Machine Learning-Based Cry Diagnostic System for Identifying Septic Newborns

Fatemeh Salehian Matikolaie, and Chakib Tadj, *Montreal, QC, Canada*

Summary: Background and Objective. Processing the newborns' cry audio signal (CAS) provides valuable information about the newborns' condition. This information can be used to diagnose the disease. This article analyzes the CASs of newborns under two months old using machine learning approaches to develop an automatic diagnostic system for identifying septic infants from healthy ones. Septic infants have not been studied in this context.

Methodology. The proposed features include Mel frequency cepstral coefficients and the prosodic features of tilt, rhythm, and intensity. The performance of each feature set was evaluated using a collection of classifiers, including Support Vector Machine (SVM), decision tree, and discriminant analysis. We also examined the majority voting method for improving the classification results and feature manipulation and multiple classifier framework, which has not previously been reported in the literature on developing an automatic diagnostic system based on the infant's CAS. We tested our methodology on two datasets of expiration and inspiration episodes of newborns' CASs.

Results and Conclusion. The framework of the concatenation of all feature sets using quadratic SVM resulted in the best F-score with 86% for the expiration dataset. Furthermore, the framework of tilt feature set with quadratic discriminant with 83.90% resulted in the best F-score for inspiration. We found out that septic infants cry differently than healthy infants through these experiments. Thus, our proposed method can be used as a noninvasive tool for identifying septic infants from healthy ones only based on their CAS.

Key Words: Sepsis−Infants’ cry−Mel frequency cepstral coefficient−Prosodic feature−Principal component analysis−Feature manipulation−Support vector machine−Decision tree−Discriminant analysis−Classifiers fusion.

INTRODUCTION

In recent years, the infant mortality rate in developed countries has decreased. However, this rate is still high in developing countries. Moreover, saving newborns' lives and promoting their health is of particular importance in the health of any nation and for further providing health services. This paper set out a Newborn Cry Diagnostic System (NCDS) to see if we can apply machine learning techniques to categorize newborns' cry audio signal (CAS) as septic or healthy. This section discuss what the CAS is, the types of NCDSs proposed by researchers, the problems they faced, and how we can apply them to sepsis pathology, which has not been studied before.

The act of crying for infants is their most prominent communication activity. Infants produce cry by pushing airflow from their lungs to the vocal tract,[1] and then airflow vibrates the vocal cords, which generates the sound. Lungs work like power and provide patterns. This explanation is called source-filter theory. In general, the CAS results from the altered sound of source (vibrating larynx) by the vocal tract. The set from vocal cords to the lips forms the vocal tract.[2] The vocal tract adjusts the vocalization and works as a filter. It attenuates or amplifies some frequencies.

Crying is the infants’ only weapon against the inconveniences like hunger, pain, discomfort, and infection that happen to them. Hence, crying is a natural warning method to call on those around to help. Not responding correctly to these warning signs can cause harm to the infant and their parents. A fair number of researchers indicated that infants' CAS holds information that, if properly analyzed, can be used to access messages sent from the newborns’ brain.[3] We also know that mothers and hospital staff who are constantly in contact with infants can distinguish several infants’ needs only based on their CAS.[4]

Further investigations on infants' CAS even revealed its reliability for diagnostic purposes.[5] In the study presented by,[3] they anecdotally explained the characteristics of the CAS of infants affiliated with specific diseases such as asphyxia, deafness, etc., versus healthy ones. There are patterns in a CAS that warn about the menacing pathology for the infant's health, which may be clueless even in physical examinations by doctors.[6]

The infants' CAS has been studied for decades.[7] Traditional popular approaches were based on visual inspections of the spectrogram of infants' CASs.[3] However, manually sorting the patterns in the CAS and categorizing accordingly are not practical for human beings due to the vast amount of information for processing.[8] Thus, this shortcoming has led to various automatic classification systems. There have been works on developing an automatic system for recognizing the infants' CASs from other surrounding

Accepted for publication December 29, 2021.
From the Department of Electrical Engineering, École De Technologie Supérieure, Montreal, QC, H3C 1K3, Canada.

