNN), long short-term memory (LSTM) neural networks, and convolutional neural networks (CNN).

In addition, we develop CAD systems based on conventional machine learning models tuned by Bayesian optimization (BO) and trained with acoustic features selected by χ^2 to be validated on the second dataset (See Fig.0.4). The set of acoustic features include mel-frequency cepstral coefficients (MFCC), auditory-inspired amplitude modulation (AAM), and prosody; and, the machine learning models included are support vector machine (SVM) with radial basis function (RBF) kernel and k-nearest neighbors (kNN).

Besides, for complexity analysis, various statistical physics will be estimated from cry signals in cepstrum domain. Specifically, we investigate whether complexity measures could discriminate between healthy and unhealthy subjects (See Fig.0.5), and whether multiscale complexity analysis-based features could discriminate between the aforementioned subjects (See Fig.0.6).

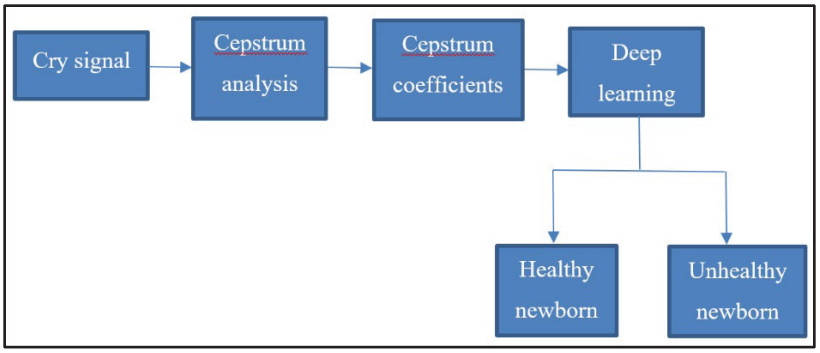

Figure 0.3 Proposed CAD system based on cepstrum analysis and deep learning

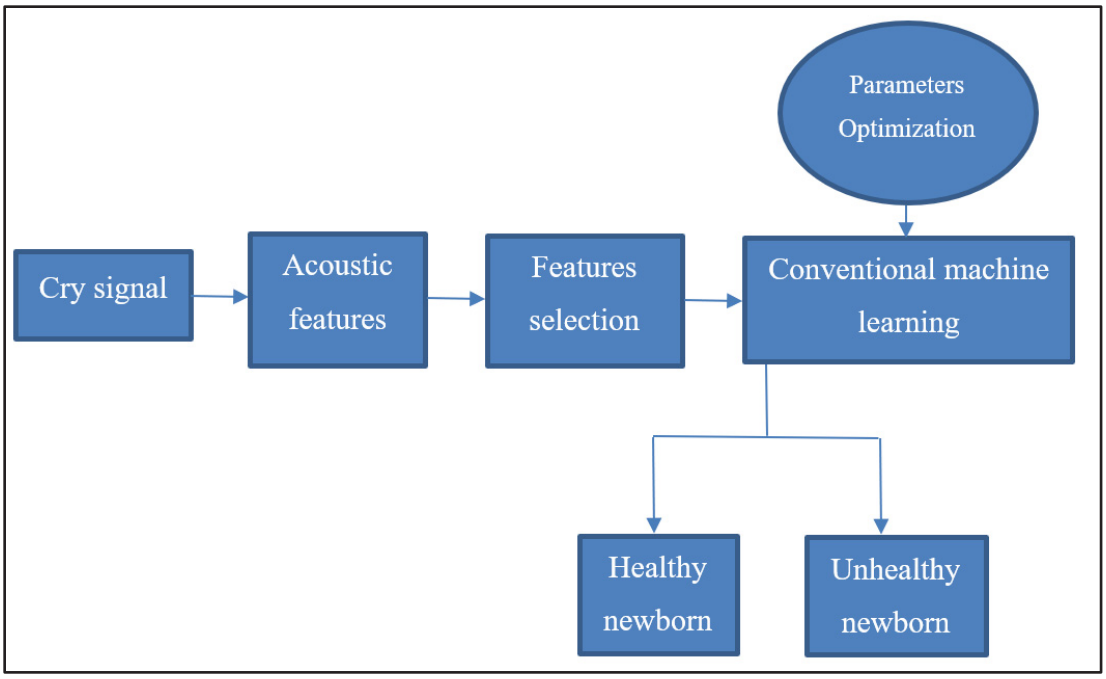

Figure 0.4 Proposed CAD system based on machine learning and acoustics

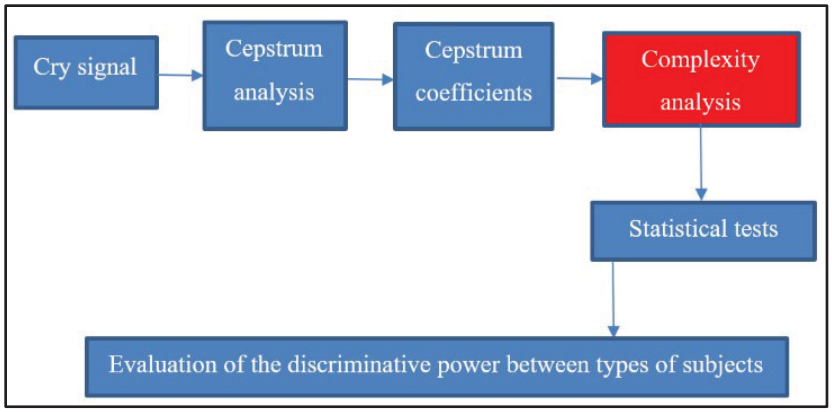

Figure 0.5 Complexity analysis by means of statistical physics measures

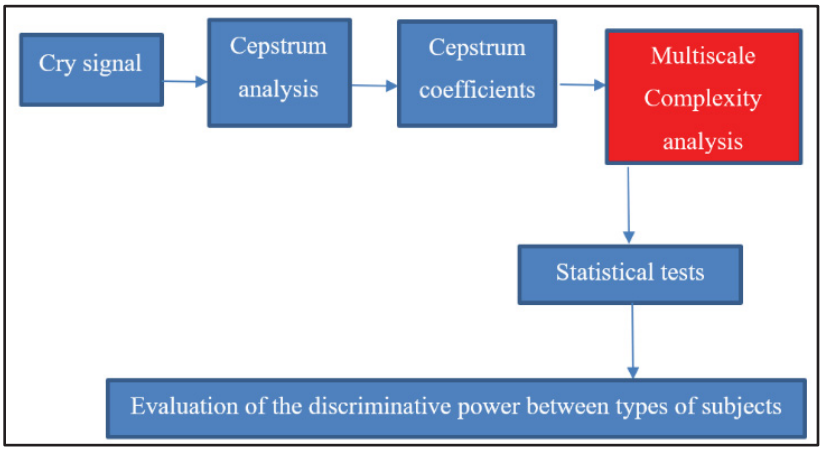

Figure 0.6 Complexity analysis by means of multiscale statistical physics measures

0.6 Organization of the thesis

The current research study is based on five published scientific journal papers. For CAD systems used to classify healthy against unhealthy newborn cry signals, we published three journal papers on deep learning and features selection and optimization of machine learning models. Besides, for assessment of complexity using statistical physics measures in newborn cry signals we published two journal papers. They are described next.

In the first journal paper, we propose a new CAD system to distinguish between healthy and unhealthy infant cry signals. The proposed CAD system is composed of four major steps. First,

the original cry signal is analyzed by cepstrum analysis to obtain its cepstrum coeffcients. Second, the obtained cepstrum coefficients are fed to a deep feedforward neural network (DFFNN) for training and classification. Third, the performance of the cepstrum-DFFNN system is evaluated by standard classification performance metrics. The results are compared to baseline classifiers including the support vector machines, Naïve Bayes, and probabilistic neural networks. (*Lahmiri, S., Tadj, C., Gargour, C. (2021). Biomedical diagnosis of infant cry signal based on analysis of cepstrum by deep feedforward artificial neural networks. IEEE Instrumentation & Measurement Magazine, 24 (2), 24-29*).

In the second journal paper, we implement deep feedforward neural networks (DFFNN), long short-term memory (LSTM) neural networks, and convolutional neural networks (CNN) all trained with cepstrum analysis-based coefficients as inputs to distinguish between healthy and unhealthy infant cry records. Then we evaluate and compare their respective performances. (*Lahmiri, S., Tadj,C., Gargour, C., Bekiros, S. (2022). Deep learning systems for automatic diagnosis of infant cry signals, Chaos, Solitons & Fractals, 154, 111700*).

In the third journal paper, we propose to use Bayesian optimization method to optimize the hyper-parameters of support vector machine with radial basis function kernel and k-nearest neighbors trained with different audio features separately or combined; namely, mel-frequency cepstral coefficients (MFCC), auditory-inspired amplitude modulation (AAM), and prosody. Particularly, the chi-square test is applied to each set of features to select the best features to train optimal classifiers. (*Lahmiri, S., Tadj, C., Gargour, C., Bekiros, S. (2023). Optimal tuning of support vector machines and k-NN algorithm by using Bayesian optimization for newborn cry signal diagnosis based on audio signal processing features, Chaos, Solitons & Fractals, 167, 112972*).

In the fourth journal paper, we characterize infant normal and pathological cry signals by studying their respective oscillations by means of approximate entropy and correlation dimension estimated from their respective cepstrums. After estimating approximate entropy and correlation dimensions from cepstrums, three standard statistical tests are applied to them

including the Student t-test, F-test, and two-sample Kolmogorov-Smirnov test. (*Lahmiri, S., Tadj, C., Gargour, C., Bekiros, S. (2021). Characterization of infant healthy and pathological cry signals in cepstrum domain based on approximate entropy and correlation dimension. Chaos, Solitons & Fractals, 143, 110639*).

In the fifth journal paper, we examine the multifractal behavior in the cepstrum representation of healthy and unhealthy infant cry signals by means of wavelet leaders and we compare their discriminative power by applying the Student t-test. (*Lahmiri, S., Tadj, C., Gargour, C. (2022). Nonlinear Statistical Analysis of Normal and Pathological Infant Cry Signals in Cepstrum Domain by Multifractal Wavelet Leaders. Entropy, 24, 1166.*).

The remaining of the thesis is organized as follows. The first chapter is dedicated to the review of literature and contributions. The second chapter to seventh chapter respectively presents the above journal papers. Finally, in the last chapter, we summarize the findings and provide future research directions.

CHAPITRE 1

REVIEW OF LITERATURE AND CONTRIBUTIONS

1.1 Review of literature

Nowadays, automatic voice pathology detection - based on non-invasive techniques such as analysis and classification of patients' speech or cry records - is receiving a growing attention in the design of computer-aided diagnosis systems. Indeed, the medical diagnosis of diseases based on voice signal analysis and classification is a cost-effective and non-invasive solution that may provide promising accurate performance. Besides, evaluation of complexity in voice signal is helpful to understand the dynamics and physiology of such records. In this subsection, we review studies on voice analysis for general pathology detection, newborn and enfant cry analysis and classification, and measurement of nonlinear statistical properties of voice records to assess regularity in their physiological fluctuations.

1.1.1 Voice analysis for pathology detection

In the problem of voice analysis for pathology detection, Lahmiri (2017) investigated the effect of different dysphonia measurements on Parkinson's disease and found that the support vector machine trained with vocal fundamental frequency measures yielded 88% accuracy, 94% accuracy when all dysphonia measurements are employed, and 97% accuracy by removing dysphonia measurements with similar variation across healthy and PD subjects. Lahmiri and Shmuel (2019) assessed the performance of eight different feature selection methods applied to voice patterns coupled with nonlinear support vector machine to distinguish between Parkinson's disease and healthy control subjects. The parameters of the radial basis function kernel of the SVM classifier were optimized by using Bayesian optimization technique and the obtained results showed that the receiver operating characteristic and the Wilcoxon-based ranking techniques provide the highest performance in terms of sensitivity and specificity. Chen and Chen (2020) trained deep neural networks based on stacked autoencoder with Mel frequency cepstral coefficients to detect pathologies in voice records. The proposed system

yielded 98.6% accuracy compared to the support vector machine (92.9%) and random forest (92.2%). In the work by Hammami et al (2020), processed voice records in variational mode decomposition domain from which the features are extracted. The support vector machine achieved 99.29% accuracy in pathology detection when trained with high order statistics and 93.1% when trained with discrete wavelet transform coefficients. Souli et al (2021) proposed a convolutional neural network trained with scatter wavelet features to detect pathologies in speech. The proposed system achieved 99.62% recognition rate. Ankışhan and İnam (2021) proposed a multi-modal features approach to train convolutional neural network and long short-term memory for pathological voice disorders. The proposed system achieved 93.05% to 99.58% correct classification rate of healthy versus pathological voices. Guo et al (2022) investigated voice quality in Mandarin-speaking children with autism spectrum disorder using multidimensional acoustic parameters. The random forest classifier achieved 78.5% accuracy rate with shimmer and jitter measures. Chaiani et al (2022) combined convolutional neural network with long short-term memory trained with spectrogram-based features and activated by using a new sinusoidal rectified unit function. The proposed system achieved an accuracy of 70.62% in the classification of four voice pathologies including vocal cordectomy, psychogenic dysphonia, pachydermia laryngis, and frontolateral partial laryngectomy. Kumar et al (2023) used 2D convolutional neural network to classify voice disorders from kymograms. The deep classifier achieved 94.237% to distinguish between healthy and disorders, 94.8% to distinguish between healthy and muscle tension dysphonia, and 93.1% to distinguish between healthy, functional, and organic disorders. Zhang et al (2023) trained the random forest classifier with features extracted based on the asymmetric fluid–structure interaction vocal cord model. The obtained accuracy was 77.18% in classification of various vocal cord pathological voices. In Shaikh et al (2023), both the 1D convolutional neural network trained with short-time Fourier transform and 2D convolutional neural network trained with Mel-frequency cepstral coefficients achieved 97.10% accuracy in classification of normal against various pathological voices. To detect dysphagia in voice signals, Kim et al (2023) designed an ensemble system where convolutional neural networks are trained with short time Fourier transform and Mel-frequency cepstral coefficients and the soft voting method is applied to distinguish between normal and dysphagia classes. The proposed ensemble system achieved

an accuracy varying between 78.20% and 81.20%. To distinguish Parkinson's disease patients from healthy controls, Zhang et al (2023) used the method of co-occurrence direction attribute topology to obtain structural features of the voice signal in the time-frequency domain. The random forest achieved 95.84% and 93.90% on Chinese and Turkish language voice dataset respectively.

1.1.2 Newborn and enfant cry signal diagnosis

In the problem of enfant cry signal diagnosis and classification, the medical diagnosis of newborn diseases based on cry signal analysis and classification is a cost-effective and non-invasive solution that may provide promising accurate performance. Indeed, newborn cry signal analysis and classification is an up-to-date approach to build computer-aided diagnosis (CAD) systems used to distinguish between healthy and unhealthy newborns. In this regard, several CAD systems have been proposed and evaluated by biomedical engineering researchers as the topic is attracting a growing interest.

Indeed, most of the existing works used standard audio processing features typically estimated in frequency, time and cepstral domains to train machine learning (ML) classifiers to distinguish between healthy and unhealthy cry signals of newborns. Very early studies include detection of hypoxia-related disorder using radial basis function neural networks with 85% accuracy (Poel and Ekkel, 2006) and detection of asphyxia with a classification accuracy of 95.86% by combining principal component analysis and support vector machines (Sahak et al, 2010).

In the problem of enfant cry signal diagnosis and classification, Rosales-Pérez et al (2015) used Mel frequency cepstral coefficients and Linear predictive coding for feature extraction and genetic selection of a fuzzy model for classification of normal and pathological cry. The proposed automatic system obtained 99.42% accuracy. Chittora et al (2016) used spectrogram-based energy features to examine ten different types of infant cry pathologies. The support vector machine classifier achieved 98.22% accuracy. Sachin et al (2017) designed an automatic system to detect neonatal asphyxia based on support vector machines trained with Mel frequency cepstrum. The proposed model yielded 92% accuracy. Lim et al (2018) extracted

energy features from the dual-tree complex wavelet packet transform representation of infant cry signal and employed different methods of features selection to identify energy features to be fed to the extreme learning machines. They obtained 97.87% accuracy in classification of asphyxia versus normal, 100% accuracy in classification of deaf versus normal, and 87.26% accuracy of hunger versus pain. Relevant studies found in the last decade were also interesting. For instance, the authors employed the Multiple Mixtures of Gaussian (MMG) algorithm to segment newborn cry records; then, Mel Frequency Cepstral Coefficients (MFCC) are estimated from the segmented records and used to train Hidden Markov Models (HMM) which achieved 83.79% accuracy (Abou-Abbas et al., 2015). Kheddache and Tadj (2015) examined the association between acoustic measurements (resonance frequencies; for instance) and pathologies to identify the most relevant ones. They found that the distributions of acoustic cry acoustics statistically and significantly vary with the pathology of newborn. In another study, dynamic and static features derived from Mel-Frequency cepstral coefficients (MFCC) of both expiratory and inspiratory cry vocalizations were used to train various classifiers; namely, multilayer perceptron (MLP) using the back-propagation algorithm, probabilistic neural networks (PNN) and a support vector machine (SVM) (Alaie et al., 2016). The MLP, SVM with MLP kernel and PNN achieved respectively 91.68%, 90.41%, and 89.93% maximum accuracy rate depending on the Gaussian Mixture Model adaptation method used in experiments. Abou-Abbas et al (2017) considered the problem of cry signal segmentation and classification to distinguish between expiratory and inspiratory phases. In this regard, the original cry signal is decomposed in time and frequency domains from which features are extracted and used to train Gaussian mixture models (GMM) and HMM. They respectively achieved 8.9% and 11.06% error classification rates. The effectiveness of acoustic features (fundamental frequency glide and resonance frequencies dysregulation) and conventional features (MFCC) on the performance of PNN when used to classify healthy and pathological cry signals was examined by Kheddache and Tadj (2019). The best result obtained is 88.71% for the correct classification of healthy preterm newborns and 82% for correct classification of unhealthy full-term newborns.

In recent studies, short-term and long-term features from different timescales were combined to train SVM to distinguish between healthy newborn and those suffering from respiratory distress syndrome to achieve 68.40% correct classification rate (Matikolaie and Tadj, 2020). In Matikolaie et al (2022), PNN and linear SVM were trained by three different feature sets including MFCC set, auditory-inspired amplitude modulation (AAM) set, and prosody set composed of tilt, intensity, and rhythm features. The linear SVM outperformed the PNN in terms of accuracy under MFCC (76.50% versus 68.90%), AAM (75.75% versus 70.70%), and prosody (61.50% versus 52.10%). The fusion of MFCC and AAM yielded to the best correct classification rate obtained by linear SVM (78.70%) and PNN (77.90%). In another very recent study, the authors (Matikolaie and Tadj, 2022) compared the performance SVM under various kernels, different models of decision trees, and variants of discriminant analysis and principal component analysis for dimension reduction of MFCC features set. The SVM with quadratic kernel and trained with all features achieved the highest F-score (86%) on expiration set whilst the quadratic discriminant analysis trained with tilt features set yield to the highest F-score (83.90%) in inspiration set.

