Using Acoustic Features and Features Fusion in Infant Cry Diagnostic System

by

Zahra KHALILZAD

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

MONTREAL, DECEMBER 19, 2023

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Chakib Tadj, Thesis Supervisor
Department of Electrical Engineering, l'École de technologie supérieure

Mr. Chamseddine Talhi, President of the board of examiners
Department of Software Engineering and IT, l'École de technologie supérieure

Mr. Christian Gargour, Member of the jury
Department of Electrical Engineering, l'École de technologie supérieure

Mr. Nahi Kandil, External examiner
Department of Electrical Engineering, l'Université du Québec en Abitibi-Témiscamingue

THIS THESIS WAS PRENSENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON NOVEMBER 9

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

This thesis is the fruit of the kindness and support of many tenderhearted souls. Nevertheless, if it was not for Professor Chakib Tadj's unconditional support, patience, inspirations, and illuminating advice, this thesis would have not been completed. I am thus grateful to him for his ongoing professional mentorship and for sharing his priceless experience and insights with me. It was an honor to work with him and I am forever indebted to him for his supportiveness and the great journey that I had during my PhD.

I also would like to express my gratitude towards the members of jury who devoted their expertise and precious time to the evaluation of my thesis, I am deeply grateful to each and every one of them.

Thanks to Professor Tadj, my PhD journey was enriched with marvelous opportunities to learn from other people and collaborate with them. In this regard, I would like to thank my colleagues Dr. Yasmina Kheddache for her collaboration on my first article and sharing the experience of her PhD journey with me which paved the way for my research, and to Dr. Ahmad Hasasneh for helping me become more confident by his attentiveness and by sharing his experience and knowledge with me throughout the course of my second article.

The writing and assembly of the articles in this thesis also benefited from the help of Mr. Prasun Lala from SARA writing group at ETS through spending long hours of his invaluable time, his feedbacks, and his thoughtful comments. I highly appreciate all his efforts.

Ultimately, I would like to dedicate this thesis to my husband, Mohammad and my Parents, without whom this journey seemed like a far-fetched dream. I send my heartfelt love and gratitude to my husband who stood by me, inspired and encouraged me, and supported me unconditionally across all these years. I also extend my appreciation to my parents to whom I turned in my hardships and struggles and in return, procured love and adoration.

Utilisation des caractéristiques acoustiques et de la fusion dans un système diagnostique des pleurs du nouveau-né

Zahra KHALILZAD

RÉSUMÉ

Les nouveau-nés communiquent leurs besoins et leurs malaises en pleurant. Au fil des années, les chercheurs ont découvert que ses pleurs donnent beaucoup d'informations sur la santé, les besoins et l'état émotionnel du nouveau-né.

Cependant, ces informations ne sont pas évidentes pour l'oreille humaine et il existe un besoin pour le développement de systèmes précis capables de percevoir les informations transmises par le signal de ces pleurs. Le fait que les pleurs des nouveau-nés ne soient pas compris entraîne de nombreuses complications car ils ne peuvent pas informer les soignants sur leurs besoins. C'est peut-être l'une des raisons des taux élevés de mortalité néonatale dans le monde. En fait, les nouveau-nés courent les risques les plus élevés parmi tous les groupes d'âge des jeunes adolescents. Par conséquent, le développement et l'introduction d'un outil automatisé susceptible de traduire les informations sous-jacentes à différents niveaux du signal des pleurs pourrait être bénéfiques pour sauver des milliers de vies.

Le signal des pleurs a été identifié comme ayant des caractéristiques particulières qui pourraient être altérées en présence d'une pathologie ou sous l'impression d'un état émotionnel tel que la peur. Les différences entre les modèles de signaux de pleurs sains et pathologiques ont favorisé l'émergence de systèmes de diagnostic des pleurs du nouveau-né (NCDS) qui aident au diagnostic et distinguent les pathologies uniquement sur la base des signaux de pleurs du nouveau-né. Plus tard, on a découvert que les pleurs au stade néonatal sont simplement dus à des rythmes biologiques intrinsèques et indépendants et à une maturation sensorimotrice, ce qui signifie que le nouveau-né n'a aucun contrôle sur la génération des pleurs. Cette découverte a conduit à la reconnaissance des signaux de pleurs comme de puissants biomarqueurs dans l'identification des nouveau-nés pathologiques.

Cette thèse visait à proposer un NCDS complet qui bénéficierait de méthodes et d'algorithmes simples mais efficaces pour produire une performance souhaitable. Cet objectif a été atteint à partir de deux perspectives : premièrement, la septicémie en tant que principale cause de décès néonatal a été ciblée, ce qui est sans précédent dans les conceptions NCDS; et deuxièmement, le NCDS a été amélioré à toutes les étapes de sa conception grâce à l'utilisation appropriée de nouvelles fonctionnalités, classificateurs, méthodes de fusion et d'optimisation.

L'étape d'extraction des caractéristiques a été améliorée avec la combinaison appropriée de caractéristiques vocales et musicales qui représentaient différents niveaux d'information. Ces caractéristiques comprenaient des caractéristiques de bas niveau du centroïde spectral et du pic, des caractéristiques de niveau moyen de MFCC, GFCC et BFCC, et enfin, des

caractéristiques de haut niveau du rapport harmonique (harmonic ratio) et des caractéristiques delta et d'accélération pour le cepstral.

Par la suite, l'espace des caractéristiques composé de diverses combinaisons de ces caractéristiques a été réduit pour éviter la redondance et la haute dimensionnalité avec une entropie floue (fuzzy) et des méthodes d'analyse des composants de voisinage (neighborhood component analysis) de la sélection des caractéristiques. Afin de consolider différents ensembles de fonctionnalités dans un espace de fonctionnalités uniforme, l'analyse de corrélation canonique (canonical correlation analysis) a été utilisée comme méthode de fusion au niveau des fonctionnalités.

La prochaine étape du NCDS comprend la classification et le réglage fin des classificateurs pour chacune des expériences. Dans cette étude, nous avons utilisé les schémas de classification la machine à vecteurs de support, le K plus proches voisins (K-nearest neighborhood), le perceptron multicouche (multilayer perceptron) et les réseaux de longue mémoire à court terme pour classer les signaux de pleurs en fonction de leurs classes correspondantes. Chacun de ces classificateurs a été réglé avec différentes méthodes d'optimisation d'hyperparamètres telles que la recherche aléatoire, la recherche de grille (grid search) et Bayesian pour s'adapter à chaque expérience.

La dernière étape de notre NCDS proposé introduit la méthode de fusion de modèles de décision (decision template fusion) pour fusionner les décisions prises par différents classificateurs qui ont été formés par diverses caractéristiques, ce qui permet l'utilisation de caractéristiques de différentes modalités et origines sans avoir besoin de mesures supplémentaires pour les combiner. Les performances du NCDS proposé ont été évaluées à l'aide de différentes mesures d'évaluation telles que la précision, l'aire sous la courbe ROC (AUC-ROC), la précision, le rappel et la F-mesure.

L'objectif principal de cette étude était le développement d'un NCDS complet tout en gardant la pathologie inexplorée de la septicémie comme point focal. En conséquence, en plus de différencier les nouveau-nés septiques des nouveaux-nés sains, le NCDS a été conçu pour distinguer pour la première fois deux pathologies étroitement liées. Succédant aux réalisations précédentes, le NCDS a franchi une étape pour détecter les nouveau-nés septiques parmi un ensemble de 32 autres pathologies. Enfin, une conception complète non intrusive et non sophistiquée a été réalisée qui peut être utilisée comme système d'alerte pour signaler les nouveau-nés à risque plus élevé d'être diagnostiqué avec un groupe de pathologies critiques telles que la septicémie.

Mots-clés: Système de diagnostic des pleurs du nouveau-né, Fusion, Modèle de décision, Analyse cepstrale, Crête, Coefficients cepstraux de fréquence d'écorce, Coefficients cepstraux de fréquence gammatone, Centroïde spectral, Crête spectrale, Rapport harmonique, Mémoire longue à court terme, Perceptron multicouche, Optimisation des hyperparamètres, Corrélation canonique Analyse, analyse des composants de voisinage, déformation du spectre psychoacoustique, fusion de décision, sélection de caractéristiques, réseaux de neurones, cri expiratoire et inspiratoire

Using Acoustic Features and Features Fusion in Infant Cry Diagnostic System

Zahra KHALILZAD

ABSTRACT

Newborns communicate their needs and discomforts through crying. Throughout the years, researchers discovered that the cry emanates opulent information about the newborn’s health, needs, and emotional state. However, this information is not evident to the human ear and there is an inevitable need for the development of precise systems capable of perceiving the information embodied in the cry signal. The abstruseness of the cry signal reveals the newborns to many complications since they cannot divulge their needs to their caregivers. This may be one of the reasons behind the high newborn mortality rates worldwide. In fact, the newborns face the highest risks among all the young adolescent age groups. Therefore, the development and introduction of an automated tool that is susceptible of translating the underlying information in different levels of the cry signal could be beneficial to saving thousands of lives.

The cry signal was discerned to hold peculiar characteristics that could be altered in the presence of a pathology or under the impression of an emotional state such as fear. The differences across the patterns of healthy and pathologic cry signals promoted the emerge of Newborn Cry Diagnostic Systems (NCDS) that facilitate diagnosis and distinguishing the pathologies only based on the cry signals of the newborns. Later on, it was discovered that the cries during the neonatal phase are merely due to intrinsic and independent biological rhythms and sensorimotor maturation, which means that the neonate has no control over the cry generation. This discovery led to recognition of the cry signals as powerful biomarkers in identifying pathologic newborns.

This thesis aimed to propose a comprehensive NCDS that would benefit from simple yet effective methods and algorithms to yield a desirable performance. This objective was realized from two perspectives: firstly, sepsis as a leading newborn mortality root was targeted which is unprecedented in NCDS designs; and secondly, the NCDS was improved across all stages of its design by the proper utilization of novel features, classifiers, fusion, and optimization methods.

The feature extraction stage was improved with the apropos combination of speech-based and music-based features that represented different levels of information. These features included low-level features of spectral centroid and crest, mid-level features of MFCC, GFCC, and BFCC, and finally, high-level features of harmonic ratio and entropy for the cepstral analysis. Subsequently, the feature space consisting of various combinations of these features was pruned against redundancy and high dimensionality with fuzzy entropy and neighborhood component analysis methods of feature selection. In order to consolidate different feature sets into one uniform feature space, the canonical correlation analysis was employed as a fusion method at feature level.

The next stage of the NCDS comprises classification and fine-tuning the classifiers for each of the experiments. In this study, we employed support vector machine, K-nearest neighborhood, multilayer perceptron, and long short-term memory classification schemes to classify the cry signals based on their corresponding classes. Each of these classifiers were tuned with different hyperparameter optimization methods such as random search, grid search, and Bayesian to fit each experiment.

The final stage of our proposed NCDS introduces the decision template fusion method for the fusion of decisions made by different classifiers that were trained by diverse features that capacitates the employment of features from different modalities and origins without the need for any extra measures to combine them. The performance of the proposed NCDS was assessed through different evaluation measures such as accuracy, area under curve of receiver operator characteristic (AUC-ROC), precision, recall and F-score.

The main target of this study was the development of a comprehensive NCDS while revolving around the unexplored pathology of sepsis as a focal point. Accordingly, in addition to identifying septic newborns from the healthy, the NCDS was designed to distinguish between two closely entangled pathologies for the first time. Succeeding the former accomplishments, the NCDS was taken one step former to detect septic newborns from an ensemble of 32 other pathologies. Finally, a comprehensive non-intrusive and unsophisticated design was attained that can be used as an alert system in marking the newborns encountering a higher risk of being diagnosed with a critical pathology group such as sepsis.

Keywords: Newborn Cry Diagnostic System, Fusion, Decision Template, Cepstral Analysis, Crest, Bark-frequency Cepstral Coefficients, Gammatone-frequency Cepstral Coefficients, Spectral Centroid, Spectral Crest, Harmonic Ratio, Long Short-term Memory, Multilayer Perceptron, Hyperparameter Optimization, Canonical Correlation Analysis, Neighborhood Component Analysis, Psychoacoustic spectrum warping, Decision Fusion, Feature Selection, Neural Networks, Expiratory and Inspiratory Cry

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBREVIATION AND ACRONYMS

AAM	Auditory-inspired Amplitude Modulation
ACR	Automatic Cry Recognition
AI	Artificial Intelligence
APGAR	Appearance, Pulse, Grimace, Activity and Respiration
ASD	Autism Spectrum Disorder
AUC-ROC	Area Under Curve of Receiver Operator Characteristic
BFCC	Bark-frequency Cepstral Coefficients
BHPO	Bayesian Hyperparameter Optimization
BSC	Bark Spectral Centroid
CCA	Canonical Correlation Analysis
CNN	Convolutional Neural Network
CNS	Central Nervous System
DCT	Discrete Cosine Transform
DFNN	Deep Feedforward Neural Network
DFT	Discrete Fourier Transform
DL	Deep Learning
DP	Decision Profile
DTF	Decision Template Fusion
EEG	Electroencephalogram
EKD	Ensemble Knowledge Distillation
ERBS	Equivalent Rectangular Bandwidth

EXP	Expiratory Cries
FE	Fuzzy Entropy
FE Selection	Fuzzy Entropy Feature Selection
FFL	Feature Fusion Learning
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
FSM	Forward Feature Selection Method
GF	Gammatone Filter
GFCC	Gammatone-frequency Cepstral Coefficients
GMM	Gaussian Mixture Models
GT	Gammatone
HIE	Hypoxic Ischemic Encephalopathy
HPO	Hyperparameter Optimization
HR	Harmonic Ratio
INSV	Voiced Inspiratory Cries
KNN	K-nearest Neighborhood
LFCC	Linear Frequency Cepstral Coefficients
LPC	Linear Prediction Coding
LPCC	Linear Prediction Cepstral Coefficients
LSTM	Long Short-term Memory
MCC	Matthews Correlation Coefficient

MDQ	Maxima Dispersion Quotient
MFCC	Mel-frequency Cepstral Coefficients
MFDWC	Mel-frequency Discrete Wavelet Coefficients
ML	Machine Learning
MLP	Multilayer Perceptron
NBC	Naïve-Bayes Classifiers
NCA	Neighborhood Component Analysis
NCDS	Newborn Cry Diagnostic System
NICU	Newborn Intensive Care Unit
NN	Neural Networks
OLS	Orthogonal Least Square
PCA	Principal Component Analysis
PLP	Perceptual Linear Prediction
PMF	Probability Mass Function
PNN	Probabilistic Neural Network
PPV	Predictive Positive Value
RBF	Radial Basis Function
RDS	Respiratory Distress Syndrome
RF	Random Forest
RMSprop	Root Mean Square Propagation
RNN	Recurrent Neural Networks
SC	Spectral Centroid

SCCC	Spectral Centroid Cepstral Coefficients
SEN	Spectral Entropy
SENCC	Spectral Entropy Cepstral Coefficients
STFT	Short-time Fourier Transform
SVM	Support Vector Machine
TN	True Negative
TP	true positive
TPR	True Positive Rate
UNICEF	United Nations Children's Fund
URTI	Upper Respiratory Tract Infection
WHO	World Health Organization
WPT	Wavelet Packet Transform
ZCR	Zero-Crossing Rate

INTRODUCTION

0.1 Research Context

The reports provided by United Nations Children's Fund (Unicef, 2022) and World Health Organization (WHO) (World Health Organization, 2021) bear poignant statistics regarding the children and more specifically, newborn mortality rates. UNICEF states that in 2021 the number of under 5 years old children that did not survive due to preventable causes and infectious diseases amounted to 5 million. Moreover, sub-Saharan Africa and southern Asia had the highest mortality rates compared to all the other regions of the world.

The reports from years 2019 to 2022 all show that the neonatal period (the first 28 days of life) is associated with the highest risk of mortality in children under 5 years old with 2.4 million deaths in 2020. It has been reported that 33% of all youth (under 25 years old) deaths occurs within the neonatal period of life which by far has the largest share of all other periods of life. This is drastic news given that there is no means for the neonates to elucidate they are suffering as they are only able to generate cries for different discomforts they encounter. Both the WHO and UNICEF mark infectious diseases such as sepsis, respiratory distress syndrome (RDS), and meningitis to be the mainspring of newborn mortality, causing more than 550000 deaths annually (World Health Organzation, 2014). The rate of possible serious bacterial infections (PSBI) in newborns aged less than two months old (0-59 days old) is 6.9 million in low- and middle-income countries (Unicef, 2020b). Moreover, the focal challenge in low-resource environments is lack of access to diagnostics laboratories and quality care during the first month of life (World Health Organization, 2021).

Newborns impart their needs and emotional state through crying. Crying can stem from physical or emotional exigencies or discomfort such as hunger, wet diaper, and sleepiness. However, unlike speech -even with very limited vocabulary and incorrect pronunciations like toddlers-, a cry is not perspicuous to the caregivers of the newborn. This alone is enough reason for the newborns to be prone to dangers and health risks. Throughout the history, parents and caregivers have tried to link certain types of cries to specific needs; or attribute the health status to various cry characteristics. Although it has been shown that even well-trained medical experts can achieve an accuracy of only 33% depending on the auditory capabilities to discriminate between different cries (Mukhopadhyay et al., 2013), the idea of considering newborn cries as a biomarker opened the door to many future studies and in-depth analysis of the cries.

The newborn cries have been studied for an extensive variety of purposes that include detecting the cry in surveillance systems and Newborn Intensive Care Units (NICUs) (Kim, Kim, Hong, & Kim, 2013; Torres, Battaglino, & Lepauloux, 2017), segmenting and decomposition of the cry signal into its episodes, e.g., expiration, inspiration, hiccups, and silence (Abou-Abbas, Alaie, & Tadj, 2015; Aucouturier, Nonaka, Katahira, & Okanoya, 2011), evaluating the emotional state of the newborn, identifying the reason that initiated the cry, e.g., pain, hunger, wet diaper, or cold (Bano & RaviKumar, 2015; Cohen, Ruinskiy, Zickfeld, IJzerman, & Lavner, 2020; Parga et al., 2020), studying the effect of physical or environmental factors on the cry nature (Bellieni, Sisto, Cordelli, & Buonocore, 2004; Maitre et al., 2017; Mijović et al., 2010) (Wasz-Hockert, Valanne, Vuorenkoski, Michelsson, & Sovijarvi, 1963) (Mampe, Friederici, Christophe, & Wermke, 2009), and finally, diagnosis of pathologies based on the cry signal (Ji, Mudiyanselage, Gao, & Pan, 2021). The focus of the presented study is the diagnostic aspect of the newborn cry signals.

It was not until the late 1800s that researchers recognized the cry signals conveyed more information than simply being a sign of newborn's discomfort and could be in fact a marker in distinguishing the morbid newborns (Bell, 1878; Skeptis, 1995). Copious studies support the fact that the healthy newborns' cry signals are disparate from morbid newborns. The researchers observed that the spectrograms of the pathologic cry recordings had certain irregularities and attributes that eminently differed from the patterns associated with normal healthy cries (Hirschberg, 1980; Vuorenkoski, Lind, Wasz-Hockert, & Partanen, 1971). For instance, chromosomal abnormalities lead to a monotonous, low-pitched, longer-lasting cries (Lind, Vuorenkoski, Rosberg, Partanen, & Wasz-Höckert, 1970); and prematurity correlates with high-pitched cries (Wasz-Hockert et al., 1963). The spectrograms alone were inefficient in capturing all the abnormalities and characteristics that would differ across pathologic cries, and thus, the automatic Newborn Cry Diagnostic Systems (NCDS) were established. During the last decades, the development of the NCDS designs has been influenced by the expansion of both Machine Learning (ML) and Deep Learning (DL) methods; however, contrasted to other audio recognition applications such as speech and music, NCDS still has a high potential for enhancement and exploring.

0.2 Statement of Research Problem

As it was mentioned earlier, the main obstacle in overcoming the newborn mortality is the inadequacy of diagnostic facilities, which necessitated the development of non-invasive, cost-efficient, versatile diagnostic systems like NCDS. The timely diagnosis is of essence for many pathologies, especially the PSBIs, since in this case the treatment could be initiated even without confirmatory clinical results. Moreover, the procedure associated with the clinical diagnosis is extensive, intrusive, and even exorbitant in some cases and regions which further highlights that the development of an automated diagnostic system is imperative.

Conversely, recent advances in both machine learning and deep learning applications and methods have highlighted the potential to achieve highly accurate designs, with discrimination power approaching 100% in some cases. The NCDS architectures hold great potential for improvement when compared to other audio recognition applications, such as speech recognition. In this domain, many features, classifiers, strategies, and optimization and fusion methods remain unexplored, despite being well-known in the domains of speech or music-based applications. While the cry signal shares interchangeable attributes with both music and speech signals, it differs in nature. Therefore, combining methods that have proven successful in either domain is of interest to the NCDS design. Existing literature has only focused on either the speech-like aspect of the cry signal or solely explored it from the musical aspect. For instance, the Mel-frequency Cepstral Coefficients (MFCC) feature set, derived from speech processing applications, is the most common in cry analysis studies. The features explored in the NCDS designs belong to six main categories: cepstral, frequency, time, time-frequency, prosodic, and image (Ji et al., 2021). Each of these domains provides information on the cry signal from a different perspective. Among these domains, the cepstral domain has gained the highest attention owing to the success of MFCCs. However, a deduction from examining other publications suggests that cepstral analysis is highly correlated with the good performance of the NCDS. Features like Linear Prediction Cepstral Coefficients (LPCCs) (Liu et al., 2018), Linear Frequency Cepstral Coefficients (LFCCs) (Jagtap et al., 2016; Patil et al., 2022), Bark Frequency Cepstral Coefficients (BFCCs) (Liu et al., 2019; Maghfira, Basaruddin, & Krisnadhi, 2020; Sriraam & Pradeep, 2019), and Gammatone Cepstral Coefficients (GTCCs) (Satar et al., 2022) all achieved fair discriminative performances in NCDS designs. Feature fusion approaches are mostly limited to concatenation, even in very recent studies (Zhang, Ting, & Choo, 2018; Ji, Jiao, Chen, & Pan, 2022).

Conversely, features from the prosodic domain, such as formants, intensity, rhythm (Matikolaie et al., 2022), unvoiced regions' detection (Abou-Abbas, Tadj, & Fersaie, 2017), and duration, have shown desirable results in the NCDS designs when implemented alone or combined with features from the same domain. The main challenge is that there are none or very few research works that combine different feature domains to form a single feature space. Besides, the implementation of the apt feature fusion techniques is seldom seen in NCDS. Not only in NCDS but in many other audio recognition applications, the role of fusion in different levels of the design is not considered. Fusion opens the door to utilizing features and classifiers from different modalities, each capable of effectively capturing a certain aspect of the cry signal. More specifically, fusion at decision level which is crucial for employing features and classifiers from diverse modalities and their different combinations. There is also meager information about hyperparameter tuning algorithms and their use in NCDS despite their beneficial impacts on many audio processing systems.

Finally, certain pathologies have not been studied or well addressed with the NCDS and the characteristics of the cry signals that are marked with that pathology stay undiscovered. Among these pathologies, there is meagre literature existing about sepsis in spite of this pathology being among the mortality roots in neonates and common. However, pathologies like asphyxia own high amounts of research across the world.

In this research, different stages of the NCDS were developed with novel methods and tools. The main focus was to take the NCDS to a comprehensive level where it would be improved from different aspects to address sepsis. This goal was achieved by the means of employing and introducing novel features, feature selection methods, fusion methods, employing classifiers from different natures, and optimizing these classifiers to be well-suited to the problem at hand.

0.3 Research Objectives

The main objective in this study was to design an inclusive NCDS with the help of machine learning techniques to identify pathologic newborns based on only their cry signal recordings.

Furthermore, this study addresses sepsis in an inclusive manner; trying to diagnose septic newborns from healthy, distinguishing between sepsis and a closely entangled pathology group (RDS), identifying sepsis amidst an ensemble of more than thirty other pathologies, and finally distinguishing between a collective of pathologies -including sepsis- and the healthy. The novelty of this work is not limited to proposing new features or the introduction of methods that were novel to the NCDS; but also, the distinguishing between two pathologies as well as identifying one pathology amongst a group of other pathologies were explored for the first time.

As discussed before, one of the base conditions concerning the NCDS design is keeping the design simplistic and unsophisticated; thus, the employment of the methods that would aid pruning the feature space in terms of dimensionality and redundancy, as well as tuning the classifier hyperparameters are inevitable. There is no prior research in the field of NCDS design that would assess the role of different hyperparameter optimization (HPO) methods on the performance comparatively; hence, performing a comparison between the HPO methods would be beneficial to the advancement of NCDS.

Principally, the framework can be enhanced in certain steps which constitute the goal of this study:

- Devise novel features from diverse levels and modalities, including short-term and long-term, music-derived and speech-derived, and their combinations.

- Find and introduce suitable feature selection and fusion criteria to combine features from aforementioned modalities.

- Employ and compare different machine learning and deep learning schemes for the classification of the cry signals depending on the goal of the NCDS.

- Assess the role of HPO in tuning the classification algorithms and compare them based on instructive evaluation measures such as Matthew's Correlation Coefficient (MCC).

- Bring the classifier fusion methods to light that would ensure simplistic fusion of different classifiers effectively, such as decision template fusion (DTF).

- Analysis of the differences across different classes of newborns (healthy versus pathologic or healthy versus septic, etc.) by the introduced features.

- Design a comprehensive NCDS that assesses a pathology from different aspects: healthy vs. septic, septic vs. Respiratory Distress Syndrome (RDS), septic vs. non-septic (ensemble of pathologies), and healthy vs. pathologic.

0.4 Methodology

To this point, the necessity, the novelties, and the goals of this study were explicated; however, the means to achieving these goals, is what constructs this work. It was mentioned that certain acoustic cues in the cry signal of an ailing newborn makes it distinguishable from the healthy newborn cry. Nevertheless, not all these acoustic cues are detectable with simple methods such as spectrograms and thresholding, which calls for attentive extraction of relevant features capable of succinctly encapsulating the characteristics of each group. Besides, the auspicious feature extraction leads to the simplification of the ensuing stages, e.g., the classification.

Analogous to other audio recognition systems, the NCDS framework incorporates three main elements: 1. Preprocessing, 2. feature extraction, fusion, and selection, and 3. Classification and classifier tuning. All these elements should be designed based on the available data, and the specifications of the dataset. Therefore, the cry database will be introduced prior to the discussion of the next steps.

In the preprocessing step, we ensured consistency with previous research in our lab by applying a Hamming window of 10 ms length with a 30% overlap to the system. Additionally, we utilized both EXP and INSV segments of the cry signal. For the feature space formation step, the primary focus of this study, we consistently incorporated features from speech processing and musical applications. Novel features, including Gammatone Frequency Cepstral Coefficients (GFCCs), Bark Spectral Centroid, Spectral Centroid Cepstral Coefficients (BSC and SCCC), spectral crest, Spectral Entropy Spectral Centroid (SENCC), as well as BFCC and harmonic ratio, were extracted and evaluated. To create a robust and efficient feature space, complementing statistical methods for dimensionality reduction, we employed fuzzy entropy and nearest neighborhood feature selection approaches to reduce redundancy and dimensions. Additionally, the fusion at the feature level was significantly enhanced through the use of canonical correlation analysis.

In the classification step, we employed Bayesian, random, and grid search hyperparameter optimization methods to enhance classifier functionality, extensively comparing them in terms of computational costs and performance details. Different machine learning and deep learning classification schemes, such as SVM, KNN, MLP, and LSTM, were utilized within our NCDS architectures. Finally, the decision template method was employed for decision-level fusion, representing a significant contribution.

0.5 Cry Database and its Demographics

Data acquisition and the establishment of the database are both the basis and the primary challenge in designing pathological research. There were several considerations that made this database stand out from the rest. To design a comprehensive system, the cry signals were collected from a diversity of races and origins. The data collection took place in two distant locations of the world, Lebanon, and Canada, and enabled us to include participants from Algeria, Canada, Haiti, Portugal, Syria, Turkey, Bangladesh and other countries from Arabic, Quebecois, African, Caucasian, Latino and other races. This unique diversity of participants ensures that the system would not be limited to or biased by certain acoustical characteristics due to mother tongue. Besides, the collection of the data was not restricted to a predefined ideal condition; the recordings were gathered in the maternity rooms, NICUs, and postnatal wards in the presence of noises like staff chatter, medical equipment beeps, and whining and crying of other newborns, which makes the database practical for solving the problems discussed earlier. The recording of the cry signals was made possible by the means of a common digital handheld voice recorder that was placed in the 10-to-30-cm locus of the newborn’s head. The recorder was a 2-channel Olympus with a 44.1 kHz sampling frequency and 16-bits resolution which is accessible and affordable even for the low-income settings. There were no conditions regarding the reason of crying and the cry signals were initiated due to different reasons such as hunger, temperature reading, reflux, discomfort, fear, intravenous injection, pain, weighting, ophthalmic exams, nebulizer, wet diaper, and shower. The participants were selected from both genders, male and female. They weighed from 520 g to 5.2 Kg and were full-term below 53 days of age. Furthermore, the Appearance, Pulse, Grimace, Activity and Respiration (APGAR) scores were recorded for all the participants. Most importantly, the newborns in this study were healthy or either diagnosed with one of the following 32 pathologies: ankyloglossia, apnea, asphyxiation, aspiration, bronchiolitis, bronchopulmonary dysplasia, choanal atresia, cleft palate and lip, complex cardio, cyanosis, down syndrome, duodenal atresia, dyspnea, fever, gastroschisis, grunting, hyperbilirubinemia, hypoglycemia, hypothermia, intrauterine growth retardation, kidney failure, meconium aspiration syndrome, meningitis, myelomeningocele, respiratory distress syndrome, retraction, seizure, sepsis, tachypnea, thrombosis in vena cava, vomit.

The cry recordings have varying durations ranging from one to four minutes with an average of one and a half minutes. There are one or more recordings (up to five recordings per newborn) from each newborn. A total of 1115 recordings from 410 newborns were included in this study where the healthy newborns have more than 73% share of the entire dataset with 300 participants and 785 recordings.

It was mentioned that the cry recordings include unwanted information and noise; this includes the environmental noise and unwanted information from the newborn such as hiccups, silence, and grunting. Moreover, as can be deduced from the statistical analysis of the database, the number of newborns diagnosed with different pathologies is far less than the healthy newborns. This is not unexpected based on several points: 1. The occurrence and the observation of a certain pathology in the prespecified duration of data acquisition phase is unpredictable. 2. Meeting the ethical and the technical criteria for the addition of the recordings from each participant is onerous. 3. The correct labeling and documentation corresponding to each cry recording requires the efforts of many individuals such as nurses, medical experts, and teams of researchers. All being said, it is indispensable to make use of every recording in full measure; therefore, all the cry signals were segmented into multiple episodes based on the natural respiratory activities such as hiccups, expiratory cry, and inspiratory cry. The cry episodes were manually annotated using WaveSurfer software. A full list of these segments and their labels is given in Table 0.1.

Table 0.1 Segments of cry signal

Unit	Definition
EXP	Voiced expiratory segment during a period of cry.
EXPN	Unvoiced expiratory segment during a period of cry.
INS	Unvoiced inspiratory segment during a period of cry.
INSV	Voiced inspiratory segment during a period of cry.
EXP2	Voiced expiratory segment during a period of pseudo-cry.
INS2	Voiced inspiratory segment during a period of pseudo-cry.
PSEUDOCRY	Any sound generated by the newborn that is not a cry, such as whimpering.
Speech	Sound of the nurse or parents talking.
Background	A low noise characterized by a low-power silence affected with little noise.
Noisy cry	Any sound heard with a cry (BIP, water, diaper changing, etc.).
Noisy pseudo-cry	Any sound heard with a pseudo-cry.
Noise	Sound originated from the surrounding environment such as the microphone movement, the diaper movement, door squeak, staff chatter, background noise, or speech with BIPs.
BIP	Sound of the medical equipment surrounding the newborn.

0.6 Proposed Framework

The foremost stage of the NCDS design is to exploit and select the pertinent features for the recognition of an unhealthy newborn and then delineate the following stages accordingly.

Therefore, in this research the features that were unexplored in NCDS architecture and the features that were new to the audio analysis applications were introduced and extracted. The NCDS takes the cry signals and assigns them to their corresponding classes of unhealthy and healthy. As already stated, the NCDS designs have some common elements regardless of the purpose, namely pre-processing, feature extraction, and classification. In the pre-processing step the signals are segmented based on the respiratory activities and the expiratory and inspiratory episodes are selected as the inputs of the NCDS which are marked as EXP and INSV in Table 1. These segments are then windowed, undergo filter banks, and are pre-emphasized corresponding to the features that will be extracted in the following steps. For example, in order to extract the Mel-frequency Cepstral Coefficients (MFCCs) the signal is windowed, pre-emphasized by a high pass filter, then transformed via the Discrete Fourier Transform (DFT), passed through a triangular filter bank, and then scaled on the Mel-frequency mapping to extract the MFCCs, while for the spectral features such as Spectral Centroid (SC) it might be only required to window the signal and transform it. The details for the preprocessing of each feature set are explained prior to detailing the features in chapters 2, 3, 4, and 5.

Following the pre-processing stage, the features on different levels of information and originating from various modalities are extracted and then the most suitable ones are selected. A wide range of features were introduced and extracted in this research that will be only shortly mentioned here and then elucidated at the following chapters 2 to 5. The Cepstral analysis was the protagonist in this study owing to the fact that it facilitates the isolation of the constituents in voice generation such as glottal pulse and the vocal tract impulse response and thus, attested success in the field of speech analysis (Oppenheim, Schafer, & Stockham, 1968). Firstly, the MFCC features were extracted as a baseline to ensure comparability of the presented work with the existing literature in chapters 2 and 3. Then, in chapter 3, it was shown that the Gammatone-frequency Cepstral Coefficients (GFCCs) are superior to the MFCCs and hence, the GFCCs were implemented in the next two chapters, 4 and 5. The Bark-frequency Cepstral Coefficients (BFCCs) were also explored to represent deeper analysis and comparison between the psychoacoustic mappings of the spectrum. Spectral Entropy Cepstral Coefficients

(SENCC) were employed to show the level of complexity in the cries of septic newborns as opposed to the healthy newborns. This study adopts the psychoacoustic mappings of the spectrum to the proposed features to further befit it to the nature of the cry which resulted in developing new acoustic features. The SC is convenient to the study of musical applications for determining the brightness of a sound and it was rendered in chapters 2 and 5 as the basis of two feature sets, SC Cepstral Coefficients (SCCC) and the Bark-SC (BSC). Another feature that was also inspired by the musical applications is Equivalent Rectangular Bandwidth-based Spectral Crest (ERBS Crest) that is mainstream in clinical audiology settings and in audio fingerprinting but new to the NCDS designs. Finally, the Harmonic Ratio (HR) feature set was studied to further emphasize the primary musical aspect of the cry signal; harmonicity, and to interpret how the cry signals of the healthy newborns are more harmonic compared to the cry signals of the unhealthy. The feature extraction is always accompanied by two concepts: the feature selection to ensure the simplicity of design and preventing the system from having high-dimensional and/or redundant features; and the combination of the features in order to form an efficient feature space and represent different aspects of the input desirably. For the feature selection, two methods of Fuzzy Entropy feature selection (FE Selection) and Neighborhood Component Analysis (NCA) were investigated which will be narrated in chapters 2 and 5, respectively. As for the feature combination, this work benefits from simple concatenation of the mentioned feature sets in different parts of this study. Additionally, a fusion technique at feature level is expounded from multiple aspects in chapter 3 through the Canonical Correlation Analysis (CCA) of the GFCC and MFCC feature sets.

The last component of the NCDS is the classification scheme. Several classifiers from ML, DL, and Neural Network (NN) domains were adopted. Support Vector Machines (SVMs) were used as the baseline in all of the presented studies owing to their high performance and prevalence. K-Nearest Neighborhood (KNN), Multilayer Perceptron (MLP), and Long Short-term Memory (LSTM) classifiers are employed here as well. Many different classifiers have been introduced and utilized in different tasks in the literature; however, the discussion around the fine-tuning of the classifiers and the effects of Hyperparameters (HP) on the performance in a comparative and inclusive manner is infrequent. Therefore, this study employs three different techniques of random search, grid search and Bayesian to amend the performance of the NCDS in consonance with each of the defined tasks. The final contribution of this study is a subset of the classification stage, fusion at the decision level which combines the outcomes of different classifiers trained on diverse features in order to yield the final decision. The DTF method was implemented as the fusion algorithm owing to its robustness to overtraining and preferable performance with fewer data. At last, the NCDS was assessed through a set of evaluation measures such as accuracy, F-score, MCC, specificity, and precision to better demonstrate different aspects of the design.

0.7 Thesis Composition

The presented manuscript is a thesis by articles which constitutes three publications and one submitted article for publication in the scientific journals. The first article assessed the identification of septic neonates from the healthy based on an entropy-based framework and addressed sepsis as a single pathology group to find out whether and how the septic cry signals differ from the healthy cry patterns and answered these questions by the illustration of differences for each of the feature sets.

After conducting the first fragment of the research, we realized not only the investigation of the septic newborn cries suffers from a paucity, but also, there was no prior research that compared different pathologies to each other. Therefore, in the second article, sepsis was dissected against RDS which could be very perplexing as a consequence of the similar attributes and entanglement of the duo. It is noteworthy to mention that though not as equitably severe, the RDS itself sustains a similar research gap to sepsis in mature newborns. The success in the separation of the septic from healthy and RDS from sepsis diagnosed newborns led us to inspect two novel ideas: 1. A design that would benefit from deep learning methods with higher number of samples to investigate the possibility of discerning healthy newborns from any pathologic newborns since the MLP method was successful, and 2. A framework that would be capable of a pinpointing a single pathology from a collection of other pathologies; which formed the foundation for the next two articles.

The third article is devoted to the study of healthy versus pathologic newborns in order to propose a non-intrusive and low-cost alert system to inform the guardians of the newborn and the medical specialists to take all the necessary measures not to overlook the newborn as being healthy. From the technical point of view, this article administers a comprehensive comparison between the GFCC and MFCC feature sets as well as comparing three different methods of HPO for two classifiers in the ML and DL domains, SVM and LSTM in terms of a set of different evaluation measures, run-times and methodology. Moreover, it appraises the role of fusion at feature level in the NCDS design by demonstrating how it would affect and homogenize the feature space and consequently, the performance of the NCDS.

Finally, the fourth article peruses the distinction between sepsis and an assemblage of other pathologies by the means of studying different novel feature sets and the DTF to fuse them. Moreover, the NCA feature selection is employed to show how the extreme downsize of the feature set would result in the functioning of the NCDS.

The thesis is composed in six chapters. The first chapter presents a chronicle of NCDS development and the tools that were utilized up to date to improve its performance as well as discussing the cry generation in the newborns, the development of the newborn vocal tract, and the findings about the cry signals associated with different pathologies in newborns. The next chapters two to five, cover the articles mentioned below:

- Khalilzad, Z., Kheddache, Y., & Tadj, C. (2022). An entropy-based architecture for detection of sepsis in newborn cry diagnostic systems. *Entropy*, *24*(9), 1194.

The above-mentioned article forms the second chapter which was published in the journal of Entropy in August 2022. The novelty of this study is several-fold: 1. The introduction of a framework that revolves around the information content of the cry signal via study of the SEN and FE selection. 2. HPO of the KNN classifier by choosing different distance measures. 3. Checking for redundant features in the feature space. 4. Combination of a music-derived feature set with the Cepstral analysis in SCCC feature set. This framework had the task of identifying septic newborns from the healthy with two different datasets of INSV and EXP.

- Khalilzad, Z., Hasasneh, A., & Tadj, C. (2022). Newborn cry-based diagnostic system to distinguish between sepsis and respiratory distress syndrome using combined acoustic features. *Diagnostics*, *12*(11), 2802.

The third chapter represents the next piece of our research, published as an article in the journal of Diagnostics in November 2022. For the first time in the history of NCDS designs, two pathology groups were separated based on a combination of the short-term and long-term features. Moreover, the HR feature set was shown to have strong discriminative power in despite of its low dimensions which suggested further investigation of the harmonic and musical components in the cry signals. The MLP and SVM classification schemes were employed to rate the performance of the individual feature sets as opposed to the combined feature set.

- Khalilzad, Z., & Tadj, C. (2023). Using CCA-fused cepstral features in a deep learning-based cry diagnostic system for detecting an ensemble of pathologies in newborns. *Diagnostics*, *13*(5), 879.

Chapter four is dedicated to gauge the performance of the GFCC feature set compared to the MFCC feature set, the impact of different HPO methods, and the potency of the CCA-fusion. Moreover, it is among the first studies to introduce a fusion algorithm at feature-level. The research is structured in a way that advancing through multiple experiments, illustrates the ramification of each of the contributions. Therefore, firstly, the performance of each feature set is evaluated from various aspects: 1. With separate datasets of INSV and EXP, 2. With two different classifiers (LSTM and SVM), 3. With multiple evaluation measures, 4. From the run-time's perspective, and 5. With three different HPO methods for each of the classifiers. Then, the role of the CCA-fusion is assessed based on all the aforementioned criteria with the addition of discussing the new feature vector size as well as comparing it to the simple concatenation of features. Furthermore, the HPO methods are contrasted in terms of evaluation measures and the run-times. Finally, two more experiments were conducted that exhaustively go through the effect of HPO to validate the achievements of the prior experiments since the results were surprisingly good and to further confirm the patterns that were concluded from the previous observations in earlier steps. Notably, this study introduced a comprehensive NCDS that can be employed as a non-intrusive alert for the medical experts and the newborn caregivers not to disregard the newborn when there are no apparent symptoms in the first check-ups.

- Khalilzad, Z., & Tadj, C. (2023) Use of psychoacoustic spectrum warping and decision template fusion in newborn cry diagnostic systems. JASA.

The final piece of the current research is unveiled in the fifth chapter organized as an article that was submitted to the journal of the acoustical society of America (JASA) in Jun 2023. In this study two disparate frameworks were propounded that explore how the cries of newborns suffering from sepsis could be discriminated from other pathology groups in order to establish a system that solely focuses on sepsis as a leading mortality cause in newborns worldwide. Both experiments profit from novel features of BSC, ERBS Crest, GFCC, and BFCC that are all manipulated by the psychoacoustic warping of the spectrum to better represent the biological nature of the cry signals. The first framework employs DTF method for the first time in NCDS designs and the second framework extraordinarily curtails the feature space dimensions while preserving acceptable performance to that of the high dimensional feature vectors and maintains the essence of each feature set by the means of NCA feature selection scheme.

Lastly, chapter six is devoted to the contribution of our research and recommendations and ideas for future researchers who wish work on this topic.

CHAPTER 1

LITERATURE REVIEW

1.1 Psychoacoustic Model of the Cry in Newborns

Crying is the outcome of intricate interdependence and harmony between distinctive muscles and organs. Crying can be originated by external or internal factors such as adjusting the homeostasis via releasing the strains or excessive energy (Brazelton, 1962), a result of the CNS development (Emde, Gaensbauer, & Harmon, 1976), and the sensorimotor maturation (Ainsworth, 1963; Konner, 1972; Zeifman, 2001). One of the best models for the cry description is the physioacoustic model described by Golub et al. (Golub & Corwin, 1985) which describes the production of the cry signals and comprises two main elements: A physiological aspect which represents the way the respiratory, laryngeal and supralaryngeal compartments are arranged and are controlled; and an acoustical aspect that describes the sound generation process in the larynx and the airways above the larynx (Laitman, 1977). Moreover, this model ascribes the cry generation to four main subsystems:

1. Subglottal or Respiratory system: produces the required pressure in the subglottic area in order to propel the vocal folds.

2. The sound source positioned at the larynx that is divided into two types of sources functioning individually or concurrently:

    a) Turbulence noise source as a result of the air turbulence from the post-closure gap left in the vocal folds.

    b) Periodic source that is originated from the reverberations of the vocal folds.

3. The nasal and vocal tracts residing above larynx that act as an acoustic filter which is characterized by several physiological factors such as the shape and the length of vocal and nasal tracts as well as the nasal coupling degree.

4. The radiation effect attributed by the sound filtering along the distance of the microphone from newborn’s mouth.

After introducing the subsystems that affect the cry generation, it should be discussed how they are controlled by the newborn’s Central Nervous Systems (CNS). There are three levels of CNS processing, namely upper, middle, and lower processors. These processors each control and respond to a certain type of stimuli. During the neonacy, the upper processor is only capable of the unconscious control in response to the auditory, visual, and extrinsic or intrinsic proprioception like full bladder, hunger, or pain. The middle processor oversees the breathing, swallowing, bowel activities, coughing, and crying. Therefore, the cry is generated in newborns due to simpler causes than later infancy where the cries are no longer reflexlike but rather volitional. As a response to the stimuli, the upper and middle processors communicate with the lower processors regarding the control of the pertinent muscles and limbs. Having different levels of control each responsible for a separate group of muscles in cry generation, suggests that they are administered autonomously (Lester & Boukydis, 1985). Therefore, the acoustical anomalies could be assigned to certain physiological or anatomical peculiarities depending on the ability to discern the cries caused by subglottal, glottal, or supraglottal defect. Figure 1.1 shows a summary of the cry production details discussed so far.

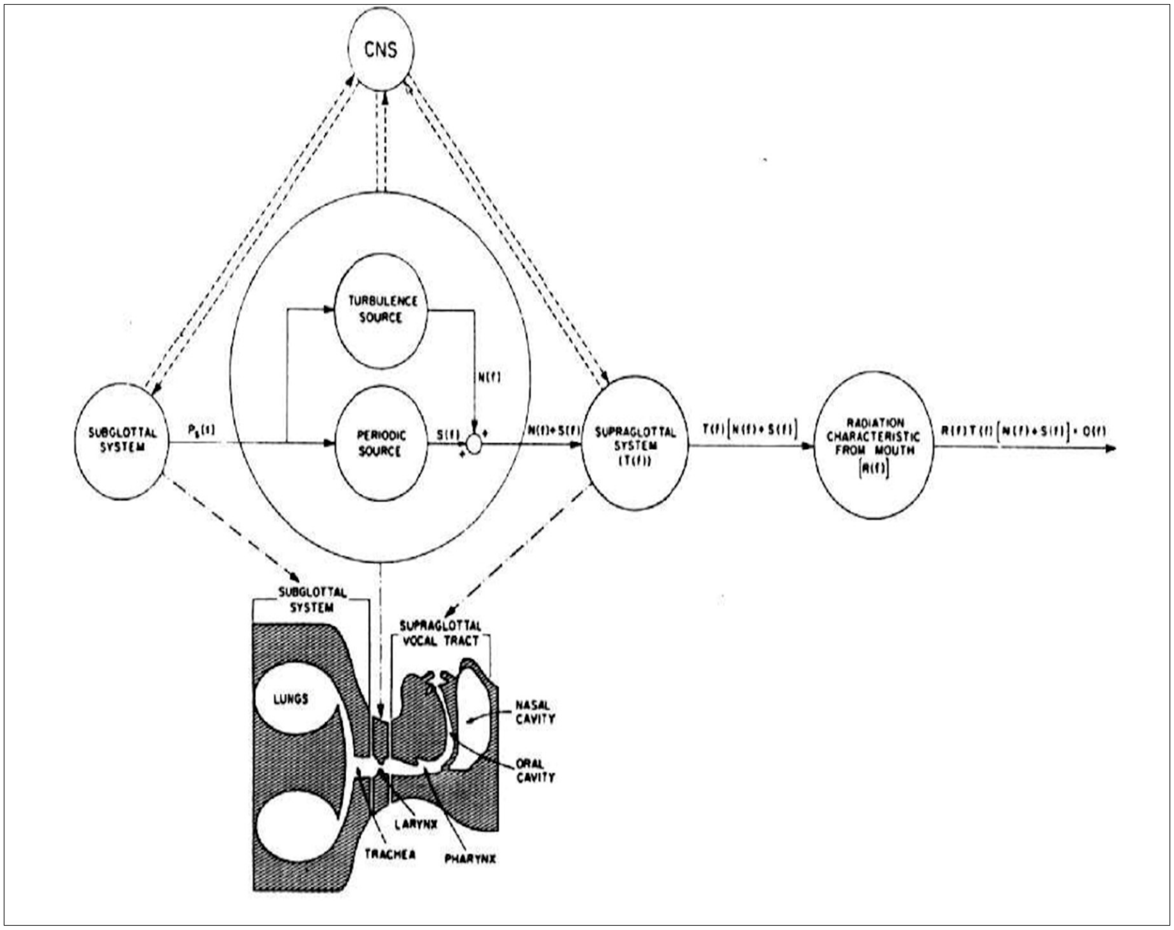

Figure 1.1 Cry production model
Taken from Lederman (2002)

1.2 Development of Vocal Tract and Acoustic Features in Children

There is no doubt that an infant's vocal tract is different from an adult's vocal tract (Nicollas et al., 2005). Hence, it is assumed that each fragment of the vocal tract may be subject to a change that is unlike the others, and that these changes may not occur simultaneously. Consequently, the developmental trajectory is not expected to be uniform (Prescott, 1975).

The fact that infants cannot breathe through their mouth even when their nose is blocked, reveals that their airways is unlike an adult's airways. The upper airways in a newborn is controlled by the neural system. The airway in newborns is sealed from the nose to the lungs. This is because of the positioning of the larynx, which is similar to other animals, located near the base of the skull. The larynx can be moved upward by the infant into the nasopharynx (Mugitani & Hiroya, 2012).

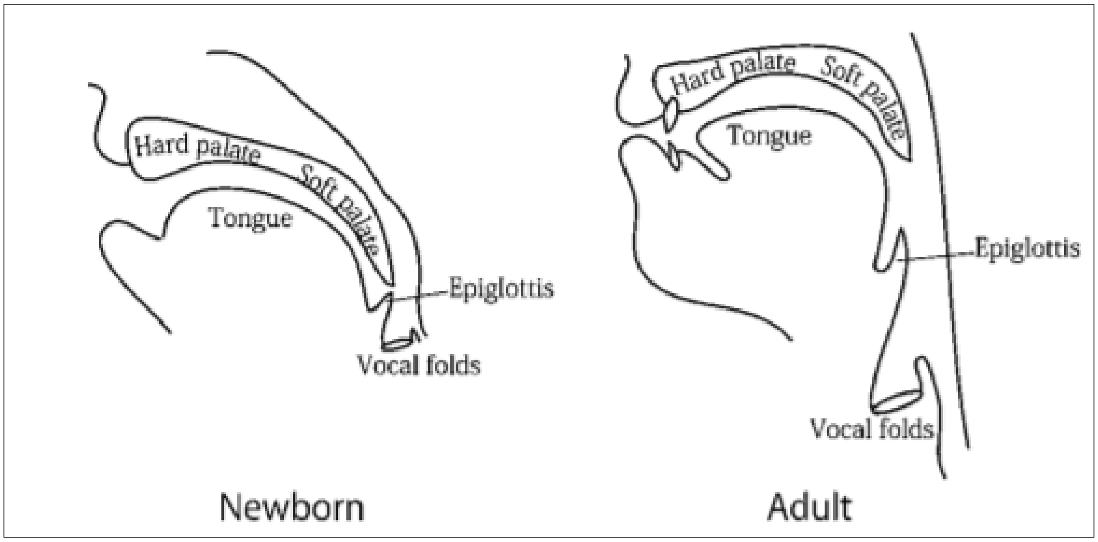

Figure 1.2 Vocal tract comparison between newborns and adults
Taken from Mugitani & Hiroya (2012)

As shown in the Figure 1.2 the epiglottis and the soft palate of an infant are placed adjacently and are capable of forming a double seal. The trajectory of the air passes through nose, then to the larynx and trachea and finally into the lungs, meanwhile the liquids flow from small larynx into the esophagus (Kent & Murray, 1982).

1.3 Acoustical Characteristics of Cry

The cry signal is segmented and analyzed based on the respiratory activities of the newborn which includes an expiratory phase, an inspiratory phase, and silence. The expiratory phase of the cry is often up to five times longer than the inspiratory phase of the cry; however, both segments have shown to convey significant information about the newborn's health and emotions. The inspiratory cries are subject to the maturity of the phonatory and respiratory systems and would eventually be replaced by the expiratory cries as the newborn matures, which further marks the importance of these cry episodes in the study of the neonate cry signals (Buder, Chorna, Oller, & Robinson, 2008).