Address correspondence and reprint requests to Fatemeh Salehian Matikolaie, Department of Electrical Engineering, École De Technologie Supérieure, Montreal, QC, H3C 1K3, Canada. E-mail: fatemeh.salehianmatikolaie.1@ens.etsmtl.ca
Journal of Voice, Vol. 38, No. 4, pp. 963.e1−963.e14
0892-1997
© 2022 The Voice Foundation. Published by Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jvoice.2021.12.021

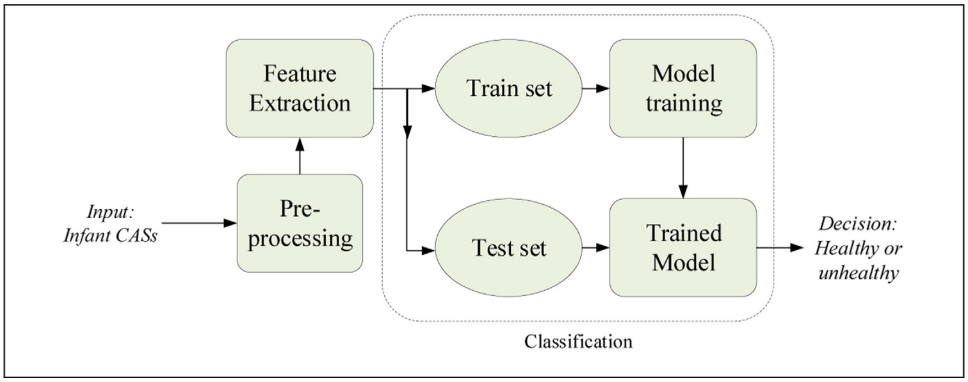

FIGURE 1. Block diagram of the NCDS.

sounds,[9] detecting different parts of CAS (such as episodes of expiration and inspiration),[8,10] identifying the need of an infant (hunger, diaper, sleepy, etc).[6,11,12] The very recent one is the pathology detection task.[13−15] In our work, we also focused on developing an automatic pathology detection task which we call it the NCDS in this reading.

Figure 1 shows the block diagram of the NCDS. The NCDS framework, like any identification system, includes the phases of pre-processing, feature extraction, and a phase of training a model based on the obtained features for classification. The pre-processing step aims to help better feature extraction. It includes applying various applications such as reemphasizing, windowing, and finding the Fundamental frequency (F_0). In the feature extraction phase, methods such as Mel Frequency Cepstral Coefficient (MFCCs),[14−19] auditory-inspired amplitude modulation,[20] Linear Prediction Coding (LPC),[16,18,19] patterns of F_0 contour,[15,17] resonance frequency[17] are the most common ones. Furthermore, extra analysis such as combining different feature sets such as MFCC and LPC[18] and MFCC, rhythm, and tilt features[15] were studied in this phase. Besides, techniques for identifying the most relevant features such as F-ratio and binary particle swarm optimization,[21] and orthogonal least square algorithm[21] for improving the classification performance were suggested.

In the classification phase a variety of pattern recognition models have been studied including Support Vector Machine (SVM),[14,15] multilayer perception neural network,[14,21] probabilistic neural network,[14,17,22] decision tree,[16] forest[16] and k-nearest neighbor algorithm.[16] The CASs of pathologies that were yet investigated by machine learning approaches to automatically identify sick infants from healthy ones includes cleft palate,[23,24] hearing disorder,[11,13,16,19,25] hyperbilirubinemia,[26] autism,[27] asphyxia,[11,16,19,21,26,28−30] hypothyroidism[31] and respiratory distress.[15,23]