There are also other studies related to the classification of infant cries. For instance, Anders et al (2020) studied the performance of the convolution neural networks in classification of infant vocalization sequences including crying, fussing, babbling, laughing, and vegetative vocalizations. The convolution neural networks yielded 72% accuracy when trained with spectrograms. Ashwini et al (2021) proposed a system to classify neonatal cries into pain, hunger, and sleepiness. First, short-time Fourier transform was applied to the neonatal cry record to obtain its spectrogram image. Second, the spectrogram is processed by deep convolutional neural network to extract its deep features. Finally, the support vector machine was employed for classification. Trained with radial basis function, the support vector machine achieved 88.89% accuracy. Ting et al (2022) investigated the use of hybrid features of MFCC, chromagram, Mel-scaled spectrogram, spectral contrast and Tonnetz and deep learning models in classifying asphyxia cry in infant subjects. They found that convolution neural networks performed better than deep neural networks when all trained with MFCC features. In addition, the deep neural network models performed better with hybrid features compared to that with single feature of MFCC. Furthermore, the deep neural networks model with multiple hidden

layers achieved an accuracy of 100% in classifying normal and asphyxia cry, and 99.96% for non-asphyxia and asphyxia cry when the hybrid features were used. Abbaskhah et al (2023) trained the support vector machine, multilayer perceptron, and convolutional neural Network classifiers with Mel frequency cepstral coefficients and entropy to distinguish between five classes of infant cries. Without synthetic minority over-sampling technique, the support vector machine, multilayer perceptron, and convolutional neural Network classifiers respectively achieved 82.3%, 87.6%, and 92.1% average accuracy. With synthetic minority over-sampling technique, they respectively obtained 86.1%, 89.2%, and 91.1% average accuracy. In an interesting study, Ozseven (2023) investigated the effect of using hand-crafted features and spectral images individually and hybrid in the classification of infant cries. The results show that texture analysis methods are insufficient and that hand-crafted feature sets and spectrogram and scalogram images provide high success. The pre-trained deep learning models ShuffleNet and ResNet-18 achieved respectively 97.57% and 95.17% accuracy by using scalogram features. The review of literature on design of CAD systems for cry diagnosis is provided in Table 1.1.

Table 1.1 Summary of literature review

Study	Detection problem	Machine learning	Findings
Poel and Ekkel (2006)	Hypoxia	Radial basis function neural	85% accuracy
Sahak et al (2010)	Asphyxia	Support vector machines	95.86% accuracy
Abou-Abbas et al (2015)	Newborn diagnosis based on acoustics	Hidden Markov models	83.79% accuracy
Kheddache and Tadj (2015)	Associative analysis between acoustics and pathologies		Distributions of acoustic cry acoustics statistically and significantly vary with the pathology of newborn
Alaie et al (2016)	Newborn diagnosis based on acoustics	Multilayer perceptron	91.68% accuracy
Abou-Abbas et al (2017)	Newborn diagnosis based on acoustics	Gaussian mixture models	8.9% error classification rate
Kheddache and Tadj (2019)	Newborn diagnosis based on acoustics	Probabilistic neural networks	88.71% correct classification of healthy preterm newborns and 82% correct classification of unhealthy full-term newborns
Matikolaie and Tadj (2020)	Newborn diagnosis based on acoustics	Linear support vector machines	68.40% accuracy
Matikolaie et al (2022)	Newborn diagnosis based on fusion of specific acoustics	Linear support vector machines	78.70%
Matikolaie and Tadj (2022)	Newborn diagnosis based on acoustics	Quadratic support vector machines	86% F-score

1.1.3 Complexity in physiological signals

The ability to discern levels of complexity within physiological signals has become increasingly important. Indeed, measuring complexity in physiological signals helps revealing their intrinsic properties and understanding their fluctuations. For instance, approximate entropy was used to characterize surface electromyographic signals (Ahmad and Chappell, 2008), cardiac synchrony (Cullen et al, 2010), gait rhythm (Wu et al, 2017), and epileptic seizure (Gao et al, 2020; Kumar et al, 2014). In addition, correlation dimension was successfully used to analyze electroencephalogram records (Acharya et al, 2005; Mekler, 2008), heart rate variability (Carvajal et al, 2005; Rawal et al, 2015), epileptic seizures (Shayegh et al, 2014), and organization of resting state cortical networks (Kalauzi et al, 2015). Furthermore, the correlation dimension was employed in electroencephalogram compression and transmission in clinical milieu (Sriraam, 2012).

Moreover, in recent years, multifractal analysis has received a growing attention in biomedical engineering. For example, it was applied in the analysis of various biophysiological signals including EEG (Zorick et al, 2013; Zhang et al, 2015; Souza et al, 2018), ECG (Gadhoumi et al, 2018; Orozco-Duque et al, 2015), magnetic resonance images and brainstem volume (Rohini, 2020), mammograms (Gerasimova-Chechkina et al, 2014; 2016), bone radiographic images (Borowska, 2019), retina digital images (Wang et al, 2020), dental implant ultrasonic signal (Scala et al, 2018), and liver tissue images (Oprić et al, 2020), to name few. Very recently, Lahmiri (2023a) employed employ multi-scale analysis (MSA) to estimate generalized Hurst exponent (GHE) from electrocardiogram (ECG) records under congestive heart failure (CHF), arrhythmia (ARR), and normal sinus rhythm (NSR). In addition, Lahmiri (2023b) proposed a wavelet leaders method with multiscale entropy measures to analyze multiscale complexities in ECG signals to characterize arrhythmia, congestive heart failure, and normal sinus rhythm.

It is worth to mention that all of the aforementioned studies showed that measures of statistical physics are useful to reveal intrinsic characteristics of physiological signals (for instance, surface electromyographic signals, cardiac synchrony, gait rhythm, epileptic seizure,

electroencephalogram records, heart rate variability, epileptic seizures, and organization of resting state cortical networks) and understand their respective fluctuations.

1.2 Contributions of the thesis

The review of literature shows that there are variety of CAD systems that have been used to distinguish between normal and abnormal cry records of infants. In this regard, moderate to high correct classification rates have been achieved by using standard machines learning such as SVM, PNN, HMM, and MMG models trained with variety of acoustic features.

The goal of the current study is to design new CAD systems to improve correct classification rate of healthy and unhealthy newborn cry signals. In addition, we aim to investigate the statistical discriminative power of complexity measured by statistical physics.

Our contributions follow:

(a) Design of CAD systems based on deep learning systems trained with cepstrum based coefficients to achieve higher classification rate compared to conventional machine learning models used in the literature. Such improvement would help physicians in the process of diagnosis of newborns and infants based on voice records following a non-invasive approach.

(b) Design of CAD systems by using standard machine learning models to distinguish between healthy and unhealthy newborns based on analysis and classification of the acoustics of their cry signals. In this regard, contrary to the literature, we employ Bayesian optimization (BO) method to optimize the hyper-parameters of the SVM and k-nearest neighbors (kNN) algorithm. In addition, we seek to investigate how various combinations of acoustics-based features could improve classification of optimally configured SVM and kNN. Furthermore, since the number of acoustic features is very large, we select the most significant ones to speed up the learning process while improving accuracy. We expect to achieve higher performance compared to previous works that employed standard machine learning models too.

(c) Investigate how complexity measures are different across normal and abnormal newborn cry signals. Specifically, we intend to characterize infant normal and

pathological cry signals by studying their respective oscillations by means of approximate entropy and correlation dimension estimated from their respective cepstrums. Such investigations allow discovering how much signal regularity and information structure are different across types of subjects.

(d) Use of multifractal analysis to obtain a full description of nonlinear dynamics in newborn cry signals to understand how much the information regarding short and long variations in the signal are different across subjects.

(e) The contributions in (c) and (d) would help understanding the physiology of cry signals across healthy and unhealthy subjects. In other words, one could understand the nonlinear dynamics and complexity in pathological signals compared to normal ones.

CHAPITRE 2

BIOMEDICAL DIAGNOSIS OF INFANT CRY SIGNAL BASED ON ANALYSIS OF CEPSTRUM BY DEEP FEEDFORWARD ARTIFICIAL NEURAL NETWROKS

Salim Lahmiri, Chakib Tadj, Christian Gargour

Electrical Engineering Department, École de Technologie Superieure, Montreal, Canada

Paper published in «IEEE Instrumentation & Measurement Magazine», April 2021

Résumé

Dans ce travail, nous proposons un nouveau système automatique pour distinguer les signaux de plusieurs nourrissons sains et malades. Le système automatique proposé est composé de quatre étapes principales. Tout d'abord, le signal de cri d'origine est prétraité pour supprimer le bruit de fond et les artefacts. Cette étape comprend également la segmentation du signal pour différencier les épisodes d'expiration et de segmentation. Deuxièmement, le signal de cri prétraité résultant est analysé pour obtenir son cepstre. Troisièmement, les coefficients de cepstre obtenus sont transmis à un réseau de neurones profond à propagation avant pour la classification des cris des nourrissons. Quatrièmement, la performance du système cepstre et réseaux profonds est évaluée par des mesures de performance de classification standard ; incluant le taux de classification correcte, la sensibilité et la spécificité. En effet, il performe mieux que les réseaux de neurones probabilistes, les machines à supports de vecteurs, et le classificateur bayésien.

Mots clés : cris de nourrissons, cepstrum, réseaux de neurones profonds à propagation avant, apprentissage machine, classification

2.1 Abstract

In this work, we propose a new CAD system to distinguish between healthy and unhealthy infant cry signals. The proposed CAD system is composed of four major steps. First, the original cry signal is pre-processed to remove background noise and artifacts. This step also

includes signal segmentation to differentiate between expiration and segmentation episodes. Second, the resulting pre-processed cry signal is analyzed to obtain its cepstrum. Third, the obtained cepstrum coefficients are fed to a deep feedforward neural network (DFFNN) for training and classification. Fourth, the performance of the cepstrum-DFFNN system is evaluated by standard classification performance metrics.

Keywords: infant cry signal, cepstrum analysis, deep feedforward neural network, machine learning, classification

2.2 Introduction

In this work, we propose a new CAD system to distinguish between healthy and unhealthy infant cry signal. The proposed CAD system is composed of four major steps. First, the original cry signal is pre-processed to remove background noise and artifacts. This step also includes signal segmentation to differentiate between expiration and segmentation episodes. Second, the resulting pre-processed cry signal is analyzed to obtain its cepstrum. Third, the obtained cepstrum coefficients are fed to a deep feedforward neural network (DFFNN) for training and classification. Fourth, the performance of the cepstrum-DFFNN system is evaluated by standard classification performance metrics.

The cepstrum is widely employed in audio signal analysis as it provides description of the spectrum envelope and spectral richness and characterizes the harmonic and noise components of the original signal (Travieso et al, 2014). Moreover, in last recent years, there has been a growing interest in deep learning in various engineering and science problems thanks to its ability to extract deep features and achieve high accuracy compared to existing machining learning techniques. However, the application of deep learning to the problem of infant cry signal classification for medical diagnosis has not been explored in the biomedical literature, except in the classification of baby cry signal under different domestic environment conditions (Cohen et al, 2020). In this work, we focus on deep feedforward neural network as it has deeper architectures compared to standard feedforward neural network, which allow the input data to be analyzed and transformed multiple times to generate the output (Luo et al, 2020). In this

regard, multiple hidden layers make the DFFNN more appropriate for comprehensive data (LeCun et al, 2015; Schmidhuber, 2015). In addition, DFFNN is faster compared to most common deep learning artificial neural network such as convolutional neural network and long short-term memory network.

The proposed CAD system for infant cry signal analysis and classification is shown in Fig. 2.1, in which the original cry signal is denoised and segmented. Then, it is processed to obtain its ceptrum signature. The latter is employed to train a DFFNN which is used to distinguish between healthy and unhealthy infant cry signal. For comparison purposes, Naïve Bayes, support vector machine, and probabilistic neural network are employed as baseline classifiers. In this regard, we seek to show that the DFFNN is effective compared to the aforementioned, and investigation of various combinations of DFFNN architectures and related functions are out of scope of the current work. The proposed CAD system for infant cry analysis and classification will be applied to a large data set obtained from a Canadian hospital located in Montreal, Quebec and two other hospitals located in Lebanon.

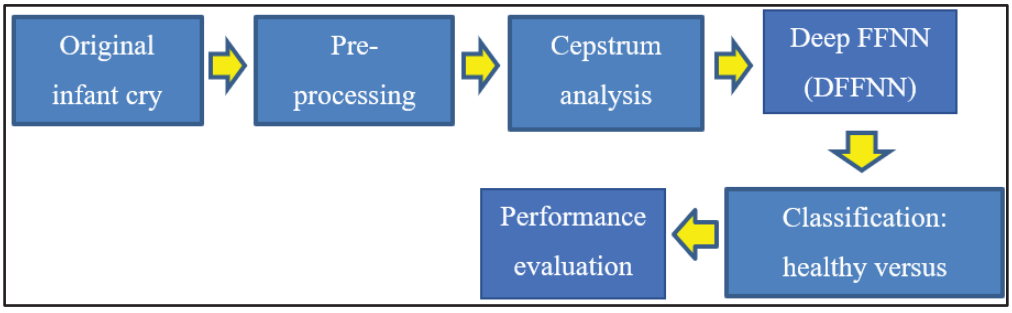

Figure 2.1 Proposed CAD system based on cepstrum analysis and DFFNN

2.3 Methods

2.3.1 Deep feedforward neural networks

The standard feedforward neural network (FFNN) is an artificial neural network with one hidden layer used to process the inputs. Besides, the deep feedforward neural network (DFFNN) has several hidden layers. Specifically, the information in DFFNN moves from the input layer through the hidden layers to the output layer, and there is no feedback or loop in the network (Ketkar, 2017) which basically makes it deep and fast.

The architecture of FFNN used for infant cry classification is presented in Fig. 2.2. Accordingly, there are one input layer, three hidden layers, and one output layer. The number of neurons is set to 1000 in the input layer and in each hidden layer. The number of neurons in the output layer is set to one to represent the class label: either a healthy cry record or an unhealthy cry record. Each hidden neuron is used to process output information of the input layer according to the following expression:

$$h(X)=wX+b \tag{2.1}$$

where X is the input vector, w is the matrix of weights, and b is the bias vector. In general, the neuron of the output is nonlinearly processed by a given activation function such as sigmoid, tanh, ReLU, and ELU functions. For instance, the input h(X) is processed by the nonlinear activation function. In our study, the sigmoid function is chosen as the activation function and it is given by:

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

Finally, the output neurone is expressed as follows:

$$y = Sigmoid\left(\sum h(X)\right) \tag{2.3}$$

Recall that the information fed to each neuron in each hidden layer is processed by a sigmoid activation function. The resilient backpropagation algorithm is chosen as the training function since it is fast and does not require large memory during computation. Finally, the number of epochs is set to 10, and the learning rate is set to 0.01.

Recall that there is no rule of thumb on how to configure the entire architecture of the DFFNN along with its parameters. In general, they are fixed, according to the experience of the user. In our study, the number of hidden layers is set to four for deeper analysis of the cepstum coefficients compared to two and three hidden layers and to obtain acceptable computation processing time. Also, the number of epochs is set to 10 for fast processing of the input signal. Finally, a learning rate of 0.01 is a good compromise between convergence and required

processing time. Indeed, our goal is to design an effective and fast DFFNN system for analysis and classification of ceptrum coefficients.

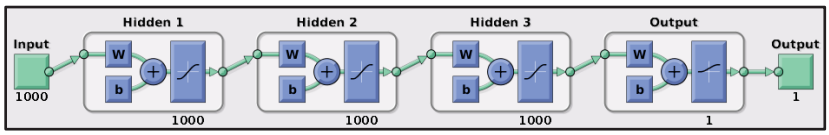

Figure 2.2 DFFNN with three hidden layers

2.3.2 Baseline classifiers: SVM, NB, and PNN

The SVM classifier (Vapnik, 1995) is a supervised learning algorithm based on statistical learning theory used to determine a hyper plane. The latter optimally splits two classes by learning a train data set. For instance, let $\{x_i, y_i\}_{i=1}^{n}$ where x is the input vector, and y is the class label. The classification decision function is expressed as follows:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j) + b\right) \tag{2.4}$$

where α_i is the Lagrange multipliers, $K(x_i,x_j)$ is a linear kernel function, and b is a constant parameter.

The Naïve Bayes classifier (Hastie et al, 2008) is based on estimation of probabilities to assign the membership of an input vector x to a particular class y. The Bayes' theorem can be used to express the conditional probability of class label y given input vector x as follows:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} \tag{2.5}$$

where $P(y|x)$ is the posterior probability of the compound class, $P(x|y)$ is the conditional probability that a compound has certain features given its class y, $P(y)$ is the prior probability estimated from the training set, and $P(x)$ is the marginal probability of observing the given features in the dataset.

The probabilistic neural network (PNN) (Specht, 1990) is composed of three layers: the input layer, pattern layer, summation layer, and output layer. The input layer has 1000 neurons, corresponding to the number of cepstrum coefficients. The second layer consists of 1000 neurons where each one is represented by a Gaussian transfer function. The pattern layer has two neurons where each one is used to represent a specific class. The pattern layer is used to perform an average operation of the outputs from the pattern layer for each class. Finally, the output layer has one neuron used to compute the maximum sum as follows:

$$y = argmax_i\{P_i(y)\} \tag{2.6}$$

where,

$$P_i(y) = \frac{1}{N_i}\sum_{j=1}^{N_i} Q_{ij}(y) \tag{2.7}$$

and,

$$Q_{ij}(y) = \frac{1}{(2\pi)^{0.5}\delta} exp\left(\frac{-(y - x_{ij})^T(y - x_{ij})}{2\delta^2}\right) \tag{2.8}$$

where $P_i(y)$ represents the probability that a test sample x belongs to class y, Q_{ij} represents the standard probability density function (PDF), δ is a smoothing parameter set to unity, and N is the number of neurons in the input layer.

2.3.3 Performance measures and protocol of experiments

The performance of each classifier is measured by computing the accuracy, sensitivity, and specificity. The accuracy is the ratio of the correct predictions to the total number of predictions, sensitivity measures the proportion of positive predictions that are correctly identified over all positive cases, and specificity measures the proportion of negative predictions that are correctly identified. Hence, accuracy, sensitivity, and specificity are expressed as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{2.9}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2.10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2.11}$$

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

To evaluate the performance of each classifier while avoiding overfitting, 10-fold cross-validation protocol is adopted in our study. For instance, under 10-fold cross-validation protocol, the data set is divided into 10 subsets. Each time, one different subset is used as the test set, and the remaining nine subsets are put together to form a training set. Then the average performance across all 10 trials is calculated. In the current work, the average and standard deviation of accuracy, sensitivity, and specificity are calculated across the 10 folds. Also, we consider an additional experimental protocol where the data is randomly separated into 50% for learning and 50% for testing.

2.4 Experimental results

We used the first dataset described in Introduction (Section 0.4.1). Fig. 2.3 compares the classification results from deep feedforward (DFFNN), linear support vector machine (SVM), Naïve Bayes (NB), and probabilistic neural network (PNN) following 10-fold cross-validation protocol when applied to EXP and INS sets, respectively. For both EXP and INS sets, the DFFNN outperforms the linear SVM, NB, and PNN in terms of accuracy, sensitivity, and specificity.

Specifically, in the problem of classifying EXP cry signals, the DFFNN, SVM, NB, and PNN achieved an accuracy of 99.92%±0.00, 61.15%±0.04, 58.11%±0.01, and 56.71%±0.01, respectively. In addition, the obtained sensitivity is 99.85%±0.00, 61.03%±0.04, 56.31%±0.01, and 57.70%±0.03, respectively, for DFFNN, SVM, NB, and PNN. In terms of specificity, the

DFFNN, SVM, NB, and PNN obtained 100%, 61.27%±0.05, and 59.93%±0.02, and 55.72%±0.02, respectively.

Moreover, in the problem of classifying INS cry signals, the DFFNN achieved perfect accuracy, sensitivity, and specificity. The linear SVM, NB, and PNN, respectively, achieved an accuracy of 59.57%±0.01, 55.46%±0.02, and 52.63%±0.05, a sensitivity of 58.82%±0.04, 55.84%±0.01, and 47.10%±0.07, and a specificity of 60.32%±0.04, 55.08%±0.02, and 58.17%±0.06.

Finally, Table 2.1 provides the performance measures of each predictive model when the data set is randomly split into 50% for training and 50% for testing. As shown, the DFFNN outperforms all classifiers on both EXP set INS sets. In summary, the DFFNN performs best in distinguishing between healthy and unhealthy infant cry signals. It is followed by the SVM, and the NB classifier performs the worst. In terms of processing time for training and classifying EXP and INS sets under 10-fold cross-validation protocol, the NB is the fast one, followed by PNN, DFFNN and SVM, respectively, as shown in Table 2.2.