The expiratory phase of the cry can be categorized into three groups: phonation or basic, Hyperphonation or shift and dysphonation or turbulence (Feng, 2020). These cry types are the outcomes of different vibrations of vocal folds and can be expressed as follows (Lester & Zeskind, 1982).

- *Phonation*: full vibration of the vocal folds with a fundamental frequency range of 250-700 Hz. This is a result of simple periodic oscillations and resonators acting as amplifiers.

- *Hyperphonation*: denotes the “falsetto” like vibration of the vocal folds in which only a few of the vocal ligament are occupied and occurs at an F0 of 1000-2000 Hz. Hyperphonation is the outcome of restricted performances of larynx and/or resonating cavities (Zeskind & Lester, 1981).

- *Dysphonation*: involves both periodic and aperiodic sound sources and occurs when turbulence noise is produced at larynx; this noise is modulated by vocal fold vibrations. Dysphonic cry may happen due to alterations of oscillations in glottis alongside supraglottic excitation (Lester & Boukydis, 1992).

On the other hand, it would be beneficial to expound the physiological aspect of each cry episode to better understand them.

- *Inspiration:* inspiration happens when the diaphragm and intercostals push the pleural space to expand. Therefore, since the product of volume and pressure is constant, the pressure will decrease. Generally, in an ACR problem, any sound made by the infant during an inhalation phase is labeled as inspiration and its fundamental frequency is usually between 367Hz and 1040Hz (Grau et al. 1997). This type of cry is rare among the healthy class of the infants (Wasz-Hockert, 1968). The inspiration is believed to contain information leading to pain and distress cries (Aucouturier et al., 2011).

- *Expiration:* the power needed for the driving the expiratory phase of a cry is stored during inspiration. Expiration can be described as a gradual decrease in the volume of the lungs. Usually cries occur during this respiratory phase, so this segment contains the main information.

- *Pause:* silent segments in a cry that could take place after either an expiratory or inspiratory phase (Abou-Abbas, Tadj, Gargour, & Montazeri, 2017).

The factors that affect the newborn cry are ample. These factors could be physiological, anatomical, environmental, and even psychological. In this section, the effect of each of these factors on the newborn cries will be addressed. However, it should be noted that in order for the proposed NCDS to be comprehensive, the following factors should not impose a restraint on choosing the participants; meaning that all the available cry samples should be considered for the development of the NCDS so that it is not limited to certain demographic attributes. Healthy newborns have a F0 (fundamental frequency) of 250 to 700 Hz, with a decreasing or increasing-decreasing melody shape with super imposed harmonics and average duration of 1-1.5s. As for premature infants, the more premature, the higher the F0 is. The cries of babies diagnosed with a pathological condition are persistent with little punctuation, reflecting high irritability and poor physiological stability (Corwin, Lester, & Golub, 1996).

The cry signal could be influenced by the health, ambient language, maturity, and prenatal substance exposure. The language spoken in the presence of an infant, influences its vocal development as soon as the 28th week of fetus' life. Besides, newborns can follow the salient F0 variation pattern of their mother tongue as well as ascending or descending F0 contours (Wermke et al., 2016). Therefore, the ambient language affects infant crying and the cry pattern could vary among neonates from different nations (Mampe et al., 2009). The cries of premature newborns were reported to have higher pitch and lower duration (Thodén, Järvenpää, &

Michelsson, 1985). Although one might suppose that this is a result of lower weight of the newborn, it was (Michelsson, Raes, Thodén, & Wasz-Hockert, 1982) observed that there is no meaningful correlation between the weight and the cry characteristics and the gestational age is the only determinant here to the extent that once the premature newborns reached the appropriate age for birth, a meaningful decline in the pitch would take place and the cries would lengthen. Moreover, it was reported that prenatal exposure to harmful substances like tobacco and alcohol would cause the cry signal to include dysphonation and hyperphonation as well as other attributes such as being shorter and faint (Blinick, Tavolga, & Antopol, 1971). Lastly, the cry patterns are majorly altered based on the health status of the newborn; pathologies each have a distinctive effect on the newborn cry signal as would be discussed in the following sections (LaGasse, Neal, & Lester, 2005).

1.3.1 Pathological Cry Attributes

As previously mentioned, there are some attributes in a cry signal associated with pathology that are different from a healthy cry and can rarely be observed in a healthy infant. Thus, abundant efforts were made to study and detect these features, both in medical and engineering fields. The medical studies date back to 19th century that address these pathologies through visual investigation of cry spectrograms and/or training of pediatricians. The acoustic characteristics of the cry may vary due to different factors such as air pressure, tension, length, thickness and shape of vocal cords and resonators (Golub, 1979). In this chapter, some of the expected features associated with each disease will be addressed shortly. Figure 1.3 presents the spectrograms of the cry signals for three different pathologies as opposed to a healthy cry.

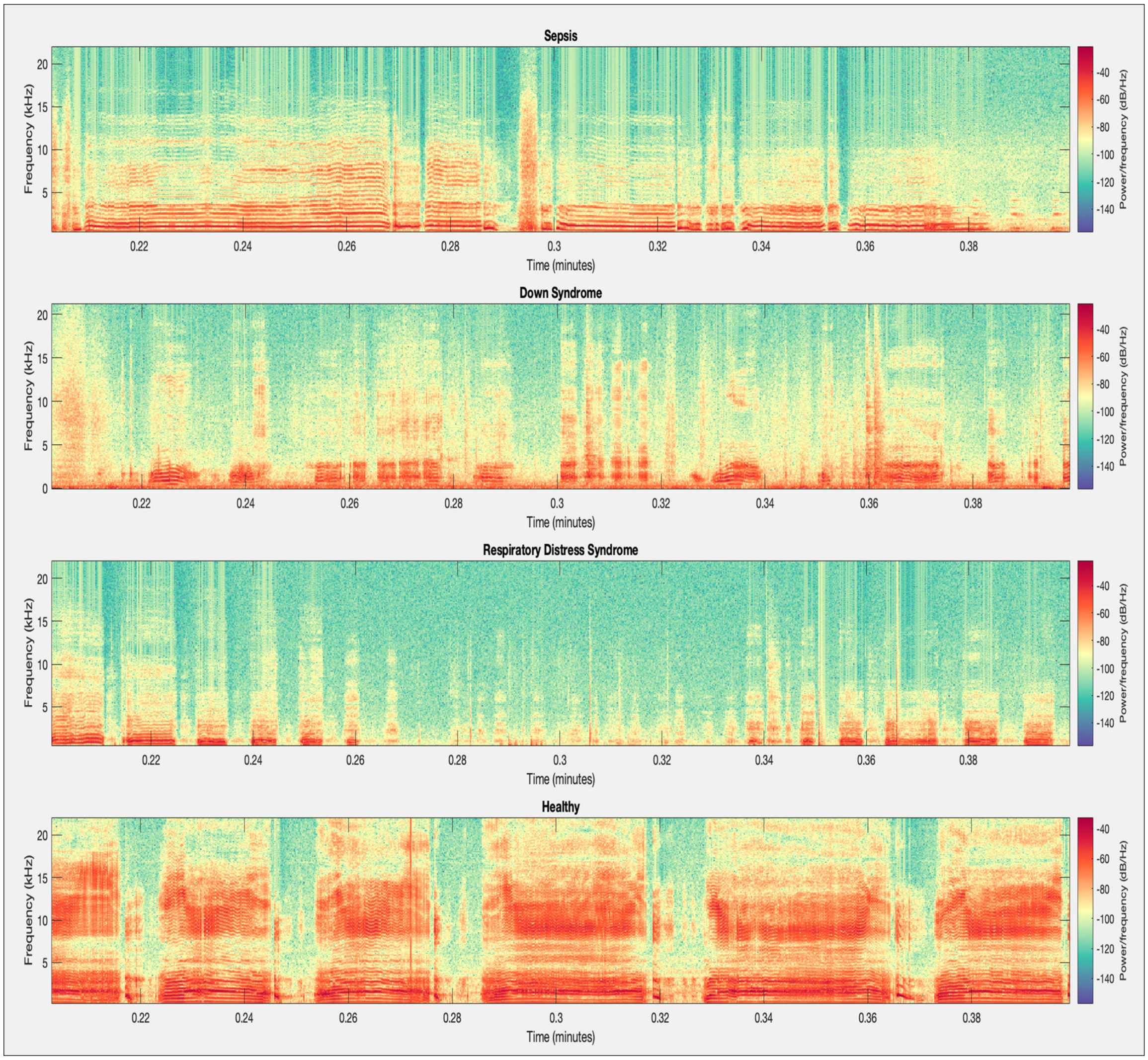

Figure 1.3 Spectrograms of newborn cry signals for pathology groups of sepsis, down syndrome, respiratory distress syndrome in contrast to the healthy

The spectrographic analysis of the cry signal has been delineated with certain characteristics that was observed in the spectrograms. These attributes help better explain the patterns of pathological and healthy cry signals. A complete list of these characteristics is given below:

- *Fundamental Frequency:* A single frequency that appears on the spectrograph of the signal as a horizontal line during the phonation. Frequency is shown as the height of this horizontal line. The lowest line in the series of parallel lines appeared on a spectrogram is the fundamental frequency.

- *Harmonics:* the consecutive higher lines or components of the fundamental frequency are called the harmonics.

- *Maximum/Minimum pitch:* The highest and the lowest voiced points marked in the fundamental frequency curve are maximum and minimum pitches of the cry signal, respectively.

- *Shift:* This term marks a rapid upward or downward occurrence in the fundamental frequency. Shift may be located at the initial point of the phonation phase. It may also be present in the middle or terminal points of the phonation course. The location of shift occurrence is significant in some of the diagnostic purposes.

- *Glide:* When the shift exceeds 500 Hz it is called a glide.

- *Continuity:* determines the cry in terms of being voiced. A cry signal can be entirely or partially voiced or completely unvoiced.

- *Glottal roll or vocal fry:* in glottal roll, the fundamental frequency and its harmonics are present but somewhat difficult to measure since their frequency range is low. Cries often end with a low intensity sound and a fall in pitch. The glottal roll sometimes is successive to a vibrato.

- *Biphonation:* An observation of other types of pitch or different kind of melody from fundamental frequency at the same instant is called biphonation. Here we can see presence of another melody type simultaneously as the fundamental frequency.

- *Vibrato:* A series of variations with the form of a wave that are observed in some pathologies, are called the vibrato.

- *Melody Type:* during the course of a phonation, the cry spectrograph has some gradual changes that are marked as the melody type. The melodies are falling, rising, falling-rising, rising-falling, flat, etc.

- *Glottal plosive:* a short turbulence of the air that is caused by the fast closing and opening of the vocal cords is called the glottal plosive. It has been also defined as the sudden pressure release at the vocal cords that results in an expiratory sound which is rather impulsive.

- *Phonation:* a phonation is generated whenever the vocal cords vibrate at the 350 to 700 Hz of the fundamental frequency and is considered as the basic oscillation mode for the glottis. Well defined harmonics with the same distance mark a phonation on a spectrograph.

- *Hyperphonation:* a hyperphonation is an abrupt modification of the basic cry. It happens when the fundamental frequency increases to the 1000 to 200 Hz range.

- *Dysphonation*: this term in fact denotes a disorder that affects several organs, namely: affecting the respiratory, cervical, costal, and abdominal muscles [12]. To determine the severity of a dysphonation, the amount of present turbulent noise is computed. It may also be characterized by the deviation of parameters like timbre, intensity, and pitch from the normal. This disorder is the result of malfunction in a fraction of expiratory phase of respiratory tract. Irregular and nonequidistant harmonics that happen due to inability of respiration control, suggest a dysphonic cry. A dysphonic cry also increases the variability of the fundamental frequency and hyperphonation. Therefore, observing any of these attributes suggests that respiratory system may be suffering from poor control. The percentage of dysphonation in a cry is more discriminative than simply observing its presence.

- *Double harmonic break:* this happens when consecutive series of double harmonics are like fundamental frequency in the melody type but have lower intensities.

- *Noise concentrations:* either for voiced or unvoiced cry signals, an audible energy peak is present at 2000-2300 Hz.

- *Furcation:* during a phonation phase the fundamental frequency curve may be split into weaker frequencies in the initial, middle or at the end point of the phonation in an unusual way. This is called furcation (Kheddache & Tadj, 2013a; Lester & Boukydis, 1985; Lewis, 2007).

Based on the aforementioned attributes, the cry signals corresponding to each pathology can be expounded and understood. As can be seen from Figure 3, the healthy infant cries follow an obvious pattern, and each expiratory cry episode has similar duration to the next and previous episodes and is followed by a short inspiratory episode. However, the cry signals for the down syndrome, sepsis, and RDS each demonstrate diverse characteristics. Different generation of researchers have reported distinctive cry attributes for each pathology that is worth mentioning here.

- **Prematurity and Intrauterine Growth Retardation**

In this case, the gestational age is more significant to the determination of cry attributes than the birth weight. Cries of a preterm infant is shorter and has a higher F0 –the fundamental frequency is higher in more preterm infants-; also, the glide is seen more frequently in this type of cry (Wasz-Hockert et al., 1963). The maximum pitch has the same behavior of fundamental frequency and biphonation was observed in a small fraction of preterm infants (Michelsson et al., 1982).

- **Prenatal Asphyxia**

Infants diagnosed with asphyxia have obvious abnormal cries. The cries marked as asphyxiated mostly exhibit higher F0 as well as higher minimum and maximum pitches. Moreover, biphonation and glide occurrence is more frequent.

Asphyxia could be divided into two main groups: Peripheral asphyxia also known as respiratory distress and central asphyxia which has neurological symptoms. The studies by Michelsson show that the cry in central asphyxia is more abnormal than peripheral asphyxia (Michelsson, Sirvio, & Wasz-Hockert, 1977).

Through spectrographic analysis, increased amounts of hyperphonation and dysphonation have been observed and reported (Kheddache & Tadj, 2013c).

- **Intracranial Infections/ Diseases of Central Nervous System**

The cry of a neonate diagnosed with bacterial meningitis is short and high-pitched. Additionally, this type of cry is associated with glides and biphonation as well as abnormal melody shape (Michelsson, SirviöM. A, & Wasz-Höckert, 1977).

- **Diseases of Peripheral Phonatory Organs**

The features that follow a cleft palate include nasality, vibrato, tonal pit, and irregularity in fundamental frequency. Glide occurred in 10% of cries (Massengill Jr, 1969). Biphonation has not been observed. However, some other studies report against this and no difference between normal and pathologic infants has been observed.

As for ankyloglossia (sometimes referred to as lingual frenulum), no cry characteristics were found in medical papers, the spectrographic analysis of cry signal mentions increased F0 irregularity, hyperphonation and dysphonation (Kupietzky & Botzer, 2005).

- **Chromosomal Disorders**

Chromosomal disorders may cause cries that are monotonous and low-pitched. These disorders include Down's syndrome, Edward's syndrome and x-chromosomal abnormalities. The cry latency in these neonates is longer as well as the duration of the cry. Lower F0 and flat melody type has been reported, along with nasality. No biphonation has been reported in chromosomal abnormalities (Lind et al., 1970).

- **Metabolic Disorders (Neonatal Hyperbilirubinemia)**

This pathology is also called jaundice. Both the minimum and maximum pitches of F0 soar highly. Biphonation has been observed in almost half of the neonates in Wasz-Hockert's observations. Furcation has also been found. Furcation is believed to be the most often among the infants diagnosed with hyperbilirubinemia. However, these two mainly happen in infants with neurological abnormalities. (Corwin & Golub, 1984) state that cry changes indicate unstable glottal functions which could be a sign of later neural toxicity caused by bilirubin.

- **Endocrine Disturbances**

Hypothyroidism could result in low-pitched cries, lower number of shifts and a frequent observance of glottal roll at the end of phonations. The cries marked with hypothyroidism have been marked as hoarse (Vuorenkoski, Vuorenkoski, & Anttolainen, 1973).

Despite the fact that the cry signals associated with a number of pathologies have been distinguished and characterized based on the spectrographic analysis, it would not be effective nor accurate to only rely on the inspection of cry spectrograms for larger number of patients or for making a prognostic decision. Therefore, the automated computer-based analysis systems were established as early as 1982 (Lounsbury & Bates, 1982). Even though NCDS' have seen many extensive improvements throughout the time, they are not as developed as the other audio recognition systems, especially compared to the speech recognition or music-related applications. Moreover, the distribution of the studied pathologies is not homogenous, which means some pathologies (e.g., asphyxia) received a lot of attention from researchers from all around the world while others -though prevalent and perilous- remain unexplored. There is an ineluctable need for a comprehensive NCDS that studies unexplored diseases and improves all the incorporated components. The following section conducts a review of the tools and methods that exist in the literature concerning NCDS designs.

1.4 Newborn Cry Diagnostic System (NCDS)

The NCDS' was developed in order to function as a non-intrusive tool to aid the diagnosis of different pathologies. All the NCDS share the same principal elements: 1. Preprocessing, 2. Feature extraction and selection, and 3. Classification. The preprocessing step encompasses all the required actions to filter, segment, clear, window, and emphasize the signal to prepare for the feature exploitation. The ultimate objective of a NCDS system is to discriminate between healthy and pathologic infant which cannot be achieved without the use of apt features. These features originate from different sources and modalities such as time domain, frequency domain, time-frequency domain, Cepstrum domain, and image domain. Following the extraction of the features, suitable algorithms for their combination, fusion, reduction, and selection is utilized to form an enhanced feature space. Finally, these features are fed to classifiers with diverse learning approaches such as SVM, KNN, LSTM, and others that are tuned and improved via HPO methods. The details regarding each component of the NCDS are determined based on the task it will serve. The succeeding sections review the methods and tools that researchers have implemented in the NCDS architectures so far. Figure 1.4 shows an outline of the NCDS and its components.

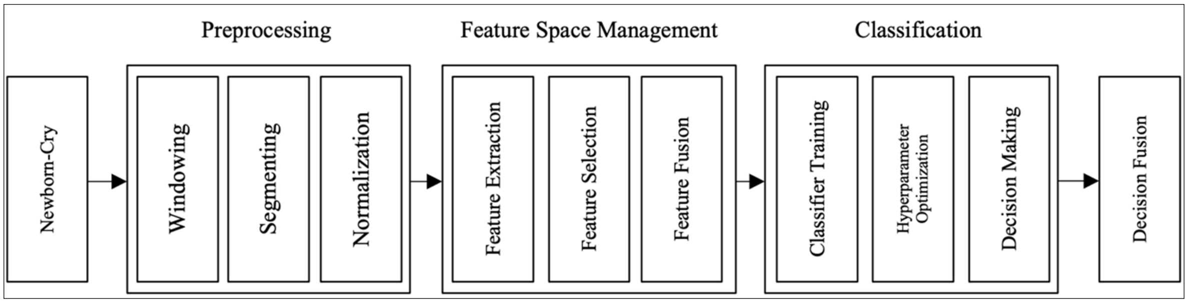

Figure 1.4 A general block diagram of the NCDS design

1.4.1 Feature Extraction

Extracting the indicative and apropos acoustic features that fittingly exhibit the characteristics of the cry signal is pivotal in developing the NCDS. In this section, we have tried to list the features that were deliberated by former researchers to the best of our ability. The features could be categorized based on different aspects: 1. Domain (frequency, time, etc.), 2. Qualitative or quantitative, 3. Short-term or long-term, and 4. Pathology association. In order to present the features with more details, we selected a combination of the domain-oriented and pathology association categorizations of the features. Table III-1 in appendix shows a list of different features studied in the history of the NCDS designs based on their categories and implementations. As seen in the table, the NCDS designs utilize a limited number of acoustic features, presenting significant potential for improvement. More specifically, spectral features that attribute the differences in the shape and the energy concentration of the cry signals, are absent in the literature. One interesting fact in the presented table was the recent attention that the image domain features such as spectrogram image (Zayed, Hasasneh, & Tadj, 2023 and Tusty; Basaruddin, & Krisnadhi, 2020) have received. This may be due to the fact that low number of samples and feature space dimensions had been imposing limits on the implementation of neural network-based classifiers and image domain features which are of acceptable dimensionality may overcome this challenge.

After discussing the features existing in the literature and their association with the pathologies, it would be fit to present the features introduced and employed for this study. Tantamount to the existing literature, this study employs features from different domains such as frequency, cepstral, prosodic, and time-frequency. The cry of the newborn is a dynamic and non-stationary signal; therefore, in order to obtain an apt representation of the signal, it is desirable to investigate features from different levels of information both individually and integrated. The MFCC, GFCC, and BFCC feature sets all conduct the short-term analysis of the cry signal; even so, their delta and delta-delta components overcome the challenge of stagnation to some extent by providing an epitome of the signal's temporal dynamics (Young et al., 2002).

On the other hand, spectral or long-term analysis of the cry signal not only provides profitable information about the signal's shape and statistical structure, but also, capacitates the comparison of the signal without the need to extract more features. For example, the interquartile range of the SC not only explains the shape and the mass of the spectrum, but also provides information about its dispersion without the need to further extend the feature space.

Among the short-term features, MFCCs are perhaps the most popular features in all audio recognition applications while being effective in many cases. MFCCs are forged from passing the windowed and Mel-mapped FFT of the cry signal through triangular filter banks followed by a Discrete Cosine Transform (DCT). Thenceforth, as a common practice (Amaro-Camargo & Reyes-García, 2007; Matikolaie & Tadj, 2020; Messaoud & Tadj, 2011) the average of the coefficients across the duration of the signal is taken to form a thirteen-element feature vector of MFCCs for each input signal. Prior to the averaging, the MFCCs embody information from only one frame of the signal; thus, in order to grasp an understanding of the fluctuations across consecutive windows of the signal, the first and second derivatives of the MFCC feature set - called delta and delta-delta, respectively- are also calculated and added to the feature vector after the computation of the average across all signal frames, yielding a thirty-nine-element feature space.

The use of psychoacoustics measures further intertwines the biological and source filter aspects of the cry signal presentation and thus, the use of GFCC and BFCC is also considered in this study. As it was prefaced in Table III-1, each of these feature sets have been formerly utilized for some tasks and proven successful; nevertheless, the propriety of pathology association as well as feature space assembling with the right combination of long-term and short-term features still has a high exploration capacity. GFCCs -as already stated- are a refinement of the MFCC feature set that were motivated from the biological study of the auditory system. It was shown that GFCCs have more robustness and efficiency (Matikolaie & Tadj, 2020) and their performance in non-speech applications has been highly desirable (Valero & Alias, 2012). The extraction of the GFCC and BFCC feature sets has a similar procedure to the MFCC feature set except that the frequency warping of the spectrum is accomplished in correspondence to the cochleagram or the Bark mappings.

Although the BFCCs were also reported to outperform MFCCs; it has also been observed that their combination with other Cepstral features such as GFCCs can form a powerful descriptor in audio recognition tasks (Liu, Li, Wu, & Zhou, 2019).

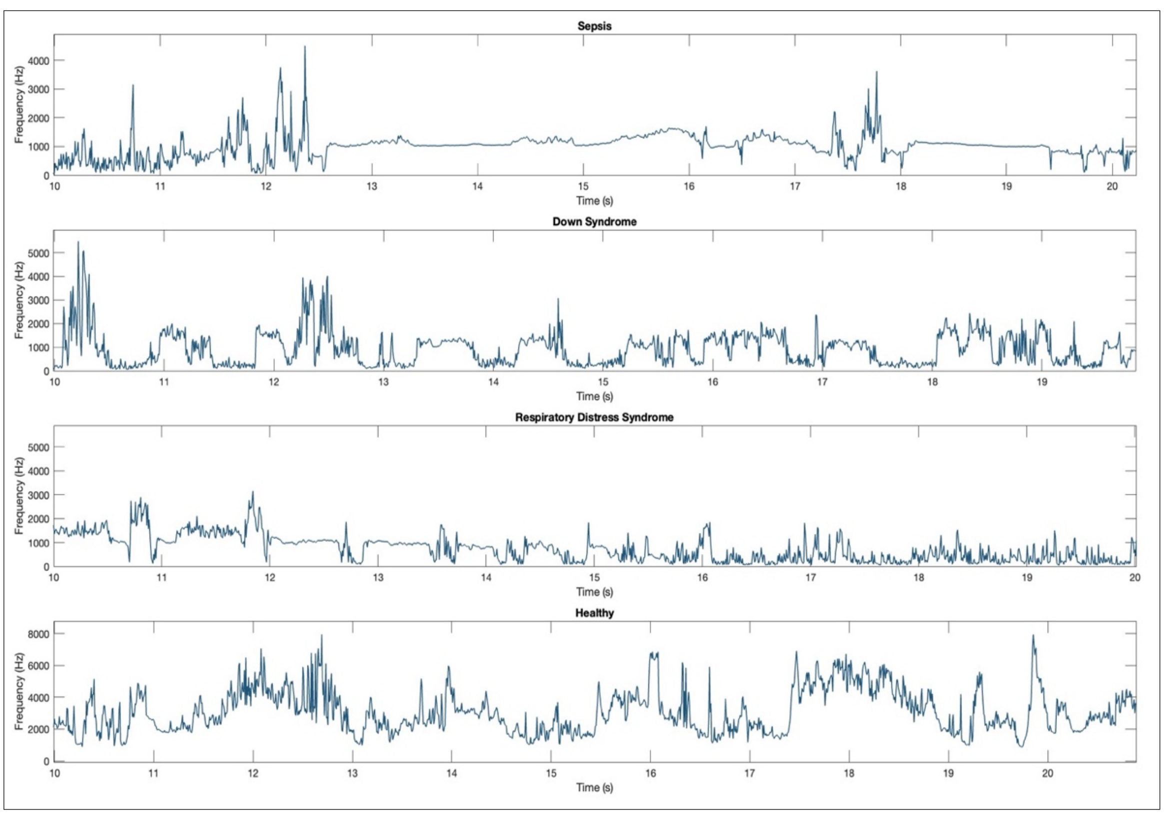

Figure 1.5 Spectral Centroids of newborn cry signals for pathology groups of sepsis, down syndrome, respiratory distress syndrome in contrast to the healthy

The long-term and spectral features in this study were also combined with the psychoacoustic warping of the spectrum as well as statistical measures and Cepstral analysis. The SC is mainstream to the musical timbre analysis to describe the "brightness" of a sound and has been used in diagnostic applications and in the detection of developmental disorders in newborns (Oren, Matzliach, Cohen, & Friedman, 2016).

Figure 1.5 shows how the SC feature varies across newborn cry signals according to different pathologies which depicts its potential for being used as a biomarker in identifying pathologies based on the cry signal. In this study, two different approaches were taken towards the analysis of the signal with the SC feature set, one is using the statistical measures of mean, median, H-speed, and standard deviation to form a four-element feature vector, and the other approach is to combine it with the cepstral analysis to extract 5 coefficients.

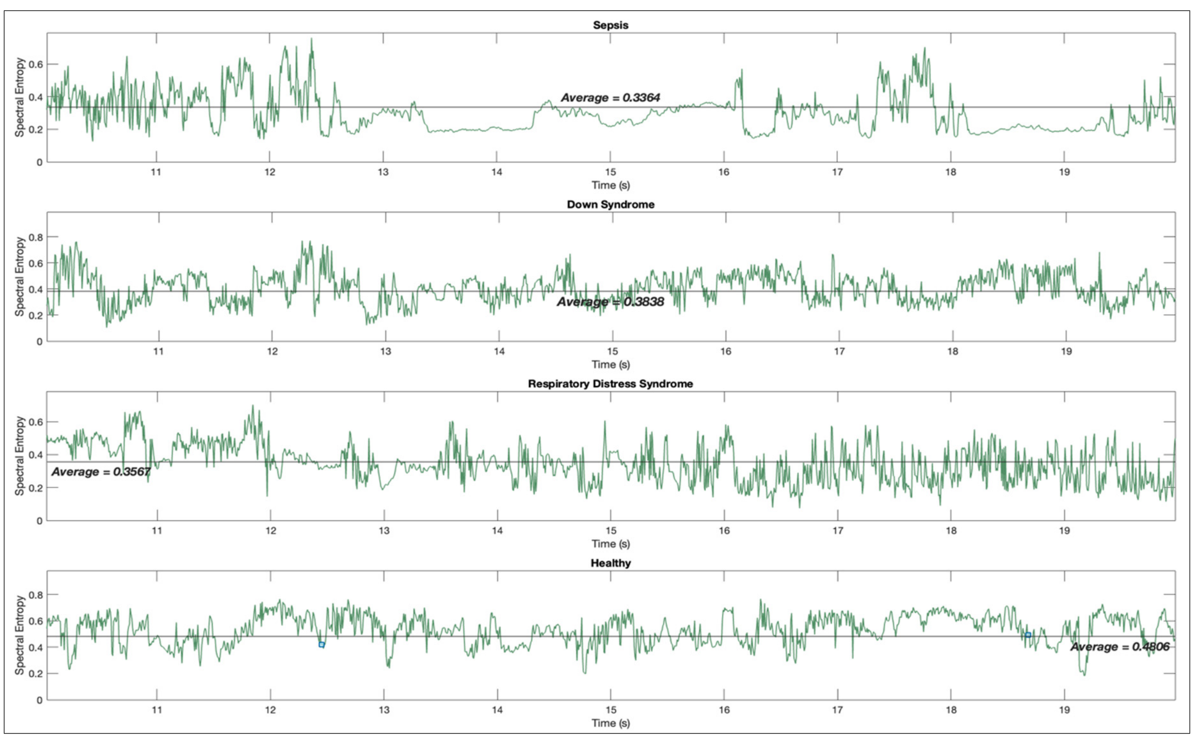

Figure 1.6 Spectral Entropies of newborn cry signals for pathology groups of sepsis, down syndrome, respiratory distress syndrome in contrast to the healthy. The vertical line denotes the average of spectral entropy across the length of each signal

The study of entropy is thought-provoking for the NCDS designs; entropy is closely associated with the information content of the signal and its complexity. It was shown that the entropy of the pathologic cry signals is generally lower than the healthy newborns cry signals, which translates to nonuniform distribution of the energy across the spectrum of the pathologic cries. This fact has been shown in Figure 1.6 where the SEN of the signal is displayed alongside its mean value across the spectrum for different pathologies in contrast to the healthy newborns. Furthermore, the average entropies of all the pathologies have a meaningful distance from the average entropy of the healthy newborns.

The HR feature set is the best representative to show the musical aspect of the cry signal. As can be deduced from Figure 1.7, the HR patterns vary significantly across different pathologies and as opposed to the healthy cry. For example, taking a brief look at Figure 1.7 shows that the HR in healthy newborns does not cross the horizontal axis which means that the healthy cries contain a consistent harmonic pattern, whereas the cry of an infant with down syndrome has a lot of zeros which translates to the absence of harmonicity in many frames of the cry signal. Furthermore, the cry of a septic newborn has near-flat patterns across long frames of time that means it would be monotonous for a majority of time. These findings align with the descriptions of the medical experts that précised the cry attributes (Weiss, Pomerantz, Torrey, & Kaplan, 2019). Finally, the crest feature set is convenient and effective to the study of audio fingerprinting (Ramalingam & Krishnan, 2005) and has been employed in medical tasks such as epileptic seizure diagnosis (Dash & Kolekar, 2020) which marks it as an interesting feature set to study the cry signals. The HR and ERBS Crest feature sets follow the same statistical approach, and each contain four elements.

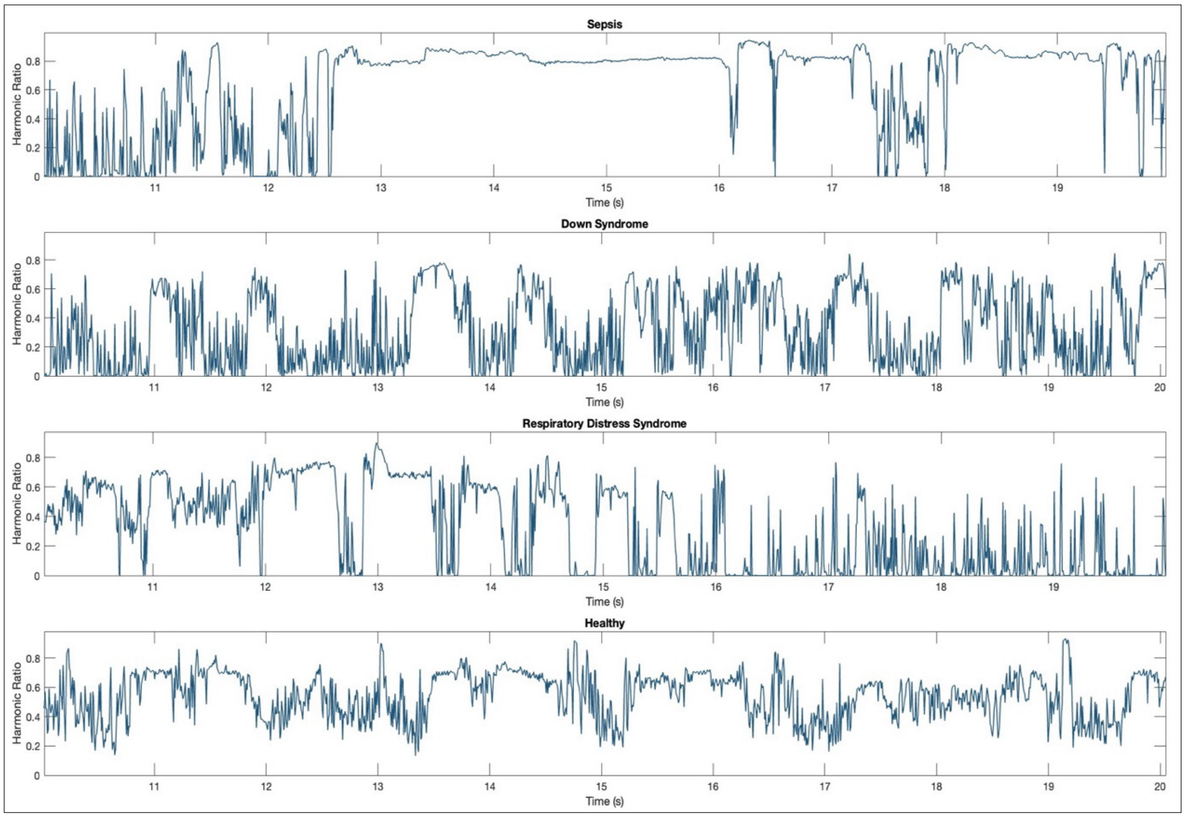

Figure 1.7 Harmonic Ratio of newborn cry signals for pathology groups of sepsis, down syndrome, respiratory distress syndrome in contrast to the healthy

So far, we discussed the features, their use, the methodology to extract them, and illustrated how they would enable the distinction between different pathology groups. There is one final point to the feature manipulation step which is feature selection and/or fusion. Feature selection comprises forming a minimal subset of the feature space that best represents the original signal and achieves the best classification performance with the features that convey the highest information. Besides, feature selection algorithms contribute to the removal of the redundant and noise-influenced features. As an example, The Fisher’s ratio is a widely used method for the feature selection. It selects the features by the computation of variance. Among other approaches, Orthogonal Least Square (OLS) and Binary Particle Swarm Optimization has been employed by (Zabidi, Mansor, Lee, Yassin, & Sahak, 2011) for the infant cries. The cross-validation is used for Forward Feature Selection Method (FSM) (Okada, Fukuta, & Nagashima, 2011). This study explores two novel feature selection methods of NCA and FE Selection. Both of these methods were chosen due to their simplicity and the fact that they would not put on further computational cost on the proposed NCDS design. The FE selection method is based on the fact that the entropy has an inverse relationship to the information content and that fuzziness would describe the degree of the membership of each element in the feature space (Khushaba, Al-Jumaily, & Al-Ani, 2007). The NCA method, however, considers the nearest neighbor classification of the feature space while trying to adjust the weights for the classification of these features with respect to a selected reference point and then reports these weights to show the contribution of each element in the feature space (Yang, Wang, & Zuo, 2012).

The features used in the design of the NCDS are extracted from multiple natures and origins, e.g., short-term and long-term features. However, adding these features to a single feature set without the proper implementation of fusion techniques can even result in performance reduction. Feature fusion can be considered in four different stages of the NCDS, that consist of: 1. Data/sensor level, 2. Feature level, 3. Matching score level and, 4. Decision level (Telgad, Deshmukh, & Siddiqui, 2014). In this study, two novel algorithms for feature and decision levels of fusion were introduced and implemented.

The basis of feature fusion consists of merging the features extracted from different sources into a single feature set by the application of germane feature normalization, conversion, and pruning strategies. As it was stated, the foremost profit of feature-level fusion is to omit redundant features which can be carried out by investigating the correlations between features and in return, being able to introduce a succinct subset including the predominant features that can enhance classification accuracy (Kim, Hyun, Chung, & Kwak, 2019). In the current study, the CCA-fusion method was implemented for the purpose of fusion at feature level where features are projected to find the maximum correlation of canonical variates by the utilization of Lagrange multipliers, and, to remove redundant information.

1.4.2 Classification

The conceptual boundary between feature extraction and classification is somewhat arbitrary: an ideal feature extractor would yield a representation that makes the job of the classifier trivial; conversely, an omnipotent classifier would not need the help of a sophisticated feature extractor. The distinction is forced upon us for practical, rather than theoretical reasons. Generally speaking, the task of feature extraction is much more problem and domain dependent than is the classification, and thus requires knowledge of the domain to build apt boundaries for the discrimination of data (Duda, Hart, & Stork, 2001). Many different families of classifiers have been studied for the NCDS architectures so far and we tried to list them in Table-A II-1. It can be deduced that a fair variety of classification approaches from different families have been used with the NCDS designs. However, there are some that are less frequently employed such as LSTM approach. Moreover, there is not enough information about tuning the classifiers based on the application at hand for the NCDS designs. Among these methods, DFFNN gained the highest distinctive performance (Lahmiri, Tadj, Gargour, et al., 2022). Other deep learning and machine learning approaches such as CNN (Lahmiri, Tadj, Gargour, & Bekiros, 2023), and SLGAN (Zhang, Ting, & Choo) also had an state-of-the-art performance. However, there were conventional classification approaches such a random forest (Pusuluri, Kachhi, & Patil, 2022) that could gain a similar performance when fine-tuned, which further highlights the role of adjusting the classifier being more significant than the classifier type.

The SVM classifier is the most convenient classifier often used as a baseline in the NCDS with a variety of kernels such as linear, cubic, and Gaussian. Therefore, we followed the same pattern as other researchers and the SVM classifier is used in all the parts of this study. Moreover, KNN, LSTM, and MLP classifiers were also employed across different constituents of the presented research. Interestingly, despite the profitable contributions of the HPO methods in exalting the performance of a majority of classifiers, the discussion and the implementation of HPO in NCDS-related applications is minute and thus, as a supplementary endeavor, this research gap was addressed so that all of the mentioned classifiers in this study benefit from one or more HPO methods. Furthermore, three HPO methods of random search, grid search, and Bayesian were compared in detail.

Ultimately, the paramount objective for a NCDS design is to attain preeminent results in identifying the pathologic newborns. In this regard, a diversity of classifiers and HPs were introduced and implemented. However, with the increment in data, feature space dimensionality, classes, and the complexities associated with drawing class boundaries, a solitary classifier would struggle to achieve the desirable performance. Therefore, a variety of methods were developed for combining the classifiers and fusing their decisions that were made based on training divergent feature sets. Not only the DF methods help enhance the system performance, but also, the final decision would be unbiased (Mangai, Samanta, Das, & Chowdhury, 2010). In summary, the employment of a fusion framework has several principal reasons:

1. Lower or distributed computational load in comparison to a single classifier and having a faster system as a result.

2. Solitary classifiers would not function proportionately if the feature space consisted of features from disparate modalities.

3. The generalization ability of the system is improved with the use of DF techniques, especially when the data is limited.

4. In some cases, the DF technique prevents overfitting.

In this study, the DTF method was selected for having three outstanding properties (Kuncheva, Bezdek, & Duin, 2001) :

1. the DTF method is among the few existing class-indifferent methods of fusion.

2. DTF is one of the simplest DF methods.

3. Does not form any presumptions (for example, about mutual independence of the features used for training each of the classifiers) in order to combine the classifiers.

DTF combines the decisions of different classifiers through conducting a comparison between the outputs with a characteristic template of each of the classes. It should be noted that unlike other DF methods that would solely consider the support of a given class to form a decision, DTF employs all the outputs from all the classifiers to form a final support for each of the classes. The DTF method was employed for a wide range of applications with the exclusion of NCDS.

1.5 Summary

This chapter expounds the motivation behind the development of a comprehensive NCDS as well as the improvements in each of its components for this research. The presented information in previous sections further highlighted the discriminative potential of the newborn cry signals as a powerful biomarker for diagnostic purposes. Moreover, the background on the designs, tasks, pathology analysis, architectures, and methods employed with existing NCDS literature was explicated to clarify the research gap. Finally, we discussed how each of the methods selected for different stages of the NCDS design in this study would benefit the system and the reason behind their selection.

In summary, our design would focus on achieving several objectives listed below. Through realization of these points, we aim to propose a comprehensive NCDS that enhances different aspects of each of the NCDS components:

1. A comprehensive design that can detect septic newborns from the RDS-diagnosed newborns, among other pathology groups, and amid the healthy.

2. A simplistic yet effective tool that would identify pathologic newborns in a generalized and comprehensive manner, regardless of the pathology group, and serve as an early alert for the caregivers of the newborn.

3. Introducing fusion at different levels to the NCDS design, more specifically, the CCA-fusion at feature level and DTF at decision level and evaluating their impact on the NCDS performance.

4. Assessing the potential of inspiratory cries compared to the expiratory cries in the identification of pathologies.

5. A detailed evaluation of how HPO affects the performance of the NCDS in terms of run-times, different metrics, and methods.

6. Outline the patterns associated with pathologic cries, especially septic cries, as opposed to the healthy newborn cries, or other pathologies.

CHAPTER 2

AN ENTROPY-BASED ARTICHECTURE FOR DETECTION OF SEPSIS IN NEWBORN CRY DIAGNOSTIC SYSTEMS

Zahra Khalilzad, Yasmina Kheddache and Chakib Tadj,

Department of Electrical Engineering, École de Technologie Supérieur, Université du Québec, Montréal, QC H3C 1K3, Canada

Paper published in *Journal of Entropy*, August 2022

2.1 Abstract

The acoustic characteristics of cries are an exhibition of an infant's health condition and these characteristics have been acknowledged as indicators for various pathologies. This study focused on the detection of infants suffering from sepsis by developing a simplified design using acoustic features and conventional classifiers. The features for the proposed framework were Mel-frequency Cepstral Coefficients (MFCC), Spectral Entropy Cepstral Coefficients (SENCC) and Spectral Centroid Cepstral Coefficients (SCCC), which were classified through K-nearest Neighborhood (KNN) and Support Vector Machine (SVM) classification methods. The performance of the different combinations of the feature sets was also evaluated based on several measures such as accuracy, F1-score and Matthews Correlation Coefficient (MCC). Bayesian Hyperparameter Optimization (BHPO) was employed to tailor the classifiers uniquely to fit each experiment. The proposed methodology was tested on two datasets of expiratory cries (EXP) and voiced inspiratory cries (INSV). The highest accuracy and F-score were 89.99% and 89.70%, respectively. This framework also implemented a novel feature selection method based on Fuzzy Entropy (FE) as a final experiment. By employing FE, the number of features was reduced by more than 40%, whereas the evaluation measures were not hindered for the EXP dataset and were even enhanced for the INSV dataset. Therefore, it was

deduced through these experiments that an entropy-based framework is successful for identifying sepsis in neonates and has the advantage of achieving high performance with conventional machine learning (ML) approaches, which makes it a reliable means for the early diagnosis of sepsis in deprived areas of the world.

Keywords: newborn cry diagnostic system; Spectral Entropy; sepsis; fuzzy entropy; Bayesian Hyperparameter Optimization

2.2 Introduction

Studies conducted by the United Nations Children's Fund (UNICEF) report that 7000 newborns die every day from mostly treatable causes, which amounts to 2.6 million neonates per year. Although neonates constitute the most vulnerable group, they are also the most difficult to interact with; in-depth examinations and medications are intricate and seldom prescribed. The main challenge in working with neonates is that their only means of communication is crying. According to UNICEF reports, newborn mortality is mainly attributable to infectious pathologies such as sepsis and meningitis. These two pathologic conditions together comprise a 15% share of all neonate death causes, especially in middle and lower-income countries (Unicef, 2020a).

Crying is the result of cooperation between numerous organs in the body, such as the respiratory system, central and peripheral nervous system, and a variety of muscles and limbs. If any organs fail to function properly, a cry different from a healthy one is expected (Fort & Manfredi, 1998). As early as the 20th century, it was observed that the cry of neonates diagnosed with certain pathologies was different from healthy neonates (Michelsson, SirviÖ, et al., 1977). This led to further investigation of cries and the use of sound spectrographic analysis. The results claimed that the cry signal conveys a significant amount of information about a newborn's health. The researchers developed a more accurate system since the spectrographs could not capture all the abnormalities and disorders in a cry signal; therefore, the automatic newborn cry diagnostic systems (NCDSs) were designed and proposed (Abou-

Abbas, Tadj, & Fersaie, 2017; Farsaie Alaie & Tadj, 2012; Kheddache & Tadj, 2013a, 2013c; Matikolaie & Tadj, 2020; Messaoud & Tadj, 2011).

NCDS architectures are designed to serve different purposes. These purposes include detecting the reason for crying in healthy infants (Bano & RaviKumar, 2015; Parga et al., 2020), such as pain, hunger, etc., segmenting the crying episodes into expiration and inspiration (Abou-Abbas et al., 2015), detection of the cry from the surrounding environment (Torres et al., 2017) and diagnosis of pathologies (Orlandi, Manfredi, Bocchi, & Scattoni, 2012; Satar, Cengizler, Hamitoglu, & Ozdemir, 2022; Zabidi, Mansor, Khuan, Yassin, & Sahak, 2009). The design proposed in this study focuses on the last category of NCDSs where the goal is to discriminate between healthy and septic infants (Kheddache & Tadj, 2019). Similar to other audio analysis systems, the NCDS consists of three main stages: pre-processing, feature extraction and classification, as seen in Figure 2.1.

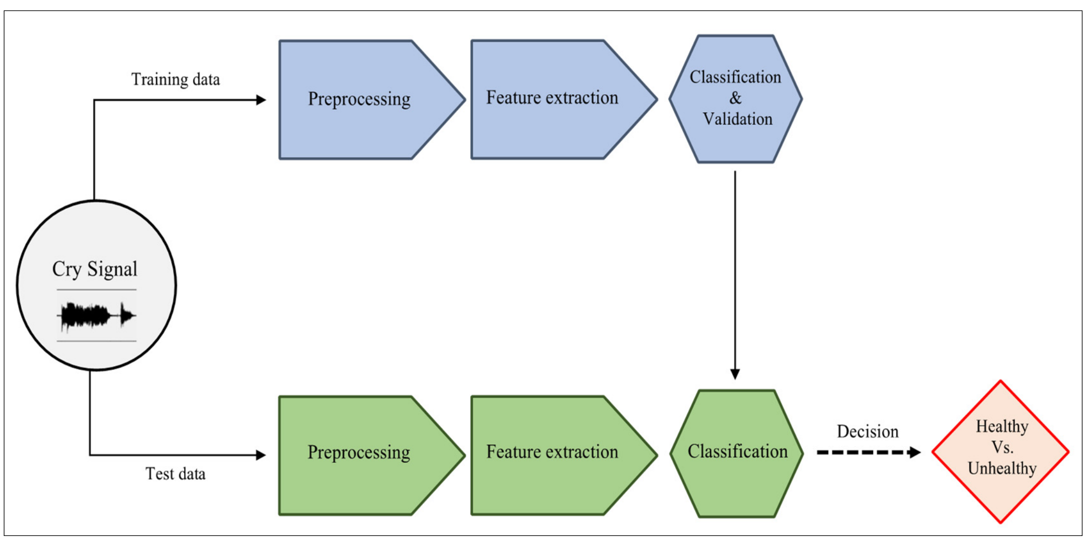

Figure 2.1 The block diagram of the NCDS

Mel-frequency Cepstral Coefficients (MFCC) are one of the most common features in the analysis of audio signals. They have been employed in the detection of many health conditions, such as cleft palate (Massengill Jr, 1969), asphyxia (Wahid, Saad, & Hariharan, 2016; Zabidi, Mansor, & Lee, 2017), respiratory distress syndrome (Matikolaie & Tadj, 2020) and hearing impairment (Jam & Sadjedi, 2009), and have demonstrated efficient performance. Other feature sets, including fundamental and resonant frequencies (Kheddache & Tadj, 2015), Linear Prediction Coding (LPC) (Liu et al., 2019) and prosodic features (Matikolaie, Kheddache, & Tadj, 2022), have been explored in the feature extraction step of other NCDS designs. Various entropy feature sets were utilized in order to identify deaf neonates from the healthy group (Jam & Sadjedi, 2009), for detection of asphyxia in newborns (Hariharan, Saraswathy, Sindhu, Khairunizam, & Yaacob, 2012) and for automated detection of the cry (Vaishnavi & Dhanaselvam, 2019). It has been reported that approximate entropy has different levels across healthy and pathologic newborns (Lahmiri, Tadj, Gargour, & Bekiros, 2021). We extracted Spectral Entropy Cepstral Coefficients (SENCC) and Spectral Centroid Cepstral Coefficients (SCCC) and combined them. The combination of these features provides more analysis for the study of septic cry signals. Finally, the feature sets are fed to a classifier and the predicted class labels are the output of the NCDS.

Spectral Centroid (SC) has been studied in order to find the reason for crying (Chang, Chang, Kathiravan, Lin, & Chen, 2017; Osmani, Hamidi, & Chibani, 2017) and to detect infants with developmental disorders (Oren et al., 2016). This feature has shown promising results in musical applications for studying timbre (Lakatos, 2000) and medical studies such as detecting Alzheimer's disease based on Electroencephalogram (EEG) signals (Kulkarni & Bairagi, 2017). To the best of our knowledge, cepstral analysis of this feature set has not been explored in NCDS designs so far. For a long time, crying has been treated similarly to the speech signal, and the features that showed potential in speech recognition tasks have been employed in cry research. This study aims to introduce the features that have been prevalent in the study of music to cry-based applications since the cry signal has harmonic components and rhythm (Kheddache & Tadj, 2015; Matikolaie et al., 2022). In the next step of NCDSs, many different classification approaches have been explored. Support Vector Machine (SVM) (Alaie, Abou-Abbas, & Tadj, 2016; Chang, Hsiao, & Chen, 2015), Probabilistic Neural Network (PNN) (Matikolaie et al., 2022), Forest (Rosales-Pérez et al., 2015), Decision Trees (Osmani et al., 2017), K-nearest Neighborhood (KNN) (Fuhr, Reetz, & Wegener, 2015), and discriminant analysis are some of the algorithms implemented in this field (Matikolaie & Tadj, 2022).

Hyperparameter Optimization (HPO) was introduced in the 1990s (King, Feng, & Sutherland, 1995; Michie, Spiegelhalter, & Taylor, 1994) when several studies reported that adjusting various hyperparameters led to better results across different datasets (Kohavi & John, 1995). HPO is employed to enhance the performance of the default settings provided by conventional machine learning (ML) architectures (Mantovani, Horváth, Cerri, Vanschoren, & de Carvalho, 2016; Olson, Cava, Mustahsan, Varik, & Moore, 2018). Moreover, Fuzzy Entropy (FE) has been studied previously for many applications in the biomedical field, such as medical database classification (Jaganathan & Kuppuchamy, 2013), and also tested on the Parkinson's database for feature selection purposes, which was able to achieve an accuracy of 98.28% (Luukka, 2011).