In this study, our contribution is twofold. One contribution is how we evaluate and manipulate features and how we use these features to make a final decision. The second contribution is to look at the unstudied pathology of sepsis. We performed four sets of experiments. We considered each expiration episode and inspiration episode of infants' CAS in the first experiment as a sample. The expiration episode and inspiration episode are perceivable sounds during exhalation and inhalation of infants during crying, and the silence episode is the soundless gap between inspiration and expiration episodes of CAS.[32,33] We refer to this experiment as the Single Episode (SE) experiment. In the second experiment, we used the predicted labels for episodes within each CAS from the SE experiment to predict each CAS label using the majority voting technique. We call this experiment the All Episode (AE) experiment. We borrowed this idea from the automatic environmental sound classification presented by.[34] Accordingly, in the SE experiment, we evaluated the performances of prosodic features of intensity, rhythm, tilt, and the commonly used feature of MFCCs using three sets of classifiers of SVM, discriminant analysis, and decision tree. Next, in the AE experiment, we used majority voting to predict the CAS group using the labels obtained from the SE experiment. In the third experiment, we examined the idea of the concatenation of all feature sets of MFCC and the three prosodic feature sets of tilt, rhythm, and intensity and then fed them to the classifiers. In the fourth experiment, we set up a framework to aggregate the prediction of the most competent classifiers for each set of features and then predict the CAS label using the majority voting technique. We explain these methods later in the methodology section.

Regarding our second contribution, according to our knowledge, despite the frequent infants' death due to sepsis, disappointingly, so far, there is no investigation on the connection between the CASs of infants with sepsis and this pathology. Previously in our research lab, sepsis pathology in newborns was investigated in a multi pathology group; however, there is no study on the unique pathology group of sepsis when investigating the symptoms in the infants' CAS. In Canada alone in 2019, among the newborns' cause of death, sepsis is reported on rank six.[35] The rank of sepsis among leading to death has increased in recent years, as shown in Table 1. Thus, it would be beneficial to have an NCDS to classify septic from healthy ones. Sepsis is a severe disease that is usually caused by bacteria. Infants under two months are more prone to sepsis because their immune systems are

TABLE 1.
The Leading Cause of Newborns' Death Related to Sepsis Pathology in Canada[35]

	2015	2016	2017	2018	2019
Sepsis rank	9	8	6	8	6
Number of infant deaths	31	32	43	47	38

not yet developed enough to fight off some sources of infection. In clinical findings, a set of specific symptoms are reported for sepsis. However, newborns have few apparent symptoms; these may vary from child to child.

Perhaps the reason that this critical pathology of septic remained unstudied is that enough data did not exist. Consequently, having this dataset available in our lab lends support to delivering this work. An NCDS is a valuable tool in saving lives and promoting newborns' health levels, specifically in developing countries suffering from the lack of pediatricians. The NCDS would address this issue as its installation cost is relatively low.[7,14] Practical applications of the NCDS include its use for infant screening,[36] infancy education,[37] robot nursing,[38] and as a medical assistant for pediatricians. Moreover, NCDS is a non-intrusive tool.

The paper is prepared as follows: Section 2 describes the collection of dataset, information of the dataset, the participants, feature sets definition, and explanations of the examined classifiers in this work; Section 3 reports the results of the four implemented experiments including the SE experiment, the AE experiment, and the effects of feature manipulation and use of multiple classifier framework. Section 4 concentrates on the discussion of the research developed, including the usefulness of each feature set, the feature reduction schemes, the majority voting technique, and the computation cost of each framework. Lastly, a complete list of acronyms introduced in this article is given in Table 14 in section Appendix.

MATERIALS AND METHODS

Dataset Description

This section describes the data collection procedure, the dataset details, the participants in our experiments, and the dataset preprocessing procedures.