Table 2.1 Experimental results from random split of the data into 50% learning and 50% testing

	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
	Expiration set			**Inspiration set**		
SVM	56.63%	59.79%	53.48%	54.62%	51.61%	57.63%
NB	57.32%	55.91%	58.73%	57.10%	55.48%	58.71%
PNN	53.83%	52.88%	54.78%	52.80%	44.73%	60.86%
DFFNN	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Table 2.2 Processing time (in seconds) of each classifier under 10-fold cross-validation

	Expiration (EXP) set	**Inspiration (INS) set**
SVM	653.14	73.57
NB	3.18	1.46
PNN	19.57	14.30
DFFNN	29.30	31.63

Figure 2.3 Bar plots of performance measures under 10-fold cross-validation

The DFFNN outperforms all classifiers and requires less than one minute to converge. Its superiority can be explained by the ability to capture complexities inherent in the data structure. Specifically, as they have high level of abstraction, the DFFNN can fully account for the complex relationships between cepstrums and their corresponding classes. Moreover, the slight improvement in performance measures observed for the INS set that yield to perfect accuracy by DFFNN could be attributed to its capability to learn and model large variability in cepstrums, which is clearly observed in Fig.0.2 in Introduction.

In addition, it is worth mentioning that the NB classifier is fast since it requires only a small training data set to estimate the probabilities. However, it performed the worst, which is most likely linked to violation of the assumption of independent predictors, for instance, cepstrum coefficients. Furthermore, it is worth mentioning that the performance of the linear SVM is moderate, possibly because its key parameters have not been optimized, and it is very slow compared to NB and DFFNN due to the complexity of the optimization process that underlies the SVM and the complexity of the data under study. The accuracy of the PNN is the lowest, and this could be explained by the fact that this kind of artificial neural network employs a Gaussian transfer function to process the inputs. In this regard, its performance depends on the distribution of the inputs and on the value of the smoothing parameter used to determine the length of the probability density function (PDF) of the transfer function.

Finally, it is worth mentioning that the DFFNN outperformed most recent studies in distinguishing between healthy and unhealthy infant cry signals in terms of accuracy, including Hidden Markov Models trained with segmented cry signals (83.79%) (Abou-Abbas et al, 2015), probabilistic neural network trained with prevalence of fundamental frequency glide, resonance frequencies dysregulation, and Mel-frequency cestrum coefficients (67.00% to 88.71%) (Kheddache and Tadj, 2019), and linear SVM trained with Mel-frequency cepstral coefficients, tilt, and rhythm features, (67.80%) (Matikolaie and Tadj, 2020). Therefore, the proposed CAD system for infant cry signal classification based on DFFNN trained with cepstrum coefficients appears to be effective and promising.

2.5 Conclusion

Infant cry signal analysis is a non-invasive acoustic evaluation that represents an important tool for physicians for pathology diagnosis. The purpose of the current work was to design a CAD system to distinguish between healthy and unhealthy infant cry signals. In this regard, we proposed to calculate the cepstrum of infant cry signal to capture its periodic patterns. Then, the obtained cepstrum coefficients were fed to deep feedforward neural network for training and classification under ten-fold cross-validation protocol.

Thanks to its ability to learn and model deep complex structures in the data, the deep feedforward neural network achieved very close to perfect accuracy when applied to expiration infant cry signals and yielded to perfect accuracy when applied to inspiration infant cry signals. In addition, it outperformed the linear SVM and the Naïve Bayes systems when tested both on the expiration and inspiration sets. Furthermore, the proposed approach outperformed very recent works found in the literature. In short, the proposed system for infant cry signal analysis and classification was found to be effective and fast.

In future work, some interesting issues will be considered. First, regarding DFFNN, various topologies and transfer functions will be evaluated. As this is out of scope of the current work, such investigation is expected to be comprehensive regarding the possible architectures, parameters, and transfer functions. Second, other categories of deep learning neural networks will be examined and compared to DFFNN, such as convolutional neural network and long short-term memory networks. Third, we will examine the performance of optimized classifiers when trained with two types of features, namely deep features and nonlinear statistical features. Such investigation will shed light on the effectiveness of other existing deep learning neural networks and on the discrimination power of deep features compared to features used to describe nonlinear dynamics in infant cry signals.

CHAPITRE 3

DEEP LEARNING SYSTEMS FOR AUTOMATIC CLASSIFICATION OF INFANT CRY SIGNALS

Salim Lahmiri[1,2], Chakib Tadj[2], Christian Gargour[2], Stelios Bekiros[3,4,5]

[1] Department of Supply Chain and Business Technology Management, John Molson School of Business, Concordia University, Montreal, Canada.
[2] Department of Electrical Engineering, École de Technologie Supérieure, Montreal, Canada.
[3] European University Institute, Florence, Italy.
[4] Wilfrid Laurier University, Waterloo, Canada.
[5] IPAG Business School, Department of Finance and Information Systems, Paris, France.

Paper published in «Chaos Soliton and Fractal», January 2022

Résumé

De nos jours, les architectures d'apprentissage profond constituent des systèmes d'intelligence artificielle prometteurs dans diverses applications du génie biomédical. Par exemple, ils peuvent être combinés avec des techniques de traitement du signal pour créer des systèmes de diagnostic assisté par ordinateur utilisés pour aider le médecin à prendre la décision appropriée liée à la tâche de diagnostic. L'objectif de la présente étude est de concevoir et de valider divers systèmes d'apprentissage profond pour améliorer le diagnostic des enregistrements des cris du nourrisson. Plus précisément, les réseaux neuronaux à action anticipée profonde (DFFNN), les réseaux neuronaux à mémoire à long terme et à court terme (LSTM) et les réseaux neuronaux convolutifs (CNN) sont conçus, mis en œuvre et entraînés avec des coefficients basés sur l'analyse cepstrale comme entrées pour distinguer les cris d'un nourrisson sain et malsain. Tous les systèmes d'apprentissage profond sont validés séparément sur les ensembles d'enregistrements de cris d'expiration et d'inspiration. Le nombre de couches convolutives et le nombre de neurones dans les couches cachées varient respectivement dans CNN et DFFNN. Il s'avère que CNN a atteint la précision et la sensibilité les plus élevées, suivi de DFFNN. Ce dernier a obtenu la plus haute spécificité. Par rapport à des travaux similaires dans la littérature, il est conclu que les systèmes d'apprentissage profond entraînés avec des coefficients basés sur l'analyse du cepstre sont des machines puissantes qui peuvent être utilisées pour un diagnostic

précis des enregistrements de cris du nourrisson afin de distinguer les signaux sains et pathologiques.

Mots clés : réseaux neuronaux à action anticipée profonde, réseaux neuronaux à mémoire à long terme et à court terme, réseaux neuronaux convolutifs, cris d'enfants, classification

3.1 Abstract

Nowadays, deep learning architectures are promising artificial intelligence systems in various applications of biomedical engineering. For instance, they can be combined with signal processing techniques to build computer-aided diagnosis systems used to help physician making appropriate decision related to the diagnosis task. The goal of the current study is to design and validate various deep learning systems to improve diagnosis of infant cry records. Specifically, deep feedforward neural networks (DFFNN), long short-term memory (LSTM) neural networks, and convolutional neural networks (CNN) are designed, implemented and trained with cepstrum analysis-based coefficients as inputs to distinguish between healthy and unhealthy infant cry records. All deep learning systems are validated on expiration and inspiration sets separately. The number of convolutional layers and number of neurons in hidden layers are respectively varied in CNN and DFFNN. It is found that CNN achieved the highest accuracy and sensitivity, followed by DFFNN. The latter, obtained the highest specificity. Compared to similar work in the literature, it is concluded that deep learning systems trained with cepstrum analysis-based coefficients are powerful machines that can be employed for accurate diagnosis of infant cry records so as to distinguish between healthy and pathological signals.

Keywords: deep learning, deep feedforward neural networks, long short-term memory neural networks, convolutional neural networks, infant cry record, classification

3.2 Introduction

To extend the literature on infant cry record classification and on applications of deep learning in biomedical engineering problems, the goal of our work is to examine the effectiveness of various deep learning neural networks when used in the diagnosis of infant cry records. Specifically, we focus exclusively on three different deep learning systems and compare their respective performances.

To sum up, the chosen deep learning neural networks to distinguish between healthy and unhealthy infant cries are deep feedforward neural networks (DFFNN), long short-term memory (LSTM) neural networks, and convolutional neural networks (CNN). The inputs to each deep learning system are cepstrum based coefficients obtained from the original cry signals. In this regard, the contributions of our paper can be summarized as follows:

i. We pioneer the evaluation of the performance of various deep learning systems in the task of infant cry record classification to distinguish between healthy and unhealthy infants. Specifically, deep feedforward neural networks (DFFNN), long short-term memory (LSTM) neural networks, and convolutional neural networks (CNN) are evaluated and compared.
ii. Various topologies of CNN and various number of hidden neurons in DFFNN are examined since the performances of these networks depend on such factors.
iii. We use cepstrum based coefficients as inputs to deep learning architectures.
iv. Our study enriches the literature on infant cry record classification for medical diagnosis purpose by applying, evaluating and comparing deep learning techniques.

3.3 Methods

In the current study, three deep learning neural networks are implemented to classify healthy and unhealthy infant cries namely, DFFNN, LSTM, and CNN. They are all trained with cepstrum based coefficients as inputs. The performance of each deep learning system is evaluated by using standard machine learning performance techniques including accuracy, sensitivity, and specificity. Fig.1 shows flowchart of the automated system for infant cry signal analysis and classification. All methods are presented next.

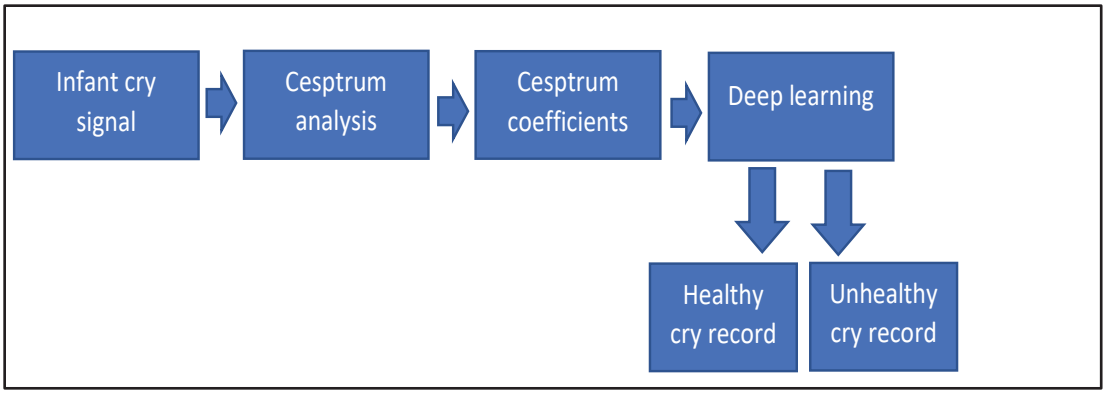

Figure 3.1 Proposed CAD system based on cepstrum analysis and deep learning

3.3.1 DFFNN

The deep feed-forward neural networks (DFFNN) (Kam and Kim, 2017) are basically feed-forward neural networks with multiple hidden layers. In this regard, the DFFNN can learn to transform from one-level description into a multi-level one to deal with complex and high non-linear problems. Recall that each layer in DFFNN utilizes a nonlinear transformation on its input and provides a representation in its output. By assuming that the DFFNN architecture is composed of N layers and each layer has j units (neurones), the output signal of the (*l*)*th* layer (l=1,2,3,…,N) is given by:

$$z_j^l = f\left(w_j^T a_j^{l-1} + b_j\right) \tag{3.1}$$

where *f* is the activation function, *w* is the weight vector, *a* is the output signal in a particular layer, and *b* is the bias parameter.

In this study, the sigmoid function is chosen to be the activation function and the resilient backpropagation algorithm is chosen to train the network to avoid limits of the standard steepest descent algorithm. For the experiments, the input layer has 1000 neurones corresponding to samples of the cepstrum and we will try multiple hidden layers (from 1 to 3) and multiple number of neurones in each layer: 250, 500, and 750.

3.3.2 LSTM

The long-short term memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) are robust to vanishing gradient problem that can be encountered by RNN. To overcome such limit, the LSTM incorporates memory cells to replace traditional nodes of recurrent neural networks (RNN). Their major role is to store information for learning. Hence, the long sequence learning process is improved under LSTM.

The memory cell contains four parts; namely, cell state (*c*), forget gate (*f*), input gate (*i*), and output gate (*o*). They mathematically expressed as follows:

$$f_t = S\left(\omega_f[x_t, h_{t-1}] + b_f\right) \tag{3.2}$$

$$i_t = S(\omega_i[x_t, h_{t-1}] + b_i) \tag{3.3}$$

$$o_t = S(\omega_0[x_t, h_{t-1}] + b_0) \tag{3.4}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot tanh(\omega_c[x_t, h_{t-1}] + b_c) \tag{3.5}$$

$$h_t = o_t \cdot tanh(c_t) \tag{3.6}$$

where x_t is the input state, h_t is the hidden state, c_t is the memory cell state; and, ω and b are their respective weight matrices and bias vectors. Here, S is the softmax function used to calculate the highest probability to perform classification. Recall that the cell state c_t and the output h_t represent short-term memory and long-term memory respectively. In this work, the LSTM has two fully connected layers, number of epochs is set to 50, and the adaptive moment estimation method is employed since it is suitable to train LSTM systems.

3.3.3 CNN

The convolutional neural networks (CNN) are common deep feed-forward artificial neural networks originally introduced by LeCun et al (1989). It is suitable for deep features extraction

as its topology makes it invariant to translation and rotation in inputs. In addition, compared common deep feed-forward artificial neural networks, CNN require less parameters to be tuned.

The general topology of CNN consists of one or more hidden convolutional layers and a fully connected layer at the top, including associated weights and pooling layers. In the convolutional layer, a convolution of each subregion of the input data with a kernel is calculated. Then, the convolution output result is added to a bias and processed by an activation function to generate a feature map for the next layer. For instance, if the cepstrum input vector is $x_i^0 = [x_1, x_2, \dots, x_n]$ where n is the number of samples in cepstrum, then output values are calculated as follows:

$$c_i^{l,j} = h\left(b_j + \sum_{m=1}^{M} w_m^j \, x_{i+m-1}^{oj}\right) \tag{3.7}$$

where l is the layer index, h is a nonlinear activation function, b is bias term associated with the jth feature map, M is the kernel (filter) size, and w is the weight.

The pooling layer is subsequent to the convolutional layer and used to apply a down-sampling operation to features vector so as to reduce its size. In our study, the max-pooling operation is adopted to compute the maximum value in a set of nearby inputs. Hence, the pooling of a feature map in a layer is expressed as follows:

$$p_i^{l,j} = \max_{r \in R} c_{i \times T + r}^{l,j} \tag{3.8}$$

where R and T are respectively the size of the pooling window and the pooling stride. Recall that several pooling layers are inserted between successive convolutional layers to finally obtain deep features used for classification purpose. Finally, the classification layer is a fully connected layer used to perform the high-level reasoning about the deep features and produce the classification output. In our experiments, the stochastic gradient descent with momentum (SGDM) algorithm is adopted to train CNN, the number of epochs is set to 50, and the number of convolutional layers varies from one to three.

3.3.4 Performance metrics and protocol of experiments

We evaluate the performance of the classification deep learning systems by using accuracy (correct classification rate), sensitivity, and specificity. They are calculated as follows:

$$Accuracy = \frac{TN+TP}{TN+FN+TP+FP} \tag{3.9}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{3.10}$$

$$Specificity = \frac{TN}{TN+FP} \tag{3.11}$$

where *TN* is number of true negatives, *TP* is number of true positives, *FN* is number of false negatives, and *FP* is number of false positives. In this study, the positive and negative cases were assigned to unhealthy and healthy cry signals. For robustness of the experimental results, 10-fold cross-validation protocol is adopted. The 10-fold validation is commonly employed in machine learning literature. The dataset is divided into 10 folds and each fold has the chance to be tested, such that all folds are used for training and testing. Accordingly, the mean and standard deviation of each performance metric are calculated.

3.4 Data and results

We used the first dataset described in Introduction (Section 0.4.1) to describe the comparative study of DFFNN, LSTM, and CNN in distinguishing between healthy and unhealthy infant cry signals. After cepstrum coefficients used to form features of all samples are obtained, the constructed deep learning systems can be employed to distinguish between healthy and unhealthy infant cry signals. For statistical robustness of the experimental results, the dataset is divided into ten folds to perform cross-validation method. Specifically, the training and testing phases are carried out by using nine folds for training and the rest is utilized for validation. This protocol is repeated ten times to assure that all folds are used for training and for testing as well. Finally, the average values and standard deviations of accuracy, sensitivity, and specificity are computed for each single deep learning system. In our experiments, cepstral

analysis and deep learning systems were implemented in Matlab©R2021b environment. The obtained performance metrics are provided in Table 3.1.

Table 3.1 Performance measures obtained under 10-fold cross-validation

	Accuracy	**Sensitivity**	**Specificity**
Expiration set			
CNN with 1 Convolutional layer	95.31%±0.0320	95.42%±0.0838	95.21%±0.1333
CNN with 2 Convolutional layers	95.28%±0.0222	99.30%±0.0813	91.26%±0.0930
CNN with 3 Convolutional layers	94.33%±0.0272	94.82%±0.1610	93.81%±0.1724
LSTM	83.89%±0.1410	89.83%±0.4254	77.98%±0.3977
DFFNN with 250 hidden neurons	80.23%±0.0000	30.88%±0.0029	99.58%±0.0029
DFFNN with 500 hidden neurons	79.70%±0.0000	98.85%±0.0000	50.55%±0.0000
DFFNN with 750 hidden neurons	80.00%±0.0000	98.48%±0.0000	31.52%±0.0000
Inspiration set			
CNN with 1 Convolutional layer	86.83%±0.0280	82.58%±0.1792	91.08%±0.1798
CNN with 2 Convolutional layers	92.04%±0.0433	91.40%±0.1384	92.69%±0.1211
CNN with 3 Convolutional layers	90.05%±0.0269	92.47%±0.1032	87.63%±0.0962
LSTM	80.18%±0.0051	90.32%±0.4779	60.00%±0.4830
DFFNN with 250 hidden neurons	80.05%±0.0000	99.14%±0.0000	50.97%±0.0000
DFFNN with 500 hidden neurons	80.01%±0.0000	30.89%±0.0000	99.12%±0.0000
DFFNN with 750 hidden neurons	79.84%±0.0000	30.75%±0.0282	98.92%±0.0279

Accordingly, CNN with 1 convolutional layer yielded to the highest accuracy on expiration set (95.31%±0.0320) followed by CNN with 2 convolutional layers (95.28%±0.0222), and CNN with 3 convolutional layers (94.33%±0.0272). The DFFNN with 750 hidden neurons achieved the lowest accuracy (80.00%±0.0000) on expiration set. Hence, CNN outperformed LSTM and DFFNN on expiration in terms of correct classification rate (accuracy). Besides, LSTM achieved the lowest accuracy (83.89%±0.1410). Also, CNN with 2 Convolutional layers yielded to the highest sensitivity (99.30%±0.0813) followed by DFFNN with 500 hidden neurons (98.85%±0.0000), and LSTM obtained 89.83%±0.4254. In terms of sensitivity performance, DFFNN with 250 hidden neurons obtained the highest value (99.58%±0.0029) followed by CNN with 1 convolutional layer (95.21%±0.1333), and LSTM yielded to 89.83%±0.4254. In short, on expiration set, CNN outperformed DFFNN and LSTM in terms of accuracy and sensitivity, while DFFNN outperformed CNN and LSTM in terms of specificity as indicated in Table 1.