The contribution in this study has several aspects: first, the identification of septic newborns using their cry signals is of great significance, which has considerable potential and has been rarely looked at so far. To the best of our knowledge, even though sepsis is taking the lives of many newborns every day, there is only one other very recent study dedicated to this pathology. The second contribution is our approach in the design of an NCDS with different feature sets, their combination, and unique HPO for each feature set and classifiers, in order to identify septic newborns. Lastly, we employed a feature selection method based on Fuzzy Entropy (FE Selection) in order to select the features with the highest information content and to reduce the feature space dimensionality (Lee, Chen, Chen, & Jou, 2001; Lohrmann, Luukka, Jablonska-Sabuka, & Kauranne, 2018); to the best of the authors' knowledge, this method has not been explored in research associated with NCDS so far. There are many other entropy-based features and methods present in the literature. FE selection was chosen for this study due to its simplicity and the fact that it does not burden the system with complex computational costs (Khushaba et al., 2007). Moreover, Lee et al. (Lee et al., 2001) stated that their FE-based feature selection method enhanced the classification rate by discarding the features that were detrimental and affected by noise. The term sepsis refers to an infection that enters the bloodstream. Medical studies suggest that major infections, including sepsis, are associated with tenacious crying, and therefore, for a neonate with persistent crying, the predominant manifestation of sepsis should be seriously considered (Ruiz-Contreras, Urquía, & Bastero, 1999). Expedient diagnosis is of utmost importance for this pathology and medical staff should be alert to the risk factors of sepsis in neonates (Singh & Gray, 2018). It should be mentioned that there are other effective approaches to the study of sepsis in newborns, which range from studying heart rate monitoring to biosensing and electrochemical detection (Balayan, Chauhan, Chandra, Kuchhal, & Jain, 2020; Moorman et al., 2011). However, we proposed this study as an early and simple alert for diagnosing sepsis without the need for any clinical equipment, or even contact with the newborn, which would be complementary in adding information regarding sepsis. The areas that suffer the most from septic mortality have a lack of pediatricians and are categorized among low-income countries. Thus, a method that is simple and has efficient performance is preferred to one benefiting from complicated architecture and high computational requirements.

This article aims to provide an automated approach for identifying septic neonates through the development of a Newborn Cry Diagnostic System (NCDS). Furthermore, our goal is to assess the performance of the existing methods in the fields of ML and speech analysis in order to provide a simple tool for early diagnosis of sepsis in infants. It is noteworthy that there are a very limited number of studies dedicated to the automatic identification of septic newborns so far, and we will address them in the following sections. Therefore, there is a lacuna in the studies regarding the automatic analysis of sepsis in neonates. The methodology section explains the data acquisition process, participants and NCDS stages with a detailed description of the features and classifiers. Next, we expound the NCDS evaluation methods and the results in terms of the evaluation metrics are presented. We will then discuss the achieved results and compare them to the work of other researchers. The final section is dedicated to the conclusion.

2.3 Methodology

2.3.1 Cry Dataset and Recording Procedure

The database used for this study was created in collaboration and cooperation with Al-Raee and Al-Sahel hospitals in Lebanon and Saint Justine Hospital in Montreal, Canada. Most of the infants chosen for this study were neonates by the definition of UNICEF, which means they were less than four weeks old. The large number of cases and the diversity of race and pathologies make this database exceptional from all the other databases. The signals were recorded in the hospital environment; they were recorded in different conditions and times, such as after birth, when infants were placed in intensive care units, in the maternity room (either public or private), etc.

The crying reasons were not the same for all the infants; for example, cries may be due to wet diapers, hunger, fear, etc. These reasons were determined according to the conditions causing the cry with the help of medical staff and the infant's guardians. They were also based on the various tests performed after birth (Abou-Abbas, Tadj, Gargour, et al., 2017). The dataset acquisition and the selection of the neonates that participated in this study were not limited to a specific cry stimulus, making our study a comprehensive one.

The recorder utilized for this database was an Olympus hand-held digital two-channel device. It had a sampling frequency of 44.1 kHz and 16 bit resolution. The recorder was placed 10 to 30 cm from the newborn's mouth. There was no well-defined procedure during the acquisition of the cry sounds. Therefore, during the data collection process, unwanted information and noises, such as staff chatter, medical instrument beeps, the cry of the other newborns, and other environmental noises and sounds, were also recorded. Hence, we consider our database a real corpus recorded in an actual clinical environment. Table 2.1 is a description of the cry database used in this study.

Table 2.1 Description of the cry database

	Septic	Healthy
Gender	11 Males and 6 Females	55 Females and 53 Males
Weight	3.03 ± 0.40 kg	3.50 ± 0.55 kg
APGAR Score	8 to 10, measured 2–3 times	9–10, measured 2–3 times
Babies' Ages	1 to 53 days old	
Prematurity	Full term	
Gestational Age	38 ± 1 week	
Origin	Canada, Haiti, Portugal, Syria, Lebanon, Algeria, Palestine, Bangladesh, Turkey	
Race	Caucasian, Arabic, Asian, Latino, African, Native Hawaiian, Quebec	
Reason for Crying	Birth cry, hunger, dirty diaper, discomfort, needs to sleep, cold, pain	

The pathology group selected for this study was sepsis. Our database includes 108 full-term healthy neonates and 17 neonates that were marked as having sepsis by the medical staff through in-depth examinations. There are 53 cry signals recorded from the septic neonates in total, which means each newborn has more than one recording in the database. In order to obtain a balanced study, the same number of samples were chosen from the full-term healthy neonates' group. The healthy samples were selected completely randomly and without any pre-specified conditions in order to maintain the proposed NCDS free of any bias towards race, reason for crying and origin. In order to have a balanced study, we randomly selected an equal number of samples from both groups. As shown in Table 2.2, the control group consisted of randomly chosen samples from the whole healthy dataset of 108 healthy newborns to match the number of samples from the septic group. We wanted our NCDS to include newborns from all races, genders and any cry stimuli. The only remaining difference in the two datasets is the number of males and females. However, it has been shown that the length of vocal cords is the factor that determines the fundamental frequency of newborn cries as well as other characteristics, and this is similar across male and female neonates and does not have any meaningful impact on the cry (Reby, Levréro, Gustafsson, & Mathevon, 2016). The average lengths of expiratory and inspiratory cries were 0.72 and 0.21 s, respectively. We set a condition to only select the samples with a length of more than two consecutive windows (17 ms = two 10 ms windows with 30% overlap) in order to achieve a reliable analysis of the dataset.

Table 2.2 Specifications of EXP and INSV datasets for healthy and pathologic cry signals

	No. of Healthy	No. of Septic	No. of Train Samples	No. of Test Samples	Available Time (s)
EXP	1132	1132	1585	679	1773.66
INSV	461	461	646	276	442.27

2.3.2 Dataset Preprocessing

Neonates have no significant control over their cries and therefore can only have a few of the respiratory maneuvers present in adults. Lester et al. (Lester & Boukydis, 1985) reported that the cry pattern of newborns often shows an expiration phase that is five times longer than the inspiration, which was confirmed by the durations of signals for the expiration and voiced inspiration in our dataset.

The process of segmenting and labeling the cry signals was manual and rather perceptive, and consequently a time-consuming one as well. The usual method was to detect the start and end of a cry unit by visual and auditory investigation of the spectrogram of the cry signal (Abou-Abbas et al., 2015).

Our team of researchers annotated the labels corresponding to various segments of cry signals for this study using WaveSurfer software, as in Figure 2.2. The recordings of our corpus have been manually annotated to mark the start and endpoints of each vocalization. A newborn cry can comprise typical cry sounds, glottal sounds, hiccups, short pause segments between cries and faint cries (Abou-Abbas, Tadj, & Fersaie, 2017). The inspiration is believed to contain information pointing to pain and distress cries (Aucouturier et al., 2011).

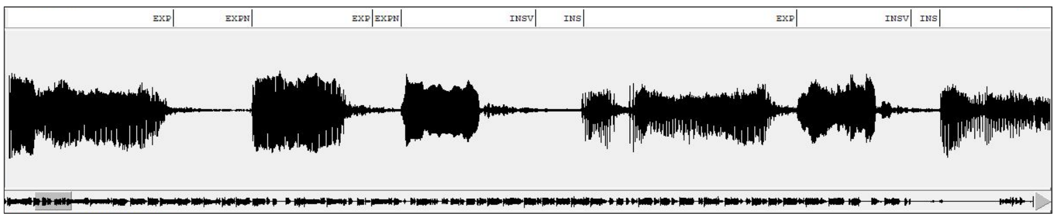

Figure 2.2 Labels annotated using WaveSurfer software for a cry signal

The power needed for driving the expiratory phase of a cry is stored during inspiration. Usually, cries occur during this respiratory phase, so this segment contains the main information (Abou-Abbas, Tadj, & Fersaie, 2017). Additionally, voiced inspiration has proven to be significant in the study of pathologic neonates (Abou-Abbas, Tadj, Gargour, et al., 2017). Therefore, INSV and EXP units are used separately for this study in order to discriminate between healthy and pathologic cries.

2.3.3 Feature Extraction

In the process of generating a cry sound, the impulses produced by the glottis pass through the vocal tract, which acts as a filter. In other words, the vocal tract filters the glottal impulses so as to produce the desired sounds (Wasz-Hockert, Lind, Partanen, Valanne, & Vuorenkoski, 1968). The Cepstrum is a homomorphic transformation that allows for the discrimination of the source and filter (Huang, Acero, Hon, & Foreword By-Reddy, 2001); therefore, cepstral analysis was employed here. Furthermore, the cry signal is non-stationary and dynamic. Hence, an entropy-based feature vector that can capture the presence of complexity in the cry signal is indispensable in the study of newborn pathology diagnosis (Brent, 2010). Our dataset was recorded in real-world conditions; therefore, the presence of noise was inevitable. In other biological signals, the noise is treated differently based on the purpose and applications (Porta, De Maria, Bari, Marchi, & Faes, 2016). In this regard, as suggested by the previous researchers in our lab (Alaie et al., 2016), we addressed this issue by studying both INSV and EXP datasets in order to be able to have a more reliable representation of the results. Alaie et al. (Alaie et al., 2016) mentioned that EXP cries are more reliable in terms of estimating the true value. Furthermore, the acquisition of the cry signals was done in the same conditions for both healthy and septic newborns, and all the steps for the analysis of both groups were similar. The biological signals are associated with nonstationarities. Maganin et al. (Magagnin et al., 2011) reported that these nonstationarities may have detrimental effects on the results. In order to overcome the difficulties in processing and the classification of the nonstationary cry signal, it is standard practice to employ filter banks and a sliding window of short length (10 ms) (Young et al., 2002). The windowing of the nonstationary signal has been introduced as a solution for achieving a locally stationary signal (Cohen, 1995). In this study, the Hamming window and Mel-filter banks were utilized before extracting the features. Each of the introduced feature sets was tested both individually and combined with other features. In the next step, these feature sets were fed to the KNN and SVM classifiers, and the hyperparameters for each of them was optimized using the BHPO method.

2.3.3.1 Mel-Frequency Cepstral Coefficients (MFCC)

Prior to the extraction of MFCC features, the cry signal needs to be pre-emphasized, which means that the signal is filtered by $H(z) = 1 - az^{-1}$ as the transfer function of the signal. This filtering allocates higher gains to higher frequencies. In this study, the value of a was selected equal to 0.97 based on previous researchers' work (Alaie et al., 2016). Extracting MFCCs consists of four main steps, which are described here (Vaishnavi & Dhanaselvam, 2019):

1- Applying a windowing criterion to the signal: The window was applied to enhance the harmonics, smooth the edges and decrease the edge effect of applying a Discrete

Fourier Transform (DFT) to the signal. Here, the Hamming window with a frame size of 10 ms and 30 percent overlap between consecutive frames was selected.

2- Implementing the DFT: In order to obtain the magnitude spectrum of each window, the DFT is applied to the cry signal. In this study, overlapping triangular filters were employed; the number of filters used varied in general between 13 and 24. The MFCC features were computed from 13 filter banks.

3- Computing the logarithm of magnitude and scaling the frequencies on a Mel scale: The magnitude spectrum was multiplied by every triangular Mel weighting filter to calculate the Mel spectrum. The Mel spectrum should be represented on a log scale to be prepared for the next step. Equation 2.1 gives the Mel scale of frequency f.

$$M(f) = 1125\ ln(1 + f/700) \tag{2.1}$$

4- Taking the inverse Discrete Cosine Transform (iDCT) of the signal: As mentioned before, the energy levels of adjacent bands tend to be correlated due to the smooth form of the vocal tract. Therefore, the transformed Mel-frequency coefficients must undergo an iDCT that results in separable cepstral coefficients. The first few MFCC coefficients might be sufficient for a robust representation of the system (Benesty, Sondhi, & Huang, 2007). Therefore, the first 13 coefficients were extracted in this study.

MFCCs often only contain the information from one window; hence, these cepstral coefficients are considered static features. In order to gain information on the temporal dynamics, cepstral coefficients' first and second derivatives should be calculated, which are known as delta and delta-delta coefficients, Equation 2.2.

$$\Delta_n = \frac{\sum_{\theta=1}^{\Theta} \theta(c_{n+\theta} - c_{n-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{2.2}$$

where Δ_n is a delta coefficient from discrete-time n computed in interval of the static coefficients $c_{n-\Theta}$ to $c_{n+\Theta}$; the value of Θ is usually set to 2 (Young et al., 2002). The delta-delta coefficients are calculated with delta coefficients in a similar manner. The dynamic features help us capture the spectral changes in the cry signal. Finally, the dynamic MFCC features are added to the feature vector, and together they form the MFCC feature set with a total of 39 features.

2.3.3.2 Spectral Entropy Cepstral Coefficients (SENCC)

Spectral Entropy (SEN) evaluates the signal's energy distribution uniformity. This measure is an indicator of the complexity of the signal. It can also be employed to capture the peakiness in a signal. Figure 2.3 illustrates the SEN of multiple episodes of expiration cry for a healthy infant as opposed to an infant diagnosed with sepsis. The entropy levels for a septic cry are lower, which was also deduced in previous works (Misra, Ikbal, Bourlard, & Hermansky, 2004).

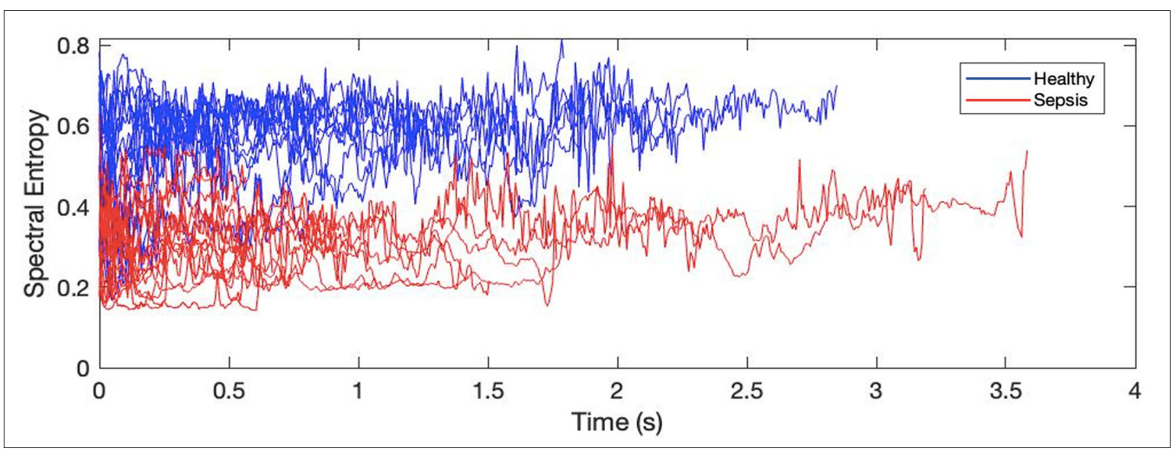

Figure 2.3 Spectral entropy for 20 EXP utterances from one healthy neonate and 20 EXP utterances from one septic neonate

In order to compute the SEN, the spectrum is written in terms of a Probability Mass Function (PMF)-like function, Equation 2.3.

$$x_i = \frac{X_i}{\sum_{i=1}^{N} X_i} \quad \text{for i = 1 to N} \tag{2.3}$$

Here, (the uppercase) X_i, appearing in the nominator and denominator, is the energy of ith frequency component of the spectrum. The PMF of the spectrum is represented by (the lowercase) $x = (x_1,\ldots,x_N)$, and the number of points in the spectrum is specified by N. The entropy of each frame was computed from Equation 2.4 (Toh, Togneri, & Nordholm, 2005).

$$H = -\sum_{i=1}^{N} x_i \, . \, \log_2 x_i \tag{2.4}$$

In order to detect the position of peakiness or flatness present in the spectrum, a process similar to the extraction of the MFCCs was employed. The fast Fourier Transform (FFT) of each frame was calculated. Following the calculation of the FFT, the achieved spectrum was mapped to the Mel-scale in order to mimic the signal based on the human sound perception model. Then, the SEN was computed from the Mel-spectrum. Finally, DCT was applied to decorrelate between the coefficients and further improve the results, and 13 SENCC coefficients were obtained.

2.3.3.3 Spectral Centroid Cepstral Coefficients (SCCC)

SC is a measure of the shape of the spectrum of the signal and the position of the mass of the spectrum. The mean value of SC was shown to be a discriminative feature (Kulkarni & Bairagi, 2018) that indicates where the major energy of the signal is concentrated. SC is expected to be higher for the "brighter sounds" and has been widely employed in the study of timbre for music applications (Brent, 2010). It is also a discriminative feature in the measurement of tone in audio signals (Almeida, Schubert, Smith, & Wolfe, 2017). Figure 2.4 presents how the cries of the neonates suffering from sepsis are associated with lower tone, as is listed as one of the red-flag listings associated with neonatal sepsis (Weiss et al., 2019).

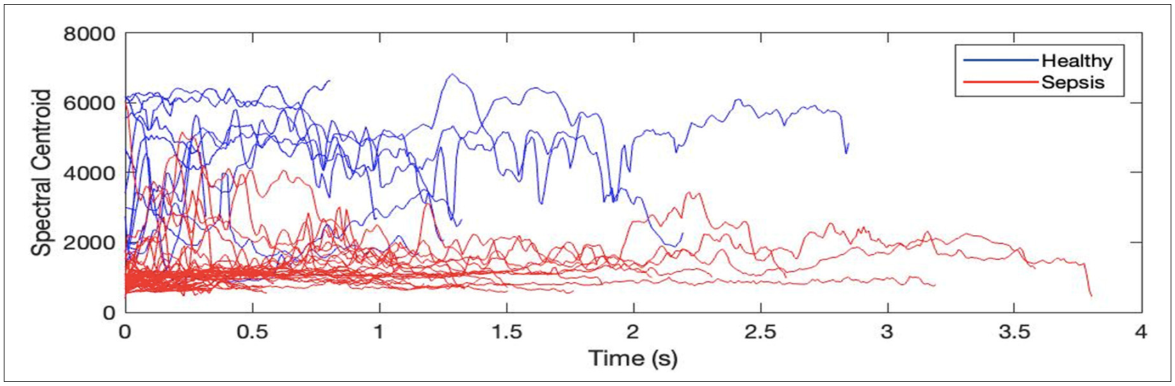

Figure 2.4 Spectral centroid for 15 EXP utterances from one healthy neonate and 15 EXP utterances from one septic neonate

SC denotes the center of the signal's gravity and is computed by taking the weighted mean of the frequency bins. The SC value, Ci of the i-th window, is computed using Equation 2.5.

$$C_i = \frac{\sum_{k=1}^{W_{f_L}} kX_i(k)}{\sum_{k=1}^{W_{f_L}} X_i(k)} \tag{2.5}$$

where $x_i(n)$ are the i-th window samples, and $X_i(k)$ are the DFT coefficients. The SC cepstral coefficients' extraction procedure is similar to what was described for MFCC and SENCC, except that for the SCCC feature vector, the first five coefficients were extracted.

2.3.4 Feature Reduction

The first and most crucial aspect of post-processing is to reduce the dimensionality of the feature vectors to decrease the storage and computational costs. Feature reduction includes all the techniques that aim to make a compact feature set out of the original sets while trying to keep as much information as possible. Camargo et al. (Amaro-Camargo & Reyes-García, 2007) suggested a simple and rapid method that reduces data through statistical operations such as minimum, maximum, average and standard deviation. Messaoud et al. (Messaoud & Tadj, 2011) also proposed an arithmetic method by averaging MFCCs over a time axis. Matikolaie et al. (Matikolaie & Tadj, 2020) further investigated the use of statistical methods in the compression of the MFCC feature set and reported that this method was effective in terms of computational costs and classification accuracy. In order to reduce the dimensionality of the MFCC feature set, the statistical approach was employed, and the mean value of each MFCC coefficient over the time axis of each signal was calculated.

2.3.5 Fuzzy Entropy Based Feature Selection

As explained in the previous sections, entropy is associated with the uncertainty of a given variable. Here, we aim to focus on the concept of fuzzy entropy, which calculates entropy through a fuzzy c-means clustering algorithm. This method is called Fuzzy Entropy Selection of the features (FE Selection). In general, fuzziness refers to a possibilistic point of view, while the aforementioned entropy measure focuses on randomness and has a probabilistic perspective. This method was chosen because it is very fast and imposes a negligible computational cost on the system (Khushaba et al., 2007).

Trivedi et al. (Trivedi & Bezdek, 1986) introduced a Fuzzy c-Partition model that computed the membership of each feature dimension and its corresponding FE. Suppose a finite set where Y = $\{y_1, y_2, \dots, y_n\}$, a set of real $c \times n$ matrices denoted by V_{cn}, and c is an integer so that $2 \leq c < n$. The fuzzy c-partition space, M_{fc}, for Y is given by Equation 2.6.

$$M_{fc} = \{U \in V_{cn} | \, u_{ik} \in [0,1], \forall i, k; \textstyle\sum_{i=1}^{c} u_{ik} = 1, \forall k;\ 0 < \sum_{i=1}^{c} u_{ik} < n, \forall i\} \qquad (2.6)$$

This means that membership values of y_j in the c subsets could be obtained from the jth column of matrix U, which is from $c \times n$ dimensions. The grade of membership of y_k in the ith fuzzy subset of Y is represented by $u_{ik} = u_i(y_k)$. Therefore, the membership of each pattern y_k in all subsets is calculated and then normalized. Instead of applying this algorithm to each pattern, it is applied to each feature similar to previous studies (Khushaba et al., 2007). The FE is calculated based on the matching degree, D_c, described by Equation 2.7, where u_c

is the membership of the feature y_d in each of our two classes, denoted by c for each class and C for the set of the two classes (Lee et al., 2001).

$$D_c = \frac{\sum_{y_d \in c} u_c\,(y_d)}{\sum_{y \in C} u_c\,(y_d)} \tag{2.7}$$

The FE of the elements of each of these classes is achieved through Equation 2.8.

$$FE_c = -D_c \, log \, D_c \tag{2.8}$$

Finally, the overall FE is given by Equation 2.9:

$$FE = \sum_{c=1}^{C} FE_c = FE_{Healthy} + FE_{Septic} \tag{2.9}$$

The main interpretation of the FE is very similar to the SEN which was described before; higher entropy translates to lower information content. We based our feature selection on the fact that smaller FE values contribute more to the recognition of septic infants. Thus, we first calculated the average FE value across the features and set this value as a threshold for our feature selection. In the next step, we imposed a condition where only the features with FE values lower than the overall average FE should be selected and formed a new feature set to be fed into the classifier. This condition secures the selection of features with minimum overlap and also will likely result in a lower misclassification possibility, which will be evaluated by the Matthews Correlation Coefficient (MCC) measure.

2.3.6 Classification

The performance of the feature sets was tested by the two classification methods of KNN and SVM in order to discriminate between the healthy and septic neonates. Each EXP or INSV cry episode was treated as a sample and the classifier assigned a label of healthy or septic to it.

Both classification methods benefit from five-fold cross-validation in order to avoid over-fitting and ensure credibility. The models were tuned with the BHPO method in order to enhance the performance of each model.

2.3.6.1 K-nearest Neighborhood (KNN)

This method is an efficient yet simple method of classifying data. As the name of this method suggests, the features with similar values belong to the same class. The KNN classifiers often use Euclidean distance for the measurement of the distance between data points. This classifier has three bases for classification: sets of labeled data, a distance measure and, finally, the number of neighbors, which is denoted by K. In other words, KNN classifies a given sample based on the majority vote of the neighborhood and the distance (Latifpour, Mosleh, & Kheyrandish, 2015; Wu, Kumar, Quinlan, et al., 2008). The number of neighbours was automatically tuned with the BHPO method in the first step, which in all of the given experiments returned K = 1 as the best choice. The other hyperparameter selected for tuning is the type of distance used with each feature set. The distance measures included in this optimization include Minkowski, Chebyshev, Euclidean, standard Euclidean, cosine, Jaccard, Manhattan and Hamming.

2.3.6.2 Support Vector Machine (SVM)

SVM has a broad application in the classification of audio signals. An SVM differentiates between two cases by implementing a hyperplane. SVM is inspired by the statistical learning theory and the Vapnik–Chervonenkis (VC) dimension. The optimal hyperplane is constructed when the distance between the hyperplane and data is considerable. The linear data can be classified by simply constructing a straight hyperplane, while the nonlinear data should be made linearly separable for the purpose of classification. It means that the data must pass through a transformation into high-dimensional space, which is known as the kernel function (Sahak, Mansor, Lee, Zabidi, & Yassin, 2013). The gaussian kernel is used in this study. The hyperparameters selected for HPO were kernel scale and box constraint. The BHPO was used for the tuning of the mentioned hyperparameters of the SVM model as well.

2.3.6.3 Bayesian Hyperparameter Optimization (BHPO)

In order to maintain the classification errors at a minimum while achieving high performance in a ML problem, HPO methods are used. A majority of ML designs include hyperparameters. With recent advances in the field of automated ML, various methods such as random search, grid search and Bayesian optimization have been introduced that no longer require human efforts for tuning these hyperparameters. More importantly, the hyperparameters are tailored to meet the requirements of each specific task and the results are reproducible. The basis of HPO is finding the optimal value for the hyperparameters in a finite set of predefined values, in order to minimize or maximize an objective function (e.g., model performance). The common challenge with these grid search and random search methods is the high number (~90 iterations) of function evaluations needed to obtain minimal error, which in turn is not cost-effective and may cause curse of dimensionality (Feurer & Hutter, 2019). BHPO is also an iterative method in which the acquisition function and the probabilistic surrogate model are the vital elements. The model is constantly updated based on the objective function evaluation, which is expressed as Equation 2.10 (Ashwini & Vincent, 2022):

$$x^* = \underset{x \in X}{argmin}\, f(x) \tag{2.10}$$

The methodology in summary is deduction of the information on the model in each iteration based on new hyperparameters and the resulting model performance. When the number of determined iterations ends, the global optimal hyperparameter configuration is reported. In order to establish the local optimal hyperparameter, the acquisition function employs the predictive information of each possible hyperparameter configuration. BHPO requires far fewer iterations when compared to the other two methods and all the experiments in this study were performed with only 30 iterations.

2.4 Evaluation and Results

The features introduced in this study were extracted and fed to the classifiers with the purpose of distinguishing between healthy and septic neonates. In order to compare their abilities to reach that goal, several experiments were conducted which were comprised of different feature sets, implementing the features individually or combined, and two classification methods with a wide range of parameters. Finally, the models were tuned to obtain the best performance. In this framework, the following feature sets were used:

- MFCC;
- SENCC;
- SCCC;
- MFCC + SENCC;
- SENCC + SCCC;
- MFCC + SCCC;
- MFCC + SENCC + SCCC.

Five-fold cross-validation was carried out after feeding each feature set to the classifier. This means that one fold of data was treated as the test data in each iteration of the training process, and the other four were the training folds. This process was repeated until all the folds had been used as the test fold. This process was repeated for both EXP and INSV datasets.

2.4.1 Evaluation Criteria

There are different approaches to evaluating a system's performance. One of the main measures for that purpose is accuracy. Accuracy is the ratio of correct decisions to the total number of cases, Equation 2.11.

$$\mathrm{Acc} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FN+FP}} \tag{2.11}$$

where N stands for negative and P stands for positive, and T and F stand for true and false. However, when the task is diagnosing a pathology, it is of utmost importance that the system does not miss a pathologic case. A confusion matrix is defined for the binary classification task where the problem is the discrimination between healthy and pathologic cries, as shown in Figure 2.5. In this study, the positive label stands for septic infants and the negative label stands for healthy (not septic).

		True Class	
		Septic (P)	Healthy (N)
Predicted Class	Septic (T)	True Positive (TP)	False Positive (FP)
	Healthy (F)	False Negative (FN)	True Negative (TN)

Figure 2.5 The confusion matrix for a binary classification

The True Positive Rate (TPR) is referred to as sensitivity, hit rate or recall. In the concept of his study, recall is also an important measure as it demonstrates how many true septic cases have been captured by the NCDS. Hence, recall owes its importance to the fact that a false healthy detection is not desirable, Equation 2.12 (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008).

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{2.12}$$

The Positive Predictive Value (PPV) is another measure and is also referred to as precision. In this framework, precision is the probability that a septic case is predicted as septic, Equation 2.13.

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{2.13}$$

The next evaluation measure is called the F1-score, which shows the balance between precision and recall and is a good measure of the system's performance. Mathematically, the F1-score is the harmonic mean of precision and recall, Equation 2.14.

$$\mathrm{F1} = \frac{\mathrm{TP}}{\mathrm{TP+0.5\ (\ FP+FN)}} = 2.\frac{\mathrm{precision\ .recall}}{\mathrm{precision+recall}} \tag{2.14}$$

Finally, the MCC considers all the information in a contingency matrix. The value of this measure belongs to the [−1, +1] interval where 0 denotes a random distribution, −1 shows complete misclassification and +1 corresponds to perfect classification (Chicco & Jurman, 2020). The MCC is computed using Equation 2.15:

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP+FN})(\mathrm{TN+FP})(\mathrm{TP+FP})(\mathrm{TN+FN})}} \tag{2.15}$$

The MCC measure is highly informative for binary classification tasks in general (Vihinen, 2012). Since we have a healthy versus septic classification problem in this study, implementing the MCC is considered beneficial and proper.

2.4.2 Results

The results of different experiments conducted in this study are given in Table 2.3 to Table 2.10. As previously mentioned, we analyzed the performance of feature sets for two separate datasets of EXP and INSV. Moreover, KNN and SVM were employed as the classifiers in this study. The feature sets were used both individually and jointly. They were concatenated so that we could compare the performance of larger feature sets as opposed to the individual feature sets. It is noteworthy that our findings regarding the behavior of feature sets were consistent with medical findings and other researchers' work, as discussed in Sections 2.3.3.2 and 2.3.3.3. Regarding the evaluation criteria discussed in the previous section, the higher the value of each measure, the better the performance of our NCDS. The results presented in this section are all in the form of average and standard deviation of five-fold cross validation values. For all the measures, the values represent percentages except for the MCC measure, which is unitless and belongs to the [−1, 1] range.

Table 2.3 Evaluation metrics for the MFCC feature set

MFCC	EXP		INSV	
	SVM	KNN	SVM	KNN
Accuracy (%)	88.07 ± 0.98	81.97 ± 0.70	89.06 ± 1.80	85.36 ± 0.49
Recall (%)	85.71 ± 1.91	91.67 ± 1.83	91.85 ± 2.96	92.74 ± 0.33
Precision (%)	90.38 ± 0.74	72.48 ± 1.93	86.38 ± 1.05	78.30 ± 0.81
Specificity (%)	89.72 ± 0.72	76.56 ± 1.01	86.58 ± 1.16	80.36 ± 0.61
F-score (%)	87.66 ± 1.13	83.42 ± 0.67	89.13 ± 1.90	86.11 ± 0.42
MCC	0.76 ± 0.02	0.65 ± 0.01	0.78 ± 0.04	0.72 ± 0.01
Distance/Kernel Scale	1.7864	Cosine	5.8165	Cosine

Table 2.4 Evaluation metrics for the SENCC feature set

SENCC	EXP		INSV	
	SVM	KNN	SVM	KNN
Accuracy (%)	71.55 ± 0.70	72.02 ± 0.82	69.20 ± 0.85	65.00 ± 1.71
Recall (%)	42.50 ± 1.42	44.88 ± 1.85	37.04 ± 1.74	58.81 ± 3.13
Precision (%)	100.00 ± 0.00	98.60 ± 0.24	100.00 ± 0.00	70.92 ± 2.01
Specificity (%)	100.00 ± 0.00	96.93 ± 0.43	100.00 ± 0.00	65.95 ± 1.82
F-score (%)	59.64 ± 1.40	61.33 ± 1.68	54.04 ± 1.84	62.15 ± 2.29
MCC	0.52 ± 0.01	0.52 ± 0.01	0.48 ± 0.01	0.30 ± 0.03
Distance/Kernel Scale	0.0116	Cosine	0.1063	Chebyshev

Table 2.5 Evaluation metrics for the SCCC feature set

SCCC	EXP		INSV	
	SVM	KNN	SVM	KNN
Accuracy (%)	71.46 ± 0.67	72.02 ± 0.77	69.06 ± 0.83	68.12 ± 0.87
Recall (%)	42.32 ± 1.35	45.60 ± 1.73	36.74 ± 1.71	37.33 ± 1.93
Precision (%)	100 ± 0.00	97.90 ± 0.24	100.00 ± 0.00	96.03 ± 0.81
Specificity (%)	100 ± 0.00	95.52 ± 0.39	100.00 ± 0.00	90.04 ± 1.72
F-score (%)	59.46 ± 1.33	61.71 ± 1.55	53.72 ± 1.82	52.75 ± 1.91
MCC	0.52 ± 0.01	0.51 ± 0.01	0.48 ± 0.01	0.41 ± 0.02
Distance/Kernel Scale	0.0089	Jaccard	0.0129	Hamming

Table 2.6 Evaluation metrics for the combination of SCCC and SENCC feature set

SCCC + SENCC	EXP		INSV	
	SVM	KNN	SVM	KNN
Accuracy (%)	71.55 ± 0.70	72.52 ± 0.89	69.06 ± 0.83	65.72 ± 1.24
Recall (%)	42.50 ± 1.42	47.80 ± 2.13	36.74 ± 1.71	58.52 ± 2.10
Precision (%)	100.00 ± 0.00	96.73 ± 0.38	100.00 ± 0.00	72.62 ± 2.28
Specificity (%)	100.00 ± 0.00	93.50 ± 0.48	100.00 ± 0.00	67.21 ± 1.68
F-score (%)	59.64 ± 1.40	63.23 ± 1.79	53.72 ± 1.82	62.54 ± 1.47
MCC	0.52 ± 0.01	0.51 ± 0.01	0.48 ± 0.01	0.31 ± 0.03
Distance/Kernel Scale	0.0951	Jaccard	0.0764	Cosine

Table 2.7 Evaluation metrics for the combination of SCCC and MFCC feature set

MFCC + SCCC	EXP		INSV	
	SVM	KNN	SVM	KNN
Accuracy (%)	81.50 ± 1.46	82.44 ± 0.65	88.41 ± 1.77	87.25 ± 0.94
Recall (%)	83.69 ± 2.40	89.05 ± 1.42	89.19 ± 3.25	92.74 ± 1.61
Precision (%)	79.36 ± 1.53	75.98 ± 0.98	87.66 ± 1.47	81.99 ± 2.28
Specificity (%)	79.89 ± 1.31	78.41 ± 0.62	87.38 ± 1.39	83.17 ± 1.62
F-score (%)	81.74 ± 1.57	83.39 ± 0.69	88.25 ± 1.92	87.68 ± 0.84
MCC	0.63 ± 0.03	0.66 ± 0.01	0.77 ± 0.04	0.75 ± 0.02
Distance/Kernel Scale	6.5705	Standard Euclidean	2.5893	Manhattan

Table 2.8 Evaluation metrics for the combination of SENCC and SENCC feature set

MFCC + SENCC	EXP		INSV	
	SVM	KNN	SVM	KNN
Accuracy (%)	89.99 ± 0.71	86.83 ± 0.44	88.19 ± 1.42	84.57 ± 0.75
Recall (%)	88.15 ± 1.75	91.07 ± 0.87	88.89 ± 3.31	90.07 ± 0.66
Precision (%)	91.78 ± 0.75	82.68 ± 1.56	87.52 ± 1.19	79.29 ± 0.92
Specificity (%)	91.31 ± 0.65	83.76 ± 1.10	87.22 ± 0.94	80.64 ± 0.79
F-score (%)	89.70 ± 0.83	87.26 ± 0.31	88.02 ± 1.61	85.10 ± 0.70
MCC	0.80 ± 0.01	0.74 ± 0.01	0.76 ± 0.03	0.70 ± 0.01
Distance/Kernel Scale	2.1612	Minkowski	4.5656	Correlation

Table 2.9 Evaluation metrics for the combination of all feature sets

All Features	EXP		INSV	
	SVM	KNN	SVM	KNN
Accuracy (%)	85.71 ± 1.17	82.77 ± 0.29	89.42 ± 1.01	85.87 ± 0.92
Recall (%)	78.75 ± 3.34	85.03 ± 1.54	91.41 ± 1.62	94.22 ± 0.97
Precision (%)	92.54 ± 1.26	80.29 ± 1.51	87.52 ± 1.63	77.87 ± 1.36
Specificity (%)	91.21 ± 1.04	80.93 ± 0.93	87.54 ± 1.42	80.31 ± 1.02
F-score (%)	84.48 ± 1.62	83.05 ± 0.38	89.42 ± 1.00	86.71 ± 0.84
MCC	0.72 ± 0.02	0.66 ± 0.01	0.79 ± 0.02	0.73 ± 0.02
Distance/Kernel Scale	2.3092	Euclidean	3.8005	Cosine

Table 2.10 Evaluation metrics after applying FE Selection to the best feature sets of previous experiments

FE Selection	EXP: MFCC + SENCC		INSV: All Features Combined	
	All	FE Selection	All	FE Selection
Accuracy (%)	89.99 ± 0.71	88.51 ± 0.77	89.42 ± 1.01	91.81 ± 0.75
Recall (%)	88.15 ± 1.75	89.11 ± 1.32	91.41 ± 1.62	93.23 ± 0.44
Precision (%)	91.78 ± 0.75	87.93 ± 0.84	87.52 ± 1.63	90.66 ± 1.18
Specificity (%)	91.31 ± 0.65	87.86 ± 0.76	87.54 ± 1.42	89.07 ± 1.25
F-score (%)	89.70 ± 0.83	88.47 ± 0.81	89.42 ± 1.00	91.10 ± 0.77
MCC	0.80 ± 0.01	0.77 ± 0.02	0.79 ± 0.02	0.84 ± 0.01
Number of Features	52	27	57	35

Table 2.3 presents the results for the evaluation of the MFCC feature set for EXP and INSV datasets. Furthermore, the MFCC feature set was evaluated with the use of the HPO method. We used BHPO for both classifiers, as mentioned in the previous sections. Finally, the performance of this feature set was tested with the KNN and SVM classifiers. The HPO led to consistent enhancement of accuracy and F-score measures across both datasets for the MFCC feature set. The SVM classifier had better performance in the evaluation of the MFCC feature set in both datasets in terms of all the evaluation measures except for recall, where the KNN classifier showed better performance. The best results achieved by this feature set are highlighted.

Overall, the highest achieved F-score and accuracy for the EXP dataset were 88.07% and 87.66%, respectively. In this regard, the performance of the NCDS with the INSV dataset was superior to the EXP dataset; the highest overall results obtained for this dataset in terms of F-score and accuracy were 89.06% and 89.13%, respectively.

As can be seen in Table 2.4 and Table 2.5, the performance of our NCDS with the SENCC and the SCCC feature sets were similar; both feature sets achieved 72.02% accuracy measures (with different standard deviations). Furthermore, the SENCC and the SCCC feature sets obtained 61.33% and 61.71%, respectively, for F-score with the KNN classifier for the EXP dataset. Also, both datasets and feature sets obtained 100% precision and specificity with the SVM classifier. In the evaluation of the INSV dataset, KNN had better performance in terms of accuracy and F-score. The best F-score for the SENCC dataset was achieved with the KNN classifier for the INSV dataset, which was equal to 62.15%. Regarding the SCCC feature set, the highest F-score was 61.71% for the EXP dataset using the KNN classification method.

In the next step, the framework of feature combination was investigated. We examined all possible combinations of these feature sets that were made possible through their concatenation. The results of these combinations are presented in

Table 2.6 to Table 2.9. It can be observed that using the SVM classification method, the combination of SENCC and SCCC was dominated by the SENCC feature set for the EXP dataset and by SCCC for the INSV method since, despite the difference in their kernel scales, there was not a change in the evaluation measures. The overall best accuracy and F-score for the combination of SCCC and SENCC belonged to the KNN classification of the EXP dataset with 72.52% and 63.23%, respectively.

The addition of the SCCC feature set to the MFCC feature set with the SVM classifier achieved the results of 88.41% and 88.25% for accuracy and F-score measures with the INSV dataset, as seen in Table 2.7. Furthermore, using the KNN classifier with the EXP dataset resulted in better performance in terms of accuracy and F-score, with 82.44% and 83.39%, respectively.

As can be interpreted from Table 2.8, the best performance in terms of accuracy and F-score measures for the EXP dataset across all the experiments was achieved by the combination of the MFCC and SENCC feature sets. The highest accuracy and F-score among all the experiments on the EXP were 89.99% and 89.70%, respectively. Regarding the EXP dataset, the accuracy and F-score measures were enhanced by 1.92% and 2.04%, respectively, compared to the MFCC feature set, which had the highest accuracy and F-score among the individual datasets.

Finally, the combination of all the individual feature sets with the SVM classification resulted in the highest accuracy and F-score across all the experiments for the INSV dataset, with 89.42% for both measures, as seen in Table 2.9. The combination of all individual feature sets enhanced these two measures by 0.36% and 3.31%, respectively, compared to the MFCC feature set, which achieved the best results among the individual feature sets.

As our final experiment, we computed the FE measure for the best two experiments discussed above and selected the most compatible features in each presented feature set. These two experiments included the combination of the MFCC and SENCC features for the EXP dataset and the combination of all features for the INSV dataset, both classified using the SVM method. Table 2.10 represents the results of applying the FE selection method to these two experiments.

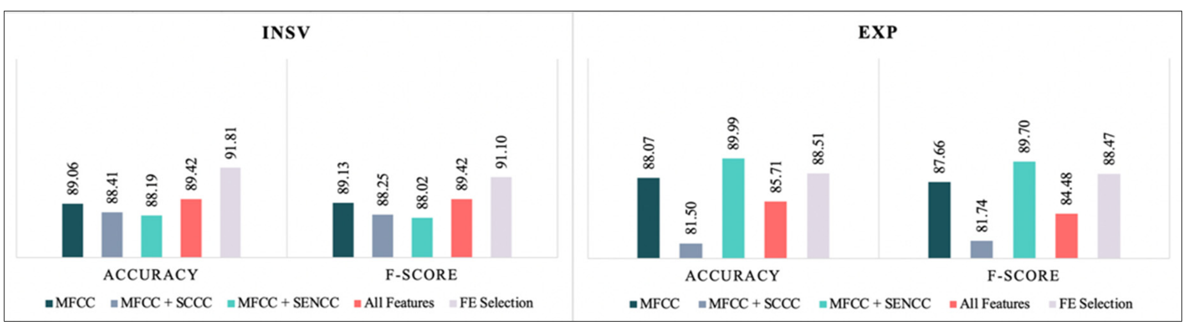

Figure 2.6 Best F-score and accuracy measures for the SVM classifier in each feature set

According to the evaluation measures studied here, the FE selection method was highly successful. Implementing fewer features resulted in a negligible decrease in the evaluation measures for the EXP dataset. As for the INSV dataset, the FE selection led to enhancement of all the evaluation measures, which marked the highest accuracy and F-score measures across all the experiments with 91.81% and 91.10%, respectively. Figure 2.6 summarizes the results of the experiments in terms of F-score and accuracy measures for the SVM classifier that yielded the best results for a clearer comparison.

2.5 Discussion

This study further explored sepsis in newborns by the means of studying their cry signal through developing an NCDS design. Even though sepsis is associated with high mortality rates in newborns, only one recent work in our lab has studied the cries of septic infants in parallel to the study presented here. The previous study in our lab did not discuss the performance of the system in terms of the accuracy measure (Matikolaie & Tadj, 2022). In this study, accuracy as well as several other evaluation measures were included to help better study the performance of NCDSs for diagnosing septic newborns. Our goal was to build upon the previous work and also design a simple model that could achieve improved or comparable performance. Moreover, it is worth highlighting this research's novelty in terms of analyzing the infant cry from the perspective of musical machine-learning applications. Most of the works addressing infant cries have treated the cry signal as a pre-speech audio. We believed that the harmonic nature of the infant cry, as well as the natural differences in the voice generation organs of infants and adults, had the potential to be analyzed with the features and methods that have shown promising results in the field of musical signal processing. There is meager information on the behaviour of pathologic cries based on analysis of the SC, and this work is the only study that combines SC with cepstral analysis in the study of pathologic newborn cries.

Nowadays, many audio recognition system designs benefit from state-of-the-art deep learning and ML methods. However, the main challenge in studying pathology-related applications is the acquisition of relevant data. The occurrence of a specific pathology in any given time interval in newborns is not predictable and meeting the ethical and technical requirements to include cry samples in a database calls for extreme measures. Therefore, this study explored different approaches to make the best use of the available data. The limitations of the data impose many challenges in NCDS design. Inspired by (Matikolaie & Tadj, 2022), we also addressed this issue by segmenting each cry signal into multiple expiratory and inspiratory episodes in order to treat each segment as a sample. Despite our efforts to make the analysis in this study unbiased towards race, origin and other factors, it should be noted that the system might still suffer from a low generalization power since it was designed based on a limited number of participants. Therefore, future research should be devoted to further investigate this matter. Moreover, the data dimensionality imposed more challenges in the process of feature extraction. It is common practice in NCDS studies to use statistical measures with extracted features to reduce computational costs (Matikolaie & Tadj, 2020; Messaoud & Tadj, 2011). The statistical method was chosen to ensure that our results are comparable to the previous studies. Furthermore, extra attention should be paid to the details in the design of conventional models because limited data may lead to overfitting of the classifiers. We addressed this challenge by using BHPO for both the SVM and KNN classification methods. As can be interpreted from Table 6, the accuracy of the NCDS was enhanced up to 89.42% for the INSV dataset. Also, we believed that the characteristics that were reported in the medical studies conducted on septic cries could be better analyzed through cepstral analysis of the SC and the SEN features, which was confirmed by our findings. Through the implementation of these features, the presented work was made capable of obtaining F-scores of 89.70% for the EXP dataset and 89.42% for the INSV dataset, which were both superior to the previous study (Matikolaie & Tadj, 2022). Therefore, we were able to show that even a single episode (as opposed to the All Episode voting scheme) analysis of the cry signal could achieve reassuring performance with careful selection of the parameters.

As mentioned, the performance of the system was tested with the two different classification approaches of SVM and KNN, and SVM showed superiority in a majority of experiments. The recall measure was an exception to this conclusion, where KNN showed better performance. The presented study also showed that elevating the number of features in a pattern recognition problem does not always enhance the system’s performance. The predictive performance of the system depends on many different factors.

As was mentioned previously, the high discriminative power of inspiratory cries in the study of pathologic newborns has been neglected in many works. However, the high values of the evaluation measures achieved for this dataset show the potential for further investigation of inspiratory cries, which was consistent with previous studies in our lab.
As discussed in Section 3, the entropy levels differ across healthy and septic infants, which is also reported by other researchers where healthy newborn cries were distinguished from pathologic cries (Lahmiri, Tadj, Gargour, et al., 2021). The same explanation applies to the SC of the infant cries, which marks these feature sets as potential biomarkers for further study of septic newborns. The SENCC measure alone could achieve 72% accuracy with the SVM classifier; it yields the highest performance in this study when combined with the MFCC feature sets.

Figure 2.7 shows the elapsed time for extracting each of our feature sets for EXP and INSV datasets. The elapsed times are rational in terms of the duration of datasets and the number of coefficients in each feature set. Nevertheless, it was validated that extracting the SENCC and SCCC features does not aggravate the system's complexity in terms of computational costs, and they have similar performance and run-times.

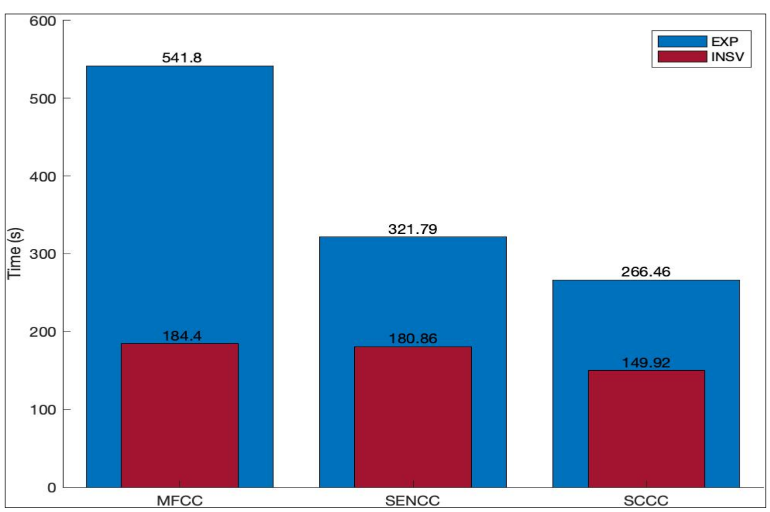

Figure 2.7 The elapsed time for the extraction of features

It has been reported that the aggregation of multiple classifiers, with the intention of having the classifiers compensate for the errors of each other, does not yield good results and only burdens the system with more complexity and computational cost (Matikolaie & Tadj, 2022). In order to overcome this issue, we utilized BHPO with only 30 iterations, which is a low-cost and fast method. We were able to outperform the mentioned model in terms of F-score by between 3–6% for both datasets.

None of the conducted experiments showed misclassification in terms of the MCC measure since they all had positive values. Moreover, all the combined feature sets for the EXP dataset yielded MCC values higher than 0.50. MCC values consider all elements from a confusion matrix; thus, their high value means prediction had satisfactory performance in terms of TP, TN, FN and FP. The same explanation applies to the INSV dataset, except for the feature set formed by the combination of the SENCC and SCCC features.

As a final contribution, we further explored the use of entropy-based measures in the framework of diagnosing pathologies in infants based on their cry signals. By calculating the FE of the combined feature sets, we were able to remove redundant features, and also identified which features yielded better information in the feature set. After calculating the average FE across all measures, we set a threshold for the selection of the features and removed all the features with a higher FE value than the average. As a result, the system's accuracy for the EXP dataset was not notably hindered by removing more than 40% of the features, and it was even enhanced in terms of the recall measure. Moreover, all of the evaluation measures were enhanced for the INSV dataset, which shows the reliability of this feature selection method in selecting the most prominent features. Figure 2.8 shows the difference in the evaluation measures for the best experiments in each dataset, after removing nearly 50% of the features based on their FE.

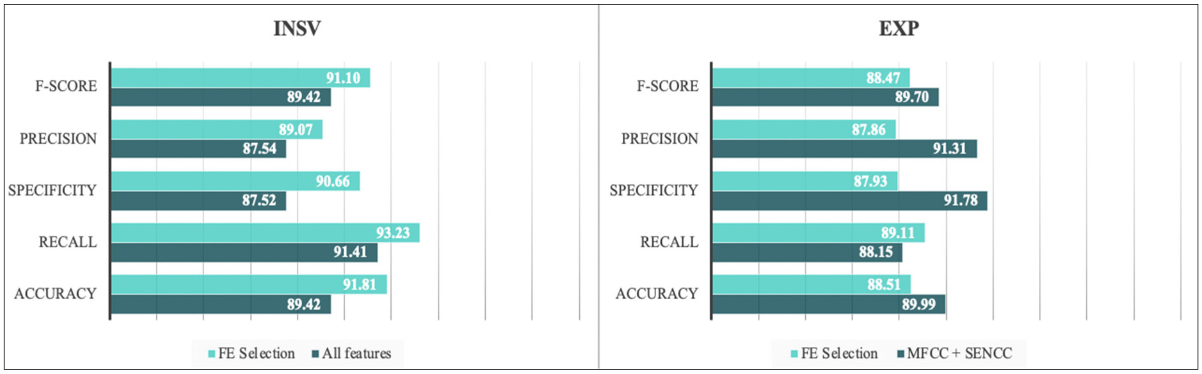

Figure 2.8 The comparison of results before and after applying the FE Selection method

The results from these experiments also highlighted the fact that incrementing the number of features may not always lead to higher accuracy or enhanced performance of the system. Furthermore, it is noteworthy that understanding the information content of the feature space and selection of the most compatible features accordingly improves the performance of the system, as seen through the INSV dataset experiments where using FE selection enhanced the system's performance by an average of 2%.

As discussed before, high recall values show the ability of the NCDS in the successful detection of septic cases. The MFCC feature set had the best performance in terms of recall among all the individual feature sets with 92.74% for the INSV dataset. The overall highest recall was obtained by combining all feature sets for the INSV dataset with 94.22%.