Data Collection and Recording

The research group in our laboratory collected the CASs of infants at Hôpital Sainte-Justine in Montréal, Canada, and hospitals of Al-Sahel and Al-Raee in Lebanon. The hospital staff of mentioned hospitals recorded the CASs in the clinical medium. A 2-channel digital hand-held WS-650M Olympus digital voice recorder was posed at 10−30 cm from the infants. The sampling frequency of the recordings is 44.1 kHz, and the sample resolution is 16 bits. In the recording procedure, careful attention was given to maintaining the surrounding noise at the minimum level. Thus, every newborn's CAS was recorded independently, and whenever the environmental noise would rise, they would stop the recording procedure.[14] For more information about the procedure of recording, the author suggests reading.[14]

Alongside the recordings phase, they collected details of infants, including the reason for crying, gestational age, birth weight, Apgar[1] result, gender, name of the hospital, type of disease, infants' age during the recording, and prematurely state of the infants.

In our study, our dataset includes the CASs of infants initiated by various reasons of crying, and were recorded at different times in a day.[39] The reason for crying includes CASs initiated by hunger, discomfort, diaper, blood tests, shower, birth, collection of urine, etc. The term "reason for crying" refers to the stimuli that caused the newborns start crying. This term "reason for crying" is irrelated to the newborns' health condition. The staff would note why the newborn started crying in the CAS collection procedure.

Participants

The age range of infants in our dataset is from one day to 208 days. However, in the current experiment, similar to our previous ones,[14,15,17] we excluded the CASs of infants whose ages were more than 53 days. This is because infants above this age can control their voices.[3]

The female and male newborns of ethnic groups of half Caucasian and half Haitian, African, Arabic, Caucasian, Latino, Native Hawaiian, and Quebecois were included in our experiments. The CASs of infants studied in this experiment are either healthy or affected by sepsis pathology. The pathology dataset consisted of 53 recordings of 17 infants with sepsis whom pediatricians diagnosed through medical examinations. Each infant in our dataset has one or more recordings. For the healthy dataset, there were 108 infants' CASs. We only used an equal number of expiration and inspiration of the CASs in our experiment to observe the balanced dataset for precise diagnosis by the classification models. Table 4 shows the number of episodes in each dataset of expiration and inspiration in our experiments.

To increase the credibility of our proposed model, we imposed criteria for our system similar to our previous work.[15,20] Our dataset was variable in conditions. First, we included all reasons for crying initiated for various reasons, while reason for crying affects the durational feature of the CAS.[15] Second, we considered a wide variety of newborns whose parents are from different linguistic groups. This is important as we know that the unborn infants start learning the prosodic features such as rhythm, intensity, and melody from the last three months of pregnancy, which affects the prosodic aspect of the CAS production as discussed in.[40] Lastly, the CASs were recorded in hospitals, including

[1]The very first test taken from newborns for measuring the newborn's general health state.

ambient noises such as human speech, the instrument's sound, etc.

The recording condition was the same for all infants, including healthy ones and those with sepsis. Moreover, to ensure that the NCDS is only learning the pathologically informed patterns, we only used the vocal segments of expiration and inspiration episodes of the CAS that do not include noise, which are explained in Table 2 as "EXP" and "INSV". Other segments of the newborns' CAS, including "Background, BIP, noisy Crying, noisy pseudo-Crying, Noise," are noisy, and thus we excluded them in our study. The segments of the CAS are listed in Table 2.

TABLE 2.
Example Description of Some CAS Labels

Label	Description
EXP	Voiced expiration segment during a period of crying
EXPN	Unvoiced expiration segment during a period of crying
INS	Unvoiced inspiration segment during a period of crying
INSV	Voiced inspiration segment during a period of crying
EXP2	Voiced expiration segment during a period of pseudo-crying
INS2	Voiced inspiration segment during a period of pseudo-crying
PSEUDOCRY	Any sound generated by the baby and it is not crying
Speech	Sound of the nurse or parents talking
Background	Low noise characterized by a very low power-silence affected with little noise
BIP	Sound of the medical instruments next to the baby
Noisy Crying	Any sound heard along with the crying: machine's beep, water, diaper, etc.
Noisy pseudo-Crying	Any sound heard along with the pseudo-crying
Noise	The sound caused by the mic being moved, diaper, a door, speech + background, speech + beep, etc.