Experimental results from inspiration set provided in Table 1 indicate that CNN with 2 convolutional layers achieved the highest accuracy (92.04%±0.0433), LSTM obtained 80.18%±0.0051, and DFFNN with 750 hidden neurons yielded to the lowest accuracy (79.84%±0.0000). Besides, DFFNN with 250 hidden neurons obtained the highest sensitivity (99.14%±0.0000) and DFFNN with 750 hidden neurons obtained the lowest sensitivity (30.75%±0.0282). Finally, DFFNN with 500 hidden neurons yielded to the highest specificity (99.12%±0.0000) followed by CNN with 2 convolutional layers (92.69%±0.1211), LSTM achieved 60.00%±0.4830, and DFFNN with 250 hidden neurons obtained the lowest specificity (50.97%±0.0000). In short, on inspiration set, CNN obtained the highest accuracy while DFFNN achieved the highest sensitivity and specificity.

To summarize, CNN outperformed both LSTM and DFFNN in terms of accuracy on both expiration and inspiration sets. Its best accuracy is obtained on expiration set. Also, its accuracy performance slightly decreases with number of convolutional layers.

To complete the 10-fold cross-validation on expiration set, the DFFNN took respectively 23.24s, 9.02s, and 15.73s to run with 250, 500, and 750 neurons in each hidden layer. The LSTM took 60.32s and CNN took 208.30s, 446.72s, and 536.03 to run with one, two, and three convolutional layers respectively. Besides, to complete the 10-fold cross-validation on inspiration set, the DFFNN took respectively 4.03s, 24.62s, and 11.17s to run with 250, 500, and 750 neurons in each hidden layer. The LSTM took 50.22s and CNN took 181.36s, 270.72s, and 342.39s respectively to run with one, two, and three convolutional layers. In this regard, the DFFNN is faster in training and testing compared to CNN and LSTM. This can be attributed to the resilient backpropagation algorithm used to train the DFFNN.

Recall that the best deep neural network is CNN with one convolutional layer as it achieved 95.31%±0.0320 accuracy, 95.42%±0.0838 sensitivity, and 95.21%±0.1333 specificity. In this regard, CNN trained with cepstrum coefficients outperformed similar models applied to the same problem; including the Hidden Markov Models trained with segmented cry signals (83.79% accuracy) (Acharya et al, 2005), the probabilistic neural network trained with prevalence of fundamental frequency glide, resonance frequencies dysregulation, and Mel-frequency cestrum coefficients (67.00% to 88.71% accuracy) (Mekler, 2008), the linear SVM trained with Mel-frequency cepstral coefficients, tilt, and rhythm features (67.80% accuracy)

(Rawal et al, 2015). However, it underperformed deep feedforward neural networks trained with cepstrum coefficients where 100% accuracy was obtained (Carvajal et al, 2005). Bring in mind that the work in Carvajal et al (2005) used the same number of neurons in each hidden number as the number of neurons in the input layer. This could improve the learning ability of the network.

The obtained experimental results suggest that deep learning architectures trained with cepstrum coefficients yield to better accuracy in terms of distinguishing between healthy and unhealthy infant cry records as opposed to existing works found in the literature) (Acharya et al, 2005; Mekler, 2008; Rawal et al, 2015). This can be explained by the statistical significance of cepstrum coefficients to characterize infant normal and abnormal cry signals) (Lahmiri et al, 2021) and the powerful learning ability of deep learning systems. Hence, such computer-aided diagnosis systems are promising in clinical milieu.

3.5 Conclusion

Deep learning systems are making major advances in the design of computer-aided diagnosis systems for healthcare. The aim of our paper is to compare the performance of three different deep learning systems in the task of distinguishing between recorded healthy and pathological infant cry signals. In this regard, cepstrum analysis is employed to describe harmonics in expiration and inspiration cry record and the resulting spectrum is fed to DFFNN, LSTM, and CNN.

Experimental results from expiration and inspiration sets indicated that CNN outperformed DFNN and LSTM in terms of accuracy and sensitivity, whilst DFNN yielded to the highest specificity. In comparison with similar studies, our findings suggest that deep learning systems trained with cepstrum descriptors obtained the highest accuracy; hence, they are promising for analysis and diagnosis of infant cry signals in clinical milieu.

CHAPITRE 4

OPTIMAL TUNING OF SUPPORT VECTOR MACHINES AND k-NN ALGORITHM BY USING BAYESIAN OPTIMIZATION FOR NEWBORN CRY SIGNAL DIAGNOSIS BASED ON AUDIO SIGNAL PROCESSING FEATURES

Salim Lahmiri[1,2], Chakib Tadj[2], Christian Gargour[2], Stelios Bekiros[3,4,5]

[1] Department of Supply Chain and Business Technology Management, John Molson School of Business, Concordia University, Montreal, Canada
[2] Department of Electrical Engineering, École de Technologie Supérieure, Montreal, Canada
[3] FEMA, University of Malta, MSD 2080 Msida, Malta
[4] LSE Health, London School of Economics and Political Science (LSE), London WC2A2AE, UK

Paper published in «Chaos Soliton and Fractal», February 2023

Résumé

Récemment, le nombre de modèles d'apprentissage automatique utilisés pour classer les signaux des cris des nouveau-nés en bonne santé et en mauvaise santé a considérablement augmenté. Divers travaux ont déjà rapporté des résultats de classification encourageants. Cependant, l'optimisation des hyper-paramètres des algorithmes d'apprentissage automatique reste un problème ouvert dans le contexte de la classification des signaux des cris du nouveau-né. Cet article propose d'utiliser la méthode d'optimisation bayésienne (BO) pour optimiser les hyper-paramètres de la machine à vecteurs de support (SVM) avec le noyau de la fonction à base radiale (RBF) et les k-plus proches voisins (kNN) formés avec différentes fonctionnalités audio séparément ou combinées; à savoir, les coefficients cepstraux de fréquence mel (MFCC), la modulation d'amplitude inspirée par l'audition (AAM) et la prosodie. En particulier, le test du chi-carré est appliqué à chaque ensemble de caractéristiques pour retenir les dix plus importantes utilisées pour entrainer les classificateurs optimaux. Le taux de classification correcte, la sensibilité et la spécificité de chaque modèle expérimental sont calculées selon le protocole standard de validation croisée à 10 fois. L'une des contributions de cette étude est l'amélioration par rapport aux travaux antérieurs sur la classification des signaux de cris du nouveau-né pour distinguer les normaux et les malades sur la même base de données, en termes

de performances. Le meilleur modèle est le SVM formé avec AAM, les dix caractéristiques les plus significatives ont atteint une précision de 83.62% ± 0.022, une sensibilité de 59.18% ±0.0469 et une spécificité de 93.87 %±0.0190, suivis par kNN entrainé par la plupart des caractéristiques de MFCC, AAM et prosodie pour obtenir une classification correcte de 82.88%±0.0144, sensibilité de 55.34%±0.0350 et spécificité de 94.42%±0.0075. Ces résultats ont surpassé les travaux existants validés sur la même base de données. De plus, les SVM et kNN réglés de manière optimale sont alimentés avec un nombre restreint de modèles sélectionnés, de sorte que le temps de traitement pour les phases de l'entrainement et des tests est considérablement limité. Cela signifie que le classificateur RBF-SVM-BO entrainé avec les dix caractéristiques les plus importantes de l'AAM est plus en mesure d'effectuer une meilleure distinction entre les nouveau-nés normaux et malades.

Mot clés : nouveau-né, coefficients cepstraux de fréquence Mel, modulation d'amplitude d'inspiration auditive, prosodie, machine support à de vecteurs, k plus proches voisins, optimisation bayésienne

4.1 Abstract

Recently, the number of machine learning models used to classify cry signals of healthy and unhealthy newborns has been significantly increasing. Various works have already reported encouraging classification results; however, fine-tuning of the hyper-parameters of machine leaning algorithms is still an open problem in the context of newborn cry signal classification. This paper proposes to use Bayesian optimization (BO) method to optimize the hyper-parameters of Support Vector Machine (SVM) with radial basis function (RBF) kernel and k-nearest neighbors (kNN) trained with different audio features separately or combined; namely, mel-frequency cepstral coefficients (MFCC), auditory-inspired amplitude modulation (AAM), and prosody. Particularly, the chi-square test is applied to each set of features to retain the ten most significant ones used to train optimal classifiers. The accuracy, sensitivity, and specificity of each experimental model are computed following the standard 10-fold cross-validation protocol. One of the contributions is an improvement over previous works on newborn cry

signal classification used to distinguish between healthy and unhealthy ones over the same database, in terms of performance. The best model is the SVM trained with AAM ten most significant features achieved 83.62%±0.022 accuracy, 59.18%±0.0469 sensitivity, and 93.87%±0.0190 specificity followed by kNN trained with ten most features from MFCC, AAM, and prosody to obtain 82.88%±0.0144 accuracy, 55.34%±0.0350 sensitivity, and 94.42%±0.0075 specificity. These results outperformed existing works validated on the same database. In addition, optimally tuned SVM and kNN are fed with a restricted number of selected patterns so as the processing time for training and testing is significantly limited. This means that the RBF-SVM-BO classifier trained with AAM ten most significant features is more able to distinguish between healthy and unhealthy newborns.

Keywords: newborn cry; mel-frequency cepstral coefficients; auditory-inspired amplitude modulation; prosody; support vector machines; k-nearest neighbors; Bayesian optimization

4.2 Introduction

The medical diagnosis of newborn diseases based on cry signal analysis and classification is a cost-effective and non-invasive solution that may provide promising accurate performance. Indeed, newborn cry signal analysis and classification is an up-to-date approach to build computer-aided diagnosis (CAD) systems used to distinguish between healthy and unhealthy newborns. In this regard, several CAD systems have been proposed and evaluated by biomedical engineering researchers as the topic is attracting a growing interest.

In this regard, the main purpose of the current study is to design various CAD systems to distinguish between healthy and unhealthy newborns based on analysis and classification of the acoustics of their cries. Specifically, we propose to use Bayesian optimization (BO) method to optimize the hyper-parameters of the SVM with radial basis function (RBF) kernel and *k*-nearest neighbors (*k*NN), all trained with different audio acoustic features separately or combined; precisely, MFCC, AAM, and prosody. More specifically, only the most significant patterns from each set are selected by a statistical filter (Chi- square; for instance) are used to train each optimal classifier.

We rely of acoustic patterns as they are good descriptors of any sound, easy to measure and to interpret, and was successfully applied in previous works dealing with newborn cry analysis and classification (Abou-Abbas et al, 2015; Kheddache and Tadj, 2015; Alaie et al, 2016; Abou-Abbas et al, 2017; Kheddache and Tadj, 2019; Matikolaie and Tadj, 2020; Matikolaie et al, 2022; Matikolaie and Tadj, 2022). Besides, we consider the SVM thanks to its ability to minimize the upper bound on the generalization error based on the structural risk minimization principle. Also, the *k*NN algorithm is considered in the current as it allows for local approximation of any function by learning non-linear decision boundaries and while being flexible. For better tuning of the SVM and *k*NN algorithm, the BO algorithm is adopted to determine their respective optimal key parameters thanks to its ability to update the prior belief in light of new information to produce an updated posterior belief to find promising minima; hence, to statistically and robustly approximate the objective function. The BO is fast and statistically efficient.

To sum up, the contributions of the current study are as follows:

i. To design, implement, and compare various optimal CAD systems for diagnosis of newborn based on automatic analysis and classification of cry audio features.
ii. Apply a statistical filter for most significant features selection to allow fast convergence of the classifier along with reduction in system complexity.
iii. To fine tune key parameters of classifiers by using Bayesian optimization. Hence, the accuracy of the optimal classifier is expected to improve.
iv. The performance of the optimal CAD system can easily be interpreted from a physiological perspective as main involved features in improvement of accuracy can be identified.
v. To test the efficacy of designed optimal CAD systems on a large data set considered in very recent studies. Hence, the performance of our best optimal CAD system can be compared to the most recent models tested on the same database.

4.3 Methods

The purpose of the current study is to design various CAD systems to distinguish between healthy and unhealthy newborns based on analysis and classification of acoustics of their respective cry signals. The acoustic features are categorized into three sets: MFCC, AAM, and prosody. The Chi-square (χ^2) is applied to each acoustic features set to determine the most 10 significant features used to train the classifiers. The SVM and kNN algorithm are chosen as main classifiers two distinguish between health conditions of the newborns. The key parameters of each classifier are optimized by using BO algorithm. Fig.4.1 shows the flowchart of the proposed CAD systems. The methods are described next.

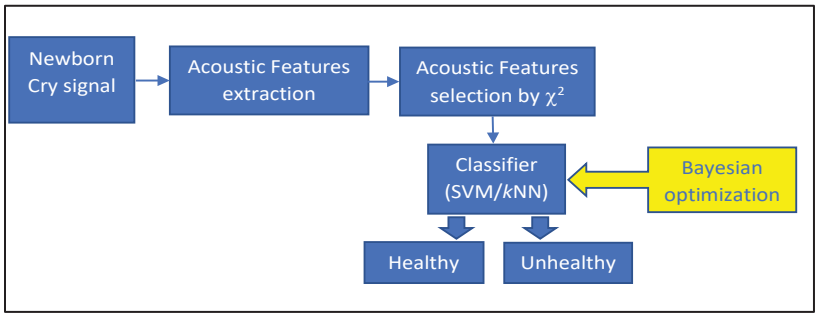

Figure 4.1 Proposed CAD system based on features selection

4.3.1 Acoustic features and selection

To characterize each newborn cry signal, four three of features are extracted; namely, mel frequency cepstral coefficients (MFCC), auditory-inspired amplitude modulation (AAM), and prosody. The MFCC have the merit to identify and track timbre fluctuations in a sound, AAM is able to characterize the rate of change of long-term speech, and prosody provides the melody and the parsing of speech. They are briefly presented afterward.

- *Mel frequency cepstral coefficients (MFCC)*

The MFCC are well-known popular short-term acoustic features useful to determine the critical bandwidth used by human auditory system to recognize a different tone based on the Mel scale. The latter is defined as follows:

$$M(f) = 1125 \times log\left(1 + \frac{f}{700}\right) \quad (4.1)$$

Here, f and $M(f)$ denote the frequency value of the signal and its corresponding Mel value respectively. To obtain MFCC, four steps should be performed: (*i*) framing (voice signal is broken down into overlapping frames), (*ii*) windowing (each frame is multiplied by a Hamming window), (*iii*) applying Fast Fourier transform (convert the signal to the frequency domain and calculate its periodogram), (*iv*) applying Mel filter banks (compute the average of each spectral power density contained in each filter and computes its logarithm), and (*v*) converting from cepstral to temporal domain by calculating the inverse discrete Fourier transform. More details are be found in Matikolaie and Tadj (2020).

- *AAM features*

First, the newborn cry signal is processed by short-time discrete Fourier transform (STDFT). Second, the squared magnitudes of the resulting acoustic frequency components are categorized into 27 sub-bands. Third, a second transform is performed across time for each of the 27 sub-band magnitude signals. Fourth, a band-pass filter is applied to the grouped squared modulation frequencies. Finally, logarithm transform is applied for compression purpose. More technical details on AAM are found in Sarria-Paja and Falk (2017).

- *Prosody features*

Prosody set is composed of tilt feature subset, intensity feature subset, and rhythm feature subset. Tilt features are estimated based on two parameters: A_t and D_t. They are defined as follows:

$$A_t = \left(\frac{|A_r| - |A_f|}{|A_r| + |A_f|}\right) \quad (4.2)$$

$$D_t = \left(\frac{|D_r| - |D_f|}{|D_r| + |D_f|}\right) \quad (4.3)$$

where A_f is the amplitude of the contours of the fundamental frequency F_0 when they are descending and A_r is their amplitude when they are ascending. Likewise, D_f is the length of the contours of F_0 when they are descending and D_r is their length when they are ascending.

Besides, the intensity represents the height of the audio signal. Specifically, the intensity of an audio signal is used to embody its height by measuring the energy of volume in a waveform. Intensity of the audio signal is given by:

$$Intensity = 10 \times log(\sum_{n=1}^{N} A^2(n)w(n)) \quad (4.4)$$

where *w* and *A* are respectively the window and the amplitude.

Finally, the rhythm feature subset includes two main parameters; namely, the raw pairwise variability index (*rPVI*) and the normalized one (*nrPVI*). Both are useful to quantify the rhythm in a given audio by expressing the level of variability in successive measurements. They expressed as follows:

$$rPVI = \left(\frac{\sum_{k=1}^{M-1}|d_k - d_{k+1}|}{m-1}\right) \quad (4.5)$$

$$nrPVI = 100 \times \left(\frac{\sum_{k=1}^{M-1}\left|2 \times \frac{d_k - d_{k+1}}{d_k - d_{k+1}}\right|}{m-1}\right) \quad (4.6)$$

In addition to *rPVI* and *nrPVI*, six other features are computed; namely, the standard deviation of the expiration signal, the standard deviation of the expiration signal divided by mean length, number of expirations in each cry signal, duration of expiration, range of expiration, average of all expirations in one signal cry signal.

- *Acoustics selection by χ^2 test*

Since each acoustic features set is large and may negatively affect the processing time and accuracy of the diagnosis, the classifiers will be trained with the ten most significant features from each features set which are selected by a statistical filter; namely, Chi-square test (χ^2). It is chosen thanks to its robustness with respect to the distribution of the data as it does not assume any distribution and is fast to compute. The feature selection process seeks to generate a score for each acoustic feature by counting its frequency in training unhealthy and healthy class samples separately and then finding a function of both. The χ^2 statistic is calculated as follows:

$$\chi^2 = \sum_{i-1} \sum_{j-1} \left(\frac{O^{ij} - E^{ij}}{E^{ij}} \right) \quad (4.7)$$

where O is the frequency that feature is observed and E is the frequency that it is expected. Larger value of χ^2 statistic suggests significant confirmation that two features are different.

4.3.2 SVM and *k*NN classifiers

The SVM (Cortes and Vapnik, 1995; Vapnik et al, 1996) seeks to find a hyperplane $w.\Phi(x)+b = 0$ to separate the features vector x from classes +1 (unhealthy) and -1 (healthy) with a maximal margin. Here, w is a weight vector, Φ is a mapping function, and b a bias. The decision frontier of classes y is written as:

$$y = sign(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b) \quad (4.8)$$

In this study, the kernel K is set to be the radial basis function (RBF) as it is a local function, which is flexible and effective in approximation of short variations in a nonlinear function. The RBF is given by:

$$K(x_i, x_j) = exp(-\gamma \|x - x_i\|^2) \quad (4.9)$$

where γ is the width of the RBF.

Besides, the *k*NN algorithm (Cover and Hart, 1967) is typically a non-parametric instance-based learning algorithm. It assigns a class to an unclassified point following a majority rule based on the *k*-nearest neighbors in the training set. As a result, the class majority among the *k*NN produces a prediction for a new point. For instance, the *k*NN of point x_i are all k nearest neighbor points x_j in a subset of datasets D defined as follows:

$$NN_k(x_i) = \{x_j \in D, d(x_i, x_j) \leq d(x_i, q)\} \quad (4.10)$$

where q is the kth nearest neighbor of point x_i and $d(x_i,x_j)$ is a distance function.

4.3.3 Bayesian optimization

The Bayesian optimization (BO) (Gelbart, et al, 2014) uses an acquisition function to find both regions where the model believes the objective function to be low and regions where uncertainty is high. Let consider $f(x)$ be the objective function and the expected improvement function $EI(x,Q)$ be the acquisition function used to evaluate the feasibility of a point x based on the posterior distribution function Q. The expected-improvement function ($EI(x,Q)$) is expressed as follows:

$$EI(x,Q) = E_Q\left[max\left(0, \mu_Q(x_{best}) - f(x)\right)\right] \tag{4.10}$$

where x_{best} is the location of the lowest posterior mean and $\mu_Q(xbest)$ is the lowest value of the posterior mean.

The BO is employed to find optimal values of the structural parameters of the SVM and the width of the RBF. Also, it is employed to find the optimal distance metric and the optimal k for the kNN algorithm. The BO technique is employed through 10-fold cross validation to find the optimal parameters. More details on BO method can be found in (Gelbart, et al, 2014).