The implementation of the FE was a successful experiment in addition to all other presented experiments on the septic newborn cry signals. Our main achievement through the study of FE was to reduce the feature space by more than 40% while keeping the same performance; however, the improvement from the FE alone was limited. This experiment was simply carried out to evaluate if the system could benefit from further simplification and to eliminate the features corrupted by noise. We tried to develop each stage of the proposed NCDS in a way that was not explored well enough or not investigated in the field of NCDS designs. This included the analysis of septic newborn cries in NCDSs for only the second time ever, introducing the use of cepstral coefficients of entropy and centroid to NCDS design, the ways we manipulated these features in order to study the newborn cries, the use of FE for feature selection, and employing BHPO for both the SVM and KNN methods, all of which, to the best of our knowledge, was unprecedented in NCDSs. We acknowledge that the study presented here cannot cover all aspects of the study of septic newborn cries and may be improved upon in many ways. There is an unceasing need for more studies in this field. The authors suggest exploring more classification schemes such as naïve Bayesian, Ensemble classifier, etc., and fusing their outcomes to form a more precise decision. There are more in-depth ideas for investigation that can assess the effect of the inevitable noise in the biological signals, as well as exploring other entropy-based measures, which could not be explored in the scope of this study.

2.6 Conclusion

In the presented study, sepsis was targeted as one of the leading mortality causes of neonates worldwide. The main goal was to develop a simple NCDS which is capable of detecting septic infants without the need for in-depth and invasive clinical tests. The recording of the cries does not need any complicated equipment, it can be done with a commercial handheld recorder, and it does not require any special conditions (our database was recorded in maternity rooms, NICUs, etc.). It does not even necessitate touching the newborn. We believed it was worth exploring how the cries of septic newborns would be different from those of healthy newborns as a complementary method to other means present in the literature. The novelty of our proposed work is in taking common tools in audio, music and speech processing, combining them, and tuning them in such a way that the final design is still simple but is able to achieve high performance in comparison to the other similar methods that are computationally expensive. The proposed NCDS could be employed as an early alarm for medical staff to detect possible pathologic neonates as soon as possible. Within this framework, entropy was utilized in various stages of the architecture, and yet it avoided complicated designs as well as any need for high-end technologies. We studied the infant cries with a musical perspective by employing SEN and SC features and their combination with cepstral analysis. These feature sets were classified using KNN and SVM classifiers that were tuned specifically for each of the feature sets and datasets by the BHPO methods. We also introduced a FE feature selection framework for the first time in the study of pathologic infant cry signals. By using this method, we further simplified our NCDS design and removed nearly half of the redundant, low-impact and noise-affected features. The performance of our design was evaluated using two separate datasets of expiratory cries (EXP) and inspiratory cries (INSV) with various evaluation measures such as accuracy, F-score and MCC. The achieved results showed promising potential in every step of the study. Each stage of the design further improved the system's performance, at least in terms of one of the evaluation metrics. The best results in terms of accuracy and F-score measures were achieved by combining all the introduced features after FE selection for the INSV dataset with the SVM classifier, and these were 91.10% and 91.81%, respectively. These results also highlight the importance of INSV cries as potential biomarkers, which has been neglected in many infant cry studies. Finally, we concluded that the framework presented here has promising potential in studying and diagnosing sepsis in newborns all around the world as a non-invasive means, especially in areas that are facing challenges with a lack of experts and specialists.

CHAPTER 3

NEWBORN CRY-BASED DIAGNOSTIC SYSTEM TO DISTINGUISH BETWEEN SEPSIS AND RESPIRATORY DISTRESS SYNDROME USING COMBINED ACOUSTIC FEATURES

Zahra Khalilzad [1], Ahmad Hasasneh [2] and Chakib Tadj[1],

[1] Department of Electrical Engineering, École de Technologie Supérieur, Université du Québec, Montréal, QC H3C 1K3, Canada

[2] Department of Natural, Engineering and Technology Sciences, Arab American University, P. O. Box 24, Ramallah, Palestine

Paper published in *Journal of Diagnostics*, November 2022

3.1 Abstract

Crying is the only means of communication for a newborn baby with its surrounding environment, but it also provides significant information about the newborn's health, emotions, and needs. The cries of newborn babies have long been known as a biomarker for the diagnosis of pathologies. However, to the best of our knowledge, exploring the discrimination of two pathology groups by means of cry signals is unprecedented. Therefore, this study aimed to identify septic newborns with Neonatal Respiratory Distress Syndrome (RDS) by employing the Machine Learning (ML) methods of Multilayer Perceptron (MLP) and Support Vector Machine (SVM). Furthermore, the cry signal was analyzed from the following two different perspectives: 1) the musical perspective by studying the spectral feature set of Harmonic Ratio (HR), and 2) the speech processing perspective using the short-term feature set of Gammatone Frequency Cepstral Coefficients (GFCCs). In order to assess the role of employing features from both short-term and spectral modalities in distinguishing the two pathology groups, they were fused in one feature set named the combined features. The hyperparameters (HPs) of the implemented ML approaches were fine-tuned to fit each experiment. Finally, by normalizing and fusing the features originating from the two modalities, the overall performance of the proposed design was improved across all evaluation measures, achieving accuracies of 92.49% and 95.3% by the MLP and SVM classifiers, respectively. The MLP classifier was outperformed in terms of all evaluation measures presented in this study, except for the Area Under Curve of Receiver Operator Characteristics (AUC-ROC), which signifies the ability of the proposed design in class separation. The achieved results highlighted the role of combining features from different levels and modalities for a more powerful analysis of the cry signals, as well as including a neural network (NN)-based classifier. Consequently, attaining a 95.3% accuracy for the separation of two entangled pathology groups of RDS and sepsis elucidated the promising potential for further studies with larger datasets and more pathology groups.

Keywords: Cepstral Features; Sepsis; RDS; SVM; MLP

3.2 Introduction

According to the World Health Organization (WHO), millions of children die every year globally. It was also indicated that the majority of deaths among children occur under the age of one month; for example, in 2020, 2.4 million children died globally in the first month of their lives, adding up to 47% of all child deaths being under-five mortality, which was 40% in 1990 (World Health Organization, 2021). This shows that the neonatal mortality rate is increasing globally. The WHO also presented the main pathological causes that may lead to neonatal death, where 75% of neonatal deaths usually occurred during the first week of life. Some of these pathological causes included Neonatal Respiratory Distress Syndrome (RDS) and sepsis.

The reason behind the RDS is unknown; however, it is often associated with surfactant deficiencies (Edwards, Kotecha, & Kotecha, 2013). From 2016 to 2020, RDS was among Canada's leading causes of post-partum mortality, and nearly 100 newborns lost their lives due to this pathology during the mentioned years (Canada Statistic, 2022). Typically, the clinical diagnosis of RDS is carried out via a series of tests which include recording echocardiography, collecting blood samples, measuring the oxygen levels in the bloodstream through pulse oximetry, and chest and lung radiography (Warley & Gairdner, 1962). RDS is, thus, identified by breathing difficulty in a newborn and red or blue color of the face and lips and should be diagnosed at an early stage since it could lead to many developmental difficulties such as vision or hearing impairment, learning challenges, and mobility problems. However, it is worth mentioning that there is no determined test for diagnosing RDS or ruling out the possibility.

On the other hand, sepsis was among the top 10 pathological causes that led to the mortality of infants in Canada between 2016 and 2020; it took the lives of more than 185 newborns (Canada Statistic, 2022). In a general sense, sepsis is an infection that entails the blood and it may lead to, or be associated with, several other pathological conditions such as hypothermia, hypotension, or even RDS (Canada Statistic, 2022; Wynn & Wong, 2010). Neonatal sepsis is clinically diagnosed based on having at least two of the following symptoms: high or low heart rates, feeding problems, lethargy, fever, hypotonia, convulsion, hemodynamic abnormalities, and apnoea that lasts for more than 20s (Mayo Clinic, 2022). Therefore, the present clinical tests for diagnosing sepsis take time and have a moderate risk of producing false negative and false positive results. Consequently, it is of great significance to promptly identify this pathology in the newborn to start the treatment procedure before the onset of symptoms.

It can thus be seen that both pathologies require intrusive and in-depth clinical tests to be diagnosed accurately, and they are associated with high mortality and morbidity rates for newborns. Furthermore, it has been shown that sepsis and RDS are closely associated and entangled (Wynn & Wong, 2010), and sepsis is one of the main causes of RDS (Mayo Clinic, 2022). Therefore, studying and analyzing these two infant pathologies by the means of a simple, automated, and non-invasive tool, such as a newborn cry-based diagnostic system (NCDS), is preeminent and essential. This system can serve as a tool for early recognition and accurate diagnosis of these infants' pathologies, which greatly contributes to acquiring the necessary treatment for the infant before the onset of symptoms and, thus, preserving the infant's life. In addition to that, the distinction between these two pathologic groups (sepsis versus RDS) will be lucrative in demonstrating that the concept of distinguishing neonates with certain pathologies from other pathological infants is an auspicious goal.

Typically, infants communicate with those around them through crying; it is a combination of vocalization, coughing, choking, and interruption, which includes a diversity of prosodic and acoustic features at different levels (Ji et al., 2021). Recently, the analysis and understanding of infant crying signals have been receiving growing attention from researchers and data scientists, with the aim of diagnosing the infant's pathology in its early stages. In this respect, it has been shown that infant cries provide important acoustic parameters or characteristics that should be taken into consideration, studied, and analyzed while monitoring the first days of an infant's life (Ji et al., 2021; Kheddache & Tadj, 2019). Furthermore, the cry signals of unhealthy infants usually contain unique features or characteristics that differ from healthy ones (Ji et al., 2021). Consequently, pathological cry signal analysis and classification can be used as a valuable tool for predicting and recognizing neonatal diseases before the onset of the symptoms.

By using the cry signals, various audio feature categories can be computed and generated, including cepstral, prosodic, and spectrograms, that have been widely used and applied to different research related to music, speech, and environmental sounds. These categories have separately been used for the identification of pathologies in newborns, and few attempts studied the combination of these features for the same purpose. In this research work, we aim to combine two feature categories, namely the cepstral domain and the prosodic domain, and then employ the combined features for training the classifiers. The ultimate goal of this research is thus to investigate the capacity of machine learning methods to discriminate between the septic and RDS cries, by using the combined feature set of the prosodic and cepstral domains. The characterization of different pathological patterns using the audio features would enable the development of an early and accurate diagnostic system that aggregates various audio feature categories to assist the early identification of abnormal acoustic behavior and link it to the early signs of a specific infant pathology. To the best of our knowledge, the question of utilizing different audio domains with a hyper-tuned machine learning model to classify infant RDS cries from infant septic cries has not been considered yet.

The presented study was proposed to address three main challenges in the field of pathological cry analysis. Firstly, despite the wide range of valuable research proving that the newborns diagnosed with a pathology cry differently than healthy newborns, there is no study where the cry signals of two pathology groups are compared to the best of authors' knowledge. Secondly, there is an inadequate number of studies that target sepsis and RDS; more specifically, the studies that target cries associated with RDS as a single pathology group (as opposed to being a part of an entire "pathologic" group) are scarce and the few existing studies never obtained an accuracy of more than 75%. Third, low-income countries suffer the most from infant mortality rates, which is due to their lack of adequate monitoring equipment, low number of pediatricians, and lack of resources. Child mortality risks in low-income countries are 16 times higher than high-income countries (Unicef, 2019), which calls for designing non-complex, fast and efficient tools for early diagnosis.

This study is the first to answer the question of pathologic versus pathologic that aimed to take the existing methods and algorithms and design a simplistic, yet efficient system, that requires only the everyday commercial tools. Our design benefits from a unique dataset owing to multiple factors. Firstly, no well-defined procedure or specific conditions were imposed during data collection phase; the data were collected in maternity rooms, Neonatal Intensive Care Units (NICUs), etc., where noise of medical equipment and staff and newborn's guardians' chatter was also present. Second, the recording was carried out by a simple handheld recorder, which can be found even in deprived areas of the world where the newborn mortality rates are at its highest. Third of all, data collection does not necessitate even as much as simply touching the newborn which makes our design a truly non-invasive method.

Despite the ever-growing use of computationally expensive tools, and also the perspective where crying is thought of as a pre-speech signal, we employed conventional tools from different fields such as musical applications, non-speech audio analysis and processing. We fused and optimized them so that the final design remains simplistic yet achieves the compatible performance of the state-of-the-art methods. The combination of the prosodic domain and cepstral domain features, which could lead to a new feature set that takes advantage of each domain and thus improves the linear separation between the two pathologies, is considered here by combining GFCCs and HR feature sets for the first time in the study of diagnostic analysis of the cry signal.

The rest of the paper is organized as follows. The related work on infant pathologies classification techniques is discussed in Section 3.3, while Section 3.4 describes the proposed methodology, including a description of the dataset and participants, features extraction, and modeling, followed by a description of the different machine learning methods that have been tuned and applied to this classification problem. Section 3.5 presents and discusses the obtained results. Finally, Section 3.6 presents conclusions and outlines future work.

3.3 Related Work

In the early years of pathological infant cry signal analysis and classification, numerous artificial intelligence (AI) and machine learning (ML) techniques were proposed and developed. Researchers can find many research works on infant pathological cry analysis and classification in (Ji et al., 2021; Saraswathy, Hariharan, Yaacob, & Khairunizam, 2012). One can see that researchers continue to apply new machine learning methods to classify infant cry signals into normal and pathological records; for example, see the recent works in (Alaie et al., 2016; Patil, Patil, & Kachhi, 2022). However, some of the current research works include identifying pathologies such as hypo-acoustic (Hariharan, Sindhu, & Yaacob, 2012), asphyxia (Badreldine, Elbeheiry, Haroon, ElShehaby, & Marzook, 2018; Ji, Xiao, Basodi, & Pan, 2019; Zabidi, Yassin, et al., 2017), hypothyroidism (Zabidi, Khuan, Mansor, Yassin, & Sahak, 2010b), septic (Khalilzad, Kheddache, & Tadj, 2022; Matikolaie & Tadj, 2022), RDS (Matikolaie & Tadj, 2020), and autism spectrum disorder (ASD) (Wu, Zhang, Wu, Wu, & Niu, 2019); additionally the authors in (Hariharan et al., 2018; Kheddache & Tadj, 2019; Lahmiri, Tadj, Gargour, et al., 2021; Matikolaie et al., 2022) have investigated different infant pathologies. In particular, the asphyxiated infant crying signals have been identified using different ML methods, including a deep feedforward neural network (DFNN) model (Ji et al., 2019), a support vector machine (SVM) model (Badreldine et al., 2018), and a convolutional neural network (CNN) approach (Zabidi, Yassin, et al., 2017), and achieved accuracy rates of 96.74%, 98.5%, and 92.8%, respectively. In addition, hypothyroidism has been studied in (Zabidi et al., 2010b) using a Multilayer Perceptron (MLP) classifier, and achieved a classification accuracy of 88.12%. Two groups of authors investigated sepsis in newborns recently; the authors in (Khalilzad, Kheddache, et al., 2022; Matikolaie & Tadj, 2022) have developed a machine learning-based CDS for identifying septic newborns and reached an accuracy of 83.9% using majority voting, while the authors in (Khalilzad, Kheddache, et al., 2022) attained 89.99% using entropy-based features. Furthermore, ASD in (Wu et al., 2019) and RDS in (Matikolaie & Tadj, 2020) have been identified based on a SVM and reached accuracies of 96% and 73.8%, respectively.

Normal and hypo-acoustic infant cry signal classification has also been proposed in (Muthusamy Hariharan et al., 2012) using general regression Neural Networks (NNs) and reached 99% accuracy. Therefore, most of the existing NCDS models have mainly focused on investigating one pathology individually versus healthy cases. The authors in (Alaie et al., 2016; Hariharan et al., 2018; Kheddache & Tadj, 2019; Lahmiri, Tadj, Gargour, et al., 2021; Matikolaie et al., 2022) have proposed to classify different pathological types of infant cry signals, namely: normal, deaf, asphyxia, hungry, pain, jaundice, and premature from the healthy group. Moreover, their proposed model is based on a combination of wavelet packet-based features and an Improved Binary Dragonfly Optimization-based feature selection method, and they conducted several classification experiments of two-class and multi-class of crying signals and achieved promising results.

As mentioned before, different audio feature categories can be extracted from infant cry signals using the following domains: cepstral domain, prosodic domain, time domain, image domain, and wavelet domain (Ji et al., 2021). Each domain represents different aspects of the infant’s cry signal and they each present specific information and characteristics. Compared to the time domain features, which are more sensitive to the background noise, the cepstral domain features have been shown to be more robust in modelling characteristics and covering variations within infant crying signals (Ji et al., 2021). These frequency-domain features can be computed using different mathematical tools, including Mel-frequency cepstral coefficients (MFCCs), Linear Prediction Cepstral Coefficients (LPCCs), Bark Frequency Cepstral Coefficients (BFCCs), Gammatone Frequency Cepstral Coefficients (GFCCs), and Linear Frequency Cepstral Coefficients (LFCCs). Indeed, cepstral features have been widely used in the field of speech processing and recognition, and the most frequently used ones to identify infant pathologies are MFCCs, LPCCs, and LFCCs, which have shown better performance compared to time domain features. In particular, MFCCs are the most used and tested features to identify infant pathologies; for example, asphyxia in (Badreldine et al., 2018) and (Zabidi, Yassin, et al., 2017), and hypothyroidism in (Zabidi et al., 2010b), and achieved promising accuracies as presented above. Liu et al. also used MFCCs along with LPCCs and BFCCs and based on a NNs model to identify infant cry reasons and the results showed that BFCCs produced the best classification rate of 76.5% (Liu, Li, & Kuo, 2018). Furthermore, the authors in (MV Varsharani Bhagatpatil & VM Sardar, 2014; Jagtap, Kadbe, & Arotale, 2016), showed that LFCC performed better than MFCC in distinguishing high-frequency audio signals such as female voice and infant crying signals. On the other hand, GFCCs have been shown to be powerful descriptors in non-speech recognition tasks, such as emotion recognition (Jiang, Jia, & Shao, 2020; Liu, 2018), understanding the reason behind the crying of infants (Kulkarni, Umarani, Diwan, Korde, & Rege, 2021), and automatic speech recognition (Tamazin, Gouda, & Khedr, 2019). There is one recent study where authors employed Gammatone Cepstral Coefficients (GTCCs) that are based on the time-representation of the signal for identifying infants suffering from Hypoxic Ischemic Encephalopathy (HIE) based on their cry signal (Satar et al., 2022). It is noteworthy to highlight that our study employs the frequency-representation by extracting GFCCs since they have proved successful in audio recognition tasks (Shao, Jin, Wang, & Srinivasan, 2009).

Prosodic domain features, which include high-level information such as formants, intensity, duration, harmonicity, and unvoiced regions, also contribute in improving the discriminative ability between the crying signals and thus identifying the type of the infant cry signal; an example of this is the identification of asphyxia in (Ji et al., 2019). It has been shown that attaching these features together with frequency domain features contributes to extracting both physiological and physical information from acoustic signals (Ji et al., 2021). Furthermore, image domain features, such as the spectrogram which is a time-frequency image representation of an audio signal and includes both acoustic and prosodic information, can be used to distinguish between healthy and unhealthy infant cries. It has been widely shown that feeding spectrograms into machine learning algorithms also plays an important role in enhancing the classification of different infant crying signals (Chang & Tsai, 2019; Felipe et al., 2019; Ji, Basodi, Xiao, & Pan, 2020; Le, Kabir, Ji, Basodi, & Pan, 2019). It is, therefore, obvious that each domain contributes to the classification of infant crying signals, and thus the mechanism of generating a combined feature set that takes advantage of different domains deserves to be considered and investigated.

Several relevant recent research works have already shown promising enhancement with combined features to the problem of infant cry signals analysis (Huckvale, 2018; Ji et al., 2020; Ji et al., 2019; Ting, Choo, & Kamar, 2022). More specifically, Ji et al. showed that combining MFCC features with weighted prosodic features contributed in improving the classification rates of the asphyxiated infant cry signals using a deep learning approach (Ji et al., 2019). In addition, a combined NNs model that combines summative and temporal features was proposed for infant cry classification and outperformed the independently-trained temporal and summative networks (Huckvale, 2018). In addition to that, the authors in (Ji et al., 2020) have shown that using hybrid features of the prosodic, spectrogram, and waveform classified by a CNN model produces better infant sound classification rates for the two different datasets. Moreover, a more recent study has investigated the use of hybrid features of MFCC, Spectral Contrast, Chromagram, Mel-scaled Spectrogram, and Tonnetz based on CNN and DFNN learning models (Ting et al., 2022). The results have shown that deep learning models performed better with hybrid features compared to the use of single feature of MFCC. It was shown that combining DCNN with RBF-SVM was capable of achieving up to 88.89% accuracy in classifying infant cries based on the reason of crying (Vincent, Srinivasan, &

Chang, 2021). Incorporating deep learning networks and combining them has shown the potential for state-of-the-art performance. For example, Khatun et al. (Khatun et al., 2022) proposed a DCNN-LSTM classifier with self-attention model, which was capable of attaining an accuracy of 99.93% for human activity recognition purposes. In another study for classifying MRI brain tumor, authors implemented CNN with PCA in the feature extraction step and fed these features to different machine learning classification algorithms, which yielded a remarkable 99.76% accuracy (Aurna, Yousuf, Taher, Azad, & Moni, 2022).

To summarize, most of the existing models focus on analyzing infant cry signals to identify one pathology by using different machine learning techniques. To the best of our knowledge, no studies have addressed classifying RDS cries from sepsis cries using machine learning methods. Moreover, we noticed a lack of studies that give attention to the question of combining cepstral domain features and prosodic domain features to be used in classifying different infant pathologies. Therefore, finding the optimal combination of cepstral and prosodic domains, followed by a fine-tuned machine learning algorithm, remains an open question and needs further research investigations. Therefore, this paper proposes to use different machine learning techniques that use a combined feature set of cepstral and prosodic. The main contributions of this research work can, thus, be summarized as follows:

- Different machine learning techniques were used to classify RDS cries from sepsis cries. In this regard, all used ML techniques were fine-tuned to give the best classification rates. Our fundamental goal is to prove the concept that a NCDS can be built, starting with these two pathologies that are most common in newborns.

- It is the first demonstration that GFCC features, and HR descriptors can be combined and used to support the diagnosis of pathologies in newborns. In this regard, we show that combining the two feature sets played an important role in improving the classification results.

- An accuracy of 95.3% with 0.95, 0.95, and 0.95 precision, recall, and F-score, respectively, were obtained using a fine-tuned SVM to distinguish between RDS and sepsis cries.

3.4 Materials and Methods

À It is well known that extracting the most significant efficient features from given data plays an important role in simplifying subsequent tasks, such as the classification process, and thus leads to more accurate results. In this proposed work, we propose a combined feature set specifically for the classification of infant pathological cries. As shown in Figure 3.1, the workflow of the proposed model involves four main stages, which can be summarized as follows: (1) signal preprocessing and segmentation, (2) features extraction, selection, and modelling (3) machine learning model, and, finally, (4) pathological cry classification.

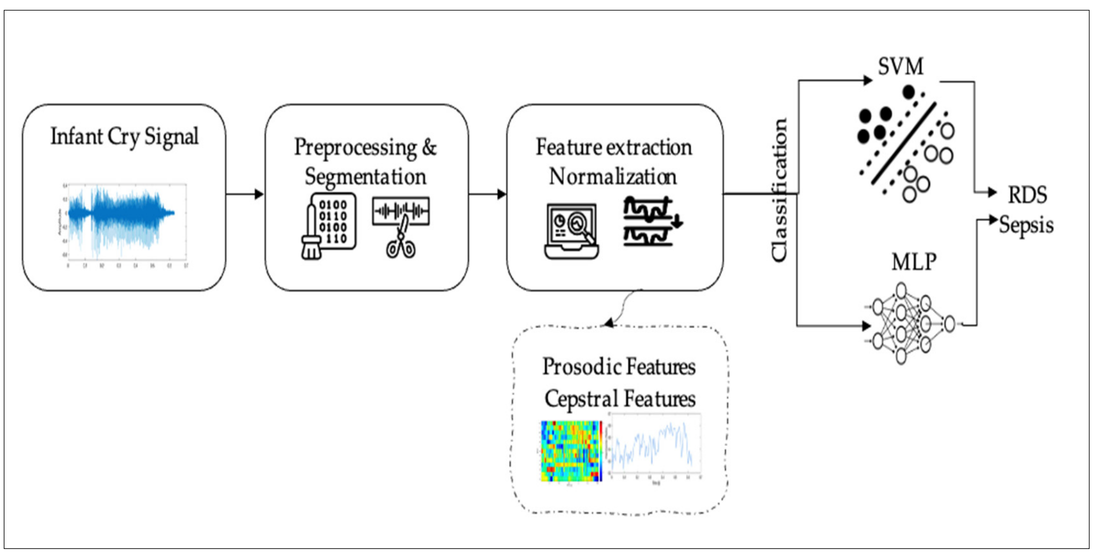

Figure 3.1 The workflow of the proposed model for different infant pathological classification and using the crying signals

3.4.1 Dataset Description

The samples included in this study were acquired as a result of collaboration between Saint-Justine children's hospital in Montreal, Canada and the Al-Raee and Al-Sahel hospitals in Lebanon. As explained in our previous works (Matikolaie et al., 2022), the cries in our dataset were collected from the newborns regardless of their race, gender, weight, or cry stimulus (pain, hunger, etc.). These cries have been collected with a common digital 2-channel Olympus handheld recorder with a 16-bit resolution and 44,100 Hz sampling frequency placed in the 10-to-30-cm vicinity of the newborn's mouth. The cries were recorded in the hospital environment including maternity rooms and NICUs with no well-defined procedure and in the presence of noise. The health status of the newborn was determined based on several screening tests performed after birth and the cry signals were labeled as healthy or with the diagnosed pathology group based on medical reports accordingly. The gestational age, race, reason of crying, babies age, weight, and APGAR score were all noted. These considerations make our dataset a real and comprehensive one that can study newborns and propose a real-world solution in designing newborn cry diagnostic systems. The age of the babies in this study ranged from 1 to 53 days old, since it is not until the end of the second month of life (53 days to be precise) when newborns gain control of the vocalizations they produce (Boukydis & Lester, 2012). Prior to this age, any vocalization is controlled by independent biological rhythms and thus it could be an indicator of newborn's health. Moreover, the restructuring of the supralaryngeal vocal tract takes place around 3 months of age (Boukydis & Lester, 2012). Therefore, this study excluded the newborns with a postnatal age more than 53 days.

It is well known that a majority of pathological studies encounter the same main obstacle, which is data acquisition. This challenge is attributable to several factors: 1) the unpredictability of whether a newborn with the targeted pathology groups will be observed during the data collection period, 2) acquiring the ethical and technical approvals to incorporate a cry sample in the database is a timely and difficult process which may result in losing some of the samples and 3) obtaining the newborns' guardians' consent to record their newborn's cry and then add it to the database is quite challenging.

Given all these obstacles, we tried to segment each recording to multiple expiration segments in order to overcome the data limitation challenge and better study the characteristics of pathological newborn cries. There was a total of 53 recordings from 17 newborns for sepsis, and 102 recording from 33 newborns for the RDS pathology groups. These recordings had an average of 90 s including silence, hiccups, inspiration cries, expiration cries, and background noise. The original newborn cries were recorded with different durations ranging from 1 to 4 min with an average of 90 s, obtaining up to 5 recordings per newborn, which was inadequate for classification purposes. As explained in data preprocessing section, multiple EXP segments were extracted from each recording, which were later treated as an individual sample; these formed the 2264 samples mentioned in Table 1 with an average length of 0.71 s for sepsis and 0.74 s for RDS.

In this study, the expiratory cries of newborns diagnosed with sepsis and RDS were included with 17 and 33 newborns in each group, respectively. In order to have a well-balanced and homogenous study, we selected the same number of samples from each pathology group, Table 3.1.

Finally, it is noteworthy to mention that despite the fact that RDS is mainly attributable to prematurity, term newborns are often misdiagnosed or not considered for RDS. Although the occurrence of RDS in term newborns is exiguous compared to the preterm newborns, several studies show that a notable number of term-born neonatal hospital admissions are still due to RDS every year (Qandalji, 2010; Qian, 2010), accounting to a total of around 8%. Another study showed that 43% of term-born respiratory failures are due to RDS, which is a serious alert to not rule out RDS in term neonates (Clark, 2005).

Table 3.1 The dataset description

	Septic	**RDS**
Gender	11 Males and 6 Females	10 Females and 23 Males
Average sample length	71 milliseconds	74 milliseconds
Babies Ages	1 to 53 days old	
Prematurity	Term	
Gestational age	38 ± 1 week	
Number of samples	2264 (1132 each)	
Origin	Canada, Haiti, Portugal, Syria, Lebanon, Algeria, Palestine, Bangladesh, Turkey	
Race	Caucasian, Arabic, Asian, Latino, African, Native Hawaiian, Quebec	
Reason of crying	Birth cry, hunger, dirty diaper, discomfort, needs to sleep, cold, pain	

3.4.2 Dataset Preprocessing

The cries of infants in our dataset have been processed by our previous colleagues in order to remove silence, filter, and segment each recording. Each recording was segmented and assigned with multiple labels. For example, the expiratory cries were marked as EXP, or the phonation during inspiration was labeled INSV, which represents a voiced inspiratory cry segment. These labels were attached by the means of WaveSurfer software. In the present study, we used the EXP segments of each cry recording and treated each segment as a sample. As has already been stated, one of the main challenges in any biomedical research is the limitation of data, especially in a problem such as this study, where the chances of observing a newborn suffering from a certain pathology are not predictable. Therefore, by segmenting each cry signal, we solved this challenge to a fair extent.

3.4.3 Features Extraction and Modelling

As mentioned before, the main focus of this study is the extraction and the study of feature sets that are capable of representing the differences in newborn cries associated with two entangled groups of pathologies, RDS and sepsis. The cry signal is non-stationary and dynamic, which calls for the study of both short-term and spectral features. Furthermore, it has been shown that although MFCCs are the most commonly used features owing to their high performance (Matikolaie & Tadj, 2020), GFCCs outperform them in terms of less computational costs and better performance (Valero & Alias, 2012). Thus, we studied GFCC features as short-term representations of the cry signal, as well as the harmonic factors that capture the spectral behaviour of resonance frequencies in newborn cries. We studied these features individually and then fused them to test the performance of the NCDS, considering both short-term and spectral features. The following sections expound on the procedure needed for the acquisition of these features.

Gammatone Frequency Cepstral Coefficients (GFCCs) are considered an alteration of the MFCC feature inspired by the biological model of the auditory system. GFCCs employ the equivalent rectangular bandwidth (ERB) bands instead of triangular bands and mimic the cochlear spectral structure in mapping the frequencies (Valero & Alias, 2012). The spectrogram representation of Gammatone-Frequency is called cochleagram. A cochleagram is expected to have fair performance with pathologic newborn cry signals since the lower frequencies can be studied with a far better resolution. This study combines the benefits of cochleagrams with Cepstral analysis. This is because, during the generation of a cry, the glottal impulses travel across the vocal tract, which then has a filtering effect on them (Wasz-Hockert, 1968). The Cepstrum facilitates distinguishing the source and the filter (Huang et al., 2001), which is desirable for identifying the region of the malfunctioning body organ. GFCCs have also shown promising potential in non-speech classification tasks such as emotion recognition (Shao et al., 2009). Regarding the computational costs associated with the extraction of GFCC features, it was shown that by cascading n 1st-order Gammatone (GT) filters, the nth-order GT filter could be well approximated. In order to attain GFCCs, the cry signal is first windowed into overlapping Hamming filters of 10 ms with 3 ms overlap length, since the performance of the feature extraction step is enhanced, and the non-stationarity of the signal could be neglected in such short frames. Next, in order to pre-emphasize the valuable signal frequencies, the signal passes the GT filters after a fast Fourier Transform (FFT) is applied. The final steps of extracting the GFCCs constitute employing the log function and then the DCT to decorrelate the compressed outputs of the previous steps. For a given frame k, the GFCCs can be computed through Equation 3.1:

$$GFCC_k = \sqrt{\frac{2}{N}} \sum_{n=1}^{N} GF[k] \cos\left(\frac{i\pi}{2N}\ (2c+1)\right) \qquad 1 \leq k \leq M, \tag{3.1}$$

where GF[k] denotes the loudness-compressed response of the Gammatone Filters (GF), and the number of filters is given by N.

Harmonic Ratio (HR) has been implemented as a powerful descriptor feature in many applications related to audio classification since it provides high accuracy (Heise, Miller, Wallace, & Galen, 2020). The newborn cry has the potential to be studied in terms of its musical aspects in addition to being treated as a pre-speech phenomenon owing to its harmonic components and rhythm and the differences in sound generator organs between newborns and adults (Kheddache & Tadj, 2015; Matikolaie et al., 2022). By definition, a sound is considered harmonic when a series of frequencies derived from the fundamental frequency as its multiples (called resonance frequencies) are observed in the sound (Chen, Gunduz, & Ozsu, 2006). Several researchers have revealed the presence of harmonics in the cry signals of newborns, and the study carried out by Kheddache et al. (Kheddache & Tadj, 2015) précised the harmonic behaviour (the behaviour of resonance frequencies) in pathologic cries, which showed different distributions and patterns among healthy and pathologic cries and among groups of pathologies. More specifically, they concluded that this behaviour depends on the pathology group. Based on these observations, this study evaluates the performance of HR as a potential biomarker for distinguishing between two pathologic groups of cries. HR determines the proportion of the energy of the harmonic segments of the cry signal to the total energy of the cry signal, and four statistical measures of mean, median, interquartile range, and standard deviation were computed based on HR in order to better represent the distribution of this feature across the spectrum of the signal (Chen et al., 2006).

Finally, it is worthwhile to discuss why we chose to fuse HR and GFCC feature sets in this study. We aimed to propose a simple yet effective design that considered both the short-term and spectral behavior of the cry signal. For this purpose, the HR was chosen because it could demonstrate the abnormalities in the cry signal with low computational costs and low feature dimensions, and in addition was shown to demonstrate a meaningful difference between infants diagnosed with RDS compared to other pathology groups. The GFCC feature set was also used as a more robust alternative to the MFCCs that are the most prevalent in the field of audio processing applications. It was shown in (Khalilzad, Kheddache, et al., 2022; Matikolaie & Tadj, 2022) that the combination of short-term and spectral features provide better classification performance for the study of RDS and sepsis. Furthermore, feature fusion was shown to enhance the performance of the diagnostic system designs for depression data (He & Cao, 2018) and artifact rejection in neuroimaging data (Hasasneh, Kampel, Sripad, Shah, & Dammers, 2018) by playing a significant role in enhancing the linear separability through constructing the apropos feature set. Thus, forming a feature vector that merges spectral and short-term and maintains simplicity, robustness, and low dimensionality, is advantageous and interesting to be explored. The feature sets were fused by the means of simple concatenation, and then normalized using a standard normalization. By implementing a fused feature set, we can expect a robust newborn pathology classification performance benefitting from a simpler classification process. Moreover, it would improve the linear separability of various pathology groups within the feature space. The individual feature sets of HR and GFCC were also normalized before being fed into the classifiers.

3.4.4 Machine Learning Classification and Tuning

In this study two classification methods were used, namely SVM and MLP, and both of them were chosen based on their common properties, which are simplicity and cost-effectiveness.

The SVM classifier is one of the prevailing algorithms when it comes to the infant cry applications, hence it is often employed as a baseline in many studies to highlight the role of other stages of the design, e.g., how successful the features are and to provide comparability to the classifiers and works of other researchers *(Badreldine et al., 2018; R. Sahak, W. Mansor, Y. Lee, A. Yassin, & A. Zabidi, 2010a; R. Sahak, W. Mansor, Y. Lee, A. M. Yassin, & A. Zabidi, 2010b)*. This is because the data in biomedical studies are often very limited and one of the main strengths of the SVM is the ability to efficiently construct complex decision boundaries from limited samples (Onu et al., 2017). Moreover, SVM is suitable for a portable and low-cost model design. The MLP classifier has a similar performance to the SVM, the samples are classified by constructing a complex decision boundary. MLP was successfully applied to several studies regarding asphyxia, which also involves the respiratory system *(Ali, Mansor, Lee, & Zabidi, 2012; Zabidi, Khuan, Mansor, Yassin, & Sahak, 2010a; Zabidi et al., 2011)*. Hence, it would be beneficial to investigate MLP in the diagnosis of RDS as well. Moreover, MLP is amongst the simplest NN classifiers. The application of MLP is lucrative to assessing the potential of more advanced NNs with more data in the future.

3.4.4.1 Support Vector Machine (SVM)

SVMs are among the most recognized classification methods implemented for the study of audio signals. Both linear and nonlinear classifications can be performed via SVMs, which are categorized as high precision supervised learning algorithms. The classification procedure of SVM consists of constructing a hyperplane that forms the farthest distance between the data points of different classes. For the case where the data points are not linearly separable, kernel functions are implemented. In this study, a Radial Basis Function (RBF) kernel was chosen, which presumes the neighboring points belong to a similar group and calculates the Euclidean distance between two given points in the feature space (Badreldine et al., 2018).

3.4.4.2 Multilayer Perceptron (MLP)

The general algorithm of a MLP consists of four steps: feeding the pattern to the network, feeding forward across the following layers, updating weights through a backpropagation method, and finally optimizing using an optimization function (Murtagh, 1991). MLP constructs a linear decision boundary for classification, and similar to SVM, a hyperplane is constructed so that the decision boundary has the minimum distance from misclassified points (James, Witten, Hastie, & Tibshirani, 2021). The Root Mean Square Propagation (RMSprop) was used as the optimization function that helps minimize this distance by tuning the backpropagation weights (Hinton, Srivastava, & Swersky, 2012). In order to evaluate the feasibility of employing neural networks for discriminating among groups of pathologies, a 7-layer MLP classifier was designed and proposed. Figure 3.2 shows the system design with the MLP classifier. With the use of HPO methods, the MLP was configured and tuned for each experiment. The input layer had the same number of neurons as the input feature vector (4, 13, and 17 neurons for HR, GFCC, and fused feature sets, respectively). Next, a 128 node fully connected layer was followed by a normalization layer and a hyperbolic tangent activation function. The activation function decided whether the neuron would fire. Next, another fully connected layer consisting of two nodes that corresponded to the number of output classes (Septic vs. RDS) was included. Finally, a sigmoid layer was used to convert the raw outputs of the previous layers into meaningful class probabilities between the range of [0, 1], and these probabilities were then fed to the classification layer where the decided label was produced. Training iterated with a learning rate of 0.001 through 120 epochs, and then validated by 15% of all the data, with 30% of the data randomly split for testing, and 55% of the data used for training.

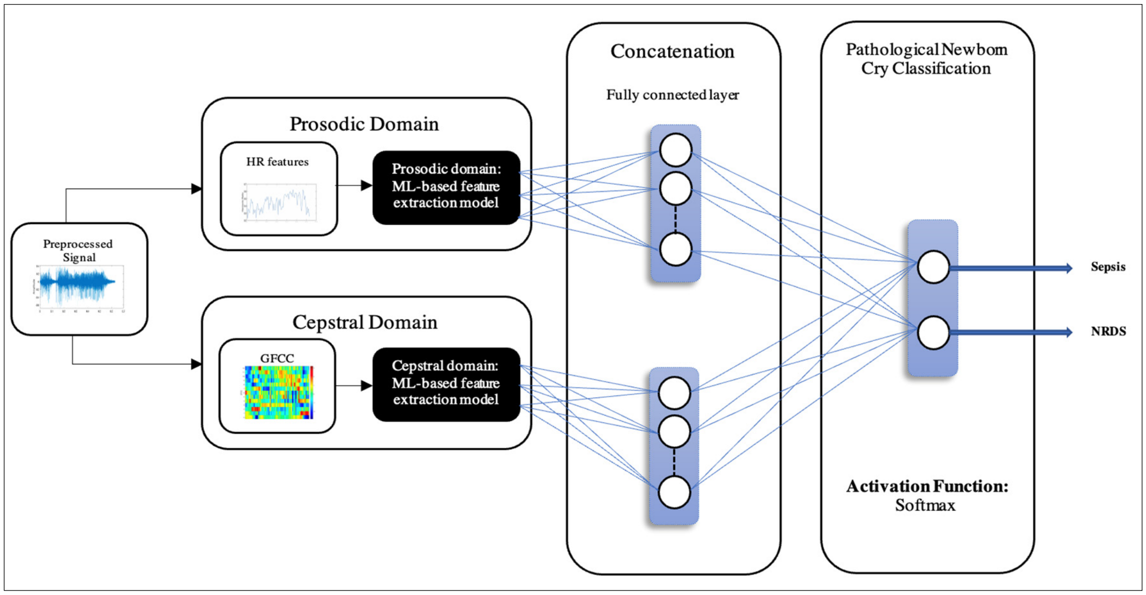

Figure 3.2 Block diagram for NCDS with MLP classifier

3.4.4.3 Hyper-parameter Fine-tuning and Evaluation Measures

Attaining desirable classification performance, as well as low error rates, is the goal and the main challenge of all classification problems; hence, the fine-tuning methods of HPs were introduced to serve this purpose.

Each experiment requires its own HP tuning since the feature matrix dimensions vary so that the classifier is tailored to fit the task. Furthermore, the HP fine-tuning methods replace human interference in determining classifier HP configuration, which includes random search, grid search, and Bayesian HPO approaches. In this study, the grid search method was used to fine-tune the classifiers' HPs, where it selected an optimum value for HPs from a limited set (Feurer & Hutter, 2019). The HPs selected for SVM fine-tuning were the γ and C, whereas the initial learn rate, L2 regularization, and the number of epochs were tuned for MLP.

In order to assess the ability of the proposed design in discriminating between the two groups of pathologies, several evaluation measures should be considered. Generally, the accuracy measure is the most prevalent measure in all systems, which is equal to the ratio of correct predictions to all the observations. The accuracy owes its prevalence to simplicity in calculation and understanding, but it is not informative in terms of class assessment and missed cases; therefore, other measures were introduced and studied. Table 3.2 presents a number of these measures used in this study (Khalilzad, Kheddache, et al., 2022).

Table 3.2 The evaluation measures and their formula

Evaluation Measure	Formula
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN} \times 100$
Sensitivity	$\frac{TP}{TP + FN} \times 100$
Precision	$\frac{TP}{TP + FP} \times 100$
F-score	$\frac{2TP}{2TP + FP + FN} \times 100$

3.5 Results and Discussion

This study targets the distinction between two entangled groups of pathologies in newborns for the first time in NCDS designs to the best of our knowledge. The aim of this study was to develop an early alert for the detection of sepsis and RDS, which are among the top newborn mortality causes around the world. Assessing the potential of analyzing acoustic features of the cry signal as a biomarker, through simple and accessible tools, was the priority of the proposed NCDS. Our dataset was recorded through a handheld recorder in the presence of noise with no prespecified conditions in maternity rooms and NICUs. Furthermore, newborns from different races, origins, genders, and various reasons of crying participated in our study which makes it comprehensive. Moreover, this study combined features that were conventional in musical applications of HR with the biologically inspired features used in speech-processing applications, and GFCCs that belonged to two levels of short-term and spectral. Additionally, with the help of HP fine-tuning, the classifiers were tailored to fit each of the presented experiments.

Various audio recognition, speech, and music processing systems benefit from sophisticated and complex deep-learning models, whereas in biomedical applications, the use of these designs depend on data availability. Data acquisition and collection are among the most significant challenges in biomedical research; when it comes to observing certain pathological groups, the probability is not deterministic in any given period of time. There is no way of knowing whether the newborns admitted to a hospital on a certain date would be diagnosed with the pathology groups subject to research. Nevertheless, obtaining the ethical and technical requirements to include data from any participant adds to the challenge of data acquisition. Therefore, this study benefits from SVM as a desirable and successful approach in NCDS designs and explores the use of a MLP neural network in order to assess the further potential for using other NN models in future works.

As mentioned in previous sections, the NCDS was designed and analyzed with the EXP dataset. The MLP and SVM classification approaches were used to identify septic newborns from RDS, and the feature sets were employed individually and also after their fusion. In order to fuse the features a simple concatenation followed by standard normalization was performed, so that the performance of the feature set implementing both modalities (short-term and spectral) would be compared to the individual feature sets. Furthermore, the classifiers were fine-tuned using the grid search hyperparameter (HP) optimization. In this case, γ and C were tuned for the SVM classifier, while the HPs of L2 regularization, initial learn rate, and number of Epochs were optimized for the MLP. In order to fully investigate the potential of HP fine-tuning, the range for each HP was determined for the optimization process, Table 3.3. Elaborating the reasons behind choosing which HPs were tuned in this study would be of essence.

Table 3.3 The pre-defined ranges for HP fine-tuning

Classifier	Parameter	Selected Range	Value Type
MLP	Initial learning rate	[0.0001, 1]	Logarithmic
	L2 Regularization	[0.0001, 0.001]	Continuous
	Number of Epochs	[50, 200]	Integer
SVM	γ	[0.1, 0.25, 0.26, 0.3, 0.5]	Categorical
	C	[0.5, 1, 2, 4, 5]	Categorical

Initial Learning Rate is the most significant HP to tune in neural networks. Following each iteration of estimating the error yielded after updating the weights, the learning rate determines how much of an adjustment the model requires.

Selecting the optimal learning rate is a trade-off between computational time and finding the optimal solution. Larger learning rates lead to the faster convergence of the model to the suboptimal solution, whereas a small learning rate calls for a higher number of epochs. Therefore, we should tune the number of epochs as well (Goodfellow, Bengio, & Courville, 2016).

Number of Epochs determines the number of changes in the weights of the network; increasing and decreasing the number of epochs may lead to the underfitting and overfitting of the model. Therefore, while tuning other HPs of the network, it is important to select the optimal number of epochs correspondingly. The optimal selection of the number of the epochs allows for the termination of the training process before the elevation of the validation error (Liu, Starzyk, & Zhu, 2008).

L2 Regularization: In order to prevent machine learning techniques from encountering overfitting, regularization methods were introduced (Nusrat & Jang, 2018) so that by adding a penalty factor to the large weights, the complexity of the overall design was reduced. L2 regularization is amongst the most prevalent methods of regularization. The value of regularization HP should be selected in such a way that both overfitting (associated with small regularization value) and underfitting (associated with large regularization value) are prevented (Han, Gondro, Reid, & Steibel, 2021).

As for the SVM classifier, both γ and C should be tuned. A higher value of the C would prioritize decreasing the support vectors count due to the fact that they each add to the optimization costs, while lower values of C lead to a higher support vector count and thus, larger margins. The γ HP determines the simplicity of a SVM model; higher values correspond to a curvier decision plane, which closely follows the data, whereas a small γ means a simpler model with flatter decision plane. γ in fact signifies the speed of lowering the domination of each point as the distance grows (Wainer & Fonseca, 2021).

We conducted three experiments to evaluate the system performance, the role of fused features, and the role of each feature set. Table 3.4 Table 3.6 present the results of the evaluation of the proposed design based on these experiments.

The results for the evaluation of the HR feature set are presented in Table 3.4. The HR feature set proved to be a successful feature in the analysis of the cry signal, since with only 4 elements, the NCDS could yield a 71.03% accuracy. However, the MLP classifier did not converge for the HR feature set. This result was unsurprising since this feature set has a low dimensionality of only 4 elements. Therefore, increasing the number of features could solve this challenge, as presented in Table 3.6.

Moreover, this feature set could also obtain fair performance in terms of recall and precision. The recall measure is of great significance in exploring the pathologies, since it demonstrates the share of true septic (or RDS) cases among all the samples. Precision shows the probability that NCDS will predict a septic (or RDS) case correctly. These two measures owe their importance to the fact that true diagnosis and timely treatment of the pathology have a considerable effect on the survival chances of the newborn.

Table 3.4 The results for the evaluation of the HR feature set

Feature Set	Classifier	Accuracy	Precision	Recall	F1-Score
HR	SVM	71.03%	0.71	0.71	0.71
	MLP	N/A	N/A	N/A	N/A

The GFCC feature set remarkably attained a high performance as an individual feature set with both classification methods, Table 3.5. Increasing the number of features resulted in the convergence of the MLP classifier as expected; however, the SVM outperformed MLP across all evaluation measures. It can also be seen that the performance of the NCDS with GFCC feature set was superior to the HR feature set by more than 10% in accuracy. Figure 3.3 Heatmap for the SVM classifier using the HR feature set depicts a more detailed look at the results of identifying the septic and RDS cases via HR feature set through presenting the heatmap for the SVM classifier.

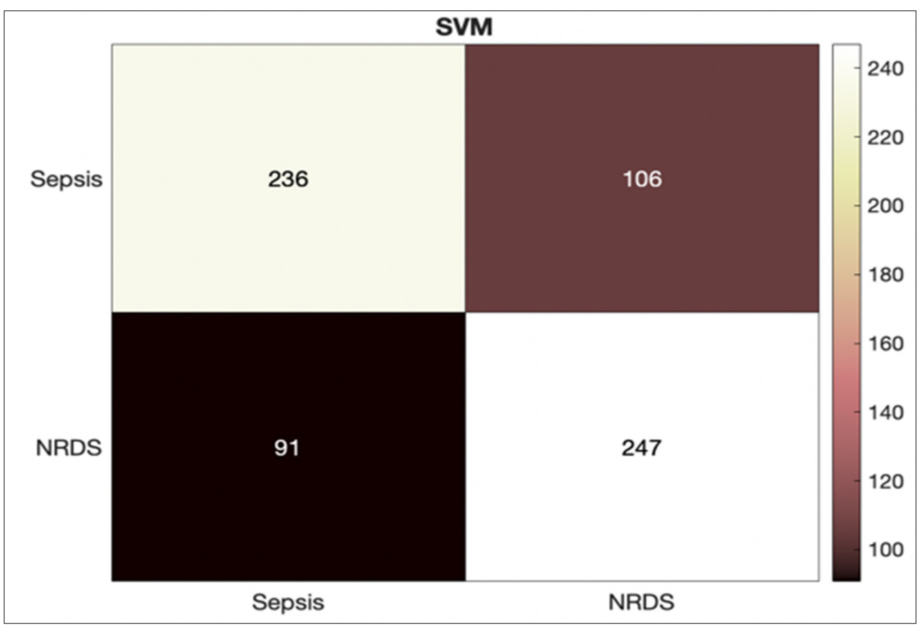

Figure 3.3 Heatmap for the SVM classifier using the HR feature set

Table 3.5 The results for the evaluation of the GFCC feature set

Feature Set	Classifier	Accuracy	Precision	Recall	F1-Score
GFCC	SVM	92.94%	0.93	0.93	0.93
	MLP	88.51%	0.88	0.89	0.89

In the final experiment, we fused the previous features to assess the performance of NCDS in discriminating between RDS and septic newborns, Table 3.6. The addition of the HR features resulted in an enhancement of more than 2% across all the evaluation measures for both classifiers compared to the GFCC feature set. These results are promising due to two main points: 1) improving the performance where the results are already at more than 90% would be difficult, and our design gained more than 2% enhancement. 2) this enhancement is consistent across all the evaluation measures investigated. Similar to the GFCC feature set, the SVM transcended the MLP throughout the evaluation measures.

Similar to the HR feature set, the detailed heatmaps for the GFCC and combined feature sets using each of the classifiers are presented in Figure 3.4 - Figure 3.5 , respectively. These heatmaps show how the data are distributed across the classes and provide a deeper look into the predictions made by the NCDS.