Methodology

This study used two datasets of expiration and inspiration episodes of infants' CAS. We extracted features from several levels in the SE experiment, including tilt, rhythm, intensity, and MFCCs from each episode of expiration and inspiration datasets. Then we fed them to different models for classification. We examined each dataset separately. Figure 2 illustrates the scheme of the SE experiment for a portion of CAS. In the AE experiment, we used all predicted labels of every single episode in the CAS from the SE experiment to indicate the label of the single CAS using the majority voting technique. Figure 3 shows the scheme of the AE experiment.

In another set of an experiment similar to the method in our previous study,[15] we concatenated all feature sets together, and in this study, we also added the intensity features.

In the last experiment, we used the best classifiers for each set of features and labeled the CAS based on the most predicted labels. The aim was to choose the framework which results in the most accurate recognition for identifying the CASs of septic infants.

The CAS Feature Description

The extracted features are in the temporal, spectral, and both domains. The feature sets include MFCC and the prosodic features of tilt, intensity, and rhythm. In the following, we bring the description of each of these feature sets and the details of the parameters we used.

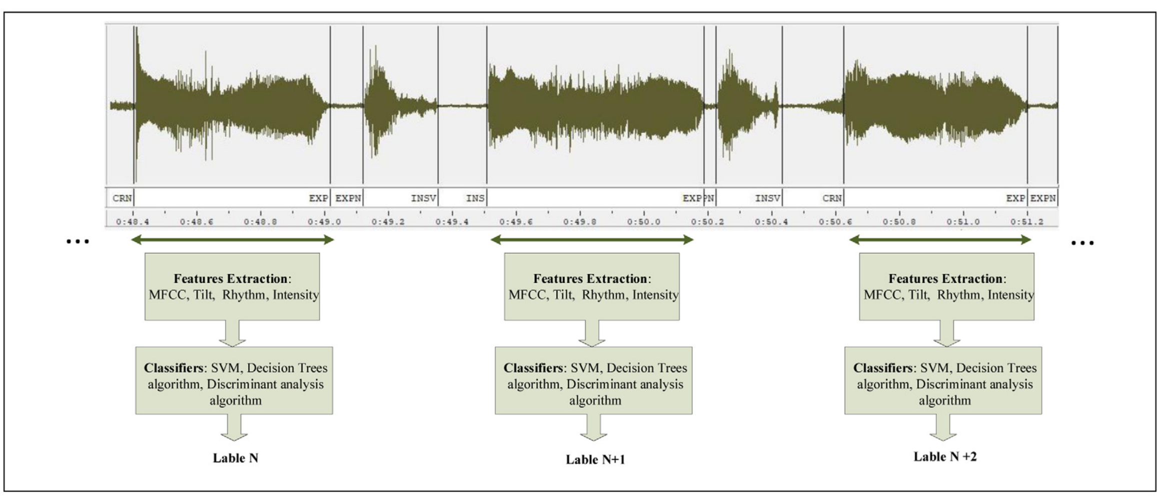

FIGURE 2. Illustration of procedure of the SE experiment in a portion of an infants' CAS visualized using WaveSurfer software.