4.4 Data and results

We used the first dataset described in Introduction (Section 0.4.1). For experiments, we consider four different vectors of features. The first one is composed of 190 MFCC features, the second is composed of 200 AAM features, the third one is composed of 38 prosody features, and the fourth is the larger one which includes all MFCC, AAM, and prosody acoustic patterns. Univariate feature ranking for classification using Chi-square test is performed to each vector of features to obtain 10 best features that explain most of variability in each one of them by using 5% statistical significance level. The ranking of features is shown in Fig.4.2 for each category of acoustic patterns. Then, each optimal classifier is trained either with best AAMF vector, best MFCC vector, best prosody vector, or best features selected from the large vector composed of AAMF, MFCC, and prosody feature set. Ten-fold cross-validation is employed and average and standard deviation of accuracy, sensitivity (correct classification

rate of unhealthy infant records) and specificity (correct classification rate of unhealthy infant cry records) are reported in Table 1.

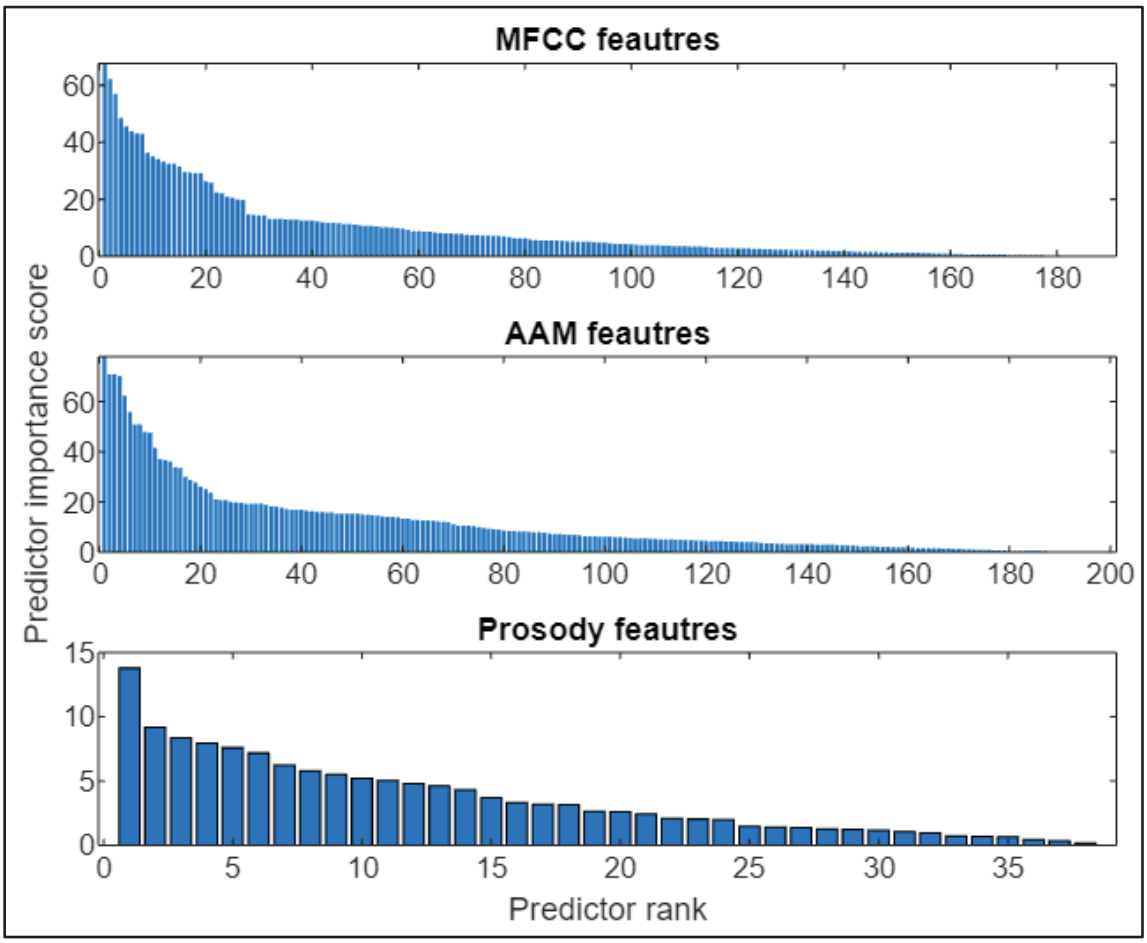

Figure 4.2 Ranking of acoustic features by type

Table 4.1 Experimental results from optimized machine learning

CAD systems	Accuracy	Sensitivity	Specificity	Processing time (sec)
AAM+SVM	83.62%±0.0229	59.18%±0.0469	93.87%±0.0190	0.5161
MFCC+SVM	72.37%±0.0091	24.49%±0.0177	92.46%±0.0164	5.1411
Prosody+SVM	70.65%±0.0025	0.0000	1.0000	0.8269
AAM+MFCC+Prosody+ SVM	81.74%±0.0183	55.33%±0.0413	92.80%±0.0096	0.7969
AAM+*k*NN	80.07%±.0162	48.80%±0.4880	93.14%±0.0151	0.4490
MFCC+*k*NN	74.07%±0.0090	33.29%±0.0156	91.17%±0.0077	0.2349
Prosody+*k*NN	70.43%±0.0007	0.0000	1.0000	0.1763
AAM+MFCC+Prosody+*k*NN	82.88%±0.0144	55.34%±0.0350	94.42%±0.0075	0.1713

Accordingly, the obtained performance by each optimal classifier trained with a specific set of features. As shown, the optimal SVM trained with AAM features yielded to the highest accuracy (correct classification rate) 83.62%±0.0229 followed by the optimal *k*NN trained with AAM, MFC and prosody (82.88%±0.0144). The least accuracy is obtained by *k*NN trained with prosody features (70.43%±0.0007) and SVM trained with prosody features (70.65%±0.0025).

Besides, the highest sensitivity (correct classification of healthy newborns) is achieved by the optimal SVM trained with AAM features (59.18%±0.0469) followed by the optimal *k*NN trained with AAM, MFC and prosody (55.34%±0.0350). Finally, the latter system achieved the highest specificity (correct classification of unhealthy newborns) 94.42%±0.0075 followed by the optimal SVM trained with AAM features (93.87%±0.0190).

It is worth to mention three interesting observations. First, integrations of all three different features sets considerably improves the accuracy of the optimal *k*NN. In addition, it improves the accuracy of the optimal SVM compared to the one trained with MFCC or prosody features. Another interesting observation is the fact that AAM features allow the optimal *k*NN to outperform the one trained with MFCC or prosody features. Finally, training and testing the best systems requires 0.5161 seconds by the optimal SVM trained with AAM and 0.1713 seconds by the optimal *k*NN trained with AAM, MFC and prosody. Hence, these two best CAD systems are fast and can be implemented for real applications.

With comparison to previous works where various CAD systems have been proposed to distinguish between healthy and unhealthy cry signals of newborns, our best CAD system achieved 83.62%±0.0229 accuracy. Hence, it outperformed a very recent study validated on the same database (Matikolaie et al, 2022) where the linear SVM and PNN achieved 76.50% and 68.90% respectively when trained with MFCC, 75.75% and 70.70% respectively when trained with AAM features, 61.50% and 52.10% respectively when trained with prosody features and 78.70% and 77.90% respectively when trained with combination of MFCC and AAM.

4.5 Conclusion

To explore the effectiveness of various automatic systems in analysis and classification of newborn cry signals to distinguish between healthy and healthy ones, our work designed and compared different CAD models involving nonlinear SVM and *k*NN classifiers optimized by using Bayesian optimization and trained by Chi-square based selected features from MFCC, AAM, prosody or combination of those selected features. This is the first study to design and compare these CAD systems for detection of unhealthy newborn cry signals. The best model is the SVM trained with AAM followed by *k*NN trained with combination of MFCC, AAM, and prosody. Our best model outperformed most existing works validated on the same database while being considerably fast to perform. As being effective and explainable, the proposed CAD system can be promising for diagnosis of newborns based on their cry signals in clinical milieu.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research is partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2016-05067].

CHAPITRE 5

CHARACTERIZATION OF INFANT HEALTHY AND PATHOLOGICAL CRY SIGNALS IN CEPSTRUM DOMAIN BASED ON APPROXIMATE ENTROPY AND CORRELATION DIMENSION

Salim Lahmiri[1,2], Chakib Tadj[2], Christian Gargour[2], Stelios Bekiros[3,4,5]

[1] Department of Supply Chain and Business Technology Management, John Molson School of Business, Concordia University, Montreal, Canada.
[2] Department of Electrical Engineering, École de Technologie Supérieure, Montreal, Canada.
[3] European University Institute, Florence, Italy.
[4] Wilfrid Laurier University, Waterloo, Canada.
[5] IPAG Business School, Department of Finance and Information Systems, Paris, France.

Paper published in «Chaos Soliton and Fractal», February 2021

Résumé

L'analyse des signaux de cris du nourrisson devient un domaine de recherche attractif en physique et en ingénierie biomédicales pour une meilleure compréhension des pathologies et un diagnostic médical approprié. L'objectif principal de l'étude actuelle est de caractériser les signaux de cris normaux et pathologiques du nourrisson en étudiant leurs oscillations respectives au moyen d'une entropie approximative et d'une dimension de corrélation estimée à partir de leurs cepstres respectifs. Nous avons analysé deux ensembles différents. Le premier est composé de 2638 signaux de cri d'expiration et le second ensemble est composé de 1860 signaux de cri d'inspiration, les deux ensembles étant équipondérés. Après avoir estimé les dimensions approximatives d'entropie et de corrélation à partir des cepstres, trois tests statistiques standard leur sont appliqués, notamment le test-*t* de Student, le test-F et le test de Kolmogorov-Smirnov. Tous les tests statistiques sont effectués à un niveau de significativité statistique de 5 %. Les résultats empiriques suivants sont obtenus. Premièrement, les mesures approximatives de l'entropie et de la dimension de corrélation présentent différentes caractéristiques statistiques entre les cris des nourrissons sains et malsains des ensembles d'expiration et d'inspiration. Deuxièmement, le niveau d'entropie approximative dans les cepstres des cris des nourrissons sains est statistiquement plus élevé que celui des cepstres des cris des nourrissons malsains. Troisièmement, le niveau de dimension de corrélation dans les cepstres de cris de nourrissons sains est statistiquement plus élevé que celui des cepstres de

cris de nourrissons malsains. En d'autres termes, les cepstres de cris de nourrissons sains présentent un caractère aléatoire et un désordre inférieurs à ceux des cepstres de pleurs de nourrissons malsains. Il est conclu que l'entropie approximative basée sur le cepstre et la dimension de corrélation discriminent les signaux de pleurs sains des nourrissons pathologiques et peuvent être utilisés comme biomarqueurs efficaces pour le diagnostic biomédical des enregistrements de cris en milieu clinique.

Mots clés : cris de nourrissons, cepstrum, entropie approximative, dimension de correlation, tests statistiques

5.1 Abstract

The analysis of infant cry signals is becoming an attractive field of research in biomedical physics and engineering for better understanding of the pathologies and appropriate medial diagnosis. The main purpose of the current study is to characterize infant normal and pathological cry signals by studying their respective oscillations by means of approximate entropy and correlation dimension estimated from their respective cepstrums. We analyzed two different sets. The first one is composed of 2638 expiration cry signals and the second set is composed of 1860 inspiration cry signals, both sets equally weighted. After estimating approximate entropy and correlation dimensions from cepstrums, three standard statistical tests are applied to them including the Student t-test, F-test, and two-sample Kolmogorov-Smirnov test. All statistical tests are performed at 5% statistical significance level. The empirical results follow. First, approximate entropy and correlation dimension measures exhibit different statistical characteristics across healthy and unhealthy infant cries from both expiration and inspiration sets. Second, the level of approximate entropy in cepstrums of healthy infant cries is statistically higher than that in cepstrums of unhealthy infant cries. Third, the level of correlation dimension in cepstrums of healthy infant cries is statistically higher than that in cepstrums of unhealthy infant cries. In other words, cepstrums of healthy infant cries show lower randomness and disorder compared to cepstrums of unhealthy infant cries. It is concluded that cepstrum-based approximate entropy and correlation dimension discriminate

healthy from pathological infant cry signals and can be employed as effective biomarkers for biomedical diagnosis of cry records in clinical milieu.

Keywords: infant cry signal, cepstrum, complexity, approximate entropy, correlation dimension, statistical tests

5.2 Introduction

In recent years, there is a growing interest in the diagnosis of cry signal to detect pathology in unhealthy infants (Abou-Abbas et al, 2015; Kheddache and Tadj, 2015; Alaie et al, 2016; Abou-Abbas et al, 2017; Kheddache and Tadj, 2019; Matikolaie and Tadj, 2020; Matikolaie et al, 2022; Matikolaie and Tadj, 2022). Indeed, the acoustical analysis of infant cry signal is independent of human intervention, helpful to assess pain in paediatric wards; and, consequently, can assist clinicians in medical decision-making.

The main purpose of the current work is to examine the statistical discriminative power of cepstrum-based complexity measures in distinguishing between healthy and pathological infant cry signals. Indeed, such investigation is expected to help understanding the nonlinear dynamics in healthy and pathological infant cry signals to better distinguish between them by physicians in clinical milieu. Without a doubt, shedding light on complexity in dynamics of cry-recorded sounds would be a benefit in the design of automatic systems for cry signal diagnosis to increase accuracy of diagnosis and reduce time of medical consultation. In short, an objective nonlinear statistical analysis can only be used as an assistive tool by clinicians who definitively have the final decision regarding the medical diagnosis.

It is worth to notice that pathological samples have a strong tendency for frequent and rapid changes in regularity (Timcke et al, 1960) and that pathological voices tend to show unusually large cycleto-cycle fluctuations (Lieberman, 1961). In this regard, the analysis of infant cry signals by means of cepstrum (Oppenheim, 2010) is suitable to detect intrinsic periodicity and rapid oscillations. Besides, approximate entropy (Pincus, 1991) and correlation dimension (Grassberger and Procaccia, 1983a; Grassberger and Procaccia, 1983b) are two complexity measures that will be estimated from the cepstrum to better characterize each infant cry signal.

In one hand, approximate entropy is a measure of regularity to quantify levels of complexity within a nonlinear signal where variability is defined through examination of the temporal variations in it. On the other hand, based on phase space reconstructions, correlation dimension, which is an example of fractal dimension, is capable to reveal structure in a nonlinear signal when compared with a random process.

The contributions of the current work follow. First, while recent works (Alaie et al, 2016; Orlandi et al, 2016; Hariharan et al, 2018; Kheddache and Tadj, 2019 ; Matikolaie and Tadj, 2020) focused on classification of infant cries, the current study conducts a deep statistical analysis on the discriminative power of two main complexity measures in distinguishing between healthy and unhealthy infant cry signals. Second, for robustness of the results, a large dataset is examined, contrary to previous works (Alaie et al, 2016; Orlandi et al, 2016; Hariharan et al, 2018; Kheddache and Tadj, 2019 ; Matikolaie and Tadj, 2020) where relatively small samples were used. Third, statistical analyses will be applied to both inspiration and expiration signals recorded from healthy and unhealthy infants. Fourth, the study will reveal the nonlinear characteristics of healthy and healthy infant inspiration and expiration signals. As a result, one would better understand the nonlinear dynamics of such signals from healthy and unhealthy infant populations. Fifth, eventually, the results are expected to help physicians in decision-making in clinical milieu.

5.3 Methods

5.3.1 Approximate entropy

The approximate entropy (Pincus, 1991) is a nonlinear statistic used which is appropriate to quantify the regularity/irregularity of a nonlinear signal. Additionally, it provides efficient statistical estimates even the original signal length is small [10]. Let consider a time series $\langle x(n) = x(1), x(2), x(3), \ldots, x(N)\rangle$ where N is the time series length. Let m be a positive integer used to represent an embedding dimension and let r be a filter factor. Then, let form the m-vectors $X(1), X(2),\ldots, X(N-m+1)$ where $X(i)=[x(i), x(i+1),\ldots, x(i+m-1)]$ and $i=1, N-m+1$. The distance between $X(i)$ and $X(j)$ is expressed as follows:

$$d[X(i),X(j)] = \max_{k=0,m-1}[|x(i+k) - x(j+k)|] \tag{5.1}$$

Then, for each i=1, (N–m+1), the $C_r^m(i)$ are computed where:

$$C_r^m(i) = \frac{number\ of\ d[X(i),X(j)] \leq r}{N-m+1} \tag{5.2}$$

Then, the quantity $\Phi^m(r)$ is computed as follows:

$$\Phi^m(i) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} log\left(C_r^m(i)\right) \tag{5.3}$$

In a similar way, the quantity $\Phi^{m+1}(r)$ is computed after increasing the dimension to m+1. Finally, the *ApEn* value of the time series can be calculated by:

$$ApEn(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{5.4}$$

Recall that a large value of *ApEn* represents strong irregularity of the current time series. In contrary, a low *ApEn* value implies regularity.

5.3.2 Correlation dimension

The Correlation Dimension (*CD*) of a given time series is effective in providing an unbiased estimator of the intrinsic dimension (Grebogi et al, 1987). In this paper, the *CD* is estimated based on the efficient Grassberger-Procaccia algorithm (Grassberger and Procaccia, 1983a; Grassberger and Procaccia, 1983b). For instance, the *CD* is measured by calculating correlations between points of a time series on the attractor (Alaie et al, 2016; Hariharan et al, 2018). First, let divide the signal x to be analyzed into m dimensions. The m dimensional vectors are expressed as follows:

$$Y_1 = \left(X_1, X_{1+\tau}, \dots, X_{1+(m-1)\tau}\right) \tag{5.5}$$

$$Y_2 = \left(X_2, X_{2+\tau}, \dots, X_{2+(m-1)\tau}\right) \tag{5.6}$$

$$Y_N = \left(X_N, X_{N+\tau}, \dots, X_{N+(m-1)\tau}\right) \tag{5.7}$$

where N is the number of vectors, m is the embedding dimension, and τ is the time delay of the reconstructed phase space. In general, Time delay is calculated by taking the first minimum of the autocorrelation function and the embedding dimension is calculated by using the nearest neighbor algorithm. Then, the correlation integral $C_m(r)$ is computed as follows:

$$C_m(r) = \frac{2}{(N(N-1))} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta\left(\left\|Y_i - Y_j\right\| - r\right) \tag{5.8}$$

where Θ is the Heaviside function given by:

$$\Theta(u) = \begin{cases} 0 & if\ u \leq 0 \\ 1 & if\ u > 0 \end{cases} \tag{5.9}$$

Here, $u = \left(\left\|Y_i - Y_j\right\| - r\right)$ and $\left\|Y_i - Y_j\right\|$ is the distance between two reconstructed vectors and r is the distance parameter. Recall that $C_m(r)$ represents the power law expressed as follows:

$$C_m(r) \sim r^{CD} \tag{5.10}$$

Hence, the correlation dimension (CD) is approximated as follows:

$$CD \equiv \lim_{r \to 0} \frac{log(C_m(r))}{log(r)} \tag{5.11}$$

Finally, the slope of the log-log plot of $C_m(r)$ versus r represents CD. In our study, the embedding dimension was set to two. It was calculated by using the false nearest neighbor algorithm (Theiler, 1987) which is an efficient computational method.

5.4 Data and results

We used the first dataset described in Introduction (Section 0.4.1). The distributions of the estimated approximate entropy measures across healthy and unhealthy infant cries for expiration and inspiration sets are exhibited in boxplots shown in Fig.5.1. Similarly, the boxplots of estimated correlation dimension measures are displayed in boxplots shown in Fig.5.2. As observed, the distributions of approximate entropy measures and correlation dimension measures are clearly different across healthy and unhealthy infants under both expiration and inspiration sets. The results from formal statistical tests in terms of calculated

probability value (*p*-value) are reported in Table 5.1 and Table 5.2 for approximate entropy and correlation dimension respectively.