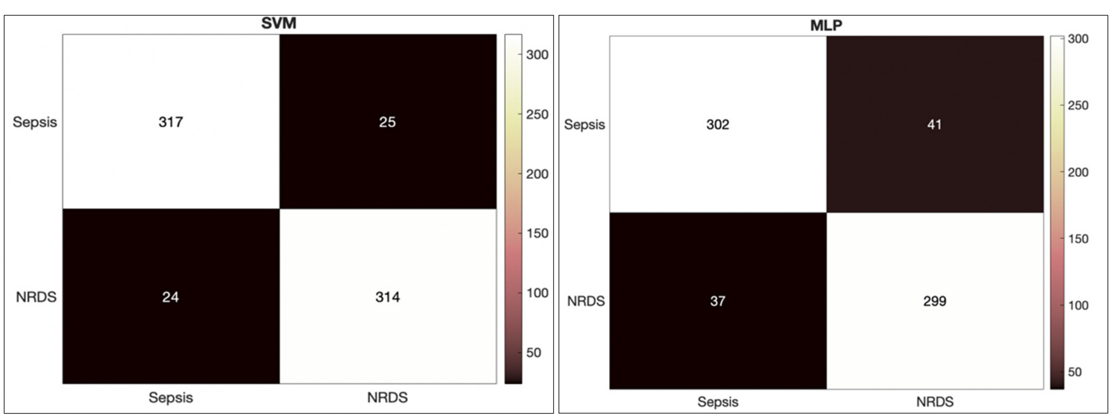

Figure 3.4 Heatmaps for the SVM and MLP classifiers using the GFCC feature set

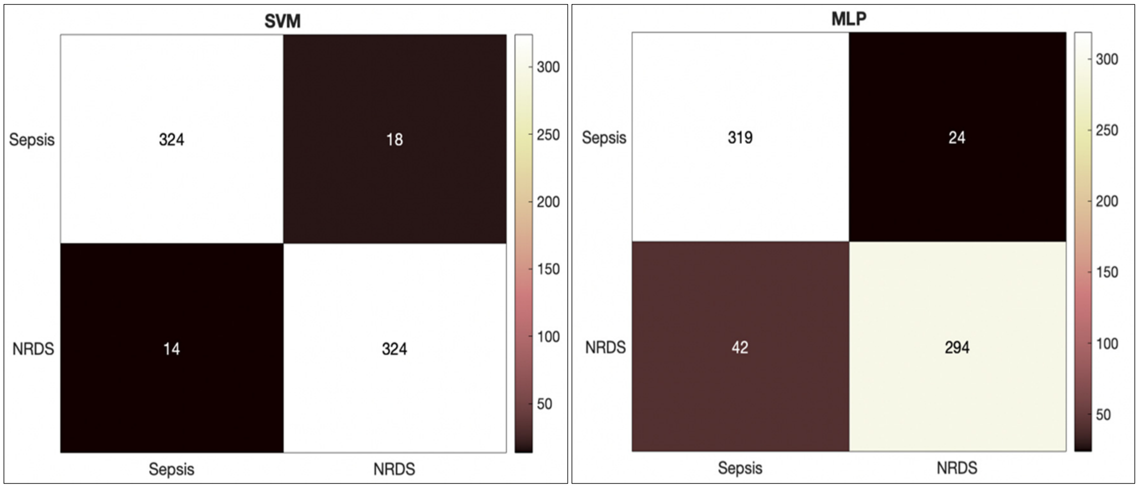

Figure 3.5 Heatmaps for the SVM and MLP classifiers using the combined feature set

Table 3.6 The results for the evaluation of the combined feature set

Feature Set	Classifier	Accuracy	Precision	Recall	F1-Score
GFCC + HR	SVM	**95.29%**	**0.95**	**0.95**	**0.95**
	MLP	92.49%	0.92	0.92	0.92

Finally, comparison of the Area Under Curve (AUC) of the Receiver Operator Characteristic (ROC) for the experiments in this study would help further assess the performance of different architectures. Figure 3.6 shows the ROC curves for the SVM classifier. The ROC curve shows the true positive rate (TPR) on the vertical axis and the false positive rate (FPR) on the horizontal axis. FPR is also an important measure, since it represents the probability of a false alert. The area under curve (AUCs) of ROCs is an indicator of model performance which will be discussed later in this section.

As can be seen through all evaluation measures, the fused feature set achieved the highest results with both classifiers. The study of the AUC is salient in terms of statistical analysis, since it demonstrates the probability of ranking any positive sample is higher than any negative sample, the same as Wilcoxon test of ranks (Hanley & McNeil, 1982) in order to compare the classifiers; the ROC curves are summarized in a single scalar, the AUC. The AUC is always between 0 and 1 since it is defined as a share of the area of the unit square (Fawcett, 2006). Any practical and acceptable classifier should have an AUC of more than 0.5 since the random guessing is equal to the diagonal line in the ROC curve that crosses (0, 0) and (1, 1); the closer values of AUC to 1 translate to better performance of the classifier. In other words, the AUC signifies the ability of the system in distinguishing between the two classes which is the main goal of this study (Bradley, 1997).

Two main goals were introduced for this study: 1.) finding the optimal feature set and study the effect of combining spectral and cepstral features. 2) finding the best classification algorithm that fits our problem/challenge.

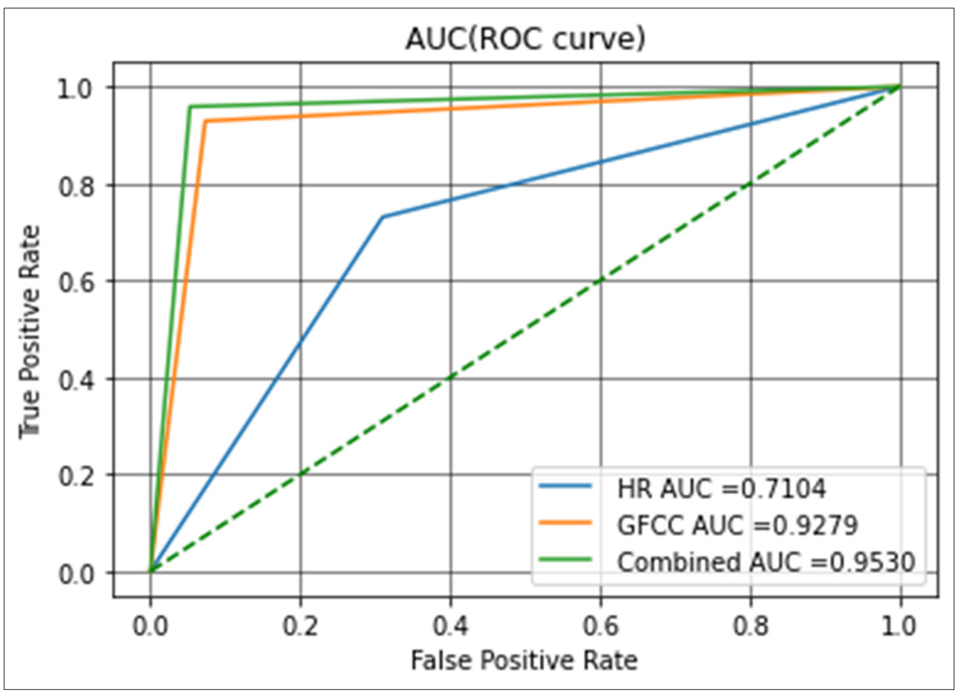

Figure 3.6 AUC-ROC for the SVM classifier using each feature set

Through comparing the AUCs resulting from analyzing the introduced feature sets in this study with the SVM classifier, the role of feature fusion in studying the pathologic infant cries became clear. It is shown that implementation and combination of different modalities can enhance the performance of the system, thus achieving the first goal. Concerning the second goal, it was shown that the MLP classifier was outperformed in terms of evaluation measures for all feature sets; therefore, as a final discussion point, we compared the AUC-ROC of the best feature sets of the SVM and MLP classifiers. Figure 3.7 illustrates the ROC curve for the MLP classifier; as can be seen from Figure 3.6 and Figure 3.7, the MLP showed better performance in terms of the AUC measure. This is an interesting result since it suggests two points: 1) the study of the ROC curve is essential for analyzing the binary classification problems since the evaluation measures might not describe all the aspects. 2) the MLP classifier shows great potential in studying the pathological infant cry signals since it has better performance in the separation of the two classes and should be considered for future studies. Finally, it can be seen that the superiority of the combined feature set is consistent across both classifiers as the MLP classifier also has a 0.17 increase in the AUC by implementing a combination of the features.

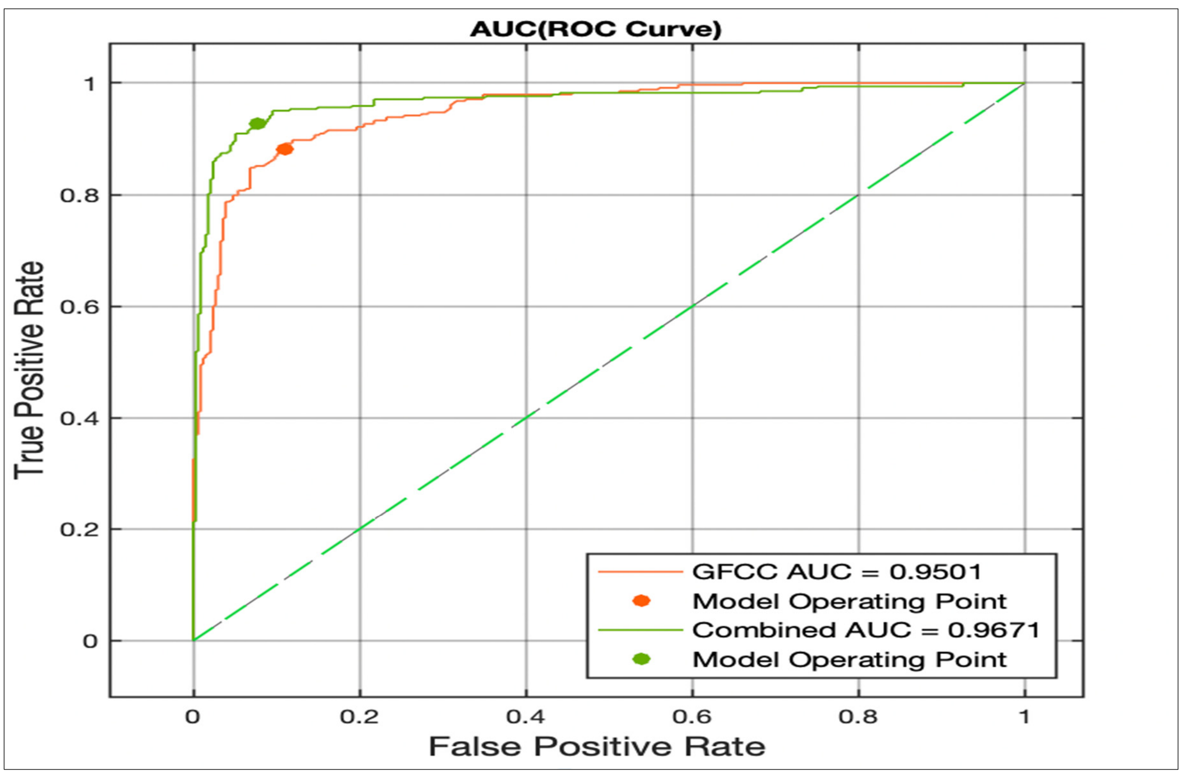

Figure 3.7 AUC-ROC for the MLP classifier using the GFCC and combined feature sets

There are few studies analyzing newborn cry signals to diagnose sepsis. Recently, two groups of researchers studied sepsis based on processing the newborn cry signals; however, they both focus on detecting septic newborns from the healthy group, whereas this study aims to target distinguishing between two pathological groups for the first time. The study presented by Matikolaie et al. (Matikolaie & Tadj, 2022) investigated the role of prosodical characterization of the cry signal in detecting sepsis which accomplished 86% as their best F-score. Furthermore, Khalilzad et al. (Khalilzad, Kheddache, et al., 2022) explored the potential of a NCDS in diagnosing sepsis by incorporating entropy-based features and fuzzy entropy feature selection, which attained 89.70% as their best F-score for the expiration cry segments. We believed that with sepsis being one of the globally leading post-partum mortality causes, there is a need for more in-depth studies that probe other perspectives of this pathology. Hence, this study could be complementary to the previous studies to give another means and modality of studying sepsis by comparing it to another cognate pathology.

The respiratory distress syndrome (RDS) suffers from a similar research gap; the existing literature on processing RDS cries is scarce. There are few studies target studying RDS as a single pathology group; Matikolaie et al. (Matikolaie & Tadj, 2020) proposed a NCDS to detect newborns suffering from RDS from the healthy and obtained 73.80% accuracy. Chittora et al. (Chittora & Patil, 2016) presented a spectrographic comparison of the RDS cries, where a double harmonic break was presented, suggesting that resonant study of the cry signal would be helpful in analyzing the RDS cries. Moreover, Lederman et al. (Lederman et al., 2002) classified the preterm infants suffering from RDS from healthy preterm infants and achieved a 63% accuracy using hidden Markov models. Finally, Alaie et al. (Alaie et al., 2016) obtained 69.59% accuracy by GMMs using the boosting mixture learning method for the detection of infants diagnosed with RDS; in another experiment, they formed a subset of pathological newborns suffering from multiple pathologies such as RDS, heart problems, blood abnormality and neurological disorders as a single pathological group to be detected from healthy newborns and gained an accuracy of 85.21%. As mentioned above, all discussed research focused on the identification of RDS/Sepsis from healthy; however, to the best of our knowledge there is no prior work on distinguishing between two (or more) pathology groups. Nevertheless, despite the entangled nature of the two pathologies studied here, our design was able to outperform all of the previous studied on sepsis and RDS cry signals by achieving 95.3% for accuracy. Similar to any other study in this field, this study also faced multiple challenges. Although we attempted to study the cry signals regardless of race, origin, and other factors such as cry stimuli, the designed NCDS has room to be further developed with more data. Furthermore, employing explainable AI, such as LIME, might help to better analyze the contribution of different features to the final result; thus, it will be considered in our future works.

This study had several achievements; it provided a proof for the concept of distinguishing between different pathology groups based on only cry signals, as well as further highlighting the benefit of combining features from different levels. Furthermore, by using proper feature manipulation, normalization, and HP fine-tuning, our machine learning design was able to achieve results similar to the more complex and resource expensive methods in the literature by attaining an accuracy and F-score of up to 95%. The high values of recall demonstrate the success of our design in detection of the true pathology group.

3.6 Conclusion and Future Work

This paper aimed to investigate RDS and sepsis as two of the pathologies associated with high mortality rates of neonates across the world through machine learning-based methods. These two pathology groups require in-depth and extensive clinical tests to be diagnosed, which calls for the development of a non-invasive tool such as the one suggested in this study. The novelty of the proposed design lies in removing the need for any extreme data collection or analysis tools by employing a commercial handheld recorder for data acquisition with no well-defined conditions, as well as using conventional machine learning techniques and combining them in such a way that the performance of the system is comparable to the highly complex and recent methods. This study proposed an early alert for detecting and discriminating two entangled groups of pathologies for the caregivers of the newborn and the medical staff in deprived areas of the world suffering from high newborn mortality rates.

The classifiers in this study were tuned for each experiment and all the feature sets were normalized before being fed to the classifiers. The cry signals were studied from a musical perspective through the HR feature set and from a speech processing aspect by means of the GFCC feature sets. Moreover, these features were from two different levels that also investigated the short-term and the spectral behaviour of the cries. The combination of these two feature sets improved the overall performance of the system, and the final accuracy and F-score were as high as 95%.

In this research work, we have noticed that training deep learning approaches requires a large size of diverse samples of infant pathologies. Therefore, increasing the number of samples is desirable for introducing deep learning models. Instead of using GFCC features modality only, we have also seen that combining HR features and GFCC features has positively contributed to improving the classification rates by 2.35% and 2.21% using SVM and MLP, respectively. Nonetheless, integrating other features from other domains that improve the linear separation ability will be further investigated in our future works. As mentioned before, it has been shown that extracting spectrogram features includes important information or characteristics in classifying infant crying signals (Chang & Tsai, 2019; Felipe et al., 2019; Ji et al., 2020; Le et al., 2019). Combining spectrogram features along with the prosodic and cepstral features will be one of our future works. Whether the features will be fused prior to training or within the learning process is also an open question.

Our next work will be based on proposing a multimodal fused model for the diagnosis of different infant pathologies leading to an accurate NCDS. This will include increasing the dataset by introducing new pathology types, extracting more robust features from different domains, fusing them with appropriate ratios, and then generating a new combined feature set that improves the discrimination ability. The feature analysis will be based on more sophisticated techniques, such as deep learning approaches. Therefore, studying and finding novel deep learning architectures, such as CNN and DFNN, with the use of combined features will also be considered.

CHAPTER 4

USING CCA-FUSED CEPSTRAL FEATURES IN A DEEP LEARNING-BASED CRY DIAGNOSTIC SYSTEM FOR DETECTING AN ENSEMBLE OF PATHOLOGIES IN NEWBORNS

Zahra Khalilzad and Chakib Tadj

Department of Electrical Engineering, École de Technologie Supérieur, Université du Québec, Montréal, QC H3C 1K3, Canada

Paper published in *Journal of Diagnostics*, February 2023

4.1 Abstract

Crying is one of the means of communication for a newborn. Newborn cry signals convey precious information about the newborn's health condition and their emotions. In this study, cry signals of healthy and pathologic newborns were analyzed for the purpose of developing an automatic, non-invasive, and comprehensive Newborn Cry Diagnostic System (NCDS) that identifies pathologic newborns from healthy infants. For this purpose, Mel-frequency Cepstral Coefficients (MFCC) and Gammatone Frequency Cepstral Coefficients (GFCC) were extracted as features. These feature sets were also combined and fused through Canonical Correlation Analysis (CCA), which provides a novel manipulation of the features that have not yet been explored in the literature on NCDS designs, to the best of our knowledge. All the mentioned feature sets were fed to the Support Vector Machine (SVM) and Long Short-term Memory (LSTM). Furthermore, two Hyperparameter optimization methods, Bayesian and grid search, were examined to enhance the system's performance. The performance of our proposed NCDS was evaluated with two different datasets of inspiratory and expiratory cries. The CCA fusion feature set using the LSTM classifier accomplished the best F-score in the study, with 99.86% for the inspiratory cry dataset. The best F-score regarding the expiratory cry dataset, 99.44%, belonged to the GFCC feature set employing the LSTM classifier. These experiments suggest the high potential and value of using the newborn cry signals in the detection of pathologies. The framework proposed in this study can be implemented as an early diagnostic tool for clinical studies and help in the identification of pathologic newborns.

Keywords: Newborn cry; Gammatone Frequency Cepstral Coefficients; Support Vector Machine; Long Short-term Memory; Hyperparameter optimization; Feature fusion

4.2 Introduction

In 2019, 6700 neonatal deaths occurred every day, and around 75% of these deaths occurred within the first 7 days after birth; this highlights the significance of expeditious diagnosis during the first few days of any neonate's life. Several pathologies associated with a neonate's mortality require invasive clinical tests and a high vigilance. Unfortunately, the regions that suffer the most from high newborn mortality rates are those deficient in the number of skilled health professionals. The World Health Organization (WHO) states that two-thirds of newborn deaths could be prevented if diagnosis and treatments took place before the second week of an infant's life. Furthermore, in most cases of pathological studies, if the treatment is initiated expeditiously, the infant may completely heal if given the right treatments (World Health Organization, 2014).

As early as the 19th century, the cry of neonates was recognized as a cue in identifying morbidity (Bell, 1878). The acoustic characteristics of a cry may vary due to various factors such as air pressure, tension, length, thickness, and shape of the vocal cords and resonators (Agrawal, 1990). Experienced parents and caregivers may distinguish types of cries only by listening; however, even trained nurses could only reach an accuracy of around 33% by relying on their auditory system (Mukhopadhyay et al., 2013). Healthy newborns have a fundamental frequency of 400–600 Hz, with an average of 450 Hz (Sulpizio et al., 2019); they also show a decreasing or increasing–decreasing melody shape with super imposed harmonics, and an average duration of 1–1.5 s (Robb & Goberman, 1997). The cries of babies suffering from a specific pathology are associated with low punctuation; they reflect high irritability and the physiological persistency is low (Corwin et al., 1996). Some of the features and attributes in infant cry signals can seldom be observed in healthy infants, though are commonly seen in pathologic ones (Michelsson et al., 1982). For example, hypothyroidism could result in low-pitched cries, a lower number of shifts, and a frequent observance of the glottal roll at the end of phonation. Cries marked with hypothyroidism have been marked as hoarse (Vuorenkoski et al., 1973). This acoustic structure has enabled us to develop a Newborn Cry Diagnostic System (NCDS) and take a deeper look into the health status of neonates.

The study of newborn cry signals unveiled that they bear abundant helpful information about the neonate's health conditions. Extensive research in this area has demanded an automatic approach and accurate analysis of the cry spectrographs; hence, newborn cry analysis systems were designed to overcome this challenge (Abou-Abbas, Tadj, & Fersaie, 2017; Farsaie Alaie & Tadj, 2012; Kheddache & Tadj, 2013a, 2013c; Matikolaie & Tadj, 2020; Messaoud & Tadj, 2011). The study of newborn cry signals has multiple goals.

There are many interesting publications in the literature that analyze cry signals from aspects other than those used in this study. These studies range from the identification of the reason for crying, e.g., hunger, pain or boredom (Bano & RaviKumar, 2015; Cohen et al., 2020; Parga et al., 2020); emotion detection (Kulkarni et al., 2021); detecting the cry in Neonatal Intensive Care Units (NICUs) and in surveillance systems (Kim et al., 2013; Torres et al., 2017); segmenting the cry signal into its episodes (Abou-Abbas et al., 2015; Aucouturier et al., 2011); diagnosis of specific pathologies (Khalilzad, Kheddache, et al., 2022) or general identification of a pathologic infant (Kheddache & Tadj, 2019; Orlandi, Garcia, Bandini, Donzelli, & Manfredi, 2016; Rosales-Pérez et al., 2015; Zabidi et al., 2010a), as well as studying how each factor would affect the cry characteristics. Some of these works have explored the roles of pain intensity (Bellieni et al., 2004; Maitre et al., 2017; Mijović et al., 2010), gender (Reby et al., 2016), gestational age (Wasz-Hockert et al., 1963), and other similar factors in cry signals. This study focuses on a different type of application, which is diagnosing pathologies in newborns based on their cry signal. What this study tried to achieve was to exploit features that could reflect the alterations in the cry signal only as a result of being unhealthy and independent of other factors. We expected these features (and their fusion) to represent attributes in the cry signals that were not obvious in simple observations of spectrograms, and also were not affected by changes in etiological factors across newborns and the emotional state of the newborns.

Every NCDS comprises three principal stages: pre-processing, feature extraction, and classification. In the pre-processing stage, the cry signal is pre-emphasized and framed; the pauses and silences are removed, filtered, and segmented to be ready for feature extraction. Following pre-processing is the feature extraction step. The features that are capable of discriminating the healthy cry signals from the pathologic ones are exploited in this stage. These features pass through dimensionality reduction techniques and are then fed as inputs into the classifier in the last stage of the NCDS. Finally, the class labels, which were predicted by the classifier, constitute the result.

The prominent features in the analysis of newborn cry signals include the Mel-frequency Cepstral Coefficients (MFCC), owing to their good performance in the diagnostic studies of cries. MFCCs are often employed as the baseline in many experiments concerning the neonate cry. The MFCC features aid the detection of multiple diseases, such as hypothyroidism, asphyxia (Wahid et al., 2016; Zabidi, Mansor, et al., 2017), hyperbilirubinemia (Kheddache & Tadj, 2019), respiratory distress syndrome (Matikolaie & Tadj, 2020), sepsis (Khalilzad, Kheddache, et al., 2022; Matikolaie & Tadj, 2022), and cleft palate (Massengill Jr, 1969).

Gammatone Frequency features (GFCC) have been employed for the purpose of emotion recognition in the study of newborn cry signals (Kulkarni et al., 2021), where they have outperformed MFCCs. GFCCs have a wide range of applications in acoustic scene classification problems, the recognition of emotions in adult speech (Garg & Bahl, 2014), and speaker identification (Admuthe & Patil). GFCCs were also employed in recent research identifying septic newborns from those diagnosed with RDS based on their cry, which proved to be successful (Khalilzad, Hasasneh, & Tadj, 2022). Among the machine learning architectures used in infant cry analysis, Support Vector Machines (SVM) is one of the most prevalent approaches. A diversity of features such as temporal, prosodic, and cepstral have functioned successfully with SVMs (Badreldine et al., 2018; Sahak et al., 2010a; Sahak et al., 2010b). Onu et al. (Onu et al., 2017) concluded that SVMs have a practical design for limited samples and data with high dimensionality, and are the most suitable for the study of asphyxiated neonates. Another classification approach employed in this work was the Long Short-term Memory (LSTM) neural network. LSTMs have been successfully paired with MFCC, GFCC, and their fusions; they showed promising performance in emotion and gender recognition applications (Kumaran, Radha Rammohan, Nagarajan, & Prathik, 2021; Verma, Agrawal, Singh, & Ansari, 2022). However, their application has been limited in NCDS designs thus far (Lahmiri, Tadj, Gargour, & Bekiros, 2022). LSTMs are one of the best choices when it comes to sequential data, such as audio signals. Nevertheless, like any other deep learning framework, LSTMs encounter the challenge of fine-tuning hyperparameters (HP) (Diaz, Fokoue-Nkoutche, Nannicini, & Samulowitz, 2017; Reimers & Gurevych, 2017). HP tuning can enhance the performance of a Neural Network (NN) from medium to state-of-the-art. Although many researchers emphasized the vital role of Hyperparameter Optimization (HPO) in NN architectures, only a few works have been published that suggest which and how many HPs should be optimized (Bergstra & Bengio, 2012; Gorgolis, Hatzilygeroudis, Istenes, & Gyenne, 2019; Nakisa, Rastgoo, Rakotonirainy, Maire, & Chandran, 2018).

This study aimed to develop a comprehensive NCDS to distinguish between healthy and morbid infants as an early alert to medical staff and the guardians of the newborn. In order to obtain a comprehensive NCDS, the cry signals were analyzed regardless of cry stimulus, region, and gender. The proposed NCDS utilized both expiratory and inspiratory cry data sets. In this regard, the priority of this work was to study the role of acoustic features of the GFCC and MFCC in assessing the acoustic structure of the cry signals. Additionally, the GFCC and MFCC feature sets were combined by means of conventional and fusion methods. To the best of the authors' knowledge, this is the first time that the Canonical Convolution Analysis (CCA) fusion of the employed feature sets has been introduced to the assessment of pathologic newborn cries. Furthermore, the discussed challenges are addressed for both classification methods through two HPO schemes, where both classifiers have been fine-tuned using the grid search and Bayesian Hyperparameter Optimization (BHPO) methods. The proposed frameworks were evaluated by several measures and the results for each one expounded and compared extensively.

This study was proposed to address multiple the challenges and shortcomings of previous studies, as represented in Table-A III-1. A majority of the NCDS designs focus on studying a certain pathology group, whereas the aim of our work is to design a comprehensive alert system to notify the guardians of the newborn and the health professionals that the infant should undergo more screening tests, as there is a high potential it might be diagnosed with one or more pathologies from the ensemble of pathologies. Furthermore, the highest infant mortality rates are unfortunately associated with lower-income countries, where the proper screening equipment is inadequate and not available to many newborns (Unicef, 2014). This calls for the design of a non-complex, efficient NCDS that can perform early diagnosis so that the newborns are examined for an ensemble of pathologies and it can be determined if they are at risk of being unhealthy. As can be seen from Table-A III-1, the studies of newborn cries, undertaken for the purpose of differentiating between healthy and pathological infants, were either performed with a less inclusive set of pathologies or included less details on how HPO would assess enhancing the NCDS design.

There are an ever-growing number of designs that trade complexity for performance; however, this study proposes that employing proper feature fusion and HPO techniques could improve an NCDS from a moderate to a highly desirable state, where all the evaluation measures are relatively high and presented. The former studies present fewer measures for the evaluation; as an example, there are a very limited number of studies that have investigated the MCC measure. Table-A 3.a-1 also shows that the use of HPO and fusion methods in the study of pathological newborn cry signals is inadequate. As an example, most of the presented studies employed the SVM classifier. However, the resulting values are far lower than those presented in this study (the same explanation applies to the LSTM classifier, where the results are around 10% lower without the use of HPO methods). The aim of this study is to highlight the effects and importance of HPO and fusion methods in all NCDS designs, by explaining run-times and comparing the results before and after fusion and employing HPO. The role of feature fusion and HP tuning could be crucial and shed light on many further applications that employ various modalities for developing a comprehensive system; thus, we tried to provide a detail-oriented study of how each step of the NCDS design contributed to enhancing or decrementing the final results, which distinguishes our study from other research in the field of cry-based diagnostic systems.

4.3 Methods and Participants

4.3.1 Cry Dataset and Participants

The first challenge in sketching a pathological study is the acquisition and collection of data. It is important to note that the priority is obtaining the consent of newborns' guardians to record the cry signal and then achieving their consent to include that cry signal in the database. Furthermore, obtaining the ethical approvals to add samples to a database is an arduous and toilsome process that might even lead to losing some of the acquired data.

The collection of data was accomplished by collaboration between Al-Sahel and Al-Raee hospitals in Lebanon and Saint Justine Hospital of Montreal, QC, Canada. All the signals have been recorded in NICUs or maternity rooms (public and private) in the hospital environment. The cry of the newborns in our dataset was initiated due to multiple reasons such as hunger, fear, and wet diapers (Abou-Abbas, Tadj, Gargour, et al., 2017). The reason for crying was resolved with the help of medical staff and newborn's caregivers regarding the conditions resulting the cry.

Cry recordings ranged from 1 to 4 min including silence, hiccups, inspiration cries, expiration cries, and background noise. They were collected using a digital 2-channel Olympus handheld recorder with a 16-bit resolution and 44,100 Hz sampling frequency. The recorder was placed in the 10-to-30 cm vicinage of the newborn's mouth with no special consideration in the acquisition process. The mean recording length is 90 s and there were up to 5 recordings from each newborn. Therefore, unwanted information such as chatter in the surrounding space, noises, instrument beeps, and cries of other newborns accompanied the signals, which makes our dataset a real corpus capable of solving the challenge of comprehensiveness. Moreover, the newborns included in our dataset represent different races, origins, genders, and weights. A summary of this dataset is represented in Table 4.1.

Table 4.1 Description of dataset and participants

Demographic Factors	Specification
Gender	Female and Male
Babies' Ages	1 to 53 days old
Weight	0.98 to 5.2 Kg
Origin	Canada, Haiti, Portugal, Syria, Lebanon, Algeria, Palestine, Bangladesh, Turkey.
Race	Caucasian, Arabic, Asian, Latino, African, Native Hawaiian, Quebec.
Cry Stimulus	Discomfort, lack of sleep, wet diaper, pain, fear, colic, reflux, birth cry, hunger.
Healthy/Pathology Group	Healthy, dyspnea, fever, gastroschisis, grunting, hyperbilirubinemia, hypoglycemia, hypothermia, intrauterine growth retardation, jaundice, kidney failure, meconium aspiration syndrome, meningitis, myelomeningocele, respiratory distress syndrome, retraction, seizure, sepsis, tachypnea, thrombosis in vena cava, vomit.

Newborns do not have any control over their vocalization before 3 months of age (more accurately, 53 days) (Boukydis & Lester, 2012). The genesis of vocalizations in advance of this age is merely affected by biological rhythms. Moreover, it was shown that the mean of fundamental frequency undergoes no increasing or decreasing trend during the first 53 days of life (Lind & Wermke, 2002). Besides this, the supralaryngeal VC is reconfigured towards a human vocal tract after the 3 months of age (Boukydis & Lester, 2012). Therefore, newborns over 53 days old were not included in the current study.

4.3.2 Pre-Processing

Corwin et al. (Khalilzad, Hasasneh, et al., 2022) described the four types of acoustic units that constitute a cry signal as expiratory phonation, expiratory hyperphonation, expiratory dysphonation, and inspiratory phonation. During the phonation, the vibrations of the newborn's vocal folds generate sound, which is also referred to as voicing. The inspiratory cries are the "gasping" inhalation after the onset of crying that has enough power to cause vibrations in the vocal folds. Since the INSV episodes of the cry represent the laryngeal straitening of the ingressive air current, these cries have the potential to be a biomarker for diagnosis purposes (Fisichelli, Karelitz, Fisichelli, & Cooper, 1974). The power needed for driving the expiratory phase of a cry is stored during the inspiratory phase. Expiration can be interpreted as a moderate decrement in the volume of the lungs (Aucouturier et al., 2011). Usually cries occur during this respiratory phase, so this segment is considered to contain the main information, while the inspiratory cries remain the less explored and cognizant type of the cry event by researchers. Although it has been reported that the restraint of the upper airway may lead to sudden infant death syndrome and apnea, and the inspiratory cry is believed to contain information leading to pain and distress (Aucouturier et al., 2011), this type of cry has been often neglected in the study of NCDS (Grau, Robb, & Cacace, 1995). Concisely, analysis of both expiratory and inspiratory cries is indispensable regarding the design of a comprehensive NCDS, and in this study, both expiratory and inspiratory phonation were included.

The cry samples in our dataset were labeled by a group of researchers. An example of the assigned cry signal units is depicted in Figure 4.1. Different segments of the cry signal have been margined, and matching labels have been attached via WaveSurfer, presented in our previous works (Abou-Abbas, Tadj, Gargour, et al., 2017; Matikolaie & Tadj, 2020).

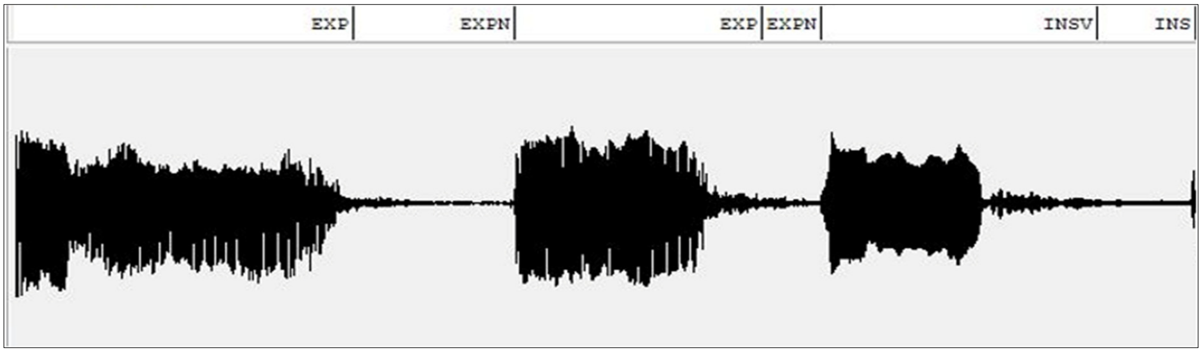

Figure 4.1 An example of a labeled cry signal via WaveSurfer. The X axis represents time, and the Y axis represents amplitude

Table 4.2 represents the number of samples in each dataset as well as the number of samples separated for the test and training. In total, 68 newborns with one of the mentioned pathologies were included in the unhealthy subsection of the data, and 300 healthy newborns participated in this study. Each of these participants yielded a different number of samples in the dataset. An equal number of samples from the healthy group were selected to ensure a balanced analysis.

Table 4.2 Number of samples in each dataset for training and test

	No. of Healthy	**No. of Pathologic**	**No. of Train Samples**	**No. of Test Samples**
EXP	3005	3005	4207	1803
INSV	3620	3620	5068	2172

4.3.3 Feature Extraction

The extraction of appropriate acoustic features capable of pertinent signal representation plays a vital role in any audio classification problem. As discussed above, for the effectuation of a cry, glottal impulses proceed through the filtering carried out by the vocal tract (Wasz-Hockert et al., 1968). With the aim of distinguishing between the source and filter of a cry, cepstral analysis employed, which enables a homomorphic transformation (Huang et al., 2001). MFCCs were derived from the Mel Filter Banks, whereas GFCCs were obtained from the Gammatone Filter Banks, which are a representation of inner and external middle ear physiological transitions (Zhao & Wang, 2013). In other words, although the two approaches are based on from the human sound perception model, the GFCCs are coordinated to comprehend the physical alterations more effectively than the MFCCs, and better delineate the auditory system (Katsiamis, Drakakis, & Lyon, 2007). Both Gammatone and Mel-frequency representations of the cry signal were mapped into the cepstrum space for the feature extraction step. Figure 4.2 illustrates our framework; the proposed steps for the acquisition of each of these features are described in the following sections.

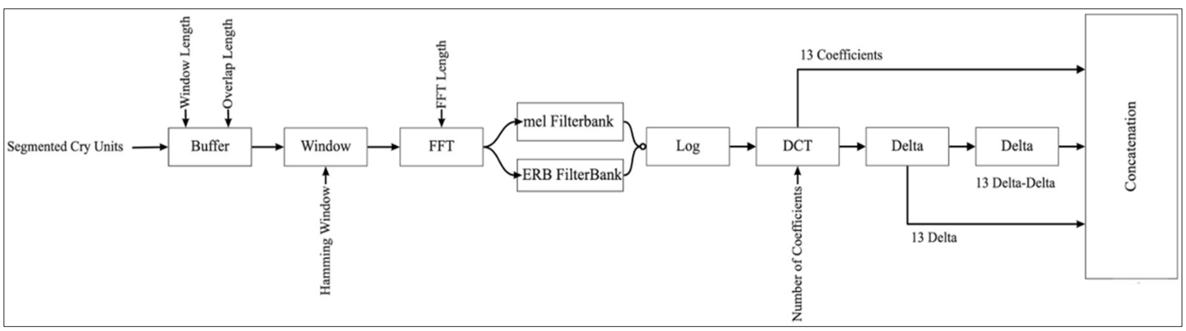

Figure 4.2 Framework of the proposed NCDS

4.3.3.1 Mel-Frequency Cepstral Coefficients

The calculation of MFCCs follows several steps. First, the preprocessed cry signal is pre-emphasized and divided into frames of 10 ms with a 30% overlap using the hamming window. The Fast Fourier Transform (FFT) then converts these frames to obtain the signal's spectrum. In the next step, the spectrum is transformed to the Mel-frequency scale, which is a representation of perceived pitch. For this purpose, a filter bank consisting of 13 triangular Mel-spaced filters is employed. As a consequence of the vocal tract's uniformity, the adjoining bands in the filter bank are inclined towards having correlated energy levels. Hence, a Discrete

Cosine Transform (DCT) is imposed to decorrelate them and yield Mel-frequency Cepstral Coefficients. The Mel-scale of the frequency f can be approximated as Equation 4.1.

$$M(f) = 1125 \ln(1 + f/700) \tag{4.1}$$

It was shown that the first thirteen coefficients could efficiently track the variations in the shape of the vocal tract during the generation of a sound by humans (Kumaran et al., 2021). A similar procedure as in the previous works (Alaie et al., 2016; Matikolaie & Tadj, 2020) was followed and the average statistical measure was used. The MFCCs hold information from one individual frame, and are therefore described as static features. In order to attain information on the fluctuations of the cry signal across multiple frames, the first and second derivatives of MFCCs are computed. Equation 4.2 gives the first derivative of the MFCCs for T consecutive frames (set equal to 2 in this study):

$$\Delta c_m(n) = \frac{\sum_{i=-T}^{T} k_i c_m(n+i)}{\sum_{i=-T}^{T} |i|} \tag{4.2}$$

Here, the *m*th feature for the *n*th frame is represented by $c_m(n)$, and k_i denotes the *i*th weight. Calculating the first-order derivative of the delta coefficients yields the delta-delta coefficients. The total number of features in the MFCC feature set equals 39, including 13 MFCCs, 13 deltas, and 13 delta-deltas (Rabiner, 1993).

4.3.3.2 Gammatone Frequency Cepstral Coefficients

Gammatone Frequency Cepstral Coefficients are a variant of MFCCs based on the biological response of the human auditory system. These features are extracted from the Gammatone filters with equivalent rectangular bandwidth (ERB) bands. Valero et al. (Valero & Alias, 2012) reported that the GFCC successfully performed non-speech audio classification tasks. It was also reported that the computation of the GFCCs was cost-efficient and has greater noise robustness compared to the MFCCs (Matikolaie & Tadj, 2020). The procedure for obtaining the GFCCs is similar to the MFCC. The non-stationary cry signal was windowed into frames of 10 ms with 30% overlap. The hamming window was applied for this purpose. The Gammatone filter banks were then applied to the FFT of the cry signals, which was done in order to amplify the perceptually meaningful sound signal frequencies. Next, the output of the last step was mapped into the logarithmic space. Finally, the Discrete Cosine Transform (DCT) was applied to decorrelate the filters' outputs and better mimic human loudness perception. The *m* coefficients from *N* Gammatone filters were then calculated via Equation 4.3.

$$GFCC_m = \sqrt{\frac{2}{N}} \sum_{n=1}^{N} log(X_n)\ cos\left[\frac{\pi n}{N}\left(m - \frac{1}{2}\right)\right] \qquad 1 \leq m \leq M \tag{4.3}$$

where X_n represents the corresponding energy of the *n*th band. Finally, the GFCC delta and delta-delta coefficients were derived, and the feature set comprised 39 coefficients matching the MFCC feature vector (Valero & Alias, 2012).

4.3.4 Features Fusion

By means of feature fusion, multiple feature sets are consolidated to create a single feature vector more robust than the individual feature vectors. Feature fusion can be undertaken in four different stages of the NCDS: 1) the data/sensor level; 2) the feature level; 3;) the matching score level, and 4) the decision level (Telgad et al., 2014).

In feature-level fusion, appropriate feature normalization, transformation, and reduction are employed in order to merge the features extracted from different sources into one feature set. The main benefit of feature-level fusion is the detection of correlated feature values generated by multiple algorithms, making it possible to introduce a new compressed set of salient features that can enhance classification accuracy. Therefore, CCA fusion at the feature level was utilized as the feature fusion strategy in this study (Kim et al., 2019).

4.3.4.1 Canonical Convolution Analysis (CCA)

Canonical convolution analysis handles the mutual statistical association between two feature sets by constructing a correlation criterion function. Subsequently, the canonical correlation regarding the criterion chosen in the last step was exploited, and discriminant vectors were forged so that the surplus information could be suppressed (Sun, Zeng, Liu, Heng, & Xia, 2005).

Suppose we take two feature sets X and Y of $p \times n$ and $q \times n$ dimensions, respectively. In other words, for each nth sample of the dataset, $p + q$ features were extracted. In order to obtain information about all the relations across the feature sets, the overall covariance matrix, S, can be written as Equation 4.4:

$$S = \begin{pmatrix} cov(x) & cov(x,y) \\ cov(y,x) & cov(y) \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix} \tag{4.4}$$

Apprehending the associations among the two sets of features may become challenging when they do not follow a steady pattern. CCA solves this challenge by finding the linear combinations, $X^* = W_x^T X$ and $Y^* = W_y^T Y$, and attaining the maximum pair-wise correlations, which is facilitated via Lagrange multipliers. The pair-wise correlation is defined as Equation 4.5.

$$corr\ (X^*, Y^*) = \frac{cov(\ X^*, Y^*)}{var(X^*)\ .\ var\ (Y^*)} \tag{4.5}$$

Finally, feature-level fusion is achieved by the concatenation of the transformed feature sets as in Equation 4.6:

$$Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^T X \\ W_y^T Y \end{pmatrix} = \begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix} \tag{4.6}$$

Z represents the Canonical Correlation Discriminant Features (CCDFs) (Haghighat, Abdel-Mottaleb, & Alhalabi, 2016). The GFCC and MFCC feature sets' fusion results, constituting 39 features each, yielded a feature vector with 60 features representing the cry signals. In this study, the performance of the fused features was compared to the individual feature sets, as well as their concatenation.

4.3.5 Classification

Classification assigns class labels to the given data points. The evaluation of the extracted feature sets was performed by classification. Two different classifiers were implemented in the current study. The first classification method was SVM, which is conventional in NCDS research. Moreover, the LSTM neural network was also employed. These classifiers and the HPO methods associated with each one are introduced in the following sections.

4.3.5.1 Support Vector Machine (SVM)

SVMs are prevalent in the analysis of audio signals. SVM is known as a supervised ML classification method that draws a hyperplane to maximize the marginal distance between the classes of data. The support vectors represent boundary feature points and form the basis for classification. A kernel function handles the nonlinearity of the data (Sahak et al., 2013). A Gaussian Kernel was implemented, which assumes that similar feature points were located in close vicinity and considers the Euclidean distance between x and x_i. In this study, the box constraint and kernel scale were tuned as the HPs of the SVM model.

4.3.5.2 Hyperparameter Optimization (HPO)

In any classification problem, the goal is to achieve high performance while keeping the errors to a minimum; therefore, HPO methods have been introduced. The several approaches to the HPO of an ML classifier include grid search, random search, and BHPO, which have omitted the need for human intervention to tune the classifier's HPs. The significance of the HPO is that each configuration is designed to fit its corresponding task. The main function of any given HPO method is to attain an optimum value for each HP from a set of finite values that minimizes the loss or maximizes the objective function. However, there are always downsides to each method, such as the high computational costs associated with the NN HPO and the probability of facing the curse of dimensionality (Feurer & Hutter, 2019).

The acquisition function and probabilistic surrogate model are the basis of the BHPO. The acquisition function enables a BHPO model to be updated in correspondence to it iteratively, defined as Equation 4.7 (Ashwini & Vincent, 2022):

$$x^* = \underset{x \in X}{argmin}\, f(x) \tag{4.7}$$

In every iteration, the model is updated based on new HPs and the corresponding model performance. Once the predefined number of iterations is reached, the best observed HPs are announced, as are the optimal observed values for the objective function. As will be seen in the following sections, BHPO often achieves better results than the other two HPO methods introduced.

4.3.5.3 Long Short-Term Memory (LSTM)

Recurrent Neural Networks (RNNs) have been shown to be propitious in the analysis of both single data points and sequential data, such as acoustic inputs. A feedback loop connects the input of RNN to its output, allowing them to model the dependencies in time series. Long Short-Term Memory (LSTM) networks are a type of RNNs with memory cells capable of learning, keeping, and forgetting data. By means of this memory cell, LSTMs function well with both short-term and long-term features (Gimeno, Viñals, Ortega, Miguel, & Lleida, 2020). Since the generation of the cry signal is intrinsically dynamic, RNNs may prove functional in their acoustic modeling. However, the challenge arises from the complexity of the training and tuning of hyperparameters of these networks. In order to overcome this challenge, the HPO methods were implemented to find and choose the optimal HPs. As mentioned above, Bayesian optimization requires fewer iterations than the grid search method to achieve the optimal values for the HPs in neural networks. The general task of HP acquisition by the BHPO is depicted in Figure 4.3.

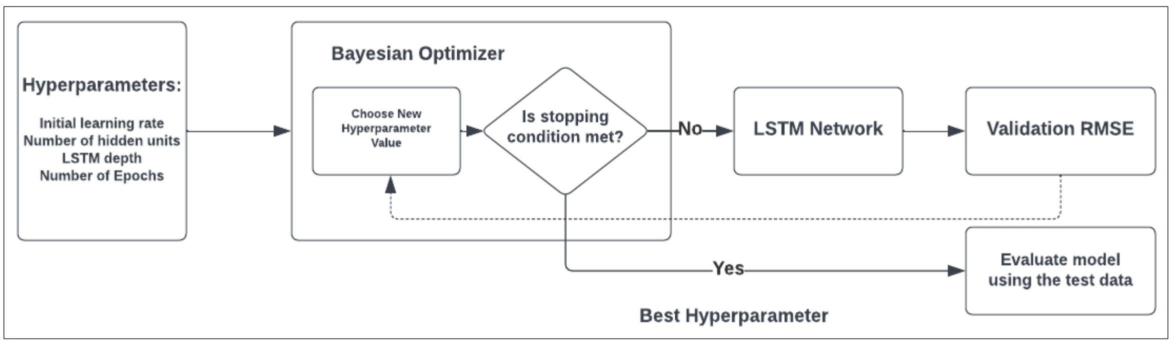

Figure 4.3 General Bayesian hyperparameter optimization process

The range of each hyperparameter was pre-determined in order to exploit the full potential of the HPO methods, as is shown in Table 4.3.

Table 4.3 Predefined ranges for the hyperparameter optimization of the LSTM

Parameter	Selected Range
Initial learning rate	[0.001, 1] (logarithmic steps)
Number of hidden units	[2, 39]
Number of Epochs	[100, 500]
Depth	[1, 3]

The activation function for this LSTM configuration is the hyperbolic tangent function (tanh). The hyperparameters included in this set of experiments were initial learning rate, number of hidden units, maximum epochs, and the depth of the LSTM architecture. Increasing the depth of the network was accompanied by higher computational costs, and all the best results were achieved when using only one layer. The loss and root mean squared errors were calculated for each run of the optimization process. The optimization was performed in 30 evaluations for each set of features, and the parameters that maximized the overall accuracy were chosen for each configuration.

4.4 Evaluation

This study aimed to differentiate pathologic infants from healthy infants and employed the GFCC and MFCC features with the LSTM and SVM classifiers. A wide range of pathologies were included in these experiments in order to achieve a comprehensive NCDS, which is able to act as an early alert given the lack of medical experts and access to expensive and extensive laboratory experiments. Different experiments were conducted with the proposed feature vectors, their combination, and their CCA fusion. Following the feature extraction step is classification. There are two approaches to validating the classifier's performance after training: holdout and cross-validation. The data were split into 70% training and 30% unseen testing data for both classifiers. For the SVM classifier, a 5-fold cross-validation was conducted on the training data, whereas for the LSTM classification, a holdout validation approach was chosen with 20% of the training data with a frequency of once every 10 iterations, because of the different natures of the classifiers. For the k-fold cross-validation, the data were split into k partitions, $k-1$ folds of which were used for training and one fold for testing in each iteration. This procedure was repeated up to the point at which each of the k folds was marked as the test fold. Finally, the results of grid search and BHPO for each architecture were compared.

The discriminatory performance of an NCDS in a binary problem can be represented by a contingency matrix, as shown in Table 4.4. The task of the NCDS in our paper was to detect the pathological neonates amid the healthy. In order to appraise how well the system performed its role, the evaluation measures were introduced and computed. Practically, the most convenient evaluation measure is the accuracy, which is equivalent to the proportion of correctly predicted samples over all the observations. The accuracy measure benefits from both calculation and apprehension simplicity; however, the lack of informativeness as well as the fewer concessions towards the minority calls for the implementation of more evaluation measures (Hossin & Sulaiman, 2015).

Table 4.4 Contingency matrix for the evaluation of NCDS

<table>
<tr><td></td><td colspan="4">True class</td></tr>
<tr><td rowspan="4">Predicted class</td><td></td><td>Pathologic</td><td>Healthy</td><td>Measures</td></tr>
<tr><td>Pathologic</td><td>True Positive</td><td>False Positive</td><td>Precision
$\frac{TP}{FP+TP}$</td></tr>
<tr><td>Healthy</td><td>False Negative</td><td>True Negative</td><td>Negative Predictive Value
$\frac{TN}{FN+TN}$</td></tr>
<tr><td>Measures</td><td>Recall
$\frac{TP}{TP+FN}$</td><td>Specificity
$\frac{TN}{FP+TN}$</td><td>Accuracy
$\frac{TP+TN}{TP+FP+FN+TN}$</td></tr>
</table>

One solution is to evaluate the NCDS performance without considering the true negative case, which will introduce a measure named precision. Precision, or Positive Predictive Value (PPV), is the ratio of true pathologic cases among the samples predicted as healthy. Another measure is recall, or sensitivity, which refers to the probability of recognizing a truly pathologic case by NCDS. The F-score and Matthews' Correlation Coefficient (MCC) were reported to be more instructive in binary classification problems. F-score is a function of both recall and precision, and indicates the inclusive performance of the system and is equal to the harmonic mean of precision and recall (Flach & Kull, 2015). The specificity measure denotes the true negative rate, and it indicates the true healthy samples correctly identified by the NCDS (Zhu, Zeng, & Wang, 2010).

The MCC is a highly informative evaluation measure when used in problems such as NCDS designs, since it accounts for all the information in a contingency matrix. The MCC, Equation 4.8, gives a value in the range of [−1, +1], where the misclassified performance results in negative values, and the higher values in the positive range signify better performance in terms of classification (Chicco & Jurman, 2020; Vihinen, 2012). In this study, a high acceptance value of +0.50 was set to evaluate the classification.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FN)(TN+FP)(TP+FP)(TN+FN)}} \tag{4.8}$$

4.5 Results

This section presents the results of evaluating different architectures with multiple measures in Tables 4.5 – 4.12. Regarding the evaluation measures introduced, higher values for each measure translate into the better performance of the system. In this study, four sets of experiments were conducted: 1. Evaluate the NCDS performance with default/random search hyperparameter configuration of the classifiers. 2. Evaluate the NCDS performance with grid search HPO. 3. Evaluate the NCDS performance with BHPO. 4. Compare the performance of the system with different iterations of HPO for each method, ranging from 30 iterations to 100 iterations for SVM and different numbers of neurons for the LSTM.

Each of the feature vectors were evaluated with the SVM classifiers, which are shown in Tables 4.5–4.8. In this step, the evaluation of system performance was undertaken by three different settings of the classifier: 1. Default settings. 2. Grid search optimization. 3. BHPO. The same procedure was repeated for the LSTM classifier, and 30 iterations of each HPO method were performed.