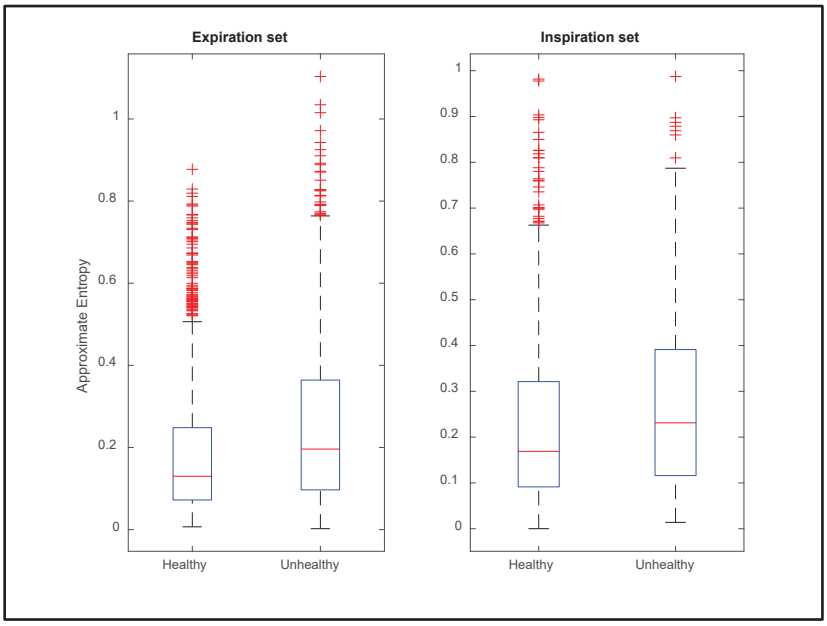

Figure 5.1 Boxplots of estimated approximate entropy measures

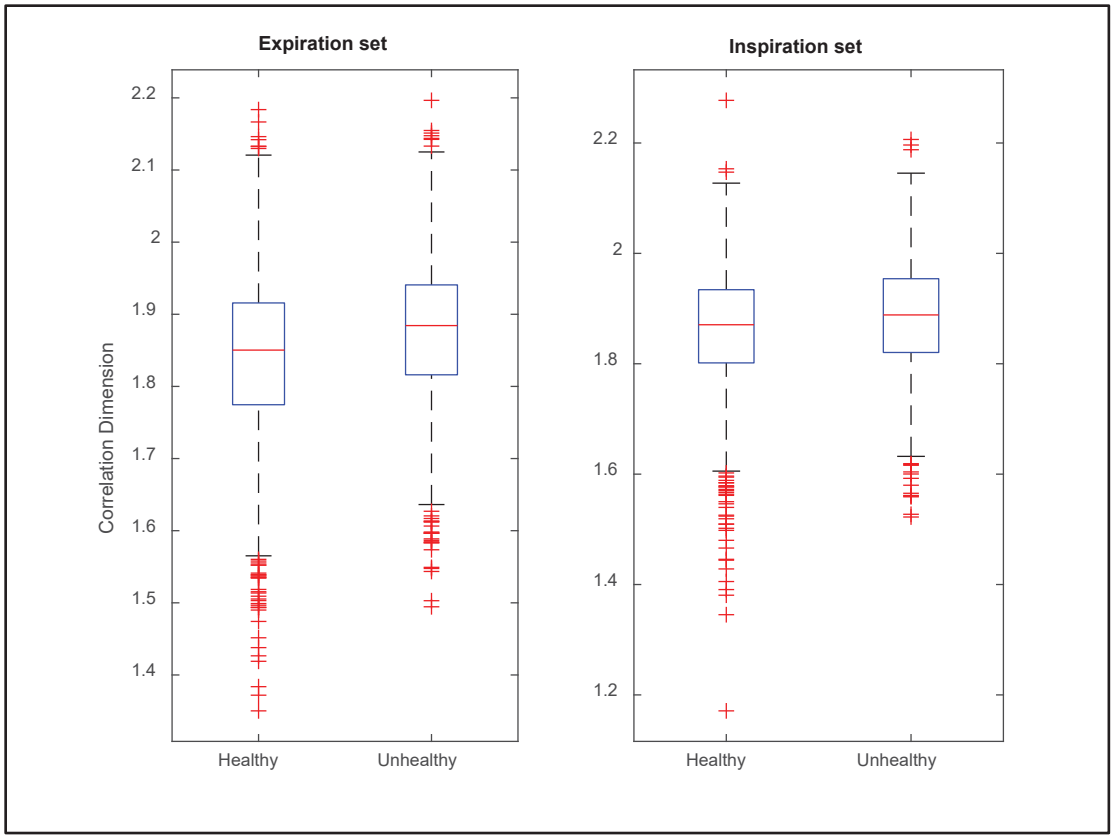

Figure 5.2 Boxplots of estimated correlation dimension measures

According to Table 5.1, the Student *t*-test strongly rejects the null hypothesis for equality of means of approximate entropy in healthy and unhealthy infants under both expiration and inspiration sets. Similarly, the Kolmogorov-Smirnov test strongly rejects the null hypothesis of similarity of distributions of approximate entropy between healthy and unhealthy infants under both expiration and inspiration sets. Besides, the *F*-test shows that variances of approximate entropy across healthy and unhealthy infants are equal in expiration set. However, they are not equal in inspiration set. Following the results of statistical tests presented in Table 5.1, the Student *t*-test strongly rejects the null hypothesis for equality of means of correlation dimension in healthy and unhealthy infants under both expiration and inspiration sets.

Similarly, according to Table 5.2, the *F*-test strongly rejects the null hypothesis of similarity of distributions of correlation dimension between healthy and unhealthy infants under both expiration and inspiration sets. In addition, the Kolmogorov-Smirnov test strongly rejects the

null hypothesis of similarity of distributions of correlation dimension between healthy and unhealthy infants under both expiration and inspiration sets.

In short, the results from statistical tests presented in Table 5.1 and Table 5.2 indicate that approximate entropy and correlation dimension measures exhibit different statistical characteristics across healthy and unhealthy infant cries from both expiration and inspiration sets. Lastly, it is interesting to check whether the level of these complexity measures varies through healthy and healthy infants. In this regard, one-tailed *t*-test is applied to estimated approximate entropy and correlation dimension measures at 5% significance level.

The one-tailed *t*-test results from expiration set rejects the null hypothesis that the mean of approximate entropy in healthy infant is larger than that in unhealthy infants (*p*-value = 1.5419×10^{-22}). Similarly, the one-tailed *t*-test results from inspiration set rejects the null hypothesis that the mean of approximate entropy in healthy infant is larger than that in unhealthy infants (*p*-value = 7.9970×10^{-8}). Therefore, in both expiration and inspiration sets, the level of approximate entropy in cepstrums of healthy infant cries is strongly and statistically lower than that in cepstrums of unhealthy infant cries.

Besides, the one-tailed *t*-test results from expiration set rejects the null hypothesis that the mean of correlation dimension in healthy infant is larger than that in unhealthy infants (*p*-value = 3.4794×10^{-19}). Similarly, the one-tailed *t*-test results from inspiration set rejects the null hypothesis that the mean of correlation dimension in healthy infant is larger than that in unhealthy infants (*p*-value = 1.3556×10^{-7}). Therefore, in both expiration and inspiration sets, the level of correlation dimension in cepstrums of healthy infant cries is strongly statistically lower than that in cepstrums of unhealthy infant cries.

In summary, our nonlinear statistical analysis of cepstrums of healthy and unhealthy infant cries reveals that:

a) Approximate entropy and correlation dimension measures of cesptrum exhibit different statistical characteristics across healthy and unhealthy infant cries from both expiration and inspiration sets.

b) The level of approximate entropy in cepstrums of healthy infant cries is statistically higher than that in cepstrums of unhealthy infant cries. Hence, cepstrums of healthy infant cries show lower randomness compared to cepstrums of unhealthy infant cries.

c) The level of correlation dimension in cepstrums of healthy infant cries is statistically higher than that in cepstrums of unhealthy infant cries. Therefore, cepstrums of healthy infant cries show lower instability compared to cepstrums of unhealthy infant cries.

It follows from the results that the characteristics of the distributions of approximate entropy and correlation dimension carry valuable information in discriminating between pathological and normal cry cepstrums. Specifically, mean, variance and distribution similarity are useful to capture a full picture of the analyzed infant cry signal. In other words, these findings show that the investigated descriptive statistics for cepstrum-based approximate entropy and correlation dimension distributions as possible indicators of cry health status. Finally, it is worth to notice that since the main purpose of the work is to extract nonlinear features from infant cry records to distinguish between healthy and unhealthy infants, performing a surrogate analysis is not required for characterization of records and for the subsequent automatic classification task.

Table 5.1 The *p*-values from statistical tests applied to approximate entropy

	Null hypothesis	**Expiration set**	**Inspiration set**
Student *t*-test	Equality of the means	3.0839×10^{-22}	1.5994×10^{-07}
***F*-test**	Equality of the variances	2.8776×10^{-15}	0.0973
Kolmogorov-Smirnov test	Similarity of distributions	1.4178×10^{-19}	8.7991×10^{-09}
All tests are performed at 5% statistical significance level. A *p*-value less than 5% yields to rejection of the null hypothesis.			

Table 5.2 The *p*-values from statistical tests applied to correlation dimension

	Null hypothesis	**Expiration set**	**Inspiration set**
Student *t*-test	Equality of the means	6.9589×10^{-19}	2.7113×10^{-07}
***F*-test**	Equality of the variances	7.7999×10^{-06}	1.4517×10^{-05}
Kolmogorov-Smirnov test	Similarity of distributions	1.8512×10^{-13}	9.2822×10^{-04}
All tests are performed at 5% statistical significance level. A *p*-value less than 5% yields to rejection of the null hypothesis.			

5.5 Conclusion

The clinical diagnosis of unhealthy infants by analysis of their cries is a non-invasive approach which is becoming attractive in biomedical physics and engineering. Indeed, voice performance is one of the several human functional operations that are affected by disease.

The aim of the current study was to investigate the ability of approximate entropy and correlation dimension estimated in cepstrum domain to characterize infant cry signal. In particular, to check whether cepstrum-based approximate entropy and correlation dimension are different across healthy and unhealthy infant cries, statistical tests were performed by using the Student t-test, F-test, rank sum test, and two-sample Kolmogorov-Smirnov test. The level of significance was set at $p \leq 0.05$. The results reveal that differences in statistical shapes of approximate entropy and correlation dimension, in most cases, are statistically different between the groups of cry cepstrums. More importantly, cepstrums of healthy infant cries show higher approximate entropy level than those of pathological infants. Similarly, cepstrums of healthy infant cries show higher correlation dimension level than those of pathological infants. To be specific, cepstrums of healthy infant cries show lower randomness and disorder compared to cepstrums of unhealthy infant cries. As a result, sample entropy and correlation dimension estimated in Fourier domain provide reliable statistical results for discriminating healthy from pathological voices.

In conclusion, this study indicated that the cepstrum-based approximate entropy and correlation dimension are statistically robust as biomarkers and could potentially be used in the diagnosis of healthy and pathological infant cries.

Acknowledgments

This research is partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2016-05067]

CHAPITRE 6

NONLINEAR STATISTICAL ANALYSIS OF NORMAL AND PATHOLOGICAL INFANT CRY SIGNALS IN CEPSTRUM DOMAIN BY MULTIFRACTAL WAVELET LEADERS

Salim Lahmiri[1,2], Chakib Tadj[2] and Christian Gargour[2]

[1]Department of Supply Chain and Business Technology Management, John Molson School of Business, Concordia University, Montreal.
[2]Department of Electrical Engineering, École de Technologie Supérieure.

Paper published in «Entropy», August 2022

Résumé

Le comportement multifractal dans la représentation cepstrale des signaux de cris d'un nourrisson sain et malsain est examiné au moyen de leaders d'ondelettes et comparé à l'aide du t-test de Student. Les résultats empiriques montrent que les signaux d'expiration et d'inspiration présentent des preuves claires de propriétés multifractales dans des conditions saines et malsaines. De plus, les signaux d'expiration et d'inspiration présentent plus de complexité dans des conditions saines que dans des conditions malsaines. En outre, les distributions des caractéristiques multifractales sont différentes selon les conditions saines et malsaines. Par conséquent, cette étude améliore la compréhension des cris du nourrisson en fournissant une description complète de sa dynamique intrinsèque afin de mieux évaluer son état de santé.

Mots clés : cris de nourrissons, expiration, inspiration, cepstrum, multifractal, wavelet leaders, spectre multifractal, cumulants

6.1 Abstract

Multifractal behavior in cepstrum representation of healthy and unhealthy infant cry signals is examined by means of wavelet leaders and compared by using Student t-test. The empirical results show that both expiration and inspiration signals exhibit clear evidence of multifractal properties under healthy and unhealthy conditions. In addition, expiration and inspiration sig-

nals exhibit more complexity under healthy condition than under unhealthy one. Furthermore, distributions of multifractal characteristics are different across healthy and unhealthy conditions. Hence, this study improves the understanding of infant cry by providing a complete description of its intrinsic dynamics to better evaluate its healthy status.

Keywords: infant cry signal, expiration, inspiration, cepstrum, multifractal wavelet leaders, multifractal spectrum, cumulants

6.2 Introduction

The complexity in various physiological signals is mainly due to the associated complex nonlinear dynamical processes (Müller, et al, 2016). Therefore, complete characterization of non-linear variations in a given physiological signal would help understanding the differences between healthy and pathological cases. For instance, this could be achieved by quantifying the changes in a physiological signal complexity due to abnormalities in terms of variations in its measures.

One of the most useful nonlinear analysis techniques for complex and dynamical systems is the multifractal analysis. Indeed, fractal structures are characterized by self-similarity measured by a scaling independent exponent obtained from a power or scaling law (Peitgen, et al, 1992). Therefore, power laws are prevailing analytical techniques to measure self-similar and the scaling properties of information contents in biological signals.

The main purpose of the current work is to measure multifractal properties in healthy and pathological infant cry signals to obtain multiscale distinctive biomarkers. The ad-vantage of multifractal formalism is to obtain more exponents. Particularly, the signal under study is divided into several fractal sets each producing a fractal dimension. The generated fractals are represented in a spectrum of exponents where each scale is associated with a specific fractal dimension. Indeed, such representation contains full description of nonlinear dynamics in the original signal, for instance, information regarding short and long variations in the signal.

In addition to measuring multifractal properties in healthy and pathological infant cry signals, we seek to investigate whether such measures of multifractal properties are statistically

different between healthy and pathological infant cry signals. In this regard, we measure multifractal properties by using multifractal wavelet leaders (MFWL) technique (Wendt and Abry, 2007) thanks to its intrinsic merits in terms of computational cost and efficiency compared to multifractal detrended fluctuation analysis (MFDFA) (Peng et al, 1994) and the generalized Hurst exponent (GHE) (Di Matteo, 2007).

Besides, we believe that measuring multifractal properties of a baby cry record is a valid approach to study its nonlinear structure so as to better understand its physiology. Hence, in the present paper, we report multifractal analysis from the cepstrum (Childers et al, 1977; Oppenheim, 2010) of infant cry where each fractal quantifies the variability in the scaling of the fluctuations in the underlying cepstrum. Indeed, the cepstrum is widely employed in audio signal analysis as it provides description of the spectrum envelope and spectral richness and characterizes the harmonic and noise components of the original signal (Travieso et al, 2014). In this regard, we make the hypothesis that the subtle variations in the cepstrum can be related to the variations in the original cry dynamics. Such hypothesis will be verified by performing a formal statistical test; for instance, the Student-t test for equality of means between multifractal spectrums of healthy and pathological infant cepstrum domain.

To the best of our knowledge, our paper is the first to report a study in the direction where measures derived from the multifractal spectrum of infant cry signal can be used as a promising tool in the analysis of abnormalities in infant cry signals. Such analysis is followed by a rigorous statistical inference to check whether or not multifractal spectrum are statically different across healthy and unhealthy infant cries. In this regard, our findings are expected to help improving our understanding of the nonlinear dynamics in healthy and pathological cries for better diagnosis.

6.3 Methods: The wavelet leaders

To evaluate the discriminative power of multifractals proprieties in distinguishing between heathy and pathological infant cry signals, we follow a methodology which is based on three steps. First, Fourier analysis is applied to each infant cry signal to obtain its corresponding cepstrum used to describe spectral richness and harmonic characteristics of the original infant

cry signal. Second, multifractal analysis will be applied to each cepstrum to extract its multiscale Hurst exponent used to describe cepstrum multiscale nonlinear dynamics. In this regard, the multifractal wavelet leaders (MFWL) method is employed to obtain multiscale spectrum that characterize cepstrum of healthy and un-healthy cry records. Finally, in the third step, the Student t-test will be performed to check whether or not multifractal descriptors are different across healthy and pathological cepstrums.

The multi-fractal analysis method based on wavelet leaders (Wendt and Abry, 2007) relies on the discrete wavelet transform to describe the characters of singularity spectrum on a full domain while having solid theoretical and mathematical foundations. For instance, the discrete wavelet transform of signal $X=\{x_k,k\in Z\}$ is given by:

$$d_x(j,k) = \int_R X(t)\, 2^{-j} \psi_0\left(2^{-j}t - k\right)dt \tag{6.1}$$

where ψ_0 is a mother function with a compact time support, t is time script, j is parameter of dilation scale, and k is parameter of translation. Then, for signal X, let $S(q,j)$ denotes the structure function and $\zeta(q)$ is the scaling exponents, where q is the order (or moment) of multi-resolution. They are expressed as follows:

$$S(q,j) = \frac{1}{n_j}\sum_{k=1}^{n_j} |L_X(j,k)|^q \tag{6.2}$$

$$\zeta(q) = \lim_{j\to 0} inf\left(\frac{log\, S(q,j)}{j}\right) \tag{6.3}$$

where L_X represents the largest wavelet coefficient calculated at all finer scales. The multi-fractal spectrum $D(h)$ is obtained by the Legendre transform of the scaling exponents. It is defined in the following way:

$$D(h) = \inf_{q\neq 0}\left(1 + qh - \zeta(q)\right) \tag{6.4}$$

The information about variability of the regularity of the signal X is described by singularity (or multi-fractal) spectrum $D(h)$ defined as the Hausdorff dimension as a function of Hölder (Hurst) exponent that takes the value h. Subsequently, the scaling exponents can be computed

as:

$$\zeta(q) = \sum_{p=1}^{\infty} c_p \frac{q^p}{p!} \tag{6.5}$$

where the log cumulants c_p satisfy $\forall p \geq 1$, $C(j,p) = c_{0,p} + c_p \log(2^j)$, and $C(j,p)$ is the cumulant of order $p \geq 1$ of the random variable $\log(L_X(j,\cdot))$.

In practice, $\zeta(q)$ and c_p can be estimated by linear regressions as:

$$\zeta(q) = \sum_j w_j \, log\big(S(j,q)\big) \tag{6.6}$$

$$c_p = log(e) \sum_j b_j C_p(j) \tag{6.7}$$

for scales j with classical linear regressions weights w (Eq.6.6) and b (Eq.6.7). In the current study, the biorthogonal wavelet is employed as the mother function, the number of scales j is set to three, and the q-moment varies from -5 to +5.

6.4 Data and results

We used the first dataset described in Introduction (Section 0.4.1). Fig.6.1 shows the plot of the average multifractal spectrum $D(h)$ for healthy and unhealthy infants following expiration and inspiration. As shown, in all situations, the average $D(h)$ is describing a nonlinear form as a function of h. Hence, cepstrums of healthy and unhealthy infant cry signals are exhibiting multifractal characteristics under both expiration and inspiration phases. Besides, Fig.6.2 exhibits the average estimated scaling exponent function $\zeta(q)$ of cesptrums of healthy and unhealthy infant cry signals. As shown, the average scaling exponents $\zeta(q)$ are a nonlinear function of the moments q for both expirations and inspirations under both healthy and unhealthy conditions. These findings confirm the multifractal behavior in cepstrums as revealed first by the average multifractal spectrum $D(h)$.