First, the results related to using the SVM classifier as a baseline to compare the results of the next steps are discussed. Table 4.5 represents the results for the MFCC feature set for the INSV and EXP datasets. The use of HPO similarly increased the evaluation measures across both datasets. Moreover, BHPO achieved a very similar or better performance except for in the recall measure. The highest accuracy and F-score for the EXP dataset were 87.37% and 86.64%, respectively; both were obtained through BHPO. This experiment yielded a better performance with the INSV dataset, and yielded 89.05% for the accuracy measure, which was again achieved through BHPO. However, grid search had a slight superiority in terms of the F-score, and achieved 89.24%.

Table 4.5 Results of evaluating the MFCC feature set classification with SVM

	MFCC	*Accuracy*	*Recall*	*Specificity*	*Precision*	*F-Score*	*MCC*
EXP	Random Search	72.16	62.93	81.26	76.80	69.17	0.45
	Grid Search	86.34	88.09	84.63	84.97	86.49	0.73
	Bayesian	87.37	82.55	92.11	91.17	86.64	0.75
INSV	Default	79.38	78.36	80.45	80.73	79.53	0.59
	Grid Search	88.90	90.09	87.65	88.40	89.24	0.78
	Bayesian	89.05	88.72	89.40	89.74	89.23	0.78

Table 4.6 presents the results of evaluating the GFCC feature set with the SVM classifier. By briefly looking at Tables 5 and 6, it can be seen that the MFCC feature set outperformed the GFCC feature set across both datasets. Similar to all the other feature sets, the best results in terms of F-score and accuracy in relation to the EXP dataset for the GFCC feature set were achieved through BHPO. In a general sense, the combination of the GFCC with the SVM yielded better results with the INSV dataset compared to the EXP dataset. The GFCC features' highest accuracy and F-score were 85.51% and 85.88%, respectively; both were achieved with BHPO and INSV dataset.

Table 4.6 Results of evaluating the GFCC feature set classification with SVM

	GFCC	*Accuracy*	*Recall*	*Specificity*	*Precision*	*F-Score*	*MCC*
EXP	Random Search	67.42	57.59	77.11	71.27	63.70	0.35
	Grid Search	83.75	86.06	81.48	82.08	84.02	0.74
	Bayesian	84.49	86.39	82.62	83.05	84.69	0.69
INSV	Random Search	76.57	73.53	79.74	79.14	76.23	0.53
	Grid Search	85.26	86.23	84.24	85.12	85.67	0.71
	Bayesian	85.51	86.25	84.73	85.52	85.88	0.71

In the next step, the GFCC and MFCC feature sets were combined to evaluate the NCDS performance under these conditions. As for the EXP dataset, the concatenated feature set could increase the accuracy and F-score measures by 1% and 1.7%, respectively, compared to the best results of the last two feature sets. The highest results for the EXP dataset were 88.41% and 88.30% for accuracy and F-score, respectively. The performance of the NCDS with this configuration for the INSV dataset was very similar to that for the MFCC feature set used individually, and there was a slight improvement in the evaluation measures (Table 4.7).

Table 4.7 Results of evaluating the concatenation feature set classification with SVM

	Concatenation	***Accuracy***	***Recall***	***Specificity***	***Precision***	***F-Score***	***MCC***
EXP	Random Search	75.49	68.58	82.29	79.25	73.53	0.51
	Grid Search	87.85	87.93	87.78	87.64	87.78	0.76
	Bayesian	88.41	88.13	88.68	88.47	88.30	0.77
INSV	Random Search	81.06	81.03	81.09	81.75	81.38	0.62
	Grid Search	88.59	88.40	88.79	89.19	88.79	0.77
	Bayesian	89.07	89.57	88.55	89.10	89.33	0.78

As a final experiment with the SVM classifier, the GFCC and MFCC feature sets—each containing 39 elements—were fused, and the feature vector was reduced to 60 elements, which was a more than 25% reduction in the size of the feature space. Since the size of the feature space was reduced, it might be expected that we see a rather small drop or a similar performance across the evaluation measures with this experiment compared to with the EXP dataset. However, as can be seen from Table 4.8, not only were the overall best results in terms of accuracy and F-score maintained, but they were also increased by about 1%. The results for the INSV dataset show the new highest accuracy and F-score across all the experiments with the SVM classifier, with 89.96% and 90.27%, respectively. For the EXP dataset, compared to the best results in terms of accuracy and F-score in previous experiments, the fusion of the features decreased the performance of the NCDS by 0.7% and 0.35%, respectively.

Table 4.8 Results of evaluating the CCA fusion feature set classification with SVM

	CCA Fusion	***Accuracy***	***Recall***	***Specificity***	***Precision***	***F-Score***	***MCC***
EXP	Random Search	81.43	81.59	81.28	81.11	81.35	0.63
	Grid Search	85.31	87.20	83.46	83.86	85.49	0.71
	Bayesian	87.71	90.35	85.11	85.68	87.95	0.76
INSV	Random Search	88.09	88.43	87.74	88.29	88.36	0.76
	Grid Search	88.28	89.14	87.38	88.07	88.60	0.77
	Bayesian	89.96	91.14	88.74	89.43	90.27	0.80

After evaluating different aspects of the NCDS with the SVM classifier, the study proceeded to design an LSTM configuration to differentiate pathologic newborns from the healthy group. The same procedure of the experiments as with the SVM classifier was followed, and the system was evaluated with each feature configuration separately. The performances of all feature sets were improved considerably by using the LSTM classification method. The MFCC feature set achieved the highest accuracy and F-score of 99.03% and 99.05%, respectively, with the LSTM classifier for the INSV dataset, which is a nearly 10% improvement compared to the SVM method. As can be seen from Table 4.9Table 4.10, the performance of the GFCC feature set was slightly better than the MFCC feature set with the LSTM classifier for the EXP dataset, and vice versa for the INSV dataset. Both HPO methods worked marvellously with the LSTM classifier; however, they were not efficient in terms of run-time, which will be compared in the Discussion section.

Table 4.9 Results of evaluating the MFCC feature set classification with LSTM

	MFCC	*Accuracy*	*Recall*	*Specificity*	*Precision*	*F-Score*	*MCC*
EXP	Random Search	76.59	95.98	57.49	69.00	80.28	0.58
	Grid Search	97.78	95.53	100.00	100.00	97.71	0.96
	Bayesian	99.33	98.66	100.00	100.00	99.33	0.99
INSV	Random Search	95.17	96.58	93.69	94.12	95.33	0.90
	Grid Search	96.09	92.34	100.00	100.00	96.02	0.92
	Bayesian	99.03	98.56	99.53	99.55	99.05	0.98

The best accuracy and F-score achieved by the GFCC feature set were 99.45% and 99.44%, respectively, whereas the MFCC obtained 99.33% for both measures. The mentioned results were accomplished for the EXP dataset. It is noteworthy to mention that both feature sets attained 100.00% for specificity and precision measures. Moreover, the MCC measure has acquired a high value for the BHPO with EXP dataset for both feature sets, which indicates close to perfect classification quality (Table 4.10).

Table 4.10 Results of evaluating the GFCC feature set classification with SVM

	GFCC	*Accuracy*	*Recall*	*Specificity*	*Precision*	*F-Score*	*MCC*
EXP	Random Search	84.53	75.20	93.72	92.19	82.83	0.70
	Grid Search	96.56	93.07	100.00	100.00	96.41	0.93
	Bayesian	99.45	98.88	100.00	100.00	99.44	0.99
INSV	Random Search	96.09	92.34	100.00	100.00	96.02	0.92
	Grid Search	97.88	97.66	98.12	98.19	97.92	0.96
	Bayesian	97.51	95.14	100.00	100.00	97.51	0.95

Even though the state-of-the-art performance of both individual feature sets through HPO methods leaves little room for improvement, it is still beneficial to study the behavior of the system by the combination of the two feature vectors to assess their efficacy compared to the SVM classifier. As can be deduced from Table 4.11, the performance of the NCDS was degraded by simply concatenating the feature sets, which may translate to lower uniformity of the feature space. The highest accuracy and F-score achieved with this experiment belonged to the INSV dataset, which reached 98.99% and 99.00%, respectively.

Table 4.11 Results of evaluating the concatenation feature set classification with SVM

	Concatenation	***Accuracy***	***Recall***	***Specificity***	***Precision***	***F-Score***	***MCC***
EXP	Random Search	89.24	97.54	81.06	83.54	90.00	0.80
	Grid Search	96.23	92.40	100.00	100.00	96.05	0.93
	Bayesian	98.34	96.76	99.89	99.88	98.30	0.97
INSV	Random Search	98.48	97.21	99.81	99.81	98.49	0.97
	Grid Search	98.85	97.75	100.00	100.00	98.86	0.98
	Bayesian	98.99	98.11	99.91	99.91	99.00	0.98

Table 4.12 constitutes the results of the next experiment with the LSTM classifier. The system's performance was better than the concatenation framework since the CCA fusion removes the redundant features and helps improve the uniformity of feature space. This experiment showed the best performance in assessing the INSV dataset among all the previous experiments for all of the evaluation measures, specifically reaching 99.86% for both F-score and accuracy and 1.00 for the MCC measure. As for the EXP dataset, the GFCC feature set outperformed both combinational feature sets in terms of all evaluation measures.

Table 4.12 Results of evaluating the CCA fusion feature set classification with SVM

	CCA Fusion	***Accuracy***	***Recall***	***Specificity***	***Precision***	***F-Score***	***MCC***
EXP	Random Search	96.73	93.74	99.67	99.64	96.60	0.94
	Grid Search	97.34	97.65	97.03	97.00	97.33	0.95
	Bayesian	99.00	98.55	99.45	99.44	98.99	0.98
INSV	Random Search	96.09	92.61	99.72	99.71	96.03	0.92
	Grid Search	98.16	97.57	98.78	98.81	98.19	0.96
	Bayesian	99.86	99.73	100.00	100.00	99.86	1.00

In the previous section, the evaluation results regarding each feature set and classifier combination were extensively discussed; now, the discussion is undertaken from the perspective of the computational cost. For this matter, the run-time was selected as an indicator. It should be noted that in the case of the joint feature sets, namely, concatenation and CCA fusion, the given run-times include the process of concatenation and fusion, and not only the time corresponding to the HPO process. The elapsed times for the extraction of the GFCC and MFCC feature sets were 558.31 and 836.16 s, respectively, which suggests the GFCC feature set requires lower computational costs; other researchers have also mentioned the same results (Valero & Alias, 2012). Figure 4.4 compares the run-times of the grid search HPO and BHPO methods for different iterations of each one when applied to the SVM classifier.

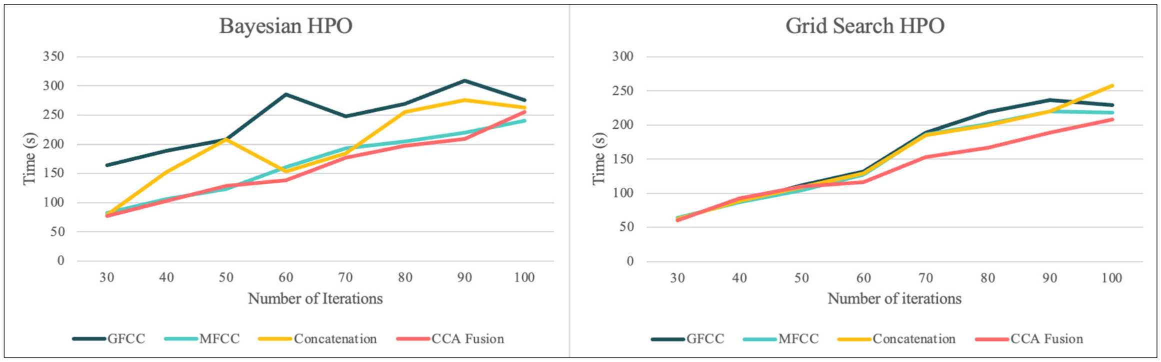

Figure 4.4 Elapsed time (seconds) regarding the different iterations of hyperparameter optimization methods for the SVM classifier

The comparison between run-times regarding each feature set firstly confirms that the CCA fusion method results in a more homogenous feature space, and reduces run-times until they are lower than the run-time for the individual feature sets, which is consistent in both HPO methods. As can be seen, BHPO resulted in the higher performance of the system and required longer run-times. In order to better illustrate this comparison, Figure 4.5 presents the average run-times of the two HPO methods for each NCDS configuration for a more detailed evaluation.

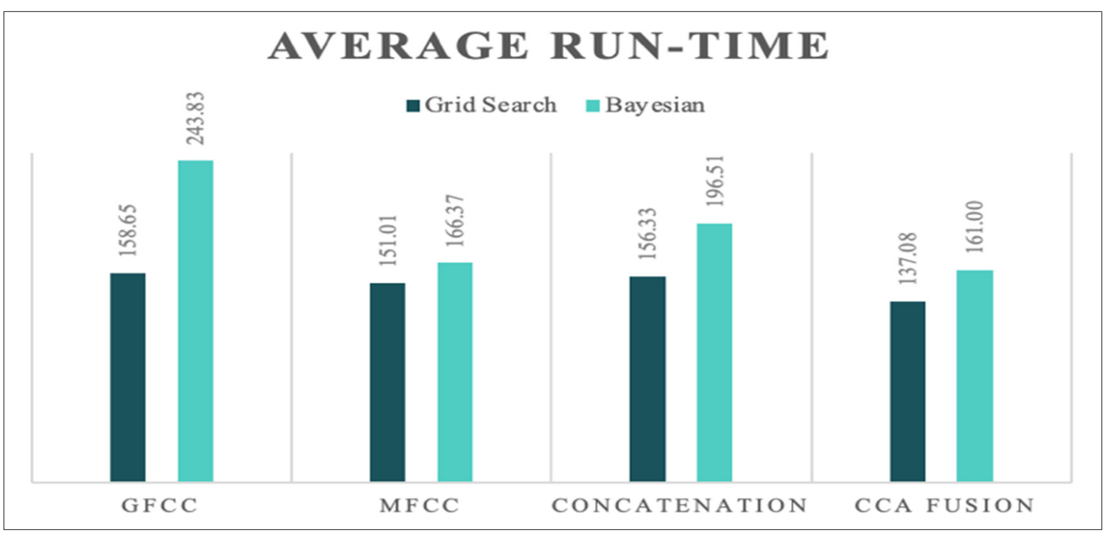

Figure 4.5 Average run-times for evaluating different iterations of hyperparameter optimization for the SVM classifier

Figure 4.6 shows the elapsed times (in seconds) for the grid search and BHPO methods for the LSTM classifier. Since the process of HPO for the NNs is highly time-consuming compared to the machine learning models, only 30 iterations of HPO were performed for this experiment. The results show that CCA fusion requires the shortest run-time out of all other feature sets for the grid search HPO, similar to the SVM HPO methods; the run-times regarding the grid search method were lower than for BHPO. It should be noted that the number of trials for both methods was limited to 30; BHPO can achieve satisfactory results with this number of iterations, whereas grid search often requires a much greater number of trials. In summary, the proposed NCDS in this study accomplished desirable results across all the experiments in terms of performance and computational costs, and the longest elapsed time was less than 1700 s simultaneously.

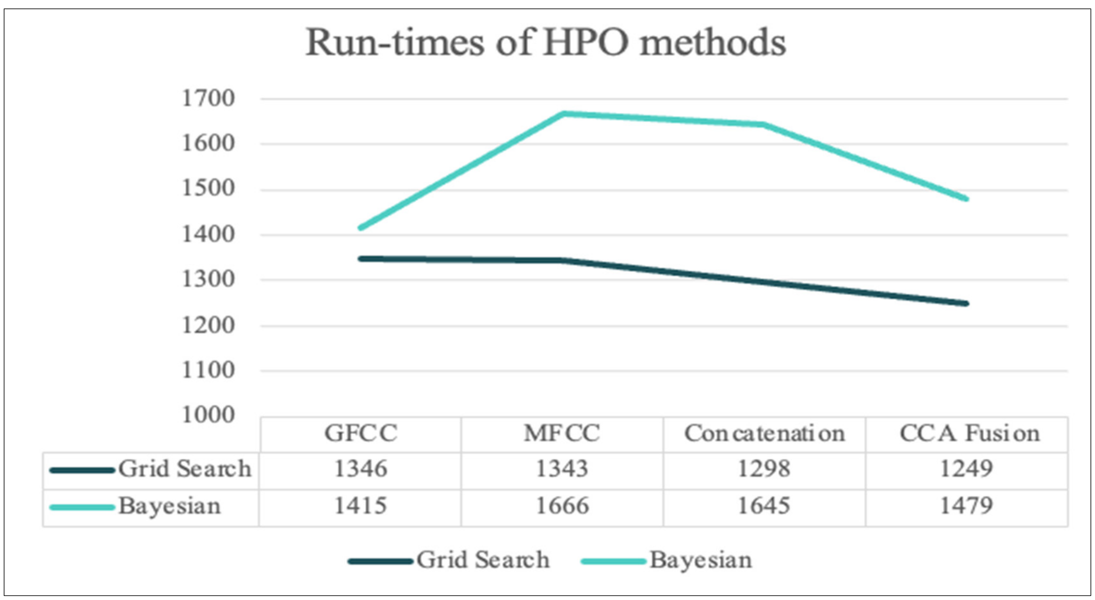

Figure 4.6 Comparing run-times for the two HPO methods of LSTM for different experiments

4.6 Discussion

The design of the NCDS is a challenging problem for every researcher aiming to study newborn cry characteristics, regardless of the purpose the NCDS aims to serve. This challenge is even more significant regarding the sensitive subject of detecting pathologic newborns. The NCDS designs are not developed enough compared to the other acoustic scene recognition systems or speech analysis applications; there is still a need for further studies in this field, which is mainly due to the fact that datasets are very limited in terms of the number of samples. This is due to certain limitations, such as the fact that the chances of having a newborn diagnosed with a specific pathology in any given duration of conducting a clinical study are not predictable. Therefore, there may not be sufficient samples from each given pathology group; ensuring the ethical and technical standards required to collect and use the cry samples in a database calls for extreme measures. In this regard, by segmenting each cry recording into multiple expiratory and inspiratory episodes, two datasets of EXP and INSV were formed. As mentioned above, the areas of the world that suffer the most from infant mortality are less developed and lack a sufficient number of expert physicians. Thus, it is vital to keep the design as simple as possible so that expensive hardware would not be required to achieve high performance.

One other aspect of the proposed study is that by employing MFCC and GFCC features, the cry signal is investigated both from the speech processing and non-speech audio processing perspectives. As was previously discussed, MFCCs have proved to be powerful discriminators, especially in speech processing tasks, while GFCCs have shown even better performance and robustness in non-speech audio applications. For the first time in newborn cry analysis a CCA fusion at the feature level was performed in order to make the feature space homogenized and omit redundant information. By looking at the results of run-times in the previous section, it can be seen that CCA fusion homogenized the feature space in a way that the fused feature vectors required less time for optimization, even when compared to the single feature vector of GFCC. This shows that although the fused vector had 60 elements, it was still optimized faster than a 39-element feature vector, even with the time required for fusion included. This is rather an interesting finding that shows potential for many further applications with the inclusion of features from various modalities in this field. Not only were the HPO run-times reduced, but also, the results were improved by reducing the number of features by about one fourth.

Since the challenge is to detect a pathologic newborn and alert the newborn caregivers and medical experts, it is worth tolerating higher run-times in order to obtain a more accurate diagnosis and benefit from the HPO methods. The other important factor here is that the NCDS cannot afford to misdiagnose a pathologic newborn as a healthy one, so the focus should be on achieving a high hit rate (recall) and F-score measure, which are the indicators of a low miss rate. This study proposed two different designs with respect to the runtime and performance trade-off. Firstly, using the SVM classification method, a simplistic design was proposed, which requires minimal run-time and could work with commercial hardware. It was shown that by implementing the HPO methods, a similar performance to the complex state-of-the-art designs with up to 90% F-score for the SVM could be achieved. Moreover, our LSTM design, which only has a one-layer depth, was able to achieve better F-scores than similar or more complex works in the literature using the proper HPO, with improvements of 99.86% and 99.45% for INSV and EXP datasets, respectively. This study also offers an extensive evaluation of the HPO factors and methods in addition to the primary goal, achieving high diagnostic power. Additionally, the powerful discriminatory role of inspiratory cries, which are neglected in most NCDS studies, is highlighted here, as is the success of our design with the EXP dataset, which worked even better with this dataset.

Finally, the high number of pathology groups included in this study makes it a comprehensive framework capable of a more reliable diagnosis, since the medical staff could suggest that the newborn does not suffer from the given list of pathologies. Figure 4.7 gives a visual summary of the best results achieved by each experiment in terms of F-score and accuracy. These results imply the similar performances of the NCDS in terms of both F-score and accuracy measures, which indicates that discussing the F-score measure alone would be sufficient.

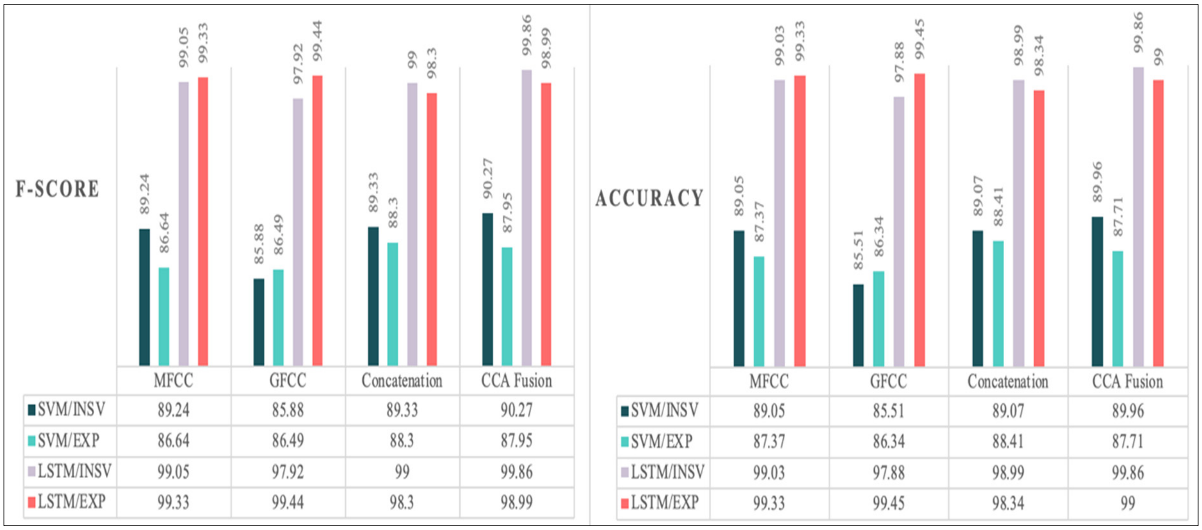

Figure 4.7 Summary of the best results achieved with the conducted experiments in terms of accuracy and F-score measures

In order to evaluate the results from another perspective and further explore the potential of both HPO methods and CCA fusion, another experiment was designed wherein the performance of the NCDS could be investigated with different HPO iterations of 30 to 100 (rising by steps of 10) on the EXP dataset. The average of all evaluations (eight experiments) across all measures is reported in Table 4.13. As can be inferred from the results, both HPO methods enhanced the system performance in terms of accuracy, recall, F-score and MCC measures. Several patterns were observed when conducting these experiments. Firstly, the performance of the BHPO method was superior to that of the grid search method across all the measures, except for the recall measure. Even though the recall measure represented an exception to the mentioned pattern, the highest recall was achieved through the BHPO method with the fusion of features, which was 90.25%. Secondly, the best performance in terms of accuracy, MCC, and F-score was achieved using the CCA fusion framework. Finally, it can be deduced that although CCA fusion slenderized feature space, the performance of the NCDS was not considerably aggravated, and was even increased in terms of the F-score measure.

Table 4.13 Results for the average of the evaluation measure for different iterations of hyperparameter optimization ranging from 30 iterations to 100 iterations

		Accuracy	***Recall***	***Specificity***	***Precision***	***F-Score***	***MCC***
GFCC	*Bayesian*	84.50	85.09	83.91	83.91	84.49	0.69
	Grid Search	83.75	86.06	81.48	82.08	84.02	0.68
MFCC	*Bayesian*	86.86	83.43	**90.24**	**89.89**	87.77	0.74
	Grid Search	86.34	88.09	84.63	84.97	86.49	0.73
Concatenation	*Bayesian*	**88.18**	87.90	88.46	88.25	88.07	0.76
	Grid Search	87.67	87.95	87.38	87.31	87.62	0.75
Fusion	*Bayesian*	87.92	**90.25**	85.63	86.10	**88.12**	**0.76**
	Grid Search	85.41	87.08	83.76	84.10	85.56	0.71

As was previously discussed in the Results section, the evaluation measures showed exceptional performance with the LSTM classifier. Therefore, to better demonstrate the power of LSTM in the NCDS design and validate the surprisingly high performance of the system, a final experiment was mapped out. In this experiment, the LSTM classifier was manually tuned for only one HP: the number of hidden neurons. For each feature set, the number of hidden neurons was changed from 2 neurons to half the size of each feature vector, e.g., 30 neurons for the fusion feature set with 60 elements for each sample. Table 4.14 presents the average of each evaluation measure used in the successful attempts with manual search methods for each feature set. Therefore, if the only parameter being tuned is the number of hidden neurons, the system's performance undergoes a bearable decline in exchange for lowering the computational costs. Moreover, as can be seen from the results, the best evaluation measures belonged to the CCA-fused feature set (except for the recall measure), which are 96.62% and 96.58% for accuracy and F-score, respectively. Therefore, by manually tuning only one HP, the system was capable of achieving up to a 96.58% average F-score, which translates to the high classification power of the LSTM classifier compared to the SVM, and the potential for an even better performance if other HPs are tuned as well.

Table 4.14 The averages of evaluation measures for the manual tuning of the hidden neurons for the LSTM classifier with each feature

	Accuracy	***Recall***	***Specificity***	***Precision***	***F-Score***	***MCC***
MFCC	89.02	94.23	83.89	89.16	90.54	0.80
GFCC	89.79	88.00	91.56	94.05	90.01	0.81
GFCC + MFCC	95.33	**95.95**	94.72	95.79	95.59	0.91
CCA Fusion	**96.62**	95.89	**97.34**	**97.34**	**96.58**	**0.93**

So far, the experiments in this study have been discussed and compared in terms of performance, classification power, and run-times. There are various tools and frameworks for the study of audio signals, which have resulted in many different applications and publications. Among these frameworks, many different machine learning and deep learning methods have been explored. It is worthwhile to compare the performance of the proposed NCDS with other similar works or architectures that analyze either newborn cry signals or other audio signals. In a recent study (Jin, Wang, & Zhan, 2022), environment sounds were classified through different models including SVM, LightGBM, XGBoost (XGB) nd CatBoost classification frameworks, employing time and frequency domain features and their combination. They were able to get 87.3% as their highest accuracy measure when using the LightGBM framework, and through the alteration of the gain factor, whereas their baseline classifiers yielded 66.7% for KNN, 67.5% for SVM, 72.7% for baseline Random Forest (RF), and 81.5% for their joint feature set with RF classifier. In another study (Bansal, Upali, & Sharma, 2022), speech signals were employed to diagnose Parkinson's with the use of RF, Decision Tree (DT), KNN, XGB, and Naïve Bayes (NBC) classifiers. The results show that the classifiers' performances ranked as follows: XGB achieved 96.61% for the accuracy measure, and KNN, RF, DT and NBC achieved accuracies of 94.91%, 88.13%, 86.44%, and 67.79%, respectively. The study of Singhal et al. (Singhal, Srivatsan, & Panda, 2022) classified music genres with Logistic Regression (LR), KNN, SVM, XGB, and RF classifiers. They also explored the effect of HPO on the RF classifier only, where the results were enhanced about 13% for accuracy and reached 98.8%. However, they did not discuss the HPO methods and trends. The highest result was achieved when using both RF and XGB classifiers—99.6% for both frameworks. The study of

Kim et al. (Kim, Oh, & Heo, 2021) explored a very similar framework to the one presented in this study for the analysis of beehive sounds through MFCCs, a Mel spectrogram and constant-Q transform features, with RF, XGB, CNN, and SVM classifiers. The highest accuracy was achieved through the combination of MFCC features with the XGB classifier, reaching 87.36%. Their VGG-13 classification showed very promising results, with 96% for the F-score measure. Lahmiri et al. (Lahmiri, Tadj, Gargour, et al., 2022) designed an NCDS for the purpose of detecting pathologic newborns with cepstrum features and multiple NN classifiers. By implementing LSTM classification, they were able to achieve an accuracy of 83.89% and 80.18% for the EXP and INSV datasets, respectively. Another work worth mentioning in this field is that by Matikolaei et al. (Matikolaie et al., 2022), wherein the proposed NCDS served the same purpose as in our study. The authors combined the MFCC with the auditory-inspired amplitude modulation features, and fed them into an SVM classifier; they attained 80.50% for the accuracy measure. Kumaran et al. (Kumaran et al., 2021) focused on the recognition of emotions; they combined the GFCC with the MFCC feature sets and employed C-RNN classification. They used a different architecture of LSTM than in our study, with the addition of a convolutional network, and the highest F-score yielded by their design was 79%. In another emotion recognition study, the MFCC features were employed with a combined CNN-LSTM architecture, and the highest accuracy of 87.4% was reported with the use of HPO methods, wherein they tuned learning rate and batch size (Verma et al., 2022). Given the results of the mentioned studies, our NCDS designs proved to be successful and introduced novelty to the study of newborn cries with the purpose of detecting pathologic infants. Our study proposed a simplistic design using the SVM classifier that benefits from BHPO; we showed it could achieve results similar to (and even better than) the state-of-the-art of NCDS employing NNs, which in the literature reached 90.27% for the F-score measure. Our second framework of LSTM classification with BHPO obtained up to 99.86% for the F-score measure, which is remarkable in the study of pathologic newborn cry signals. However, our system was outperformed by a design that implements DFFNN, since it was able to achieve 100% for both datasets of EXP and INSV (Lahmiri, Tadj, & Gargour, 2021).

In the design of the LSTM, the main concern was to prevent the model from becoming complex, and it employed only one hidden layer with a low number of hidden units. Both these achievements owe their success to the CCA fusion of the GFCC and MFCC feature sets, which not only enhances the overall performance, but also lowers the run-time by homogenizing the feature space and marks out the optimal feature set.

In summary, the presented results of this study suggest that a fusion of MFCC and GFCC features fed to deep and machine learning classifiers attains a higher performance compared to previous studies on detecting pathologic newborns. This framework is proposed as a non-invasive tool for aiding the expeditious detection of pathologic infants. There is still a vast ocean of unexplored ideas and architectures to be implemented in the study of pathological newborn cry signals, which is beyond the scope of this study. In future works, exploring more deep learning and machine learning designs such as DCNNs, and further exploring fusion techniques, especially at the decision level, such as the matching score method, would be of interest. Furthermore, studying more acoustic features and combining them with different classifiers would be worthwhile in order to highlight the efficacy of existing research on pathologic newborn cry signals.

4.7 Conclusion

The cry of infants has been recognized as a biomarker in the detection of pathologies for the purpose of early diagnosis. The presented study aimed to propose a comprehensive NCDS that distinguishes between healthy and pathologic cries regardless of the reason for crying, race, and gender. Our proposed system outlines the feature of the GFCC and its delta coefficients, which efficiently capture the dynamic nature of the cry signal and its periodic pattern. Moreover, the feature set used in this study includes the MFCC, which is well-known for its strong performance in many acoustic applications. These features were fed individually and fused into the LSTM and SVM classifiers, which belong to two different families of classifiers. In the next step, an extensive study of HPO methods for grid searching and BHPO for both classifiers was performed in order to improve the performance of NCDS. The LSTM was able to achieve a very high performance metric of 99.86% when applied to inspiratory cry signals in terms of both accuracy and F-score, owing to its capability to learn from sequential data. Furthermore, LSTM outperformed the optimized SVM when applied to both the studied datasets. All of the results obtained by the two proposed classifiers show potential for use in the investigation of pathologic infant cries.

This study contributed to the development of NCDS with the aim of designing a first alert for medical experts; it showed that healthy and pathologic infants have different cry patterns, which can be used as biomarkers. Regarding the results of this study, the proposed framework can be used as a non-invasive diagnostic tool without the need for high-end hardware and technologies.

CHAPTER 5

USE OF PSYCHOACOUSTIC SPECTRUM WARPING AND DECSISION TEMPLATE FUSION IN NEWBORN CRY DIAGNOSTIC SYSTEMS

Zahra Khalilzad and Chakib Tadj,

Department of Electrical Engineering, École de Technologie Supérieur, Université du Québec, Montréal, QC H3C 1K3, Canada

Paper submitted for publication, Jun 2023

5.1 Abstract

Dealing with newborns' health and needs is a highly delicate matter since they cannot express their needs through words, and crying does not reflect their exact condition or demand. However, during the last decades, the researchers found the cry signals to be rich in information about the health conditions, emotions, and the needs of the baby. Although the newborn cries have been used and studied for many applications and purposes, there is no prior research on distinguishing a certain pathology among other pathologies so far. In this study, an unsophisticated framework is proposed for the study of septic newborns amid a collective of other pathologies with the use of Decision Template (DT) fusion technique. Moreover, the cry signal was analyzed with both music inspired and speech processing inspired features. Furthermore, a Neighborhood Component Analysis (NCA) feature selection technique was employed with two goals: 1. Exploring how the elements of each feature set contributed to classification outcome, and 2. Investigating to what extent the feature space could be compacted and how will the outcome be affected. The attained results showed the success of both experiments introduced in this study with 88.66% for the DT fusion technique and a consistent enhancement in comparison to all feature sets in terms of accuracy and 86.22% for

the NCA feature selection method by drastically downsizing the feature space from 86 elements to only 6 elements. The achieved results showed great potential for identifying a certain pathology from other pathologies that may have similar effects on the cry patterns as well as proving the success of the proposed framework.

Keywords: Acoustic Features, Decision Templates, Decision Fusion, Sepsis, Newborn Cry Diagnostic System, Neighborhood Component Analysis

5.2 Introduction

Thanks to the advances in both engineering and medical research, it is now known that pathologic newborns cry diversely than healthy newborns and the cry characteristics could differ across different pathologies (Alaie et al., 2016). These observations sparked the idea for the design of various Newborn Cry Diagnostic Systems (NCDS). So far, NCDS architectures served the purpose of diagnosing newborns with a certain pathology from the healthy (Massengill Jr, 1969; Matikolaie & Tadj, 2020, 2022), detected the healthy newborns from a collective of pathologies (Alaie et al., 2016; Lahmiri, Tadj, & Gargour, 2021; Lahmiri, Tadj, Gargour, et al., 2022; Matikolaie et al., 2022) , and very recently, differentiated between two pathology groups (Khalilzad, Hasasneh, et al., 2022). In this study, the NCDS was taken one step further to detect a certain group of pathologies amidst an ensemble of other pathologies.

NCDS' benefit from a vast range of tools that help enhance their final diagnostic performance by improving different stages of the NCDS design. A NCDS design entails three main components which are namely preprocessing, feature extraction and manipulation, and finally classification that is vital to all the audio classification applications.

Crying is a manifestation of the newborn's health since it is the product of an extensive number of organs working together in harmony and malfunction in any of these organs would be reflected in the generated cry signal (Fort & Manfredi, 1998). The early studies showed that spectrograms of the healthy newborns followed consistent patterns whereas the spectrograms of the newborns diagnosed with pathologies would have certain acoustic attributes that makes them distinguishable (Michelsson, SirviÖ, et al., 1977). In this regard, many NCDS designs focused on the extraction and selection of the features that would effectively capture and represent these attributes. In the feature extraction step of NCDS, a wide range of features from time, frequency, and time-frequency domains were employed, which include but are not limited to Mel frequency Cepstral Coefficients (MFCCs), Gammatone-frequency Cepstral Coefficients (GFCCs), Linear Predictive Coding (LPC), F0 contour, auditory amplitude modulation, resonance frequency, prosodic features such as rhythm and tilt, entropy-based features e.g., spectral and approximate entropy, and features inspired by analysis of music such as harmonic ratio, Spectral Centroid (SC), and spectral flux.

Among these features, Mel-frequency based features, more specifically, MFCCs, are the most prevalent and often used as a baseline in many designs to ensure comparability with the works of other researchers. The reason behind the success and prevalence of MFCCs is because of their good discriminative performance (Khalilzad & Tadj, 2023); however, the role of Cepstral analysis is often not emphasized enough. Cepstral analysis facilitates the discrimination between the source and the filtering in audio analysis tasks since it is a homomorphic transformation. In speech analysis the basic study of the speech components includes the segregation of how each components affects the final outcome, which translates to declaring the functions of vocal tract impulse response, glottal pulse, and the vocal cord timing. In the study of cry signal analysis for different applications, Cepstral analysis was proven highly successful for the mentioned reasons. MFCCs were used to detect asphyxia (Wahid et al., 2016; Zabidi, Mansor, et al., 2017), hearing impairment (Jam & Sadjedi, 2009), sepsis (Khalilzad, Kheddache, et al., 2022; Matikolaie & Tadj, 2022) cleft palate (Massengill Jr, 1969), respiratory distress syndrome (Matikolaie & Tadj, 2020), and hypothyroidism (Zabidi et al., 2009). There are also a number of studies that focus on separating the healthy infants from a collective of pathologies where MFCCs have served successfully as well (Alaie et al., 2016). Another cepstral feature set that has recently gained attention are the GFCCs, owing to their better noise robustness, cost-efficiency, and better discriminative performance (Khalilzad & Tadj, 2023; Valero & Alias, 2012). GFCCs were utilized for speaker identification (Admuthe

& Patil, 2015), emotion recognition based on both newborn cry signals and adult speech (Garg & Bahl, 2014; Kumaran et al., 2021), and finally detection/discrimination of pathologies based on newborn cry signals (Khalilzad & Tadj, 2023). Inspired by this pattern for combining the psychoacoustic frequency warping with cepstral analysis, the idea of using another scale named Bark along with the Gammatone was worth exploring. Bark-frequency Cepstral Coefficients (BFCCs) were employed to identify the reason of crying in newborns (Liu et al., 2019; Maghfira, Basaruddin, & Krisnadhi, 2020; Sriraam & Pradeep, 2019), detection of high-risk prematurity in newborns (Tejaswini, Sriraam, & Pradeep, 2020), enhanced emotion detection from speech signal (Lalitha, Tripathi, & Gupta, 2019), and automatic speech recognition (Kamińska, Sapiński, & Anbarjafari, 2017) and other audio analysis applications. SC is derived from the study of timbre in musical application and tone measurement in audio signals and utilized for the detection of Alzheimer's from Electroencephalogram (EEG) (Kulkarni & Bairagi, 2017), and in NCDS applications to detect pathologies (Khalilzad, Kheddache, et al., 2022), developmental disorders (Oren et al., 2016) and understanding the reason of crying (Osmani et al., 2017). Spectral crest is often utilized in feature sets along other spectral features in several studies, with the purpose of visualizing music emotion (Kim, Van Ho, & Lim, 2017), detection of hunger from stomach sounds (Maria & Jeyaseelan, 2021), epileptic seizure detection (Dash & Kolekar, 2020), and audio fingerprinting (Ramalingam & Krishnan, 2005).

The next step of the NCDS design, is the classification where similar to feature extraction step, has been developed with many different methods and classifiers. Support Vector Machine (SVM) (Vincent et al., 2021), Multilayer Perceptron (MLP) (Khalilzad, Hasasneh, et al., 2022), K-nearest Neighborhood (KNN) (Khalilzad, Kheddache, et al., 2022), Random Forest (RF) (Chang, Bhattacharya, Raj Vincent, Lakshmanna, & Srinivasan, 2021), Decision Trees (Matikolaie & Tadj, 2022), Probabilistic Neural Network (PNN) (Matikolaie et al., 2022), Deep Feedforward Neural network (DFFNN), Convolutional Neural Networks (CNN) (Lahmiri, Tadj, Gargour, et al., 2022), Long Short-term Memory (LSTM) (Khalilzad & Tadj, 2023), and many other classification approaches are amidst the employed means in NCDS design. Although some of these studies compared the results of the mentioned classifiers, very few focused on combining the outcome of the classifiers to form a final decision, and to the best of the authors' knowledge, there is no prior study in the field of NCDS designs that fulfills this purpose. Decision Fusion (DF) has a wide range of applications in healthcare (Niemeijer, Abramoff, & Van Ginneken, 2009; Terveen & Hill, 2001), signal processing (Li, Porter, Santorelli, Popović, & Coates, 2017), image analysis (Velikova, Lucas, Samulski, & Karssemeijer, 2012), biological system activities (Synnergren, Olsson, & Gamalielsson, 2009), disease monitoring (Fung, Chen, & Chen, 2017), drug-target interactions (Peng, Liao, Zhu, Li, & Li, 2015) and many more. DF is lucrative for its role in enhancing different data sources combination and non-uniform data (Fontana, Farooq, & Sazonov, 2014), enhanced decision making (O'Regan & Marnane, 2013), better performance (Acharya, Rajasekar, Shender, Hrebien, & Kam, 2016), and finally, diminution of noise, cost, information drop, and ambiguity (Bossé & Solaiman, 2016; Lelandais, Ruan, Denœux, Vera, & Gardin, 2014; Zhong & Xiao, 2017). There are multiple approaches for combining the outcome of different classifiers and feature sets, among them Decision Template (DT) was selected for this study since it proved to have better performance in experimental studies, especially on lower sample sizes. It was shown that DF by the employment of DTs is independent of uncertain surmise and more immune to overtraining (Kuncheva et al., 2001).

The contribution of this study can be seen from three main aspects: 1. Distinguishing a certain group of pathology from a conglomeration of other pathologies which are closely related is unprecedented in the study of NCDS. Here, we distinguished sepsis from 31 other pathology groups such as Respiratory Distress Syndrome (RDS), meningitis, etc., 2. Extracting the crest and SC from Bark and Equivalent Rectangular Bandwidth (ERB) spectrum and combining them with cepstral analysis is novel in NCDS designs, and 3. Employment of DT in NCDS to combine the result of a Neural Network (NN) classifier, and SVM and KNN which are all trained on different features, is novel in the decision-making stage of NCDS designs.

The importance of newborn sepsis is several-fold; it was amidst the top 10 mortality causes of newborns worldwide that accounted for around 3 million deaths in under-five years old children (World Health Organization, 2021). The diagnosis of sepsis is complex and based on studying different medical cues: feeding difficulty, fever, convulsion, hemodynamic aberrations, apnoea lasting longer than 20 seconds, and lethargy (Mayo Clinic, 2022; Wynn & Wong, 2010). The study and monitoring of these cues requires time and medical equipment; however, time is the most crucial element in treating sepsis to the extent that once a newborn is suspected of sepsis, antibiotic treatment could be started even without validatory results (Ruiz-contreras, Urquia, & Bastero, 1999). Furthermore, availability of the medical monitoring and test equipment is not evenly distributed throughout the world and sadly, the areas that suffer from higher newborn mortality rates are struggling with lack of sufficient professionals and equipment (World Health Organization, 2021). Therefore, the design of a diagnostic system that is non-intrusive and non-complex while being time-efficient and not requiring high computational power or state-of-the-art hardware is of high importance. This study presents a NCDS that while delivering acceptable performance, maintains simplicity and non-invasiveness.

This study is composed in four main sections: an introduction is presented where the problem is highlighted and a short review of literature presents the novelty of the proposed study, the next section expounds the dataset and the proposed methodology including description of features, classifiers, fusion technique, feature selection model, and evaluation measures. After that, the discussion section provides the results of the experiments and compares and discusses these results. Finally, the last section is dedicated to the conclusion.

5.3 Data and Methodology

5.3.1 Dataset Description

Before presenting the details of dataset, it should be highlighted that the most challenging factor in developing the NCDS is data collection and then taking the measures to gain the ethical approvals regarding that data. Even after meeting all the requirements, the occurrence of any pathology could not be anticipated in any time span which means, for example, over a two-year data collection phase one might or might not encounter a newborn diagnosed with meningitis. Therefore, any acquired data is priceless and should be considered for in-depth analysis.

The dataset in this study was collected from newborns with various origins, races, gestational ages (less than 3 months old), cry stimuli, weights, genders, and pathologies. This was made possible with the collaboration of Saint Justine hospital of Montreal, Canada and Al-Raee and Al-Sahel hospitals located in Lebanon. The cry signals were recorded in the presence of noise in both public and private maternity rooms and NICUs in the hospital environment with no predefined conditions. These signals have different lengths from 1 to 4 minutes with an average of 90 seconds including both useful and unwanted information like staff chatter, equipment beeps and crying from other newborns. The equipment used for data collection was a digital 2-channel handheld recorder with 44.1 kHz sampling frequency and 16-bit resolution. The recorder was positioned 10 to 30 centimeters away from the newborn's mouth for the recording process. Up to 5 recordings were collected from each participant. A detailed overview of the database is represented in Table 5.1.

The reason behind putting a limit on the age of newborns is due to the fact that a cry utterance below 53 days of age is only effectuated due to biological rhythms and the newborn has no control over it (Lester & Boukydis, 1985). This may be related to the development of the vocal tract which takes place after 3 months of age when the supralaryngeal reconfiguration occurs, and therefore, no specific incrementing or decrementing pattern was observed in the average fundamental frequency of the cry signals (Fisichelli et al., 1974; Lind & Wermke, 2002).

Table 5.1 An overview of the database Participants

Gender	Female and Male
Babies Ages	Less than 53 days old
Weight	0.98 to 5.2 Kg
Origin	Canada, Haiti, Portugal, Syria, Lebanon, Algeria, Palestine, Bangladesh, Turkey.
Race	Caucasian, Arabic, Asian, Latino, African, Native Hawaiian, Quebec.
Cry stimulus	Discomfort, sleepiness, wet diaper, pain, fear, colic, reflux, birth cry, hunger.
Pathology Group	Ankyloglossia, apnea, asphyxiation, aspiration, bronchiolitis, choanal atresia, cleft palate and lip, complex cardio, cyanosis, down syndrome, duodenal atresia, dyspnea, fever, gastroschisis, grunting, hyperbilirubinemia, hypoglycemia, hypothermia, intrauterine growth retardation, kidney failure, meconium aspiration syndrome, meningitis, myelomeningocele, respiratory distress syndrome, retraction, seizure, sepsis, tachypnea, thrombosis in vena cava, vomit.

5.3.2 Pre-processing

In the previous section, it was mentioned that there is no guarantee that a newborn diagnosed with a certain pathology group would be observed over a prespecified time span. Therefore, upon acquiring data from a certain pathology group, it is desirable to make deeper use of the data by any possible means. The cry signals in our dataset were segmented based on the physiological differences of the acoustic activities during a cry utterance and different labels were assigned to each segment. The two main acoustic activities for example are EXP which refers to an expiratory cry and INSV which refers to a voiced inspiratory cry unit. These labels for the bounded segments were appointed by the means of WaveSurfer (version 1.8.8, Stockholm, Sweden) software by the group of researchers in our lab. Furthermore, the outlier samples and those with a length of less than 17 ms (equal to the length of two overlapping windows of 10 ms with a 30% overlap) were omitted. This condition was applied to ensure having a reliable analysis of the dataset.

The number of samples in each group of our study is presented in Table 5.2, it should be noted that the numbers in Table 5.2 denote the values before selecting an equal number of samples in order to have an approximately balanced dataset. In total, 2264 samples (1132 from each class) were selected which formed a training dataset with 1585 (789 septic and 796 healthy) samples and a test dataset with 679 (343 septic and 336 healthy) samples.

Table 5.2 Specifications of the dataset

	No. of participants	**Available time (s)**	**Average duration of samples (s)**	**No. of samples selected**
Septic	17	1773.66	0.71	1132
Other pathologies	110	10712.28	0.52	1132

5.3.3 Feature Extraction

Feature extraction has the highest significance in the design of a NCDS framework as it can change the course of following steps and affect the final decision. Moreover, the nature of a cry signal is dynamic, non-stationary, and disparate from both speech and music to some extent, while including noise. Therefore, the extraction of features that can represent the cry signal both from the spectral and short-term perspective and originate from the domains of speech processing and music analysis would be of the essence. Moreover, as it was aforementioned, since the cry signal is emanated in the nature of speech generation, employing the cepstral analysis would be inevitable. Consequently, this study combines psychoacoustic-based warping of the spectrum with cepstral analysis for the short-term analysis of the signal and studies the dynamic nature of the signal through delta and delta-delta coefficients of the Bark and Gammatone scales. Additionally, in order to capture the spectral properties of the cry signal and explore it from the musical perspective, SC and crest features were extracted.

The bark and Equivalent Rectangular Bandwidth (ERB) or Gammatone scales were developed as psychoacoustic-based spectral measures. The Bark scale for frequency f is given by Equation 5.1 and Equation 5.2, respectively:

$$Bark(f) = 6\, ln\left[\frac{f}{600} + [(f/600)^2 + 1]^{0.5}\right] \tag{5.1}$$

$$ERB(f) = 21.4\, \log(0.00437f + 1) \tag{5.2}$$

The ERB scale was chosen since it assists the study of lower frequencies with higher resolution. Besides, it was shown in several studies that ERB scaling resulted in better performance of non-speech classification problems, which is accompanied by more robustness and lower computational costs when compared to the triangular bands that are conventionally employed in MFCC feature extraction (Gulzar, Singh, & Sharma, 2014; Smith & Abel, 1999). In order to attain GFCCs, the cry signal is first windowed into overlapping Hamming filters of 10 ms with 3 ms overlap length; since the performance of feature extraction step is enhanced and that the non-stationarity of the signal could be neglected in such short frames. Then, in order to pre-emphasize the valuable signal frequencies, the signal passes the GT filters after a fast Fourier Transform (FFT) was applied. The final steps of extracting the GFCCs constitute employing the log function and then the DCT to decorrelate the compressed outputs of the previous steps. For a given frame k, the GFCCs can be computed through Equation 5.3:

$$GFCC_k = \sqrt{\frac{2}{N}} \sum_{n=1}^{N} GF[k]\, cos\left(\frac{i\pi}{2N}(2c + 1)\right), \; 1 \le k \le M \tag{5.3}$$

where GF[k] denotes the loudness-compressed response of the Gammatone Filters (GF), and the number of filters is given by N (Valero & Alias, 2012).

The width of the critical bands of the human auditory system equals 1 bark and hence, a more direct correlation with the spectral information processing of the human auditory system is achieved when the spectral energy is warped over the Bark scale (Smith & Abel, 1999). The process of extracting BFCCs is identical to GFCC feature set with only the Bark scale being the difference. Similar to GFCCs, the BFCC feature set constitutes of 39 elements.

SC is an indicator of how the signal's spectrum looks and where the majority of its mass lies. The average of SC is shown to be a powerful discriminator in audio signals, especially in the field of musical applications (Kulkarni, 2018). In order to calculate the SC of a given window i, we should take the weighted average of the frequency bins as shown in Equation 5.4:

$$\mathrm{BSC} = \frac{\sum_{\mathrm{k=1}}^{\mathrm{H/2}} f(k)\,|A(k)|}{\sum_{\mathrm{k=0}}^{\mathrm{H/2}} |A(k)|} \tag{5.4}$$

Where $|A(k)|$ is the amplitude at the corresponding bin k, H is the number of points in the Fourier transform, and *f(k)* is the frequency at the kth bin (Brent, 2010). Note that the frequencies have been mapped to the Bark scale prior to the computation of the SC, therefore, we name this feature set Bark Spectral Centroid or BSC.

Finally, we extracted the ERB-based Spectral Crest (ERBS Crest) which points out the level of peakiness in the spectrum of the signal. A higher value of crest signifies the presence of more peaks in the signal. It is calculated as the peak amplitude in each window divided by its RMS value as written in Equation 5.5:

$$crest = \frac{max\,(y_{q \in [f_1, f_2]})}{\frac{\sum_{k=f_1}^{f_2} y_q}{f_2 - f_1}} \tag{5.5}$$

The spectral value at a given bin is shown by y_q whereas f_1 and f_2 mark the edges of the corresponding window (Hosseinzadeh & Krishnan, 2007; Peeters, 2004).

5.3.4 Classification

In this step, all the feature sets were fed to the three distinguished classifiers so that their performances would be tested and, also the best classifier + feature sets would be selected for the fusion step. All the classifiers benefitted from validation. SVM and KNN were validated using 5-fold cross-validation and the MLP network was validated iteratively during the learning process. The validation secures the classifiers against over-fitting and increases their reliability. Finally, all of the classifiers were optimized using random search.