For statistical analyses, we display the box plot of the average multifractal spectrums in Fig.6.3. We performed one-tailed Student t-test to verify the null hypothesis that the mean of $D(h)$ in a healthy set is larger than that in an unhealthy set. Accordingly, we found that the

mean of $D(h)$ in healthy expirations is larger than the mean of $D(h)$ in unhealthy expirations (p-value = 0.4990). Similarly, we found that the mean of $D(h)$ in healthy inspirations is larger than the mean of $D(h)$ in unhealthy inspirations (p-value = 0.4990). This finding suggests that the mean of spectrum $D(h)$ is larger under healthy conditions than under unhealthy conditions. In addition, the width of the average multifractal spectrum ($D(h)$) is 0.9816 for healthy expirations and 0.8820 for unhealthy expirations. Thus, cepstrums of healthy expirations show larger degree of multifractality than unhealthy expirations. Besides, the width of the average multifractal spectrum ($D(h)$) is 1.0694 for healthy inspirations and 0.9956 for unhealthy inspirations. Hence, cepstrums of healthy inspirations show larger degree of multifractality than unhealthy inspirations. In summary, expiration and inspiration signals exhibit more complexity under healthy condition than under unhealthy condition.

Finally, we statistically examine the cumulants to characterize the cepstrums of infant cries. For instance, Fig.6.4 displays first cumulant boxplots. The Student t-test accepts the null hypothesis of equality of the means when applied to first cumulant data from healthy and unhealthy expirations as the associated p-value is 0.2058. Similarly, The Student t-test accepts the null hypothesis of equality of the means when applied to first cumulant data from healthy and unhealthy inspirations as the associated p-value is 0.8869. Fig.6.5 exhibits the second cumulant boxplot. The Student t-test rejects the null hypothesis of equality of the means when applied to second cumulant data from healthy and unhealthy expirations as the associated p-value is 1.4661×10^{-07}. Similarly, The Student t-test rejects the null hypothesis of equality of the means when applied to second cumulant data from healthy and unhealthy inspirations as the associated p-value is 1.0421×10^{-06}. Fig.6.6 shows the third cumulant boxplot. The Student t-test rejects the null hypothesis of equality of the means when applied to third cumulant data from healthy and unhealthy expirations as the associated p-value is 0.0053. Similarly, The Student t-test rejects the null hypothesis of equality of the means when applied to third cumulant data from healthy and unhealthy inspirations as the associated p-value is 2.1317×10^{-06}. In short, the results from statistical tests suggest that multifractal characteristics of expiration and inspiration signals are different across healthy and unhealthy conditions. In addition, linear and nonlinear components of cepstrums are statistically different across healthy and unhealthy infant cry signals (Fig.6.4 to Fig.6.7).

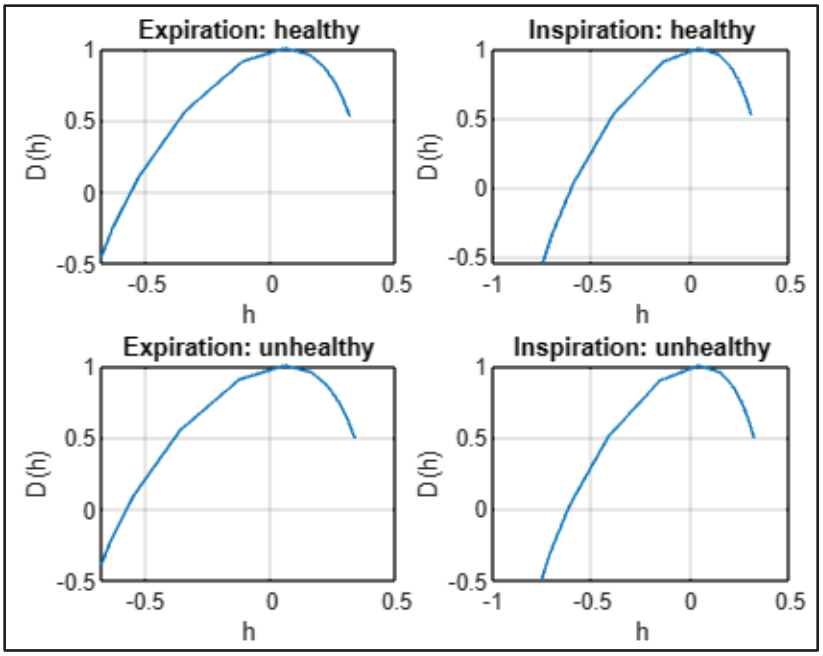

Figure 6.1 Average multifractal spectrums $D(h)$ of cepstrums

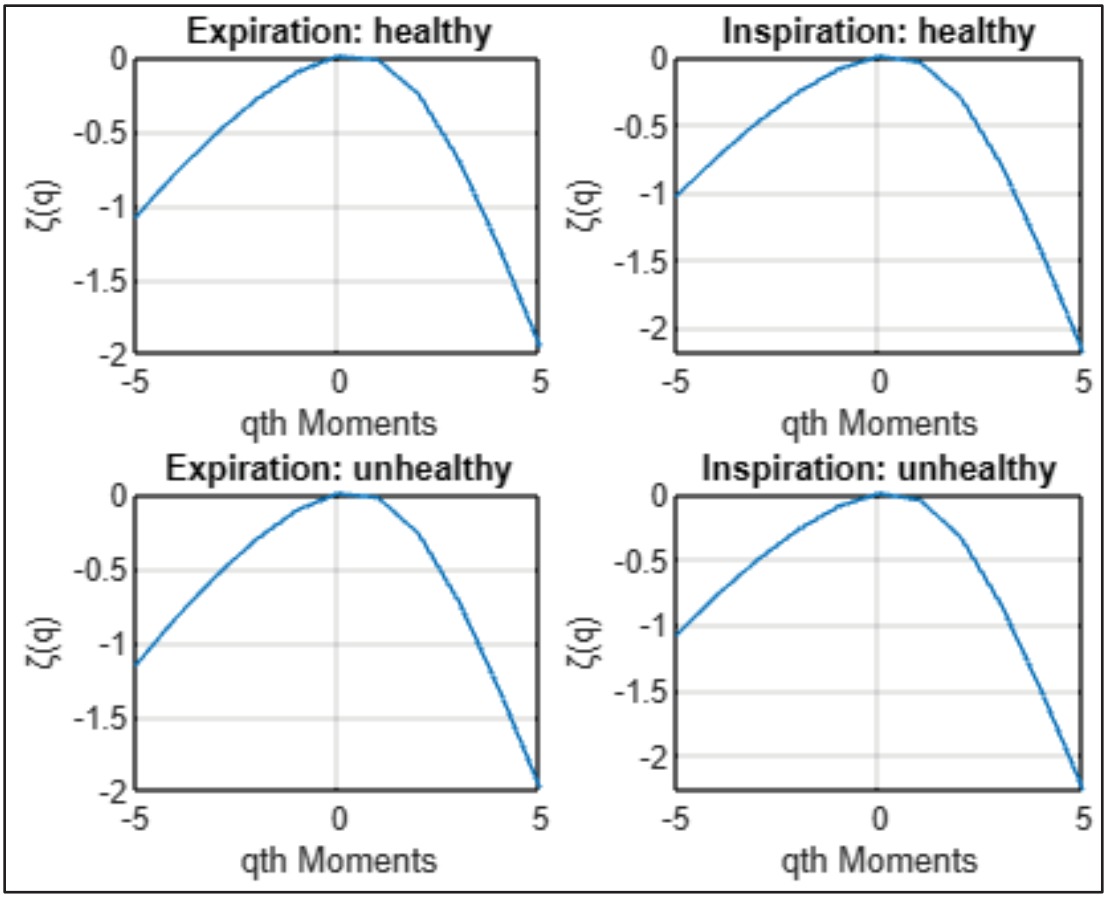

Figure 6.2 Average scaling exponent function $\zeta(q)$ of cepstrums

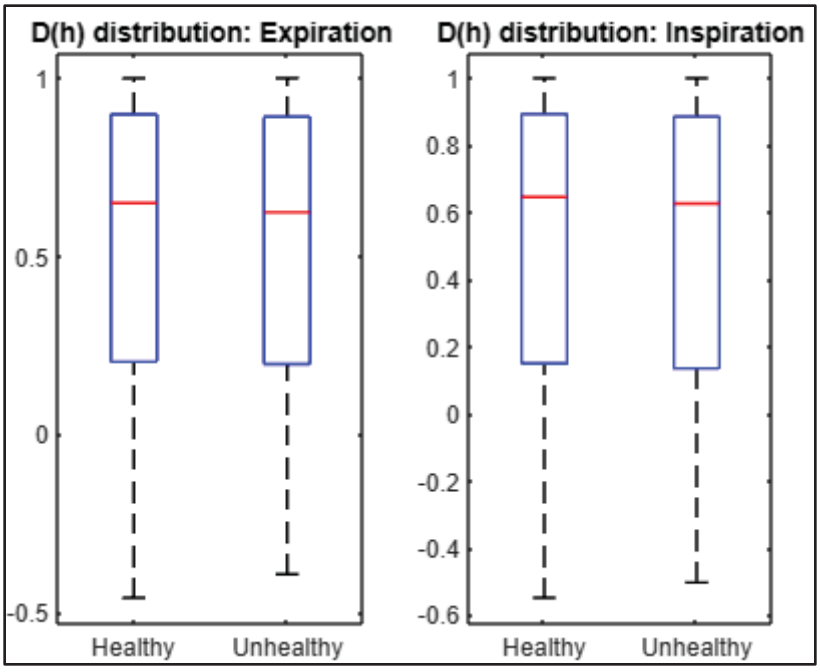

Figure 6.3 Boxplots of the average multifractal spectrums $D(h)$

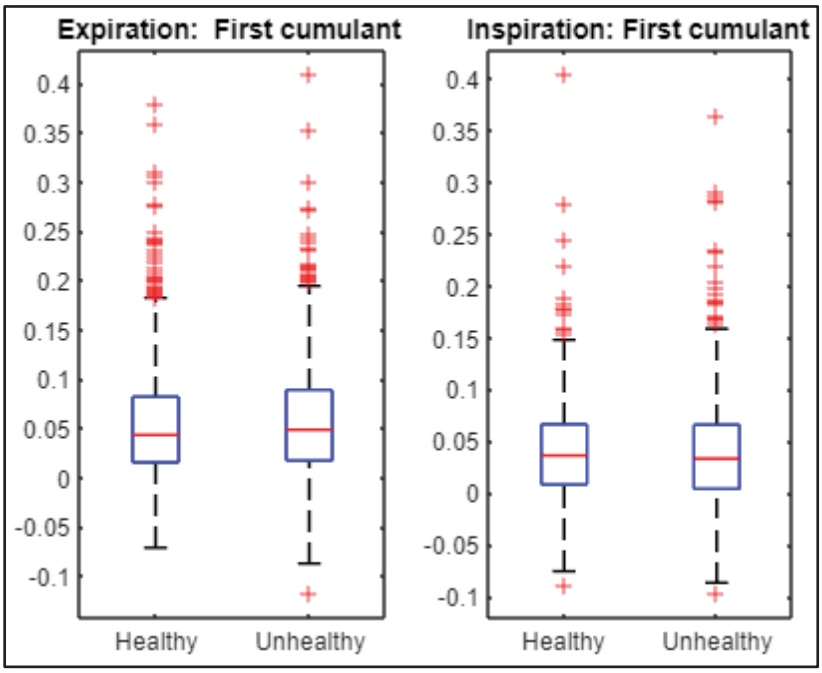

Figure 6.4 Multifractal wavelet leaders: First cumulant boxplots.

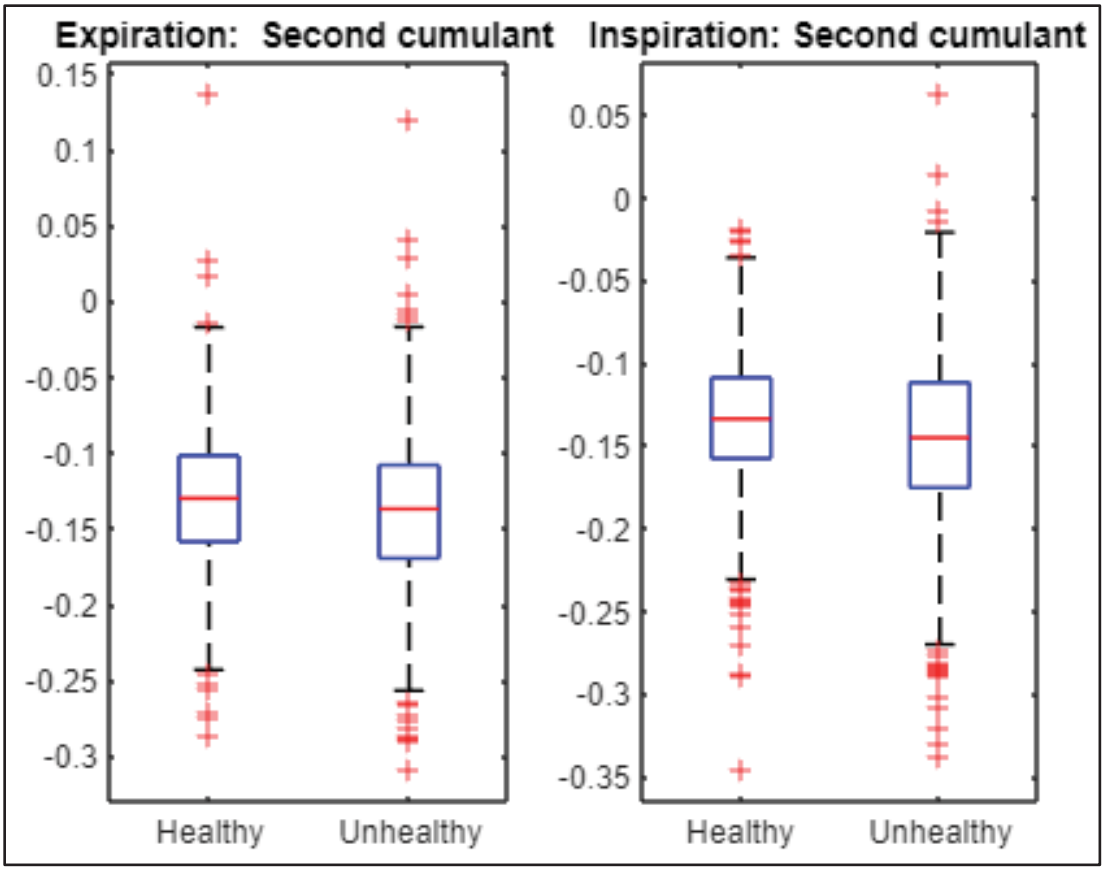

Figure 6.5 Multifractal wavelet leaders: Second cumulant boxplot

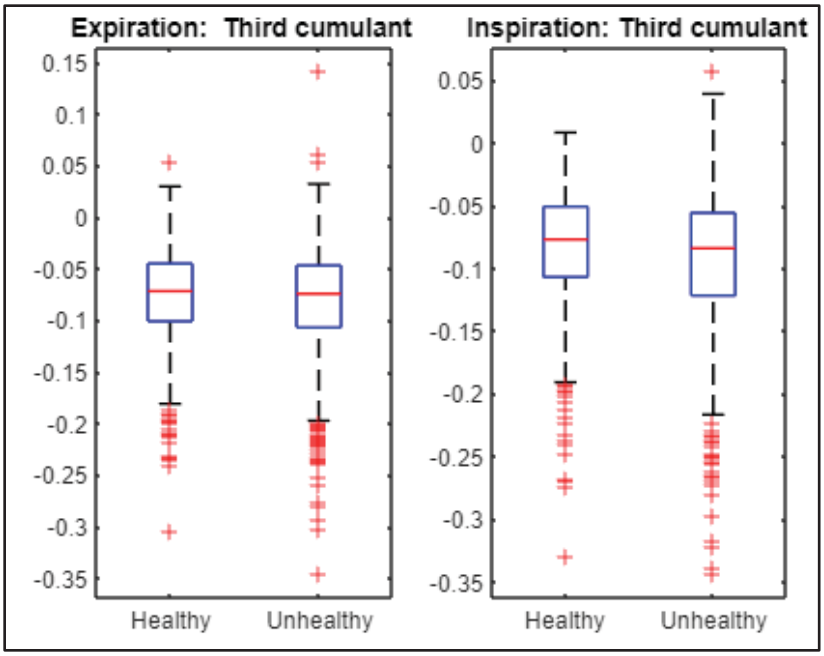

Figure 6.6 Multifractal wavelet leaders: Third cumulant boxplot

In summary, we examined the multifractal structure in cepstrums of cries in healthy and unhealthy infants during expiration and inspiration. The results can be presented as follows:

1. All infant cry records exhibit evidence of multifractal properties according to estimated multifractal spectrums $D(h)$ and scaling exponent functions $\zeta(q)$.
2. The mean of spectrum $D(h)$ is larger under healthy conditions than under unhealthy conditions. Then, expiration and inspiration signals exhibit more complexity under healthy condition than under unhealthy condition.
3. Multifractal characteristics as represented by first, second and third cumulants in expiration and inspiration signals are statistically different across healthy and unhealthy conditions.

Our statistical analysis of baby cry records by means of wavelet leaders revealed interesting findings that might be promising in the medical milieu. Indeed, the findings are statistically significant. The bottom line is that both healthy and unhealthy infant cry records are characterized by multifractal structures. More importantly, the cepstra of healthy records exhibit more fractality and complexity than the cepstra of unhealthy records. Recall that some

recent studies have shown that healthy biomedical signals are characterized by larger complexity compared to unhealthy ones. For instance, the multifractal spectrum of electrocardiograms was found to be small during periods of atrial fibrillation than during various rhythms, including normal ones (Gadhoumi et al, 2018). Likewise, normal electrocardiograms exhibit larger complexity than abnormal ones (Shekatkar et al, 2017). In this regard, the reduced complexity in abnormal biomedical signals can be explained by the presence of malfunctions in the organs and systems of the body (Shekatkar et al, 2017). Such differences can help to understand the complex system of human physiology, specifically in terms of infant cries in healthy and unhealthy conditions. In this regard, multifractal analyses using wavelet leaders may be integrated within a computer to improve the correctness of medical diagnosis for better appropriate treatment.

6.5 Conclusion

It is well known that multifractal properties exist in various biomedical signals, including EEG, ECG, magnetic resonance images and brainstem volume, mammograms, bone radiographic images, retina digital images, dental implant ultrasonic signal, and liver tissue images. However, no work has been devoted to examining the presence of multifractals in baby cry records. Indeed, this investigation could help to understand the physiological aspects of such biomedical signals across healthy and unhealthy babies.

We studied the multifractal behavior in the cepstrum representation of healthy and unhealthy infant cry signals by applying the technique of wavelet leaders. The empirical results demonstrate that both expiration and inspiration signals show strong evidence of multifractal properties under healthy and unhealthy conditions. In addition, for both expiration and inspiration sets, healthy signals exhibit a higher degree of multifractality than unhealthy ones. Our findings could help to understand the multifractal nature in cepstra of healthy and pathological infant cry signals for the better characterization of pathologies. We have not explored the sources of multifractality in the cepstrum domain. However, such investigation is left for future work and will be applied to original infant cry signals. In addition, our future

work will also consider the combination of multifractal characteristics and machine learning for the classification of baby cry records.

Funding:

This research is partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2016- 05067].

Conflicts of Interest:

The authors declare no conflict of interest.

CONCLUSION AND FUTURE WORKS

The goal of the current research study was to design various CAD systems (*i*) to improve the classification of healthy versus unhealthy newborn cry signals and (*ii*) to verify if these signals exhibit different complexity patterns measured by various statistical physics methods. In this regard, our findings can be summarised as follows:

In the problem of classification of healthy versus unhealthy newborn cry signals based on cepstrum coefficients:

(a) The deep feedforward neural network (DFNN) achieved very close to perfect accuracy when applied to expiration infant cry signals and yielded to perfect accuracy when applied to inspiration infant cry signals.

(b) The DFNN outperformed the linear SVM and the Naïve Bayes systems when tested both on the expiration and inspiration sets.

(c) The DFNN outperformed very recent works found in the literature.

(d) The convolution neural networks (CNN) outperformed DFNN and long short term memory (LSTM).

(e) In comparison with similar studies, deep learning systems trained with cepstrum descriptors obtained the highest accuracy.