5.3.4.1 Multilayer Perceptron (MLP)

A MLP classifier has four main components; firstly, the extracted features are fed to the input of the network, and then they are conveyed forward across the layers. A backpropagation method is employed in order to update the weights of the network and an optimization function assists the tuning of the weights' update (Murtagh, 1991). The decision in a MLP network is made based on having the minimum distance from the decision boundary hyperplane (James, Witten, Hastie, & Tibshirani, 2013). In this study, Root Mean Square Propagation (RMSprop) optimization function updates the backpropagation weights by the means of minimizing the distance to the decision boundary hyperplane (Hinton et al., 2012). In order to further improve this classifier, random search hyperparameter optimization was employed. The number of input layer neurons was set according to the feature vectors' sizes. The hidden layer consisted of 128 fully connected neurons which is accompanied by a normalization layer. In order to specify whether the neurons would fire during the process of learning, a hyperbolic tangent activation function was added to the layers. The output layer was made of a fully connected layer with two nodes that represent the two classes of septic versus other pathologies, and a sigmoid function that is in charge of translating the raw outputs of all the layers into class probabilities. Finally, the classification layer generates the final decision of class labels based on the class probabilities. Other details regarding the learning process are the learning rate of 0.001 along 120 epochs, validation data which included a 15% random share of all data. 30%

of the data was randomly selected for testing and separated from the dataset and finally, 55% of the data was randomly chosen for training.

5.3.4.2 Support Vector Machine (SVM)

As mentioned before, SVMs are one of the most well-known classifiers and have a wide range of applications, especially in the analysis of the audio signals. SVMs are precise, versatile and capable of dealing with linear and non-linear data. In order to classify data points, SVM attempts to build a hyperplane that is able to separate the data points of the two classes as far as possible and if the data is not divisible linearly, the Radial Basis Function (RBF) kernel which computes the Euclidean distance is chosen (Winters-Hilt, Yelundur, McChesney, & Landry, 2006).

5.3.4.3 K-Nearest Neighborhood (KNN)

KNNs are known for their simplicity and effectiveness. As the name suggests, the basis of classifying the data points is measuring the distance from the neighbors where each point would be placed in the same class as its neighbors with the lowest distance. There are three elements in a KNN: the distance measure (which can be Minkowski, standard Euclidean, Euclidean, Jaccard, Hamming, cosine, Chebyshev, and Manhattan); the number of neighboring data points K; and sets of labeled data for training and testing (Wu, Kumar, Ross Quinlan, et al., 2008).

5.3.5 Fusion Using Decision Templates

Suppose that in a classification problem with l classifiers $C = \{C_1, C_2, \dots, C_l\}$, $X = [x_1, x_2, \dots, x_n]^T$ denotes the n-dimensional input feature vector, which corresponds to the *m* class labels $W = \{w_1, w_2, \dots, w_m\}$. Each *i*th classifier will produce an output where $C_i(X) = \left[c_{i,1}(X),\ c_{i,2}(X),\ \dots,\ c_{i,m}(X)\right]^T$. Here, $c_{i,j}(X)$ represents the posterior probability that the *i*th classifier suggests that X belongs to the class ω_j.

In order to fuse the outputs of the classifiers, a $l \times m$ Decision Profile (DP) is constructed as shown in Equation 5.6:

$$DP(X) = \begin{bmatrix} c_{1,1}(X) & \cdots & c_{1,m}(X) \\ \vdots & \ddots & \vdots \\ c_{l,1}(X) & \cdots & c_{l,m}(X) \end{bmatrix} \quad (5.6)$$

Each column j shows the possibility that a collective of l classifiers declare that X corresponds to the class label ω_j. Finally, the result of fusion would be in the form of a vector of length m as shown in Equation 5.7:

$$C(X) = [d_1(X), d_2(X), \dots, d_m(X)]^T \quad (5.7)$$

For $d_i(X)$ denotes the possibility that the result of fusion declares the input X to belong to class ω_i The final decision is made based on a certain rule of fusion such as min, max, median, product, and sum operating each corresponding column of the DP matrix to yield the Decision Templates (DT). Here, the min rule was chosen which is given in Equation 5.8:

$$d_j(X) = \min_{i=1:l} c_{i,j}(X)\,,\; j = 1,2,\dots,m. \quad (5.8)$$

Thereafter, the DTs are calculated as shown in Equation 5.9, where z_j denotes the samples that are from class ω_i in the training set Z, and the number of z_j is given by N_z.

$$d_j(X) = \min_{i=1:l} c_{i,j}(X)\,,\; j = 1,2,\dots,m. \quad (5.9)$$

The input's labels are decided based on a similarity measure between the DP and different DTs. In this study, the Euclidean distance was selected as the similarity measure. Equation 5.10 shows the calculation for determining the output labels based for a given sample P.

$$d_j(X) = \sum_{j=1}^{m} \sum_{k=1}^{l} \left(c_{k,j}(P) \; - \; dt_i(k,j) \right) \tag{5.10}$$

$dt_i(k,j)$ stands for the element marking the intersection of the column j and row k (Kuncheva et al., 2001; Mi, Wang, & Qi, 2016). Figure 5.1 shows the design of our NCDS employing the DT fusion technique.

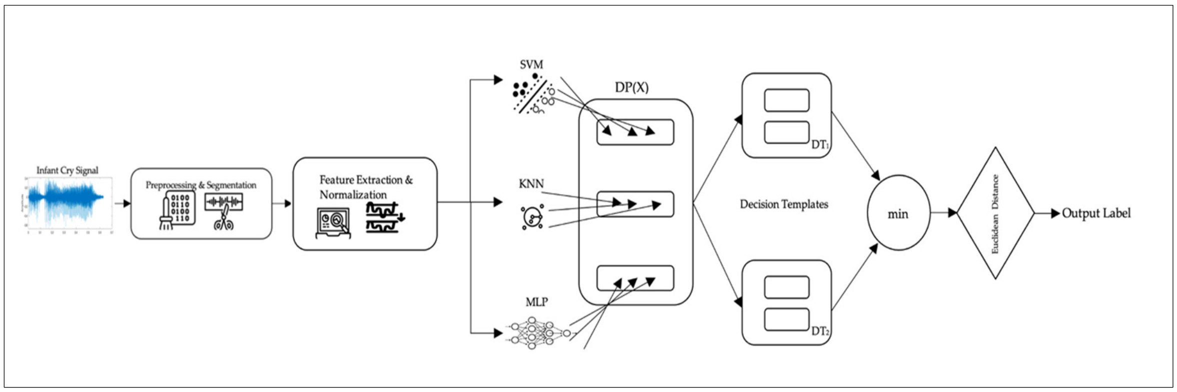

Figure 5.1 Design of the NCDS employing DT fusion

5.3.6 Neighborhood Component Analysis Feature Selection

As a final experiment, the Neighborhood Component Analysis (NCA) was implemented to determine which elements of the feature sets contributed the most to the final classification results. NCA is non-parametric and aims to enhance the accuracy of the classification to its peak performance. The general performance of the NCA can be explained as a KNN classifier where K=1 and neighbours are chosen randomly so that there is a probability for each point in the feature space to be chosen as the reference point. The goal is to learn a classifier that predicts the true label y of x based on the features fed to the input by selecting a random point, Ref(x) from the training set as the reference point and deciding the label of the point x based on this reference point Ref(x).

The chance of any given point x_j to be picked as the reference point is evaluated based on a weighted distance function, d_w, which is given by Equation 5.11:

$$d_w(x_i, x_j) = \sum_{r=1}^{p} w_r^2 \left|x_{ir} - x_{jr}\right| \tag{5.11}$$

where w_r denotes the weight for the rth feature and p denotes the feature dimension of x_i. In order for the nearest neighbor classifier to perform desirably, one suitable way is to maximize its leave-one-out accuracy. This would not be practical since the selection of the nearest neighbor as the reference point in the leave-one-out accuracy would result in a non-differentiable function. Therefore, the approximation where the reference point is in the form of a probability distribution would be effective. Equation 5.12 presents the probability p_{ij} that a given point x_i gets x_j as the reference point.

$$p_{ij} = \begin{cases} \frac{\kappa\left(d_w(x_i, x_j)\right)}{\sum_{k \neq i} \kappa\left(d_w(x_i, x_k)\right)}, & if\ i \neq j \\ 0, & otherwise \end{cases} \tag{5.12}$$

where κ is a kernel function $\kappa(z) = exp\left(-\frac{z}{\sigma}\right)$ that affects the probability of any given point being selected as the reference point through the kernel width σ which is determined by an input. In other words, if $\sigma \longrightarrow \infty$ all the points in the training set have equal probability to be chosen as the reference point; and, if the $\sigma \longrightarrow 0$ only the nearest neighbor has the chance of being the reference point. Hence, the probability of the correct classification of the query point x_i is given by Equation 5.13:

$$p_i = \sum_j p_{ij} y_{ij} \tag{5.13a}$$

where,

$$y_{ij} = \begin{cases} 1, & y_i = y_j \\ 0, & otherwise \end{cases} \tag{5.13b}$$

Now the leave-one-out classifier's accuracy can be approximated by Equation 5.14:

$$\mathrm{F(w)} = \frac{1}{N}\sum_i p_i = \frac{1}{N}\sum_i \sum_j p_{ij} y_{ij} \tag{5.14}$$

In order to prevent the classifier from overfitting, a positive-valued regularization term λ is added to the object function that can affect the influence of the weights and will be tuned via cross-validation. The objective function can now be written as Equation 5.15.

$$F(w) = \sum_i \sum_j p_{ij} y_{ij} - \lambda \sum_{l=1}^{p} w_r^2 \tag{5.15}$$

The $\frac{1}{N}$ term is ignored since it does not affect the solution vector. Finally, in order to maximize the objective function, its derivative with respect to the feature weights is taken as shown in Equation 5.16:

$$\begin{aligned} \frac{\partial F(w)}{\partial w_r} &= \sum_i \sum_j y_{ij} \left[\frac{2}{\sigma} p_{ij} \left(\sum_{k \neq i} p_{ik} |x_{ir} - x_{kr}| - |x_{ir} - x_{jr}|\right) w_r\right] - 2\lambda w_r \\ &= 2 \left(\frac{1}{\sigma} \sum_i (p_i \sum_{k \neq i} p_{ik} |x_{ir} - x_{kr}| - \sum_j p_{ij} y_{ij} |x_{ir} - x_{jr}|) - \lambda\right) w_r \\ &= 2 \left(\frac{1}{\sigma} \sum_i (p_i \sum_{k \neq j} p_{ij} |x_{ir} - x_{jr}| - \sum_j p_{ij} y_{ij} |x_{ir} - x_{jr}|) - \lambda\right) w_r \end{aligned} \tag{5.16}$$

The above equation is the basis of the NCA features selection (Yang et al., 2012). In this study, in each of the feature sets, the features that accounted for more than 80% of the final classification results were extracted from the set. Then, these features were concatenated in a single vector and fed to the classifier to determine the role of NCA.

5.3.7 Evaluation Measures

The framework of this study was designed and developed with the goal of identifying septic infants from a collective of several other pathologies for the first time. The features and classifiers are used from diverse natures and origins, and in order to compare their performance several evaluation measures are introduced in this study. The first measure that is used in any binary classification problem is accuracy due to its simplicity of calculation and being straightforward. Accuracy is computed from the ratio of correct predictions over all the samples. However, this measure is not illuminating enough to cover all aspects of system performance and more measures are needed to study other aspects of the problem (Hossin & Sulaiman, 2015). Therefore, two other measures, namely precision and specificity were studied alongside accuracy. Specificity shows the rate of true negative cases which translates of to the number of the cases that were correctly marked as non-septic, and precision or Positive Predictive Value (PPV) denotes the proportion of true septic samples which were marked as non-septic by the NCDS (Zhu et al., 2010). F-score measure is highly instructive as it summarizes these measures into one single value from calculating the harmonic mean of PPV and True Positive Rate (TPR) (Flach & Kull, 2015).

These measures evaluate the performance of the system from the problem-solving perspective; however, the system can also be assessed with regards to its classification performance. Therefore, one final evaluation measure is added to our evaluation criteria which is Matthews' Correlation Coefficient (MCC). MCC helps elucidate all the information from a contingency matrix (true negative or TN, true positive or TP, False Negative or FN, and False Positive or FP) as they are all taken into account for the calculation of MCC, Equation 5.17. The value of MCC can be anything in the range of [-1, +1] where the negative value represents a misclassification, zero signifies random classification and the higher positive values translate into better classification performance (Chicco & Jurman, 2020; Vihinen, 2012).

$$F(w) = \sum_i \sum_j p_{ij} y_{ij} - \lambda \sum_{l=1}^{p} w_r^2 \tag{5.17}$$

5.4 Results and Discussion

The NCDS in this study was designed and developed with the purpose of identifying the septic newborns among an ensemble of other pathologies for the first time in NCDS designs. The features employed for the NCDS were ERBS Crest, BSC, GFCC, and finally, BFCC. These features were fed to three classifiers namely SVM, KNN, and MLP. As it was discussed before, in this section, the result of classifying each feature set with different classifiers will be presented firstly, Table 5.3 to Table 5.6. In order to fuse the results of different features fed to various classifiers, the set of feature + classifier that resulted in the highest accuracy measure was selected to form the DPs and DTs. These sets are highlighted in each of the Tables below. As for the feature sets of BSC and ERBS Crest, the low dimensionality of the feature vectors prevented the MLP classifier from converging which was expected and MLP was more suitable for BFCC and GFCC features sets that had a larger size (Khalilzad, Hasasneh, et al., 2022).

Table 5.3 Results for the classification of the BFCC feature set with KNN, SVM, and MLP classifiers

Classifier	Accuracy	Specificity	Precision	F1-Score	MCC
KNN	67.01	72.01	68.42	65.00	0.34
SVM	70.99	70.26	70.26	70.99	0.42
MLP	83.51	87.46	86.13	82.66	0.67

Table 5.4 Results for the classification of the GFCC feature set with KNN, SVM, and MLP classifiers

Classifier	Accuracy	Specificity	Precision	F1-Score	MCC
KNN	84.09	88.63	87.25	83.18	0.68
SVM	76.14	76.97	76.20	75.75	0.52
MLP	82.92	84.55	83.74	82.48	0.66

Table 5.5 Results for the classification of the BSC feature set with KNN and SVM classifiers

Classifier	Accuracy	Specificity	Precision	F1-Score	MCC
KNN	63.18	40.48	85.42	73.12	0.29
SVM	53.61	63.69	43.73	52.58	0.08

Table 5.6 Results for the classification of the ERBS Crest feature set with KNN and SVM classifiers

Classifier	Accuracy	Specificity	Precision	F1-Score	MCC
KNN	61.56	86.30	72.19	48.32	0.26
SVM	77.91	56.27	69.14	81.75	0.62

The results for the classification of data with BFCC feature set is given in Table 5.3. The BFCC feature set showed great potential with the highest accuracy of 83.51% and an MCC of 0.67 with the MLP classifier. By taking a look at Table 5.4, it can be seen that the combination of GFCC + MLP was outperformed by the BFCC feature set with the same classifier. Moreover, Table 5.3 shows that all of the classifiers had positive values for MCC and hence, the classification was performed successfully. The SVM and KNN classifiers were less efficient in terms of all evaluation criteria. Therefore, the set of BFCC + MLP was chosen for the DT and DP calculations of the next experiment. Table 5.4 presents the results for the GFCC feature set. The best evaluation measures were achieved through the combination of the GFCC and KNN classifier with 84.09% for the accuracy and 0.68 for the MCC measure. It is also worth mentioning that this set remarkably achieved the highest values across all the experiments from the first part. Another point worth highlighting is that not only GFCC has the overall best performance among feature + classifier combinations, but also, it has shown interestingly higher performance with simpler classifiers (SVM and KNN) compared to all of the other combinations in this study. This could signify higher efficiency of the GFCC compared to other feature sets.

The results of evaluating the NCDS with the BSC and ERBS Crest feature sets are given in Table 5.5 and Table 5.6, respectively. It should be highlighted that even though both feature sets had lower performances compared to the GFCC and BFCC feature sets, they only have low dimensions of four elements which proves their favourable outcomes. The ERBS Crest feature set had better performance than the BSC feature set overall; however, each of these feature sets responded better to different classifiers. The highest results achieved for the BSC feature set was via KNN classifier with 63.18% and 0.29 for the MCC which was the combination selected for the next step. The SVM + ERBS Crest had the highest values across evaluation measures of accuracy and MCC with 77.91% and 0.62, respectively.

In order to fuse the outputs of the highlighted feature + classifier sets, the corresponding posterior probabilities of training and test datasets, as well as the training labels were recorded to form the DPs and DTs. The result of the DT technique for fusion is presented in Table 5.7 along with the best feature + classifier sets for a clearer interpretation.

Table 5.7 Results for the DT fusion technique showing the best feature + classifier sets selected

Classifier	**Feature Set**	**Accuracy**	**Specificity**	**Precision**	**F1-Score**	**MCC**
SVM	ERBS Crest	77.91	56.27	69.14	81.75	0.62
MLP	BFCC	83.51	87.46	86.13	82.66	0.67
KNN	BSC	63.18	85.42	73.12	52.11	0.29
KNN	GFCC	84.09	88.63	87.25	83.18	0.68
	Fusion	**88.66**	**88.34**	**88.20**	**88.59**	**0.77**

Figure 5.2 and Figure 5.3 illustrate a comparison of how each feature set and its corresponding evaluation measures were impacted by the fusion. As it can be interpreted from Table 5.7, Figure 5.2, and Figure 5.3, the result of the fusion framework enhanced the results in all cases with an average of 11.49% for accuracy and 13.67% for the F-score. There is only one exception to this conclusion where the specificity measure for GFCC + KNN set was higher by 0.29%, which is negligible. Even the best results of the first step of these experiments had a 4.57% and 0.09 enhancement in the accuracy and MCC, respectively, with the DF method. The DF method imposes negligible computational cost on the system and is very fast since its calculations only take less than a second. The above results prove the high potential of this method for the design of multimodal NCDS as presented in this study where both spectral and short-term features were extracted and employed from musical and speech processing origins. Moreover, as it was mentioned before, the enhancement is consistent across different evaluation measures.

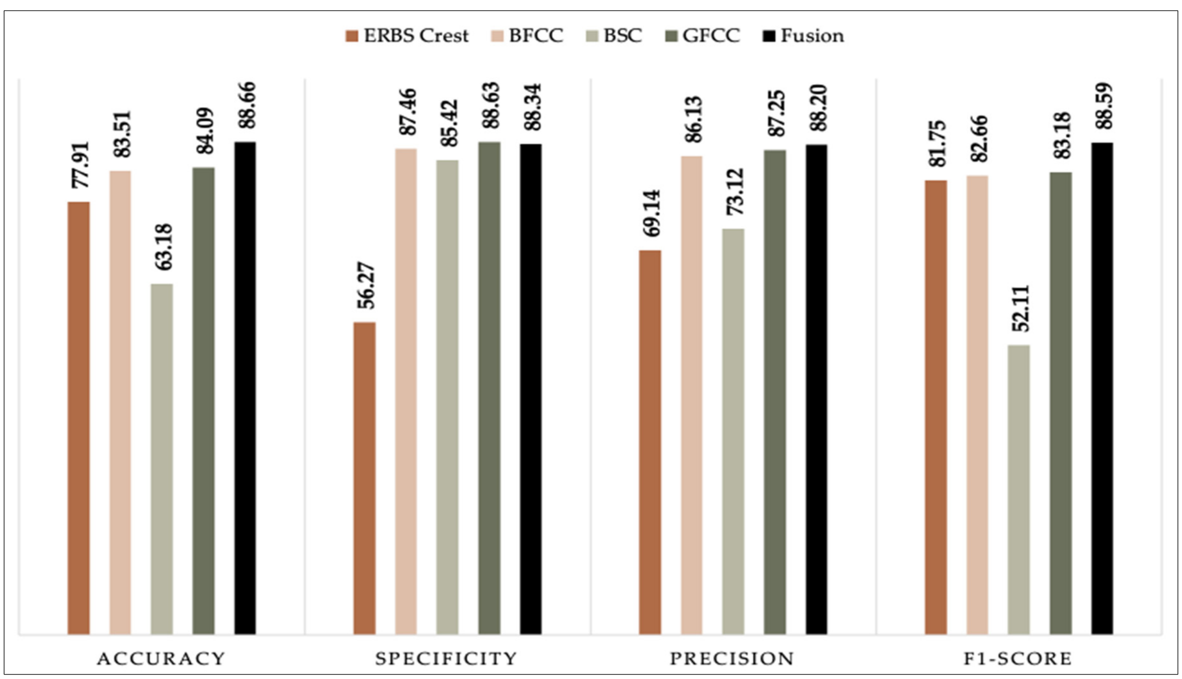

Figure 5.2 Evaluation measures for the best feature + classifier sets and the DT fusion framework

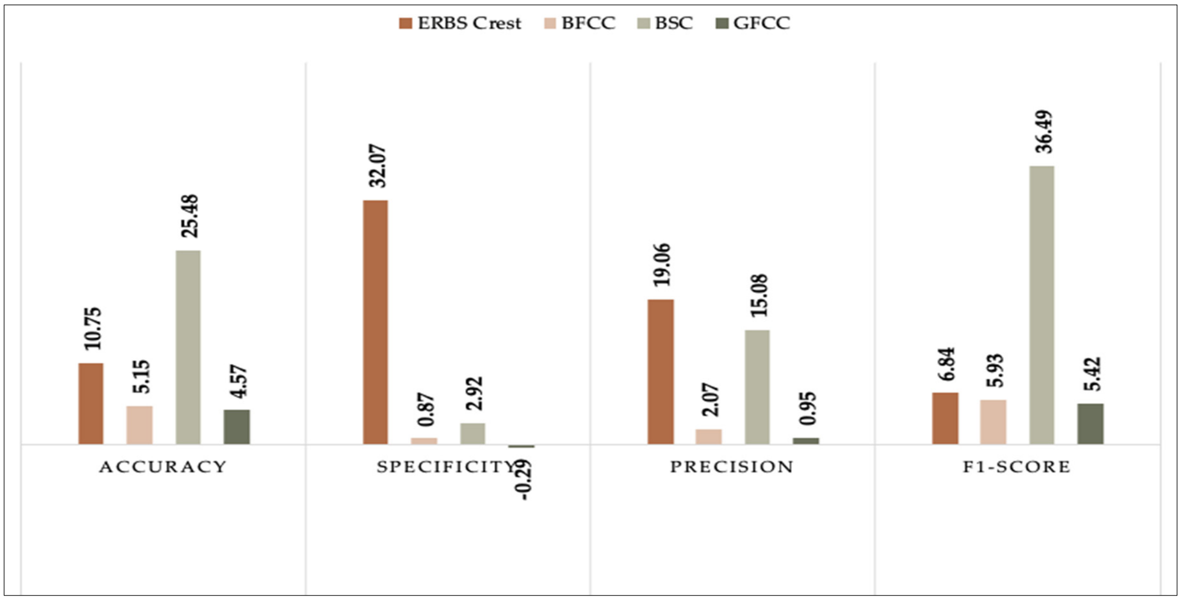

Figure 5.3 Comparison of the evaluation measures of the selected feature + classifier sets to the DT fusion technique

Another experiment was carried out to study the role of feature selection and to evaluate to what extent the feature space could be compacted. Each feature set was analyzed with the NCA method and the features that had the highest contribution to the final classification results were selected. Figure 5.4 shows the feature indices that were selected in each feature set.

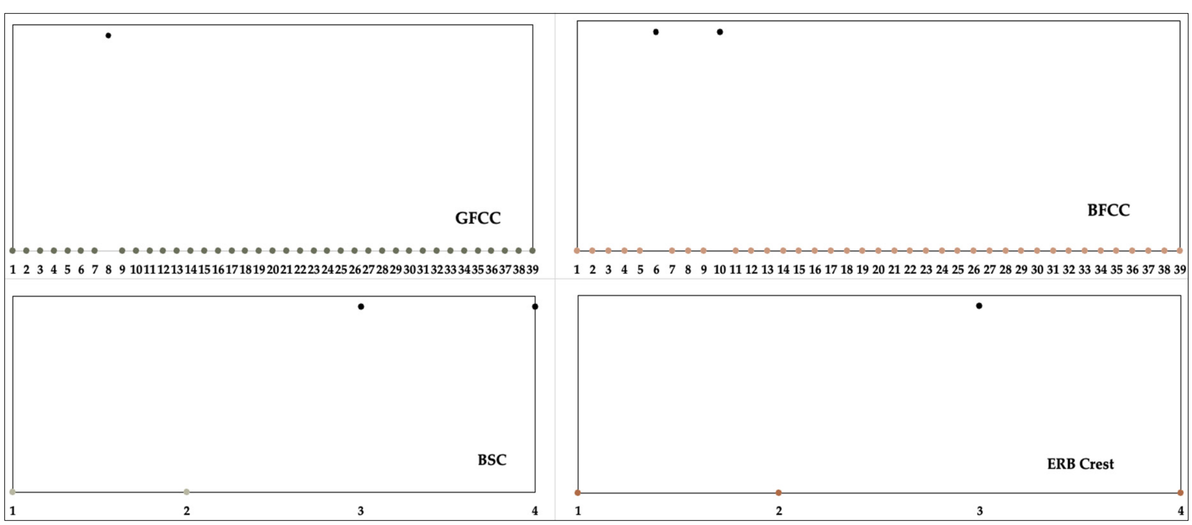

Figure 5.4 NCA feature selection for each feature set, showing which indices were selected to form the final feature vector

The NCA revealed some details worth explaining here. The study of both BFCC and GFCC showed that the most significant information in these feature sets belongs to the first 13 coefficients and not their deltas, this was shown in another study on a similar subject (Khalilzad, Hasasneh, et al., 2022), where only 13 GFCC coefficients resulted in around 93% accuracy of classification. Furthermore, the BSC and ERBS Crest feature sets both had their 3rd elements selected. The 3rd element in both feature sets belonged to the H-speed or interquartile range. It can be deduced that this statistical measure has high potential in representing and summarizing spectral data for the infant cries. As for the BSC feature set, the 4th element was also selected which denotes the median statistical measure.

In order to evaluate the system performance with these feature sets, the selected elements from each vector were all concatenated in a single vector and fed to the classifiers of this study. Once more, due to the low dimensionality of the feature vector, the MLP classifier did not converge. The result of the classification of the NCA-selected feature vector with SVM and KNN classifiers is presented in Table 5.8.

Table 5.8 Results for the NCA feature selection method with KNN and SVM classifiers

Classifier	Accuracy	Specificity	Precision	F1-Score	MCC
KNN	86.22	92.19	90.94	85.18	0.73
SVM	78.20	62.10	70.98	81.12	0.60

Although the results are lower than DF method, they still represent a high potential and the success of the NCA. In comparison to the GFCC feature set, the accuracy was enhanced by 2.13% and the MCC by 0.05. The enhancement suggests several points: firstly, the use of NCA will not have a detrimental effect on the final performance of the NCA. Moreover, it has shown that combination of the features from the domains of speech and music would improve the NCDS. Finally, the NCA feature set has only six elements and could obtain an accuracy of more than 86% which is exceptional. Figure 5.5 shows a comparison between the results of the two frameworks represented in this study, DF and NCA feature selection. As can be seen from the graphs, not only the results of the two frameworks are compatible, but also the results for specificity and precision measures shows better performance with the NCA feature selection method. It can be discussed that the feature selection leads to a more uniform structure of the feature space, extracting the essence of what each feature set represented and combining these elements together formed a more powerful indicator in terms of these two measures.

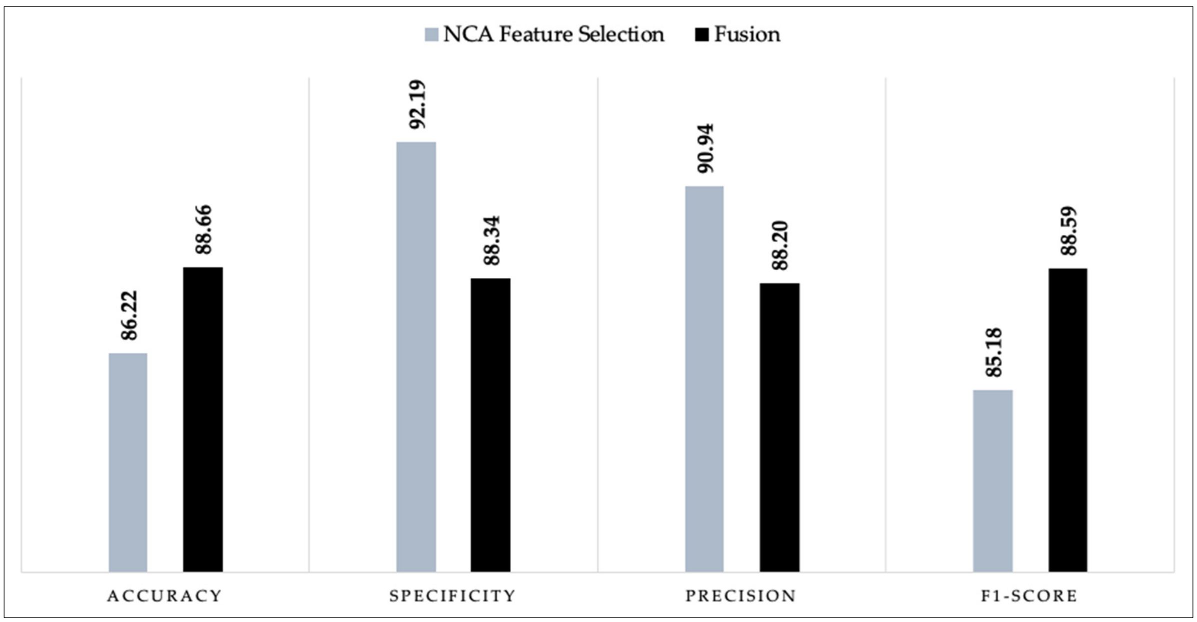

Figure 5.5 Comparison of the evaluation measure for fusion framework and the NCA feature selection method

This study served three purposes: 1. Assessing the role of decision-level fusion in NCDS designs for the first time, 2. Assessing the role of NCA feature selection in forming a highly compacted feature set while keeping acceptable performance which is novel in NCDS designs, and 3. Distinguishing a certain pathology (sepsis) amid a collective of other pathologies that is unprecedented in cry analysis studies.

In addition to the fact that many pathologies remain unexplored or not well-studied in the field of cry diagnostic applications, the NCDS itself has a great potential for further development compared to other audio recognition applications. One of the main areas that could play a part in this development is investigating whether a fusion of different modalities would contribute to the enhancement of the final decision made by the NCDS, which was the purpose of this study. This framework opens the door for employing features and classifiers from various modalities without the need for complicated designs and advanced technology. The importance of keeping the design simple arises from the fact that unfortunately, the regions that are reported as having higher newborn mortality rates suffer from lack of adequate medical equipment and professionals and are listed among low-income and middle-income areas. Therefore, if the NCDS can candidate the newborns with higher risk of suffering from certain pathologies, especially sepsis, and rule out the others, the existing equipment and experts can tend to the newborns marked with higher risk.

Although sepsis is closely entangled with newborn mortality rates (World Health Organization, 2021), the number of newborn cry studies targeting sepsis is scant. In order to address this research gap, the researchers in our lab made efforts to study sepsis from different perspectives. Matikolaie et al. (Matikolaie & Tadj, 2022) utilized prosodic features to distinguish between healthy and septic infants and attained 86% for the best F-score. Khalilzad et al. (Khalilzad, Kheddache, et al., 2022) introduced an entropy-based framework by extracting the spectral entropy cepstral coefficients and then having a fuzzy entropy as the feature selection means for the identification of septic infants from the healthy group, obtaining 88.51% for the accuracy regarding the expiratory cries. In another study, Khalilzad et al. (Khalilzad, Hasasneh, et al., 2022) differentiated between RDS and septic cries through the combination of music-derived features of Harmonic Ratio (HR) and GFCC features that yielded 95.29% for the accuracy.

Up to this point, the NCDS performance was compared regarding its performance with different features and classifiers, before and after applying the DF method and NCA feature selection. Also, the studies that scrutinized sepsis via cry signals were compared in terms of the methods, their purposes, and their performance with the accuracy or F-Score measures. This framework could also be compared to the existing literature in terms of the methods employed here and in other NCDS designs. Table 5.9 presents a short comparison of the proposed framework with other similar works in the literature.

Table 5.9 Comparison of different works employing fusion and feature selection techniques

Study	Goal	Features	Fusion/Feature selection	Machine/ Deep Learning Methods	Best Outcome
Dar et al. (Dar, Srivastava, & Lone, 2022)	Detecting pulmonary abnormalities from the respiration sound	BFCC, SC, and spectral flux	Simple concatenation of features.	Hierarchical Attention Network, CNN, RF	92.4% accuracy by HAN.
Ebrahimpour et al. (Ebrahimpour & Hamedi, 2009)	Recognition of hand-written digits in Persian and English.	Characteristic Loci.	DT Fusion and PCA.	MLP, decision tree, RBF.	Accuracy: 97.99%.
Fernandes et al. (Fernandes & Apolinário Jr, 2020)	Identifying underwater targets based on acoustical recordings from a hydrophone.	GTCC, LPC, and MFCC	NCA	KNN.	Accuracy: 83.3%.

Study	Goal	Features	Fusion/Feature selection	Machine/ Deep Learning Methods	Best Outcome
Li et al. (Li et al., 2017)	Detecting breast cancer via microwave breast screening.	PCA scores.	DT fusion, Concatenation, PCA.	SVM.	Average error: 0.01.
Khalilzad et al. (Khalilzad, Kheddache, et al., 2022)	Detecting septic newborns from healthy newborns via their cry signals.	MFCC, spectral entropy cepstral coefficients, SC cepstral coefficients.	Fuzzy entropy feature selection.	SVM, KNN.	Accuracy: 91.81%.
Khalilzad et al. (Khalilzad & Tadj, 2023)	Detecting pathologic newborns based on their cry signal.	MFCC, GFCC.	Canonical Correlation Analysis-based feature fusion	LSTM, SVM.	Accuracy: 99.86%.
This Study	Detecting septic newborns among other pathologic newborns	ERBS Crest, BSC, GFCC, BFCC.	NCA feature selection, DT fusion, concatenation.	SVM, KNN, MLP.	Accuracy: 88.66%

As can be interpreted through all the aforementioned studies, each of the tools that was implemented in the proposed framework, has shown great performance with different applications. The results of our framework also suggest the promising potential of studying DT fusion and NCA feature selection methods for further studies in NCDS development. In future, it would be great to explore the role of features from different modalities, and more classifiers with the proposed framework here. Moreover, it would be fruitful to investigate how fusion at each level would affect the outcome of the system.

5.5 Conclusion

The cry signal is a powerful biomarker for studying the physical health and needs of a newborn. This study aimed to introduce a simple yet effective framework that was capable of capturing different aspects of the septic cry in comparison to a variety of other pathologies. The study of cry signals was performed independent of newborns' race, gender, weight, and the reason of their crying. The cry signal is different from both speech and music yet shares so many common attributes with both. Via implementing features from different modalities and properties, both aspects of the cry were studied and each of the four introduced feature sets of BSC, ERBS Crest, GFCC, and BFCC showed desirable performance individually. Then, through the DT fusion technique these feature sets were fused and the outcome surpassed the results of all the individual feature sets by an average of 11.49% for the accuracy measure, reaching up to 88.66% which marks a notable increment and potential.

In order to achieve a more simplistic design and take a deeper look into each of the introduced feature sets, the NCA feature selection method was employed where each of the feature sets were analyzed and the indices that contributed the most to the final result were chosen. Next, all of the selected indices were concatenated to form a single feature vector that achieved 86.22% for the accuracy measure.

This study aimed to design an unsophisticated NCDS that served as an alert system to the medical experts for prioritizing the newborns with higher risk of being diagnosed with the fatal pathology of sepsis. The proposed framework showed that septic newborns could be effectively distinguished among a collective of other pathologies only based on their cries. Therefore, this framework could be employed as a non-invasive tool for diagnosis of sepsis and other pathologies.

CONCLUSION

This thesis proposed a comprehensive and non-intrusive NCDS that highlights the cry signals as a biomarker in the diagnosis of the pathologies that would otherwise require invasive tests and precise -and exorbitant- medical equipment, or pathologies and complications subject to the prompt detection and treatment. Sepsis is one of the leading post-partum pathologies contributing to the mortality rates worldwide and inevitably, is the focal point in this study. The proposed NCDS here attempts to comprehensively address sepsis by studying the cry signals of septic newborns thoroughly as opposed to three different groups of newborns: 1. Healthy newborns, 2. Newborns suffering from RDS, and 3. A collective of newborns diagnosed with other pathology groups. Furthermore, to advance the NCDS and proffer it as an assistant to the medical experts and guardians of the newborns, the NCDS was designed to distinguish the healthy newborns from an assemblage of pathologies -including sepsis-. Sadly, the geographic distribution of the mortality rate is proportionate to the national income of the countries around the world, which means that low- and middle-income regions suffer from the highest newborn mortality rates. As a matter of course, this translates to lower number of pediatricians and medical equipment and, ergo, lower chances of post-partum screening and care for the newborns in these regions. However, if there would be a simple and swift method to rank the newborns as being high-risk or mark them as potentially suffering from a specific pathology, the existing apparatus could be employed for validation followed by all the necessary procedures to minister the treatment of the newborn. NCDS would conclusively fulfill this task since it benefits from accessible low-cost equipment for the collection of data, as well as the compatibility to be launched on ordinary computers. The proposed NCDS possesses several specifications that makes it stand out from the existing designs:

1. Since the cry signals were acquired with no predefined conditions and in the presence of noise, the NCDS design is practical and trained on the real-world conditions.

2. The participants are from a diversity of races, origins, and languages and all of these participants were equally treated and were randomly selected to form our training and testing databases.

3. Despite the fact that the cry signals could be classified based on newborn's emotional state, they were not limited to any specific cry stimulus for the purpose of pathology recognition in this study and thus, any cry could be employed for the diagnostic purposes.

4. The physical attributes of the newborns, e.g., weight, gender, etc., was not considered as a restricting factor and both male and female newborns with any weight were included in this study.

From the developmental point of view, this thesis strived to perform extensive amelioration of various stages in NCDS architecture. Two different segments of the cry signals based on the respiratory activities of the newborn were exploited to form datasets of INSV and EXP which refer to voiced activity during inspiration and expiration, respectively. It was shown that the INSV cry dataset conveys valuable information about the newborn's health. Starting with the feature extraction step, several novel features such as SENCC, SCCC, BSC, HR, and GFCC were explored that each contributed to extending our knowledge of the pathologic newborn cry patterns and comportment. Based on our observations during different studies and experiments, several points were deduced based on the feature extraction step:

1. Although the MFCC feature set is the most convenient for the audio recognition tasks, it can be replaced with GFCC feature set owing to the lower computational cost and run-time, higher robustness, and better performance.

2. The use of arithmetic and statistical measures such as mean, standard deviation, and H-spread can effectively reduce the feature space dimensionality and provide compact yet informative feature sets that represent the spectrum.

3. Increasing the feature space dimensions does not always lead to better performance of the NCDS; therefore, the propitious avail of apt selection and fusion methods is indispensable in the NCDS architectures.

4. The cepstral analysis proved beneficial in the study of the pathologic cry signals, even inn differentiating between two pathology groups.

5. The addition of the spectral and music-derived features to the cepstral analysis led to performance improvement in a majority of experiments.

6. The SENCC feature set effectively represented the information content of the pathologic cry signals and its combination with different feature sets was advantageous which is interesting given its simplicity.

Accordingly, the next step of the NCDS development was feature selection and fusion which was enriched with the investigation of NCA and FE selection and CCA-fusion methods for the first time in NCDS designs. More especially, the number of NCDS-related studies exploring fusion at feature-level is rudimentary in spite of being essential for further development of NCDS. Through the experiments in this thesis, we deduced that not only the feature selection methods employed here did not have a counterproductive effect, but also, they led to an increment despite considerable reduction of the feature space dimensions. Furthermore, it was demonstrated through different evaluation measures that the use of CCA-fusion method homogenized the feature space so that in addition to the augmented performance of NCDS, the run-times for the HPO of the classifiers was markedly decreased to the extent that it was lower than or comparable to the run-times of HPO for the individual feature sets.

In the next step of the NCDS, a heterogeneous collective of classifiers was employed corresponding to each of the experiments. The classifiers used in this study include SVM, KNN, MLP, and LSTM. The experiments were conducive to the superiority of DL method of

LSTM across all the experiments; nevertheless, few factors should be pointed out in this regard. The DL methods hinge on the availability of sufficient data for training and testing purposes, which in biomedical applications, tends to be the main challenge and limits the implementation of state-of-the-art classification approaches. Besides, the congruous employment of the HPO methods with low-dimensional data (e.g., single pathology investigation) is crucial and can increase the performance of simple ML methods to be compatible with more sophisticated DL methods. Although the performance of the DL methods is notably preeminent compared to ML models, the HPO of the DL classifier studied here required higher run-times and deliberate selection of HPs. The simplest and the swiftest classification method taking the HPO run-time into account was the KNN classifier which in some cases outperformed both SVM and MLP. Furthermore, among the HPO methods, BHPO yielded the best results substantially.

The final contribution of this thesis revolves around the fusion at decision level, which facilitates merging the outputs of disparate classifiers that were trained on distinct feature sets by the means of uncomplicated calculations. In this study, the DTF method was selected owing to its simplicity, robustness, and compatibility with limited data. The accomplishment of the DTF opens the door for lots of diverse feature sets and classifiers.

At last, the foremost question that any NCDS endeavors to answer is that whether the morbid newborns cry differently than the healthy and how these pathologies affect the cry signal, and this thesis was no exception. Henceforth, we mention our findings regarding the cry patterns of the studied pathologies.

The septic newborn cries are associated with lower spectral entropy, long durations, monotonous or with notably reduced tonality, and inconsistent ratio of the expiration-inspiration episodes during a cry utterance. Regarding the RDS-related cry signals, it was observed that they are weaker, shorter than most cry signals -even among other pathologies-, having increased dysphonation, and significantly deeper. This is interesting since the RDS is majorly seen in preterm newborns, and prematurity is correlated with high-pitched cries, hence, the mature RDS-diagnosed newborns may have higher chances of being disregarded because of their low-pitched cries.

RECOMMENDATIONS

In this section we provide some recommendations for further improvements regarding the NCDS architectures.

In the current study, the focal point was formation of the apropos acoustic feature space that combined features from diverse applications such as speech recognition and music classification. Moreover, the features were extracted from different levels of information: high-level such as harmonic ratio, mid-level such as MFCCs and its deltas, and low-level such as energy or spectral centroid. As a consequence, the combination and fusion of these features led to higher performance of the NCDS. Besides, the overall best results among all experiments were attained through the implementation of the LSTM classifier.

The NCDS can always benefit from new features such as exploring the wavelet-based and chroma-based features, and more importantly, the Maxima Dispersion Quotient (MDQ) feature that was introduced recently to enable the voice quality comparisons. We recommend the MDQ feature set since the Cepstral analysis led to suitable results across all experiments and MDQ works in a similar manner by filtering the glottal innervation signal by a wavelet function. Moreover, in order to extract the MDQs, the glottal closure instants are calculated which might be useful in the detection of pathologies like RDS or asphyxia.

The employment of merged classifiers such as graph convolutional NN would be interesting in the development of the NCDS.

It was shown that the most important point in this study was the use of proper feature or decision fusion algorithms. In this regard, given that the fusion of different classifiers yielded exceptional improvement, we recommend exploring the Feature Fusion Learning (FFL) method in the future. This method trains a classifier efficiently by using a fusion module that can employ and mix feature maps generated by parallel neural networks.

The main idea is to divide the principal NN into parallel sub-networks and train them. In the next step, by using a fusion module the feature maps are mixed in order to achieve a meaningful map of features. FFL is versatile and can be implemented in any architecture of networks.

In FFL, a diversity of sub-networks contribute to an ensemble classifier and this ensemble classifier distills its knowledge to the fused classifier with the purpose of training it. This procedure is called Ensemble Knowledge Distillation (EKD). Its loss is modeled as the KL-divergence between softened distributions of the ensemble and the fused classifier. The inputs to the fusion module are the last layer of sub-networks' concatenated feature maps. The fused classifier trains each sub-network by distillation of its knowledge in the fusion module, which is called Fusion Knowledge Distillation (FKD). Sub-networks and the fusion module in the FFL are trained simultaneously in order to produce a final decision.

Finally, two levels of fusion were studied here; we suggest conducting a comparative study that investigates the role of fusion at different levels such as feature level, matching score level, and decision level and different methods for each level.

APPENDIX I

LITERATURE REVIEW OF FEATURES AND THEIR PATHOLOGY ASSOCIATION

Table-A I-1 Literature review of features and their pathology association

Feature Domains	**Feature Sets**	**Implementation/Associated Pathologies**	**Studies**
Cepstral Domain	MFCC	RDS, Pathology Ensemble vs. Healthy, Asphyxia, Cleft palate, Hyperbilirubinemia, Hearing Impairment, Hypothyroidism, Reason of crying.	(Jam & Sadjedi, 2009; Kheddache & Tadj, 2019; Massengill Jr, 1969; Matikolaie et al., 2022; Matikolaie & Tadj, 2020; Wahid et al., 2016; Zabidi et al., 2009; Zabidi, Mansor, et al., 2017)
	BFCC	Reason of crying, High-risk Prematurity.	(Liu et al., 2019; Maghfira et al., 2020; Sriraam & Pradeep, 2019; Tejaswini et al., 2020)
	GTCC	Hypoxic Ischemic Encephalopathy (HIE).	(Khalilzad & Tadj, 2023)

Feature Domains	Feature Sets	Implementation/Associated Pathologies	Studies
Cepstral Domain	Perceptual Linear Prediction Cepstral Coefficients (PLPCC)	Pathology Ensemble vs. Healthy.	(Chittora & Patil, 2018)
	Linear Prediction Cepstral Coefficients (LPCCs)	Reason of crying.	(Liu et al., 2018)
	Linear Frequency Cepstral Coefficients (LFCCs)	Reason of crying, Pathology Ensemble vs. Healthy.	(MV Varsharani Bhagatpatil & V Sardar, 2014; Jagtap et al., 2016; Patil et al., 2022)
	Chroma-related features	Reason of crying, Asphyxia.	(Felipe et al., 2019; Ting et al., 2022)
	Constant-Q Cepstral Coefficients	Pathology Ensemble vs. Healthy	(Patil et al., 2022)

Feature Domains	Feature Sets	Implementation/Associated Pathologies	Studies
Prosodic Domain	Rhythm	RDS, sepsis, Pathology Ensemble vs. Healthy.	(Matikolaie et al., 2022; Matikolaie & Tadj, 2020, 2022)
	Tilt		
	Intensity		
	Harmonic Factor	Reason of crying, Pathology Ensemble vs. Healthy.	(Bano & RaviKumar, 2015; Kheddache & Tadj, 2015)
	Unvoiced regions	Pathology Ensemble vs. Healthy.	(Chittora & Patil, 2015b)
Time Domain	Zero-Crossing Rate (ZCR)	Reason of crying, Pathology Ensemble vs. Healthy.	(Abou-Abbas, Tadj, & Fersaie, 2017; Kuo, 2010)
	Amplitude-based features	Reason of crying, Pathology Ensemble vs. Healthy.	(Bano & RaviKumar, 2015)
	Linear Predictive Coding (LPC)	Reason of crying.	(Liu et al., 2019; Rosales-Pérez et al., 2015)
	Duration	Reason of crying, Pathology Ensemble vs. Healthy	(Kheddache & Tadj, 2015; Osmani et al., 2017)
	Auditory-inspired Amplitude Modulation (AAM)	Pathology Ensemble vs. Healthy	(Matikolaie et al., 2022)
	Energy-based features	Reason of crying.	(Bano & RaviKumar, 2015; Rosita & Junaedi, 2016)

Feature Domains	Feature Sets	Implementation/Associated Pathologies	Studies
Frequency Domain	Bispectrum	Preterm Infants, URTI, Epilepsy, Diarrhea, Hydrocephalus, Hypocalcium, Congenital Heart Disease, Jaundice, Bronchitis, Thalassemia	(Chittora & Patil, 2015a)
Frequency Domain	Fundamental frequency-based features	Pathology Ensemble vs. Healthy	(Kheddache & Tadj, 2013a, 2013b, 2015, 2019)
	Resonance frequencies	Pathology Ensemble vs. Healthy	(Kheddache & Tadj, 2015)
	Mel-frequency Entropy Coefficients	Hearing Impairment vs. Healthy	(Jam & Sadjedi, 2009)
	Spectral shape features	Pathology Ensemble vs. Healthy	(Kheddache & Tadj, 2013a, 2013b, 2019)
	Spectrogram-derived features		

Feature Domains	Feature Sets	Implementation/Associated Pathologies	Studies
Image Domain	Waveform Image	Cry detection, Jaundice, Prematurity.	(Hariharan et al., 2018; Moharir, Sachin, Nagaraj, Samiksha, & Rao, 2017; Zhang, Zou, & Liu, 2018)
	Face Image	Asphyxia vs. Healthy.	(Cha & Bae, 2022)
	Spectrogram Image	Reason of crying.	(Bănică, Cucu, Buzo, Burileanu, & Burileanu, 2016; Franti, Ispas, & Dascalu, 2018; Tusty, Basaruddin, & Krisnadhi, 2020)
Time-Frequency Domain	Wavelet Packet Transform	Asphyxia vs. Healthy.	(Hariharan, Yaacob, & Awang, 2011)
	Mel-frequency Discrete Wavelet Coefficients (MFDWC)	Hearing Impairment vs. Healthy	(Mansouri Jam & Sadjedi, 2013)
	STFT-based Features	Hearing Impairment vs. Healthy	(Muthusamy Hariharan et al., 2012)
	Fast Fourier Transform (FFT)-based Features	Cry detection.	(Abou-Abbas, Tadj, Gargour, et al., 2017)

APPENDIX II

LITERATURE REVIEW OF CLASSIFICATION METHODS

Table-A II-1 Literature review of classification methods

Infant Cry Classification Algorithms			**Studies**
Neural Networks	MLP		(Alaie, Abou-Abbas, & Tadj, 2016)
	LSTM		(Huckvale, 2018; Lahmiri, Tadj, Gargour, et al., 2022)
	Convolutional		(Lahmiri, Tadj, Gargour, & Bekiros, 2023)
	Deep Feet-Forward		(Lahmiri, Tadj, Gargour, et al., 2022)
	Recurrent		(Sharma & Malhotra, 2020)
	Probabilistic		(Matikolaie et al., 2022)
	Reservior Network		(Ntalampiras, 2015)
	General Regression		(Saraswathy, Hariharan, Vijean, Yaacob, & Khairunizam, 2012)
	Capsule Network		(Sabour, Frosst, & Hinton, 2017)
	Convolutional Recurrent		(Maghfira et al., 2020)
Random Forest	Bagging Trees		(Tuduce, Cucu, & Burileanu, 2018)
	Boosted Trees	XGBoost	(Chang et al., 2021)
	Decision Trees	Quest	(Fuhr et al., 2015)
		C&R	
		CHAID	
		J48	
		C0.5	

Infant Cry Classification Algorithms		Studies
Logistic Regression		(Lavner, Cohen, Ruinskiy, & IJzerman, 2016; Orlandi et al., 2016)
K-Nearest Neighborhood		(Fuhr et al., 2015)
Linear Discriminant Analysis		(Martinez-Cañete, Cano-Ortiz, Lombardía-Legrá, Rodríguez-Fernández, & Veranes-Vicet, 2018)
Naive Bayes		(Tuduce et al., 2018)
Fuzzy	Neuro-Fuzzy Network	(Santiago-Sánchez, Reyes-García, & Gómez-Gil, 2009)
	Fuzzy KNN	(Rosales-Pérez et al., 2015)
	Fuzzy Decision Forest	
	Fuzzy Relational NN	
Gaussian Mixture Models (GMM)	GMM-UBM	(Alaie et al., 2016)
SVM		(Alaie et al., 2016; Chang et al., 2015)

APPENDIX III

REVIEW OF SELECTED NCDS DESIGNS

Table-A- III-1 A comparative overview of selected NCDS designs for pathological newborn cry diagnostics

Studies	Pathologies	Evaluation Measures	HPO	Feature Fusion	Machine/Deep Learning Methods	Outcome
Matikolaie et al. (Matikolaie et al., 2022; Matikolaie & Tadj, 2020, 2022)	RDS vs. healthy, septic vs. healthy, collection of pathologies vs. healthy.	Recall, precision, F-score, accuracy.	---	Concatenation	PNN, SVM, DT, Discriminant Analysis.	F-score Sepsis: 86%. Multi-Pathology: 80%. RDS: 68.40%
Lahmiri et al. (Lahmiri, Tadj, & Gargour, 2021, 2022; Lahmiri, Tadj, Gargour, et al., 2022)	Central nervous system complications, chromosomal abnormalities, congenital cardiac anomalies, blood disorders.	Accuracy, recall, specificity.	---	Concatenation	CNN, LSTM, DFNN, SVM, NBC.	Accuracy: CNN: 95.28%, DFFNN: 100%, LSTM: 83.89%.
Pusuluri et al. (Pusuluri, Kachhi, & Patil, 2022)	Asphyxia and deaf vs. healthy.	Accuracy, false positive count	Grid search	--	SVM, KNN, RF.	Accuracy: 98.48%.