In the problem of classification of healthy versus unhealthy newborn cry signals based on acoustic features:

(f) The SVM trained with AAM features achieved the highest accuracy followed by *k*NN trained with combination of MFCC, AAM, and prosody.

(g) It outperformed most existing works validated on the same database while being considerably fast to perform.

In the problem of characterization healthy and unhealthy newborn cry signals by using complexity measures:

(h) There are significant differences in approximate entropy and correlation dimension across two categories of subjects.

(i) Healthy infant cry signals show higher approximate entropy level than those of pathological infants.
(j) Healthy infant cry signals show higher correlation dimension level than those of pathological infants.
(k) Healthy signals exhibit a higher degree of multifractality than unhealthy ones.

In short, deep learning systems trained with cepstrum descriptors are promising for analysis and diagnosis of infant cry signals in clinical milieu. Besides, as being effective and explainable, the nonlinear SVM optimized by using Bayesian optimization and trained by Chi-square based selected features from MFCC, AAM, prosody or combination of those selected features, can be promising for diagnosis of newborns based on their cry signals in clinical milieu. However, deep learning allows achieving the highest performance.

Besides, the cepstrum-based approximate entropy and correlation dimension can be considered as biomarkers and could potentially be used in the diagnosis of healthy and pathological infant cries. Moreover, increased multifractal characteristics in healthy subjects suggests can be explained by the absence of malfunctions in the organs and systems of the body.
For future works, we seek to consider various machine learning models, apply different features selection methods, and tune them by using various optimization methods. In addition, another interesting approach could be training deep learning models with acoustic features. Furthermore, we will investigate how other multiscale complexity measures can enhance the discriminative power between healthy and unhealthy cry signals. Moreover, one could consider multimodal physiological features to understand healthy and unhealthy cry signals.

LIST OF REFERENCES

Abbaskhah, A., Sedighi, H. & Marvi, H. (2023). Infant cry classification by MFCC feature extraction with MLP and CNN structures. *Biomedical Signal Processing and Control*, 86, Part B, 105261.

Abou-Abbas, L., Alaie, H.F. & Tadj, C. (2015). Automatic detection of the expiratory and inspiratory phases in newborn cry signals. *Biomedical Signal Processing and Control*, vol. 19, 35-43.

Abou-Abbas, L., Alaie, H.F. & Tadj, C. (2015). Automatic detection of the expiratory and inspiratory phases in newborn cry signals. *Biomedical Signal Processing and Control*, 19, 35-43.

Abou-Abbas, L., Tadj, C., Gargour, C. & Montazeri, L. (2017). Expiratory and inspiratory cries detection using different signals' decomposition techniques. *Journal of Voice*, 31, 259.e13-259.e28.

Acharya, U.R., Faust, O., Kannathal, N., Chua, T. & Laxminarayan, S. (2005). Non-linear analy- sis of EEG signals at various sleep stages. *Computer Methods and Programs in Biomedicine*, 80, 37-45.

Ahmad, S.A. & Chappell, P.H. (2008). Moving approximate entropy applied to surface electromyographic signals. *Biomedical Signal Processing and Control*, 3, 88-93.

Alaie, H.F., Abou-Abbas, L. & Tadj, C. (2016). Cry-based infant pathology classification using GMMs. *Speech Communication*, 77, 28-52.

Anders, F., Hlawitschka, M. & Fuchs, M. (2020). Automatic classification of infant vocalization sequences with convolutional neural networks. *Speech Communication*, 119, 36-45.

Ankışhan, H. & İnam, S.Ç. (2021). Voice pathology detection by using the deep network architecture. *Applied Soft Computing*, 106, 107310.

Ashwini, K., Durai Raj, P.M., Srinivasan, K. & Chang, C.-Y. (2021). Deep learning assisted neonatal cry classification via support vector machine models. *Frontiers in Public Health*, 9, https://doi.org/10.3389/fpubh.2021.670352.

Borowska, M., Bębas, E., Szarmach, J. & Oczeretko, E. (2019). Multifractal characterization of healing process after bone loss. *Biomedical Signal Processing and Control*, 52, 179-186.

Carvajal, R., Wessel, N., Vallverdú, M., Caminal, P. & Voss, A. (2005). Correlation dimension analysis of heart rate variability in patients with dilated cardiomyopathy. Computer Methods and Programs in Biomedicine, 78,133-140.

Chaiani, M., Selouani, S.A., Boudraa, M. & Sidi Yakoub, M. (2022). Voice disorder classification using speech enhancement and deep learning models. *Biocybernetics and Biomedical Engineering*, 42, 463-480.

Chen, L. & Chen, J. (2020). Deep neural network for automatic classification of pathological voice signals. *Journal of Voice*, 36, 288.e15−288.e24.

Childers, D.G., Skinner, D.P., & Kemerait, R.C. (1977). The Cepstrum: A Guide to Processing. *Proceedings of the IEEE*, 65, 1428-1443.

Chittora, A. & Patil, H.A. (2016). Spectral analysis of infant cries and adult speech. *International Journal of Speech Technology*, 19, 841-856.

Cohen, R., Ruinskiy, D., Zickfeld, J., IJzerman, H. & Lavner, Y. (2020). Baby cry detection: deep learning and classical approaches. In *Development and Analysis of Deep Learning Architectures, Studies in Computational Intelligence*, W. Pedrycz and S. M. Chen, Eds. Cham, Switzerland: Springer, vol. 867.

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-97.

Cover T.M. & Hart, P.E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13, 21-7.

Cullen, J., Saleem, A., Swindell, R., Burt, P. & Moore, C. (2010). Measurement of cardiac synchrony using approximate entropy applied to nuclear medicine scans. *Biomedical Signal Processing and Control*, 5, 32-36.

Di Matteo, T. (2007). Multi-scaling in finance. *Quantitative finance*, 7, 21-36.

Gadhoumi, K, Do, D, Badilini, F, Pelter, M.M. & Hu, X. (2018). Wavelet leader multifractal analysis of heart rate variability in atrial fibrillation. *Journal of Electrocardiology*, 51, S83-S87.

Gao, X., Yan, X., Gao, P., Gao, X. & Zhang, S. (2020). Automatic detection of epileptic seizure based on approximate entropy, recurrence quantification analysis and convolutional neural networks. *Artificial Intelligence in Medicine*, 102, Article 101711.

Gelbart, M., Snoek, J. & Adams, R.P. (2014). Bayesian optimization with unknown constraints. https://arxiv.org/abs/1403.5607.

Gerasimova-Chechkina, E., Toner, B., Marin, Z., Audit, B., Roux, S.G., Argoul, F., Khalil, A., Gileva, O., Naimark, O. & Arneodo, A. (2016). Comparative multifractal analysis of dynamic infrared thermograms and X-ray mammograms enlightens changes in the environment of malignant tumors. *Frontiers in Physiology*, 7, 336.

Gerasimova-Chechkina, E., Toner, B., Marin, Z., Audit, B., Roux, S.G., Argoul, F., Khalil, A., Gileva, O., Naimark, O. & Arneodo, A. (2014). Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis. *Frontiers in Physiology*, 5, 176.

Grassberger, P. & Procaccia, I. (1983a). Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, 9, 189-208.

Grassberger, P. & Procaccia, I. (1983b). Characterization of strange attractors. *Physical Review Letters*, 50, 346-349.

Grebogi, C., Ott, E. & Yorke, J.A. (1987). Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics. *Science*, 238, 632-638.

Guo, C., Chen, F., Chang, Y. & Yan, J. (2022). Applying Random Forest classification to diagnose autism using acoustical voice-quality parameters during lexical tone production. *Biomedical Signal Processing and Control*, 77, 103811.

Hammami, I., Salhi, L. & Labidi, S. (2020). Voice pathologies classification and detection using EMD-DWT analysis based on higher order statistic features. *IRBM*, 41, 161-171.

Hariharan, M., Sindhu, R., Vijean, V., Yazid, H., Nadarajaw, T., Yaacob, S. & Polat, K. (2018). Improved binary dragonfly optimization algorithm and wavelet packet based non-linear features for infant cry classification. *Computer Methods Programs in Biomedicine*, 155, 39-51.

Hastie, T., Tibshirani, R. & Friedman, J. (2008). The elements of statistical learning. New York, NY, USA: *Springer*.

Kalauzi, A., Vuckovic, A. & Bojić, T. (2015). Topographic distribution of EEG alpha attractor correlation dimension values in wake and drowsy states in humans. *International Journal of Psychophysiology*, 95, 278-291.

Ketkar, N. (2017). Feed forward neural networks. In deep learning with Python. Berkeley, CA, USA: Apress, pp. 17-33.

Kheddache, Y. & Tadj, C. (2015). Resonance frequencies behavior in pathologic cries of newborns. *Journal of Voice*, 29, 1-12.

Kheddache, Y. & Tadj, C. (2019). Identification of diseases in newborns using advanced acoustic features of cry signals. *Biomedical Signal Processing and Control*, 50, 35-44.

Kim, H., Park, H.-Y., Park, D., Im, S. & Lee, S. (2023). Non-invasive way to diagnose dysphagia by training deep learning model with voice spectrograms. *Biomedical Signal Processing and Control*, 86, Part B, 105259.

Kumar, S.P., Narayanan, N., Ramachandran, J. & Thangavel, B. (2023). Convolutional neural network for voice disorders classification using kymograms. *Biomedical Signal Processing and Control*, 86, Part A, 105159.

Kumar, Y., Dewal, M.L. & Anand, R.S. (2014). Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine. *Neurocomputing*, 133, 271-279.

Lahmiri, S. & A. (2019). Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. *Biomedical Signal Processing and Control*, 49, 427-433.

Lahmiri, S. (2017). Parkinson's disease detection based on dysphonia measurements. *Physica A*, 471, 98-105.

Lahmiri, S. (2023a). A nonlinear analysis of cardiovascular diseases using multi-scale analysis and generalized hurst exponent. Healthcare Analytics, Vol. 3, 100142.

Lahmiri, S. (2023b). A wavelet leaders model with multiscale entropy measures for diagnosing arrhythmia and congestive heart failure. Healthcare Analytics, Vol. 3, 100171.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, vol. 321, 436-444.

Lieberman, P. (1961). Perturbations in vocal pitch. *The Journal of the Acoustical Society of America*, 33, 597-603.

Lim, W.J., Muthusamy, H., Vijean, V., Yazid, H., Nadarajaw, T. & Yaacob, S. (2018). Dual-tree complex wavelet packet transform and feature selection techniques for infant cry classification. *Journal of Telecommunication, Electronic and Computer Engineering*, 10, 75-79.

Luo, X.J., Oyedele, L.O., Ajayi, A.O., Akinade, O.O., Davila Delgado, J.M., Owolabi, H.A. & Ahmed, A. (2020). Genetic algorithm-determined deep feedforward neural network architecture for predicting electricity consumption in real buildings. *Energy and AI*, vol. 2, 100015.

Matikolaie, F.S. & Tadj, C. (2020). On the use of long-term features in a newborn cry diagnostic system. *Biomedical Signal Processing and Control*, 59, 101889.

Matikolaie, F.S. & Tadj, C. (2022). Machine learning-based cry diagnostic system for identifying septic newborns. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2021.12.021

Matikolaie, F.S. & Tadj, C. (2022). Machine learning-based cry diagnostic system for identifying septic newborns. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2021.12.021.

Matikolaie, F.S., Kheddache, Y. & Tadj, C. (2022). Automated newborn cry diagnostic system using machine learning approach. *Biomedical Signal Processing and Control*, 73, 103434.

Mekler, A. (2008). Calculation of EEG correlation dimension: large massifs of experimental data. Computer Methods and Programs in Biomedicine, 92, 154-160.

Müller, A., Kraemer, J.F., Penzel, T., Bonnemeier, H., Kurths, J. & Wessel, N. (2016). Causality in physiological signals. *Physiological measurement*, 37, R46.

Oppenheim, A.W. (2010). Discrete-time signal processing. *Pearson*.

Oprić, D., Stankovich, A.D., Nenadović, A., Kovačević, S., Obradović, D.D., de Luka, S., Nešović-Ostojić, J., Milašin, J., Ilić, A.Ž. & Trbovich, A.M. (2020). Fractal analysis tools for early assessment of liver inflammation induced by chronic consumption of linseed, palm and sunflower oils. *Biomedical Signal Processing and Control*, 61, 101959.

Orlandi, S., Reyes Garcia, C.A., Bandini, A., Donzelli, G. & Manfredi, C. (2016). Application of pattern recognition techniques to the classification of full-term and preterm infant cry. *Journal of Voice*, 30, 656-663.

Orozco-Duque, A., Novak, D., Kremen, V. & Bustamante, J. (2015). Multifractal analysis for grading complex fractionated electro-grams in atrial fibrillation. *Physiology Measurement*, 36, 2269-2284.

Ozseven, T. (2023). Infant cry classification by using different deep neural network models and hand-crafted features. *Biomedical Signal Processing and Control*, 83, 104648.

Peitgen, H-O, Jürgen, H. & Saupe, D. (1992). Chaos and fractals: New frontiers of science. *Springer-Verlag*: New York Ins.

Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E. & Goldberger, A.L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49, 1685-1689.

Pincus, S.M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88, 2297-2301.

Poel, M. & Ekkel, T. (2006). Analyzing infant cries using a committee of neural networks in order to detect hypoxia related disorder. *International Journal on Artificial Intelligence Tools*, 15, 397-410.

Rawal, K., Saini, B.S. & Saini, I. (2015). Adaptive correlation dimension method for analysing heart rate variability during the menstrual cycle. *Physical and Engineering Sciences in Medicine*, 38, 509-523.

Rohini, P., Sundar, S. & Ramakrishnan, S. (2020). Differentiation of early mild cognitive impairment in brainstem MR images using multifractal detrended moving average singularity spectral features. *Biomedical Signal Processing and Control*, 57, 101780.

Rosales-Pérez, A., Reyes-García, C.A., Gonzalez, J.A., Reyes-Galaviz, O.F., Escalante, H.E. & Orlandi, S. (2015). Classifying infant cry patterns by the Genetic Selection of a Fuzzy Model. *Biomedical Signal Processing and Control*, 17, 38-46.

Sachin, M.U., Nagaraj, R., Samiksha, M., Rao, S. & Moharir, M. (2017). GPU based Deep Learning to Detect Asphyxia in Neonates. *Indian Journal of Science and Technology*, 10, 1-5. https://dx.doi.org/10.17485/ijst/2017/v10i3/110617.

Sahak, R., Mansor, W., Lee, Y.K., Yassin, A.I.M. & Zabidi, A. (2010). Performance of combined support vector machine and principal component analysis in recognizing infant cry with asphyxia. *Proceeding of the 32nd IEEE EMBS International Conference*, 6292-6295.

Sarria-Paja, M. & Falk, T.H. (2017). Fusion of auditory inspired amplitude modulation spectrum and cepstral features for whispered and normal speech speaker verification. *Computer Speech and Language*, 45, 437-56.

Satar, M., Cengizler, C., Hamitoglu, S. & Ozdemir, M. (2022). Investigation of relation between hypoxic-ischemic encephalopathy and spectral features of infant cry audio. *Journal of Voice*, S0892-1997, 00147-3.

Scala, I., Rosi, G., Nguyen, V.-H., Vayron, R., Haiat, G., Seuret, S., Jaffard, S. & Naili, S. (2018). Ultrasonic characterization and multiscale analysis for the evaluation of dental implant stability: A sensitivity study. *Biomedical Signal Processing and Control*, 42, 37-44.

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Networks*, 61, 85-117.

Shaikh, A.A.S., Bhargavi, M.S. & Naik, G.R. (2023). Unraveling the complexities of pathological voice through saliency analysis. *Computers in Biology and Medicine*, 166, 107566.

Shayegh, F., Sadri, S., Amirfattahi, R. & Ansari-Asl, K. (2014). A model-based method for computation of correlation dimension, Lyapunov exponents and synchronization from depth-EEG signals. *Computer Methods and Programs in Biomedicine*, 113, 323-337.

Shekatkar, S.M., Kotriwar, Y., Harikrishnan, K.P. & Ambika, G. (2017). Detecting abnormality in heart dynamics from multifractal analysis of ECG signals. *Scientific Reports*, 7, 15127.

Souli, S., Amami, R. & Ben Yahia, S. (2021). A robust pathological voices recognition system based on DCNN and scattering transform. *Applied Acoustics*, 177, 107854.

Souza França, L.G, Vivas Miranda, J.G., Leite, M., Sharma, N.K., Walker, M.C., Lemieux, L. & Wang, Y. (2018). Fractal and multifractal properties of electrographic recordings of human brain activity: Toward its use as a signal feature for machine learning in clinical applications. *Frontiers in Physiology*, 9, Article 1767.

Specht, D.F. (1990). Probabilistic neural networks. *Neural Networks*, 3, 109-118.

Sriraam, N. (2012). Correlation dimension based lossless compression of EEG signals. *Biomedical Signal Processing and Control*, 7, 379-388.

Theiler, J. (1987). Efficient algorithm for estimating the correlation dimension from a set of discrete points. *Physical Review A*, 36, 44-56.

Timcke, R., von Leden, H. & Moore, P. (1960). Laryngeal vibrations: Measurements of the glottic wave-Part III. The pathologic larynx. *Archives of Otolaryngology-head & Neck Surgery*, 71, 16-35.

Ting, H.-N., Choo, Y.-M. & Kamar, A.A. (2022). Classification of asphyxia infant cry using hybrid speech features and deep learning models. *Expert Systems with Applications*, 208, 118064.

Travieso, C.M., Alonso, J.B., del Pozo, M., Ticay, J.R. & Castellanos-Dominguez, G. (2014). Building a Cepstrum-HMM kernel for Apnea identification. *Neurocomputing*, 132, 159-165.

Valdes, M.M., Reyes-Galaviz, O.F., Ortiz, S.D.C. & Reyes-García, C.A. (2022). Chapter 17 - Analysis and processing of infant cry for diagnosis purposes, Editor(s): Alejandro A. Torres-García, Carlos A. Reyes-García, Luis Villaseñor-Pineda, Omar Mendoza-Montoya. *Biosignal Processing and Classification Using Computational Learning and Intelligence*. Academic Press, 2022, Pages 351-374.

Vapnik, V., Golowich, S. & Smola, A. (1996). Support vector machine for function approximation, regression estimation, and signal processing. *Neural Information Processing Systems*, 9, 281-287.

Vapnik, V.N. (1995). The Nature of Statistical Learning Theory. *Springer New York*, NY.

Wang, J., Shao, W. & Kim, J. (2020). Combining MF-DFA and LSSVM for retina images classification. *Biomedical Signal Processing and Control*, 60, 101943.

Wendt, H. & Abry, P. (2007). Multifractality Tests Using Bootstrapped Wavelet Leaders. *IEEE Transactions on Signal Processing*, 55, 4811-4820.

WHO, 19 September 2020, https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality. Accessed on 1 November 2023.

Wu, Y., Chen, P., Luo, X., Wu, M., Liao, L., Yang, S. & Rangayyan, R.M. (2017). Measuring signal fluctuations in gait rhythm time series of patients with Parkinson's disease using entropy parameters. *Biomedical Signal Processing and Control*, 31, 265-271.

Zhang, T., Lin, L., Tian, J., Xue, Z. & Guo, X. (2023). Voice feature description of Parkinson's disease based on co-occurrence direction attribute topology. *Engineering Applications of Artificial Intelligence*, 122, 106097.

Zhang, X., Zhu, X., Zhou, C., Tao, Z. & Zhao, H. (2023). Pathological voice classification based on the features of an asymmetric fluid-structure interaction vocal cord model. *Applied Acoustics*, 207, 109348.

Zhang, Y., Zhou, W. & Yuan, S. (2015). Multifractal Analysis and Relevance Vector Machine-Based Automatic Seizure Detection in Intracranial EEG. *International Journal of Neural Systems*, 25, 1550020.

Zorick, T. & Mandelkern, M.A. (2013). Multifractal detrended fluctuation analysis of human EEG: preliminary investigation and comparison with the wavelet transform modulus maxima technique. *PLoSOne*, 8, e68360.