Studies	Pathologies	Evaluation Measures	HPO	Feature Fusion	Machine/Deep Learning Methods	Outcome
Kheddache et al. (Kheddache & Tadj, 2019)	Hyperbilirubinemia, vena cava thrombosis, meningitis, peritonitis, asphyxia, lingual frenum, IUGR-microcephaly, tetralogy of fallot, gastroschisis, IUGR-asphyxia, RDS.	Accuracy.	---	Concatenation	PNN.	Accuracy: 88.71%.
Farsaie et al. (Alaie et al., 2016)	Central nervous system complications, chromosomal abnormalities, congenital cardiac anomalies, blood disorders.	Average accuracy, recall, specificity, equal error rate, classification error rate.	--	--	SVM, PNN, MLP.	Accuracy: MLP: 91.68%. PNN: 89.93%. SVM: 89.85%.

Studies	**Pathologies**	**Evaluation Measures**	**HPO**	**Feature Fusion**	**Machine/Deep Learning Methods**	**Outcome**
Onu et al. (Onu, Lebensold, Hamilton, & Precup, 2019)	Prenatal asphyxia vs. healthy.	Recall, specificity, unweighted average recall.	Random search.	--	ResNet, SVM.	Specificity: 88.9%

BIBLIOGRAPHY

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., & Fersaie, H. A. (2017). A fully automated approach for baby cry signal segmentation and boundary detection of expiratory and inspiratory episodes. *The Journal of the Acoustical Society of America*, *142*(3), 1318-1331.

Abou-Abbas, L., Tadj, C., Gargour, C., & Montazeri, L. (2017). Expiratory and inspiratory cries detection using different signals' decomposition techniques. *Journal of voice*, *31*(2), 259. e213-259. e228.

Acharya, S., Rajasekar, A., Shender, B. S., Hrebien, L., & Kam, M. (2016). Real-time hypoxia prediction using decision fusion. *IEEE journal of biomedical and health informatics*, *21*(3), 696-707.

Admuthe, S., & Patil, P. H. (2015). Feature extraction method-MFCC and GFCC used for Speaker Identification. *Int. J. Sci. Res. Dev.*, *3*(04), 1261-1264.

Agrawal. (1990). The infant's cry in health and disease. *The National medical journal of India*, *3*(5), 223.

Ainsworth, M. D. (1963). The development of infant-mother interaction among the Ganda. *The determinants of infant behaviour II*.

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

Ali, M. M., Mansor, W., Lee, Y., & Zabidi, A. (2012). Asphyxiated infant cry classification using Simulink model. 2012 IEEE 8th International Colloquium on Signal Processing and its Applications.

Almeida, A., Schubert, E., Smith, J., & Wolfe, J. (2017). Brightness scaling of periodic tones. *Attention, Perception, & Psychophysics*, *79*(7), 1892-1896.

Amaro-Camargo, E., & Reyes-García, C. A. (2007). Applying statistical vectors of acoustic characteristics for the automatic classification of infant cry. International Conference on Intelligent Computing.

Ashwini, K., & Vincent, P. D. R. (2022). A deep convolutional neural network based approach for effective neonatal cry classification. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, *15*(2), 229-239.

Aucouturier, J.-J., Nonaka, Y., Katahira, K., & Okanoya, K. (2011). Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models. *The Journal of the Acoustical Society of America*, *130*(5), 2969-2977.

Aurna, N. F., Yousuf, M. A., Taher, K. A., Azad, A., & Moni, M. A. (2022). A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Computers in Biology and Medicine*, *146*, 105539.

Badreldine, O. M., Elbeheiry, N. A., Haroon, A. N. M., ElShehaby, S., & Marzook, E. M. (2018). Automatic diagnosis of asphyxia infant cry signals using wavelet based mel frequency cepstrum features. 2018 14th International Computer Engineering Conference (ICENCO).

Balayan, S., Chauhan, N., Chandra, R., Kuchhal, N. K., & Jain, U. (2020). Recent advances in developing biosensing based platforms for neonatal sepsis. *Biosensors and Bioelectronics*, *169*, 112552.

Bănică, I.-A., Cucu, H., Buzo, A., Burileanu, D., & Burileanu, C. (2016). Automatic methods for infant cry classification. 2016 International conference on communications (COMM).

Bano, S., & RaviKumar, K. (2015). Decoding baby talk: A novel approach for normal infant cry signal classification. 2015 International Conference on Soft-Computing and Networks Security (ICSNS).

Bansal, M., Upali, S. J. R., & Sharma, S. (2022). Early parkinson disease detection using audio signal processing. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2022, Volume 1* (pp. 243-250). Springer.

Bell, C. (1878). Practical Observations on Some of the More Common Diseases of Early Life. *Edinburgh medical journal*, *24*(6), 534.

Bellieni, C. V., Sisto, R., Cordelli, D. M., & Buonocore, G. (2004). Cry features reflect pain intensity in term newborns: an alarm threshold. *Pediatric Research*, *55*(1), 142-146.

Benesty, J., Sondhi, M. M., & Huang, Y. (2007). *Springer handbook of speech processing*. Springer.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, *13*(2).

Bhagatpatil, M. V., & Sardar, V. (2014). An automatic infant's cry detection using linear frequency cepstrum coefficients (LFCC). *International Journal of Scientific & Engineering Research*, *5*(5), 1379-1383.

Blinick, G., Tavolga, W. N., & Antopol, W. (1971). Variations in birth cries of newborn infants from narcotic-addicted and normal mothers. *American Journal of Obstetrics and Gynecology*, *110*(7), 948-958.

Bossé, É., & Solaiman, B. (2016). *Information fusion and analytics for big data and IoT*. Artech House.

Boukydis, C. Z., & Lester, B. M. (2012). Infant crying: Theoretical and research perspectives.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, *30*(7), 1145-1159.

Brazelton, T. B. (1962). Crying in infancy. *Pediatrics*, *29*(4), 579-588.

Brent, W. (2010). *Physical and perceptual aspects of percussive timbre*. University of California, San Diego.

Buder, E. H., Chorna, L. B., Oller, D. K., & Robinson, R. B. (2008). Vibratory regime classification of infant phonation. *Journal of voice*, *22*(5), 553-564.

Canada, Statistic. (2022). *"Leading causes of death, infants.,"* Governtment of Canada, https://doi.org/10.25318/1310039501-eng.

Cha, J., & Bae, G. (2022). Deep Learning Based Infant Cry Analysis Utilizing Computer Vision. *International Journal of Applied Engineering Research*, *17*(1), 30-35.

Chang, C.-Y., Bhattacharya, S., Raj Vincent, P., Lakshmanna, K., & Srinivasan, K. (2021). An efficient classification of neonates cry using extreme gradient boosting-assisted grouped-support-vector network. *Journal of healthcare engineering*, *2021*.

Chang, C.-Y., Chang, C.-W., Kathiravan, S., Lin, C., & Chen, S.-T. (2017). DAG-SVM based infant cry classification system using sequential forward floating feature selection. *Multidimensional Systems and Signal Processing*, *28*(3), 961-976.

Chang, C.-Y., Hsiao, Y.-C., & Chen, S.-T. (2015). Application of incremental SVM learning for infant cries recognition. 2015 18th International Conference on Network-Based Information Systems.

Chang, C.-Y., & Tsai, L.-Y. (2019). A CNN-based method for infant cry detection and recognition. Workshops of the International Conference on Advanced Information Networking and Applications.

Chen, L., Gunduz, S., & Ozsu, M. T. (2006). Mixed type audio classification with support vector machine. 2006 IEEE International Conference on Multimedia and Expo.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, *21*(1), 1-13.

Chittora, A., & Patil, H. A. (2015a). Classification of normal and pathological infant cries using bispectrum features. 2015 23rd European Signal Processing Conference (EUSIPCO).

Chittora, A., & Patil, H. A. (2015b). Significance of unvoiced segments and fundamental frequency in infant cry analysis. Text, Speech, and Dialogue: 18th International Conference, TSD 2015, Pilsen, Czech Republic, September 14-17, 2015, Proceedings.

Chittora, A., & Patil, H. A. (2016). Spectral analysis of infant cries and adult speech. *International Journal of Speech Technology*, *19*(4), 841-856.

Chittora, A., & Patil, H. A. (2018). Significance of higher-order spectral analysis in infant cry classification. *Circuits, Systems, and Signal Processing*, *37*, 232-254.

Clark, R. H. (2005). The epidemiology of respiratory failure in neonates born at an estimated gestational age of 34 weeks or more. *Journal of Perinatology*, *25*(4), 251-257.

Clinic. ARDS—Symptomps and Causes—Mayo. (2022). Available online: https://www.mayoclinic.org/diseases-conditions/ards/symptoms-causes/syc-20355576.

Cohen, L. (1995). *Time-frequency analysis* (Vol. 778). Prentice hall New Jersey.

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* (pp. 171-196). Springer.

Corwin, M. J., & Golub, H. L. (1984). Spectral analysis of a cry is abnormal in infants who have moderate hyperbilirubinemia. *Pediatric Research*, *18*(S4), 102A.

Corwin, M. J., Lester, B. M., & Golub, H. L. (1996). The infant cry: what can it tell us? *Current Problems in Pediatrics*, *26*(9), 313-334.

Dar, J. A., Srivastava, K. K., & Lone, S. A. (2022). Spectral features and optimal hierarchical attention networks for pulmonary abnormality detection from the respiratory sound signals. *Biomedical Signal Processing and Control*, *78*, 103905.

Dash, D. P., & Kolekar, M. H. (2020). EEG-based epileptic seizure detection using least square SVM with spectral and multiscale key point energy features. Soft Computing for Problem Solving 2019: Proceedings of SocProS 2019, Volume 1.

Diaz, G. I., Fokoue-Nkoutche, A., Nannicini, G., & Samulowitz, H. (2017). An effective algorithm for hyperparameter optimization of neural networks. *IBM Journal of Research and Development*, *61*(4/5), 9: 1-9: 11.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification. *International Journal of Computational Intelligence and Applications*, *1*, 335-339.

Ebrahimpour, R., & Hamedi, S. (2009). Hand written digit recognition by multiple classifier fusion based on decision templates approach. *World Academy of Science, Engineering and Technology*, *57*, 560-565.

Edwards, M. O., Kotecha, S. J., & Kotecha, S. (2013). Respiratory distress of the term newborn infant. *Paediatric respiratory reviews*, *14*(1), 29-37.

Emde, R. N., Gaensbauer, T. J., & Harmon, R. J. (1976). Emotional expression in infancy; a biobehavioral study. *Psychological issues*, *10*(01), 1-200.

Farsaie Alaie, H., & Tadj, C. (2012). Cry-based classification of healthy and sick infants using adapted boosting mixture learning method for gaussian mixture models. *Modelling and simulation in engineering*, *2012*.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874.

Felipe, G. Z., Aguiar, R. L., Costa, Y. M., Silla, C. N., Brahnam, S., Nanni, L., & McMurtrey, S. (2019). Identification of infants' cry motivation using spectrograms. 2019 International Conference on Systems, Signals and Image Processing (IWSSIP).

Feng, C. (2020). Effects of Liquid Viscosity and Food Texture on Swallowing Sounds.

Fernandes, R. P., & Apolinário Jr, J. A. (2020). Underwater target classification with optimized feature selection based on Genetic Algorithms. Proc. Simpósio Brasileiro de Telecomunicações e Processamento De Sinais.

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In *Automated machine learning* (pp. 3-33). Springer, Cham.

Fisichelli, V., Karelitz, S., Fisichelli, R., & Cooper, J. (1974). The course of induced crying activity in the first year of life. *Pediatric Research*, *8*(12), 921-928.

Flach, P., & Kull, M. (2015). Precision-recall-gain curves: PR analysis done right. *Advances in neural information processing systems*, *28*.

Fontana, J. M., Farooq, M., & Sazonov, E. (2014). Automatic ingestion monitor: a novel wearable device for monitoring of ingestive behavior. *IEEE Transactions on Biomedical Engineering*, *61*(6), 1772-1779.

Fort, A., & Manfredi, C. (1998). Acoustic analysis of newborn infant cry signals. *Medical engineering & physics*, *20*(6), 432-442.

Franti, E., Ispas, I., & Dascalu, M. (2018). Testing the universal baby language hypothesis-automatic infant speech recognition with cnns. 2018 41st International Conference on Telecommunications and Signal Processing (TSP).

Fuhr, T., Reetz, H., & Wegener, C. (2015). Comparison of supervised-learning models for infant cry classification/vergleich von klassifikationsmodellen zur säuglingsschreianalyse. *International Journal of Health Professions*, *2*(1), 4-15.

Fung, M. L., Chen, M. Z., & Chen, Y. H. (2017). Sensor fusion: A review of methods and applications. 2017 29th Chinese Control And Decision Conference (CCDC).

Garg, E., & Bahl, M. (2014). Emotion recognition in speech using gammatone cepstral coefficients. *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, *3*(10), 285-291.

Gimeno, P., Viñals, I., Ortega, A., Miguel, A., & Lleida, E. (2020). Multiclass audio segmentation based on recurrent neural networks for broadcast domain data. *EURASIP Journal on Audio, Speech, and Music Processing*, *2020*(1), 1-19.

Golub, H. L. (1979). A physioacoustic model of the infant cry and its use for medical diagnosis and prognosis. *The Journal of the Acoustical Society of America*, *65*(S1), S25-S26.

Golub, H. L., & Corwin, M. J. (1985). A physioacoustic model of the infant cry. *Infant crying: Theoretical and research perspectives*, 59-82.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Gorgolis, N., Hatzilygeroudis, I., Istenes, Z., & Gyenne, L. G. (2019). Hyperparameter optimization of LSTM network models through genetic algorithm. 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA).

Grau, S. M., Robb, M. P., & Cacace, A. T. (1995). Acoustic correlates of inspiratory phonation during infant cry. *Journal of Speech, Language, and Hearing Research*, *38*(2), 373-381.

Gulzar, T., Singh, A., & Sharma, S. (2014). Comparative analysis of LPCC, MFCC and BFCC for the recognition of Hindi words using artificial neural networks. *International Journal of Computer Applications*, *101*(12), 22-27.

Haghighat, M., Abdel-Mottaleb, M., & Alhalabi, W. (2016). Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Systems with Applications*, *47*, 23-34.

Han, J., Gondro, C., Reid, K., & Steibel, J. P. (2021). Heuristic hyperparameter optimization of deep learning models for genomic prediction. *G3*, *11*(7), jkab032.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36.

Hariharan, M., Saraswathy, J., Sindhu, R., Khairunizam, W., & Yaacob, S. (2012). Infant cry classification to identify asphyxia using time-frequency analysis and radial basis neural networks. *Expert Systems with Applications*, *39*(10), 9515-9523.

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 and programs in biomedicine*, *155*, 39-51.

Hariharan, M., Sindhu, R., & Yaacob, S. (2012). Normal and hypoacoustic infant cry signal classification using time–frequency analysis and general regression neural network. *Computer methods and programs in biomedicine*, *108*(2), 559-569.

Hariharan, M., Yaacob, S., & Awang, S. A. (2011). Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. *Expert Systems with Applications*, *38*(12), 15377-15382.

Hasasneh, A., Kampel, N., Sripad, P., Shah, N. J., & Dammers, J. (2018). Deep learning approach for automatic classification of ocular and cardiac artifacts in meg data. *Journal of Engineering*, *2018*.

He, L., & Cao, C. (2018). Automated depression analysis using convolutional neural networks from speech. *Journal of biomedical informatics*, *83*, 103-111.

Heise, D., Miller, Z., Wallace, M., & Galen, C. (2020). Bumble bee traffic monitoring using acoustics. 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC).

Hinton, G., Srivastava, N., & Swersky, K. (2012). Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Cited on*, *14*(8), 2.

Hirschberg, J. (1980). Acoustic analysis of pathological cries, stridors and coughing sounds in infancy. *International journal of pediatric otorhinolaryngology*, *2*(4), 287-300.

Hosseinzadeh, D., & Krishnan, S. (2007). Combining vocal source and MFCC features for enhanced speaker recognition performance using GMMs. 2007 IEEE 9th Workshop on Multimedia Signal Processing.

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, *5*(2), 1.

Huang, X., Acero, A., Hon, H.-W., & Foreword By-Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice hall PTR.

Huckvale, M. (2018). Neural network architecture that combines temporal and summative features for infant cry classification in the interspeech 2018 computational paralinguistics challenge. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.

Jaganathan, P., & Kuppuchamy, R. (2013). A threshold fuzzy entropy based feature selection for medical database classification. *Computers in Biology and Medicine*, *43*(12), 2222-2229.

Jagtap, S. S., Kadbe, P. K., & Arotale, P. N. (2016). System propose for Be acquainted with newborn cry emotion using linear frequency cepstral coefficient. 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT).

Jam, M. M., & Sadjedi, H. (2009). Identification of hearing disorder by multi-band entropy cepstrum extraction from infant's cry. 2009 International Conference on Biomedical and Pharmaceutical Engineering.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). Statistical learning. In *An introduction to statistical learning* (pp. 15-57). Springer.

Ji, C., Basodi, S., Xiao, X., & Pan, Y. (2020). Infant sound classification on multi-stage cnns with hybrid features and prior knowledge. International Conference on AI and Mobile Services.

Ji, C., Jiao, Y., Chen, M., & Pan, Y. (2022). Infant cry classification based-on feature fusion and mel-spectrogram decomposition with CNNs. International Conference on AI and Mobile Services, 126-134.

Ji, C., Mudiyanselage, T. B., Gao, Y., & Pan, Y. (2021). A review of infant cry analysis and classification. *EURASIP Journal on Audio, Speech, and Music Processing*, *2021*(1), 1-17.

Ji, C., Xiao, X., Basodi, S., & Pan, Y. (2019). Deep learning for asphyxiated infant cry classification based on acoustic features and weighted prosodic features. 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData).

Jiang, N., Jia, J., & Shao, D. (2020). Comparative study of speech emotion recognition based on CNN and CRNN. 2020 International Conference on Machine Learning and Cybernetics (ICMLC).

Jin, W., Wang, X., & Zhan, Y. (2022). Environmental Sound Classification Algorithm Based on Region Joint Signal Analysis Feature and Boosting Ensemble Learning. *Electronics*, *11*(22), 3743.

Kamińska, D., Sapiński, T., & Anbarjafari, G. (2017). Efficiency of chosen speech descriptors in relation to emotion recognition. *EURASIP Journal on Audio, Speech, and Music Processing*, *2017*(1), 1-9.

Katsiamis, A. G., Drakakis, E. M., & Lyon, R. F. (2007). Practical gammatone-like filters for auditory processing. *EURASIP Journal on Audio, Speech, and Music Processing*, *2007*, 1-15.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, *72*(2), 353-365.

Khalilzad, Z., Hasasneh, A., & Tadj, C. (2022). Newborn cry-based diagnostic system to distinguish between sepsis and respiratory distress syndrome using combined acoustic features. *Diagnostics*, *12*(11), 2802.

Khalilzad, Z., Kheddache, Y., & Tadj, C. (2022). An Entropy-Based Architecture for Detection of Sepsis in Newborn Cry Diagnostic Systems. *Entropy*, *24*(9), 1194.

Khalilzad, Z., & Tadj, C. (2023). Using CCA-fused cepstral features in a deep learning-based cry diagnostic system for detecting an ensemble of pathologies in newborns. *Diagnostics*, *13*(5), 879.

Khatun, M. A., Yousuf, M. A., Ahmed, S., Uddin, M. Z., Alyami, S. A., Al-Ashhab, S., . . . Moni, M. A. (2022). Deep CNN-LSTM with self-Attention model for human activity recognition using wearable sensor. *IEEE Journal of Translational Engineering in Health and Medicine*, *10*, 1-16.

Kheddache, Y., & Tadj, C. (2013a). Acoustic measures of the cry characteristics of healthy newborns and newborns with pathologies. *Journal of Biomedical Science and Engineering*, *6*(08), 796.

Kheddache, Y., & Tadj, C. (2013b). Characterization of pathologic cries of newborns based on fundamental frequency estimation. *Engineering*, *5*(10), 272.

Kheddache, Y., & Tadj, C. (2013c). Frequential characterization of healthy and pathologic newborn cries. *American Journal of Biomedical Engineering*, *3*(6), 182-193.

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

Khushaba, R. N., Al-Jumaily, A., & Al-Ani, A. (2007). Novel feature extraction method based on fuzzy entropy and wavelet packet transform for myoelectric control. 2007 International Symposium on Communications and Information Technologies.

Kim, D., Van Ho, P., & Lim, Y. (2017). A new recognition method for visualizing music emotion. *International Journal of Electrical and Computer Engineering*, *7*(3), 1246.

Kim, J., Hyun, M., Chung, I., & Kwak, N. (2019). Feature Fusion for Online Mutual Knowledge Distillation. *arXiv preprint arXiv:1904.09058*.

Kim, J., Oh, J., & Heo, T.-Y. (2021). Acoustic scene classification and visualization of beehive sounds using machine learning algorithms and Grad-CAM. *Mathematical Problems in Engineering*, *2021*, 1-13.

Kim, M. J., Kim, Y., Hong, S., & Kim, H. (2013). ROBUST detection of infant crying in adverse environments using weighted segmental two-dimensional linear frequency cepstral coefficients. 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW).

King, R. D., Feng, C., & Sutherland, A. (1995). Statlog: comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence an International Journal*, *9*(3), 289-333.

Kohavi, R., & John, G. H. (1995). Automatic parameter selection by minimizing estimated error. In *Machine Learning Proceedings 1995* (pp. 304-312). Elsevier.

Konner, M. J. (1972). Aspects of the developmental ethology of a foraging people. *Ethological studies of child behaviour*, 285-304.

Kulkarni, N. (2018). Use of complexity based features in diagnosis of mild Alzheimer disease using EEG signals. *International Journal of Information Technology*, *10*(1), 59-64.

Kulkarni, N., & Bairagi, V. (2017). Extracting salient features for EEG-based diagnosis of Alzheimer's disease using support vector machine classifier. *IETE Journal of Research*, *63*(1), 11-22.

Kulkarni, N., & Bairagi, V. (2018). *EEG-based diagnosis of alzheimer disease: a review and novel approaches for feature extraction and classification techniques*. Academic Press.

Kulkarni, P., Umarani, S., Diwan, V., Korde, V., & Rege, P. P. (2021). Child cry classification-an analysis of features and models. 2021 6th International Conference for Convergence in Technology (I2CT).

Kumaran, U., Radha Rammohan, S., Nagarajan, S. M., & Prathik, A. (2021). Fusion of mel and gammatone frequency cepstral coefficients for speech emotion recognition using deep C-RNN. *International Journal of Speech Technology*, *24*(2), 303-314.

Kuncheva, L. I., Bezdek, J. C., & Duin, R. P. (2001). Decision templates for multiple classifier fusion: an experimental comparison. *Pattern recognition*, *34*(2), 299-314.

Kuo, K. (2010). Feature extraction and recognition of infant cries. 2010 IEEE International Conference on Electro/Information Technology.

Kupietzky, A., & Botzer, E. (2005). Ankyloglossia in the infant and young child: clinical suggestions for diagnosis and management. *Pediatric dentistry*, *27*(1), 40-46.

LaGasse, L. L., Neal, A. R., & Lester, B. M. (2005). Assessment of infant cry: acoustic cry analysis and parental perception. *Mental retardation and developmental disabilities research reviews*, *11*(1), 83-93.

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.

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*(8), 1166.

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.

Lahmiri, S., Tadj, C., Gargour, C., & Bekiros, S. (2022). Deep learning systems for automatic diagnosis of infant cry signals. *Chaos, Solitons & Fractals*, *154*, 111700.

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.

Laitman, J. T. (1977). The ontogentic and phylogenetic development of the upper respiratory system and basicranium in man. Yale University.

Lakatos, S. (2000). A common perceptual space for harmonic and percussive timbres. *Perception & psychophysics*, *62*(7), 1426-1439.

Lalitha, S., Tripathi, S., & Gupta, D. (2019). Enhanced speech emotion detection using deep neural networks. *International Journal of Speech Technology*, *22*, 497-510.

Latifpour, H., Mosleh, M., & Kheyrandish, M. (2015). An intelligent audio watermarking based on KNN learning algorithm. *International Journal of Speech Technology*, *18*(4), 697-706.

Lavner, Y., Cohen, R., Ruinskiy, D., & IJzerman, H. (2016). Baby cry detection in domestic environment using deep learning. 2016 IEEE international conference on the science of electrical engineering (ICSEE).

Le, L., Kabir, A. N. M., Ji, C., Basodi, S., & Pan, Y. (2019). Using transfer learning, SVM, and ensemble classification to classify baby cries based on their spectrogram images. 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW).

Lederman, D. (2002). *Automatic classification of infants' cry*. Citeseer.

Lederman, D., Cohen, A., Zmora, E., Wermke, K., Hauschildt, S., & Stellzig-Eisenhauer, A. (2002). On the use of hidden Markov models in infants' cry classification. The 22nd Convention on Electrical and Electronics Engineers in Israel, 2002.

Lee, H.-M., Chen, C.-M., Chen, J.-M., & Jou, Y.-L. (2001). An efficient fuzzy classifier with feature selection based on fuzzy entropy. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *31*(3), 426-432.

Lelandais, B., Ruan, S., Denœux, T., Vera, P., & Gardin, I. (2014). Fusion of multi-tracer PET images for dose painting. *Medical Image Analysis*, *18*(7), 1247-1259.

Lester, B. M., & Boukydis, C. Z. (1985). *Infant crying: Theoretical and research perspectives*. Springer.

Lester, B. M., & Boukydis, C. Z. (1992). No language but a cry. *Nonverbal vocal communication. Comparative and developmental approaches*, 145-173.

Lester, B. M., & Zeskind, P. S. (1982). A biobehavioral perspective on crying in early infancy. In *Theory and research in behavioral pediatrics* (pp. 133-180). Springer.

Lewis, F. R. (2007). *Focus on nonverbal communication research*. Nova Publishers.

Li, Y., Porter, E., Santorelli, A., Popović, M., & Coates, M. (2017). Microwave breast cancer detection via cost-sensitive ensemble classifiers: Phantom and patient investigation. *Biomedical Signal Processing and Control*, *31*, 366-376.

Lind, J., Vuorenkoski, V., Rosberg, G., Partanen, T., & Wasz‑Höckert, O. (1970). Spectographic analysis of vocal response to pain stimuli in infants with Down's syndrome. *Developmental medicine & child neurology*, *12*(4), 478-486.

Lind, K., & Wermke, K. (2002). Development of the vocal fundamental frequency of spontaneous cries during the first 3 months. *International journal of pediatric otorhinolaryngology*, *64*(2), 97-104.

Liu, G. K. (2018). Evaluating gammatone frequency cepstral coefficients with neural networks for emotion recognition from speech. *arXiv preprint arXiv:1806.09010*.

Liu, L., Li, W., Wu, X., & Zhou, B. X. (2019). Infant cry language analysis and recognition: an experimental approach. *IEEE/CAA Journal of Automatica Sinica*, *6*(3), 778-788.

Liu, L., Li, Y., & Kuo, K. (2018). Infant cry signal detection, pattern extraction and recognition. 2018 International Conference on Information and Computer Technologies (ICICT).

Liu, Y., Starzyk, J. A., & Zhu, Z. (2008). Optimized approximation algorithm in neural networks without overfitting. *IEEE transactions on neural networks*, *19*(6), 983-995.

Lohrmann, C., Luukka, P., Jablonska-Sabuka, M., & Kauranne, T. (2018). A combination of fuzzy similarity measures and fuzzy entropy measures for supervised feature selection. *Expert Systems with Applications*, *110*, 216-236.

Lounsbury, M. L., & Bates, J. E. (1982). The cries of infants of differing levels of perceived temperamental difficultness: Acoustic properties and effects on listeners. *Child Development*, 677-686.

Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*, *38*(4), 4600-4607.

Magagnin, V., Bassani, T., Bari, V., Turiel, M., Maestri, R., Pinna, G. D., & Porta, A. (2011). Non-stationarities significantly distort short-term spectral, symbolic and entropy heart rate variability indices. *Physiological measurement*, *32*(11), 1775.

Maghfira, T. N., Basaruddin, T., & Krisnadhi, A. (2020). Infant cry classification using cnn–rnn. Journal of Physics: Conference Series.

Maitre, N. L., Stark, A. R., Menser, C. C. M., Chorna, O. D., France, D. J., Key, A. F., . . . Bruehl, S. (2017). Cry presence and amplitude do not reflect cortical processing of painful stimuli in newborns with distinct responses to touch or cold. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *102*(5), F428-F433.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current biology*, *19*(23), 1994-1997.

Mangai, U. G., Samanta, S., Das, S., & Chowdhury, P. R. (2010). A survey of decision fusion and feature fusion strategies for pattern classification. *IETE Technical review*, *27*(4), 293-307.

Mansouri Jam, M., & Sadjedi, H. (2013). Wavelet - based automatic cry recognition system for detecting infants with hearing - loss from normal infants. *The Journal of Engineering*, *2013*(11), 63-64.

Mantovani, R. G., Horváth, T., Cerri, R., Vanschoren, J., & de Carvalho, A. C. (2016). Hyper-parameter tuning of a decision tree induction algorithm. 2016 5th Brazilian Conference on Intelligent Systems (BRACIS).

Maria, A., & Jeyaseelan, A. S. (2021). Development of optimal feature selection and deep learning toward hungry stomach detection using audio signals. *Journal of Control, Automation and Electrical Systems*, *32*(4), 853-874.

Martinez-Cañete, Y., Cano-Ortiz, S. D., Lombardía-Legrá, L., Rodríguez-Fernández, E., & Veranes-Vicet, L. (2018). Data mining techniques in normal or pathological infant cry. Progress in Artificial Intelligence and Pattern Recognition: 6th International Workshop, IWAIPR 2018, Havana, Cuba, September 24–26, 2018, Proceedings 6.

Massengill Jr, R. M. (1969). Cry Characteristics in Cleft - Palate Neonates. *The Journal of the Acoustical Society of America*, *45*(3), 782-784.

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

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*.

Messaoud, A., & Tadj, C. (2011). Analysis of acoustic features of infant cry for classification purposes. 2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE).

Mi, A., Wang, L., & Qi, J. (2016). A multiple classifier fusion algorithm using weighted decision templates. *Scientific Programming*, *2016*.

Michelsson, K., Raes, J., Thodén, C.-J., & Wasz-Hockert, O. (1982). Sound spectrographic cry analysis in neonatal diagnostics. An evaluative study. *Journal of Phonetics*, *10*(1), 79-88.

Michelsson, K., SirviöM. A, P., & Wasz-Höckert, O. (1977). Pain cry in full‐term asphyxiated newborn infants correlated with late findings. *Acta Pædiatrica*, *66*(5), 611-616.

Michelsson, K., SirviöM. A, P., & Wasz‐Höckert, O. (1977). Sound spectrographic cry analysis of infants with bacterial meningitis. *Developmental medicine & child neurology*, *19*(3), 309-315.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning, neural and statistical classification.

Mijović, B., Silva, M., Van den BRH, B., Allegaert, K., Aerts, J.-M., Berckmans, D., & Huffel, V. S. (2010). Assessment of pain expression in infant cry signals using empirical mode decomposition. *Methods of information in medicine*, *49*(05), 448-452.

Misra, H., Ikbal, S., Bourlard, H., & Hermansky, H. (2004). Spectral entropy based feature for robust ASR. 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing.

Moharir, M., Sachin, M., Nagaraj, R., Samiksha, M., & Rao, S. (2017). Identification of asphyxia in newborns using gpu for deep learning. 2017 2nd International Conference for Convergence in Technology (I2CT).

Moorman, J. R., Delos, J. B., Flower, A. A., Cao, H., Kovatchev, B. P., Richman, J. S., & Lake, D. E. (2011). Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring. *Physiological measurement*, *32*(11), 1821.

Mugitani, R., & Hiroya, S. (2012). Development of vocal tract and acoustic features in children. *Acoustical Science and Technology*, *33*(4), 215-220.

Mukhopadhyay, J., Saha, B., Majumdar, B., Majumdar, A., Gorain, S., Arya, B. K., . . . Singh, A. (2013). An evaluation of human perception for neonatal cry using a database of cry and underlying cause. 2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT).

Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, *2*(5-6), 183-197.

Nakisa, B., Rastgoo, M. N., Rakotonirainy, A., Maire, F., & Chandran, V. (2018). Long short term memory hyperparameter optimization for a neural network based emotion recognition framework. *IEEE Access*, *6*, 49325-49338.

Nicollas, R., Giordano, J., Francius, L., Vicente, J., Burtschell, Y., Medale, M., . . . Giovanni, A. (2005). Aerodynamical Model of Human Newborn Larynx: An Approach of the First Cry. Fourth International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications.

Niemeijer, M., Abramoff, M. D., & Van Ginneken, B. (2009). Information fusion for diabetic retinopathy CAD in digital color fundus photographs. *IEEE transactions on medical imaging*, *28*(5), 775-785.

Ntalampiras, S. (2015). Audio pattern recognition of baby crying sound events. *Journal of the Audio Engineering Society*, *63*(5), 358-369.

Nusrat, I., & Jang, S.-B. (2018). A comparison of regularization techniques in deep neural networks. *Symmetry*, *10*(11), 648.

O'Regan, S., & Marnane, W. (2013). Multimodal detection of head-movement artefacts in EEG. *Journal of Neuroscience Methods*, *218*(1), 110-120.

Okada, Y., Fukuta, K., & Nagashima, T. (2011). Iterative forward on cross-validation approach and its application to infant cry classification. Proceedings of the International Multi Conference of Engineers and Computer Scientists.

Olson, R. S., Cava, W. L., Mustahsan, Z., Varik, A., & Moore, J. H. (2018). Data-driven advice for applying machine learning to bioinformatics problems. Pacific Symposium on Biocomputing 2018: Proceedings of the Pacific Symposium.

Onu, C. C., Lebensold, J., Hamilton, W. L., & Precup, D. (2019). Neural transfer learning for cry-based diagnosis of perinatal asphyxia. *arXiv preprint arXiv:1906.10199*.

Onu, C. C., Udeogu, I., Ndiomu, E., Kengni, U., Precup, D., Sant'Anna, G. M., . . . Opara, P. (2017). Ubenwa: Cry-based diagnosis of birth asphyxia. *arXiv preprint arXiv:1711.06405*.

Oppenheim, A. v., Schafer, R., & Stockham, T. (1968). Nonlinear filtering of multiplied and convolved signals. *IEEE transactions on audio and electroacoustics*, *16*(3), 437-466.

Oren, A., Matzliach, A., Cohen, R., & Friedman, H. (2016). Cry-based detection of developmental disorders in infants. 2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE).

Orlandi, S., Garcia, C. A. R., 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*(6), 656-663.

Orlandi, S., Manfredi, C., Bocchi, L., & Scattoni, M. L. (2012). Automatic newborn cry analysis: a non-invasive tool to help autism early diagnosis. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Osmani, A., Hamidi, M., & Chibani, A. (2017). Machine learning approach for infant cry interpretation. 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI).

Parga, J. J., Lewin, S., Lewis, J., Montoya-Williams, D., Alwan, A., Shaul, B., . . . Dapretto, M. (2020). Defining and distinguishing infant behavioral states using acoustic cry analysis: is colic painful? *Pediatric research*, *87*(3), 576-580.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian journal of ophthalmology*, *56*(1), 45.

Patil, H. A., Patil, A. T., & Kachhi, A. (2022). Constant Q Cepstral coefficients for classification of normal vs. Pathological infant cry. ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO Ist Project Report*, *54*(0), 1-25.

Peng, L., Liao, B., Zhu, W., Li, Z., & Li, K. (2015). Predicting drug–target interactions with multi-information fusion. *IEEE journal of biomedical and health informatics*, *21*(2), 561-572.

Porta, A., De Maria, B., Bari, V., Marchi, A., & Faes, L. (2016). Are nonlinear model-free conditional entropy approaches for the assessment of cardiac control complexity superior to the linear model-based one? *IEEE Transactions on Biomedical Engineering*, *64*(6), 1287-1296.

Prescott, R. (1975). Infant cry sound; developmental features. *The Journal of the Acoustical Society of America*, *57*(5), 1186-1191.

Pusuluri, A., Kachhi, A., & Patil, H. A. (2022). Analysis of Time-Averaged Feature Extraction Techniques on Infant Cry Classification. Speech and Computer: 24th International Conference, SPECOM 2022, Gurugram, India, November 14–16, 2022, Proceedings.

Qandalji, B. (2010). PP-293. Full term neonatal admissions. *Early Human Development*(86), S133.

Qian, L.-l. (2010). Current status of neonatal acute respiratory disorders: a one-year prospective survey from a Chinese neonatal network. *Chinese medical journal*, *123*(20), 2769-2775.

Rabiner, L. (1993). Fundamentals of speech recognition. *Fundamentals of speech recognition*.

Ramalingam, A., & Krishnan, S. (2005). Gaussian mixture modeling using short time fourier transform features for audio fingerprinting. 2005 IEEE International Conference on Multimedia and Expo.

Reby, D., Levréro, F., Gustafsson, E., & Mathevon, N. (2016). Sex stereotypes influence adults' perception of babies' cries. *BMC psychology*, *4*(1), 1-12.

Reimers, N., & Gurevych, I. (2017). Optimal hyperparameters for deep lstm-networks for sequence labeling tasks. *arXiv preprint arXiv:1707.06799*.

Robb, M. P., & Goberman, A. M. (1997). Application of an acoustic cry template to evaluate at-risk newborns: Preliminary findings. *Neonatology*, *71*(2), 131-136.

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

Rosita, Y. D., & Junaedi, H. (2016). Infant's cry sound classification using Mel-Frequency Cepstrum Coefficients feature extraction and Backpropagation Neural Network. 2016 2nd International Conference on Science and Technology-Computer (ICST).

Ruiz-contreras, J., Urquia, L., & Bastero, R. (1999). Persistent crying as predominant manifestation of sepsis in infants and newborns. *Pediatric emergency care*, *15*(2), 113-115.

Ruiz-Contreras, J., Urquía, L., & Bastero, R. (1999). Persistent crying as predominant manifestation of sepsis in infants and newborns. *Pediatric emergency care*, *15*(2), 113-115.

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. *Advances in neural information processing systems*, *30*.

Sahak, R., Mansor, W., Lee, K. Y., Zabidi, A., & Yassin, A. I. (2013). Optimization of principal component analysis and support vector machine for the recognition of infant cry with asphyxia. *International Journal of Computers and Applications*, *35*(3), 99-107.

Sahak, R., Mansor, W., Lee, Y., Yassin, A., & Zabidi, A. (2010a). Performance of combined support vector machine and principal component analysis in recognizing infant cry with asphyxia. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology.

Sahak, R., Mansor, W., Lee, Y., Yassin, A. M., & Zabidi, A. (2010b). Orthogonal least square based support vector machine for the classification of infant cry with asphyxia. 2010 3rd International Conference on Biomedical Engineering and Informatics.

Santiago-Sánchez, K., Reyes-García, C. A., & Gómez-Gil, P. (2009). Type-2 fuzzy sets applied to pattern matching for the classification of cries of infants under neurological risk. Emerging Intelligent Computing Technology and Applications: 5th International Conference on Intelligent Computing, ICIC 2009, Ulsan, South Korea, September 16-19, 2009. Proceedings 5.

Saraswathy, J., Hariharan, M., Vijean, V., Yaacob, S., & Khairunizam, W. (2012). Performance comparison of Daubechies wavelet family in infant cry classification. 2012 IEEE 8th International Colloquium on Signal Processing and its Applications.

Saraswathy, J., Hariharan, M., Yaacob, S., & Khairunizam, W. (2012). Automatic classification of infant cry: A review. 2012 International Conference on Biomedical Engineering (ICoBE).

Satar, M., Cengizler, C., Hamitoglu, S., & Ozdemir, M. (2022). Audio Analysis Based Diagnosis of Hypoxic Ischemic Encephalopathy in Newborns.

Shao, Y., Jin, Z., Wang, D., & Srinivasan, S. (2009). An auditory-based feature for robust speech recognition. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing.

Sharma, A., & Malhotra, D. (2020). Speech recognition based iicc-intelligent infant cry classifier. 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT).

Singh, M., & Gray, C. P. (2018). Neonatal sepsis.

Singhal, R., Srivatsan, S., & Panda, P. (2022). Classification of Music Genres using Feature Selection and Hyperparameter Tuning. *Journal of Artificial Intelligence and Capsule Networks*, *4*(3), 167-178.

Skeptis. (1995). Contagious folly: an adventure and its skeptics *The Female Thermometer: Eighteenth-century Culture and the Invention of the Uncanny*, 190.

Smith, J. O., & Abel, J. S. (1999). Bark and ERB bilinear transforms. *IEEE Transactions on speech and Audio Processing*, *7*(6), 697-708.

Sriraam, T. S. N., & Pradeep, G. (2019). Pre-term Neonates Cry Pattern Recognition Using Bark Frequency Cepstral Coefficients. 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE).

Sulpizio, S., Esposito, G., Katou, M., Nishina, E., Iriguchi, M., Honda, M., . . . Shinohara, K. (2019). Inaudible components of the human infant cry influence haemodynamic responses in the breast region of mothers. *The Journal of Physiological Sciences*, *69*(6), 1085-1096.

Sun, Q.-S., Zeng, S.-G., Liu, Y., Heng, P.-A., & Xia, D.-S. (2005). A new method of feature fusion and its application in image recognition. *Pattern Recognition*, *38*(12), 2437-2448.

Synnergren, J., Olsson, B., & Gamalielsson, J. (2009). Classification of information fusion methods in systems biology. *In silico biology*, *9*(3), 65-76.

Tamazin, M., Gouda, A., & Khedr, M. (2019). Enhanced automatic speech recognition system based on enhancing power-normalized cepstral coefficients. *Applied Sciences*, *9*(10), 2166.

Tejaswini, S., Sriraam, N., & Pradeep, G. (2020). Identification of High Risk and Low Risk Preterm Neonates in NICU: Pattern Recognition Approach. In *Biomedical and Clinical Engineering for Healthcare Advancement* (pp. 119-140). IGI Global.

Telgad, R. L., Deshmukh, P., & Siddiqui, A. M. (2014). Combination approach to score level fusion for Multimodal Biometric system by using face and fingerprint. International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014).

Terveen, L., & Hill, W. (2001). Beyond recommender systems: Helping people help each other. *HCI in the New Millennium*, *1*(2001), 487-509.

Thodén, C.-J., Järvenpää, A.-L., & Michelsson, K. (1985). Sound spectrographic cry analysis of pain cry in prematures. *Infant crying: Theoretical and research perspectives*, 105-117.

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.

Toh, A. M., Togneri, R., & Nordholm, S. (2005). Spectral entropy as speech features for speech recognition. *Proceedings of PEECS*, *1*, 92.

Torres, R., Battaglino, D., & Lepauloux, L. (2017). Baby cry sound detection: A comparison of hand crafted features and deep learning approach. International Conference on Engineering Applications of Neural Networks.

Trivedi, M. M., & Bezdek, J. C. (1986). Low-level segmentation of aerial images with fuzzy clustering. *IEEE Transactions on Systems, Man, and Cybernetics*, *16*(4), 589-598.

Tuduce, R. I., Cucu, H., & Burileanu, C. (2018). Why is my baby crying? An in-depth analysis of paralinguistic features and classical machine learning algorithms for baby cry classification. 2018 41st international conference on telecommunications and signal processing (TSP).

Tusty, N. M., Basaruddin, T., & Krisnadhi, A. (2020). Infant cry classification using CNN–RNN. Journal of Physics: Conference Series.

Unicef. (2014). *Levels and trends in child mortality*. https://data.unicef.org/resources/levels-trends-child-mortality-report-2014/.

Unicef. (2019). *Levels and trends in child mortality* https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2019/.

Unicef. (2020a). *Levels and trends in child mortality*. https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2020/.

Unicef. (2020b). *Sepsis Diagnostic-Infection Prevention and Control*. https://www.unicef.org/supply/media/2946/file/sepsis-diagnostic-use-cases.pdf.

Unicef. (2022). *Levels and trends in child mortality*. https://data.unicef.org/resources/levels-and-trends-in-child-mortality/.

Vaishnavi, V., & Dhanaselvam, P. S. (2019). An Automatic Approach to Extract Features from the Infant's Cry Signals.

Valero, X., & Alias, F. (2012). Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. *IEEE Transactions on Multimedia*, *14*(6), 1684-1689.

Velikova, M., Lucas, P. J., Samulski, M., & Karssemeijer, N. (2012). A probabilistic framework for image information fusion with an application to mammographic analysis. *Medical Image Analysis*, *16*(4), 865-875.

Verma, A., Agrawal, R., Singh, P. K., & Ansari, N. A. (2022). An Acoustic Analysis of Speech for Emotion Recognition using Deep Learning. 2022 1st International Conference on the Paradigm Shifts in Communication, Embedded Systems, Machine Learning and Signal Processing (PCEMS).

Vihinen, M. (2012). How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. BMC genomics.

Vincent, P. D. R., Srinivasan, K., & Chang, C.-Y. (2021). Deep learning assisted neonatal cry classification via support vector machine models. *Frontiers in Public Health*, *9*, 670352.

Vuorenkoski, L., Vuorenkoski, V., & Anttolainen, I. (1973). 21. Cry analysis in congenital hypothyroidism: an aid to diagnosis and clinical evaluation *Acta Pædiatrica*, *62*, 27-28.

Vuorenkoski, V., Lind, J., Wasz-Hockert, O., & Partanen, T. (1971). Cry score. A method for evaluating the degree of abnormality in the pain cry response of the newborn and young infant. *Speech Transmission Laboratory—Quarterly Progress and Status Report (STL–QPSR)(Stockholm)*, *12*, 68-75.

Wahid, N., Saad, P., & Hariharan, M. (2016). Automatic infant cry classification using radial basis function network. *Journal of advanced research in applied sciences and engineering technology*, *4*(1), 12-28.

Wainer, J., & Fonseca, P. (2021). How to tune the RBF SVM hyperparameters? An empirical evaluation of 18 search algorithms. *Artificial Intelligence Review*, *54*(6), 4771-4797.

Warley, M., & Gairdner, D. (1962). Respiratory distress syndrome of the newborn—principles in treatment. *Archives of Disease in Childhood*, *37*(195), 455.

Wasz-Hockert, O. (1968). The infant cry: A spectrographic and auditory analysis. *Clinics in developmental medicine*, 1-42.

Wasz-Hockert, O., Lind, J., Partanen, T., Valanne, E., & Vuorenkoski, V. (1968). *The infant cry: A spectrographic and auditory analysis* (Vol. 29). Heinemann London.

Wasz-Hockert, O., Valanne, E., Vuorenkoski, V., Michelsson, K., & Sovijarvi, A. (1963). Analysis of some types of vocalization in the newborn and in early infancy. Annales Paediatriae Fenniae.

Weiss, S. L., Pomerantz, W. J., Torrey, S. B., & Kaplan, S. L. (2019). Septic shock in children: Rapid recognition and initial resuscitation (first hour). *U: UpToDate, Randolph AG, Torrey SB, Kaplan SL ed. UpToDate [Internet]. Waltham, MA: UpToDate*.

Wermke, K., Teiser, J., Yovsi, E., Kohlenberg, P. J., Wermke, P., Robb, M., . . . Lamm, B. (2016). Fundamental frequency variation within neonatal crying: Does ambient language matter? *Speech, Language and Hearing*, *19*(4), 211-217.

World Helath Organization (WHO). (2014). *Every newborn: an action plan to end preventable deaths* (9241507446). https://www.who.int/initiatives/every-newborn-action-plan.

World Helath Organization (WHO). (2014). *Newborn infections*. https://www.who.int/teams/maternal-newborn-child-adolescent-health-and-ageing/newborn-health/newborn-infections.

World Helath Organization (WHO) (2021). *Newborn Mortality-levels and trends in child mortality report*. https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-mortality-report-2021.

Winters-Hilt, S., Yelundur, A., McChesney, C., & Landry, M. (2006). Support vector machine implementations for classification & clustering. BMC bioinformatics.

Wu, K., Zhang, C., Wu, X., Wu, D., & Niu, X. (2019). Research on acoustic feature extraction of crying for early screening of children with autism. 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC).

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., . . . Philip, S. Y. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*(1), 1-37.

Wynn, J. L., & Wong, H. R. (2010). Pathophysiology and treatment of septic shock in neonates. *Clinics in perinatology*, *37*(2), 439-479.

Yang, W., Wang, K., & Zuo, W. (2012). Neighborhood component feature selection for high-dimensional data. *J. Comput.*, *7*(1), 161-168.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., . . . Povey, D. (2002). The HTK book. *Cambridge university engineering department*, *3*(175), 12.

Zabidi, A., Khuan, L. Y., Mansor, W., Yassin, I. M., & Sahak, R. (2010a). Classification of infant cries with asphyxia using multilayer perceptron neural network. 2010 Second International Conference on Computer Engineering and Applications.

Zabidi, A., Khuan, L. Y., Mansor, W., Yassin, I. M., & Sahak, R. (2010b). Detection of infant hypothyroidism with mel frequency cepstrum analysis and multi-layer perceptron classification. 2010 6th International Colloquium on Signal Processing & its Applications.

Zabidi, A., Mansor, W., Khuan, L. Y., Yassin, I. M., & Sahak, R. (2009). Classification of infant cries with hypothyroidism using multilayer perceptron neural network. 2009 IEEE international conference on signal and image processing applications.

Zabidi, A., Mansor, W., & Lee, K. Y. (2017). Optimal Feature Selection Technique for Mel Frequency Cepstral Coefficient Feature Extraction in Classifying Infant Cry with Asphyxia. *Indonesian Journal of Electrical Engineering and Computer Science*, *6*(3), 646-655.

Zabidi, A., Mansor, W., Lee, Y. K., Yassin, I. M., & Sahak, R. (2011). Binary particle swarm optimization for selection of features in the recognition of infants cries with asphyxia. 2011 IEEE 7th International Colloquium on Signal Processing and its Applications.

Zabidi, A., Yassin, I., Hassan, H., Ismail, N., Hamzah, M., Rizman, Z., & Abidin, H. Z. (2017). Detection of asphyxia in infants using deep learning convolutional neural network (CNN) trained on Mel frequency cepstrum coefficient (MFCC) features extracted from cry sounds. *Journal of Fundamental and Applied Sciences*, *9*(3S), 768-778.

Zayedi, Y., Hasasneh, A., & Tadj, Ch. (2023). Infant cry signal diagnostic system using deep learning and fused features. Jornal of Diagnostics, 12, 2107.

Zeifman, D. M. (2001). An ethological analysis of human infant crying: answering Tinbergen's four questions. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *39*(4), 265-285.

Zeskind, P. S., & Lester, B. M. (1981). Analysis of cry features in newborns with differential fetal growth. *Child Development*, 207-212.

Zhang, K., Ting, H., & Choo, Y. (2024). Baby cry recognition based SLGAN model data generation and deep future fusion. Jornal of Expert System with Applications, 242, 122681.

Zhang, X., Zou, Y., & Liu, Y. (2018). AICDS: An infant crying detection system based on lightweight convolutional neural network. Artificial Intelligence and Mobile Services–AIMS 2018: 7th International Conference, Held as Part of the Services Conference Federation, SCF 2018, Seattle, WA, USA, June 25-30, 2018, Proceedings 7.

Zhao, X., & Wang, D. (2013). Analyzing noise robustness of MFCC and GFCC features in speaker identification. 2013 IEEE international conference on acoustics, speech and signal processing.

Zhong, H., & Xiao, J. (2017). Enhancing health risk prediction with deep learning on big data and revised fusion node paradigm. *Scientific Programming*, *2017*.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland*, *19*, 67.