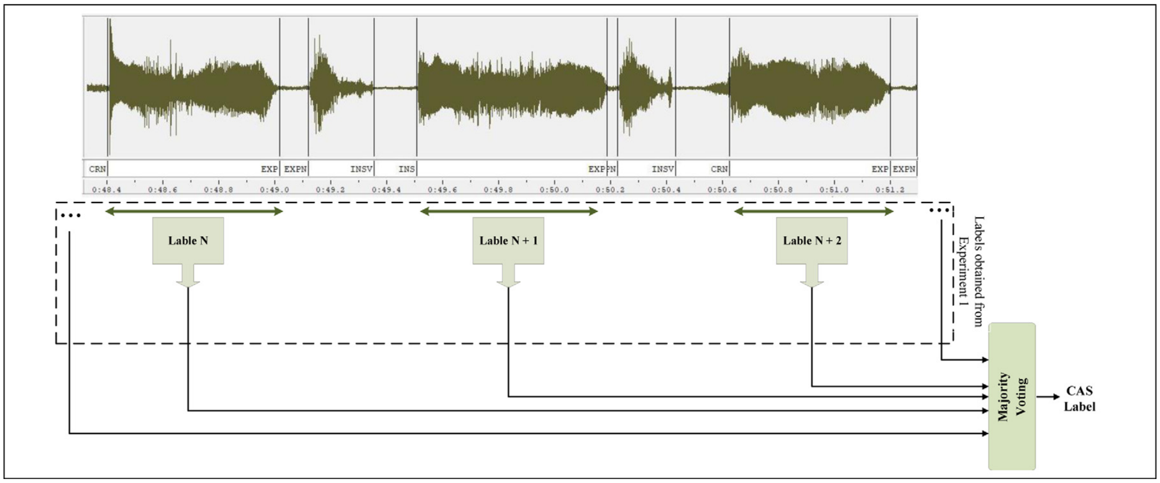

FIGURE 3. Illustration of procedure of the AE experiment in a portion of an infants' CAS visualized using WaveSurfer software.

Mel Frequency Cepstral Coefficients (MFCC)

Among several algorithms introduced in speech processing for characterization, the MFCC feature set is the most widely used method in adult and infant voice processing.[15]

Mel frequency cepstrum shows the power spectrum of an audio signal using the linear cosine transform of the power spectrum logarithm at the Mel scale. The Mel scale is defined as Equation 1.

$$M(f) = 1125\ ln\left(1 + \frac{f}{700}\right) \tag{1}$$

Where f is the frequency value, and $M\ (f)$ is the corresponding Mel value. The MFCC coefficients can be defined as the logarithmic cosine conversion of the energy obtained by applying the Mel Bank filter to the windowed signal spectrum. The steps for calculating the MFCC coefficients are shown in Figure 4.

The coefficients extracted from each frame contain only the static information of the frame, which causes the effect of adjacent structures not to be considered, and due to the nonstationarity of the newborns' CAS, the feature vector of each frame should also reflect changes in spectral characteristics. Thus, the feature vector of each frame also includes the time derivatives of the extraction coefficients. For further information on MFCC, the authors suggest reading.[41]

In this study, we only analyzed the information less than the frequency of 4 kHz according to the result of our experiment in[14] for infants' CASs. In the windowing stage, we used a hamming window with a frame size of 10 ms, with a 30% overlap between each consecutive frame. Our previous work[14] showed that the frame length of 10 ms performs better than 30 ms. Moreover, we set the number of filter bank channels to 24. These adjustments that suit infants' CAS processing are based on our previous experiments.[14,15]

Tilt Feature

The F_0 is defined as the harmony of the oscillation of the vocal folds.[3,7] The pattern of changes in F_0 repeatedly has been described as relevant with some pathology[3] in newborns. The tilt feature represents changes in F_0 of the voice. The tilt feature is based on the F_0 and was initially presented by[42] in an automatic speech recognition system and also was successfully used in our previous study[15] for developing the NCDS for infants with RDS. Tilt parameters capture the changes of the F_0 using parameters called A_t and D_t. In the present study, we followed the method provided by.[42] The parameters A_t and D_t are presented respectively by Equations 2 and 3:

$$A_t = \left(\frac{\left|A_r\right| - \left|A_f\right|}{\left|A_r\right| + \left|A_f\right|}\right) \tag{2}$$

$$D_t = \left(\frac{\left|D_r\right| - \left|D_f\right|}{\left|D_r\right| + \left|D_f\right|}\right) \tag{3}$$

Considering the contour of F_0 in a portion of CAS, A_r is the amplitude of the F_0 when it rises to reach the peak of F_0, and A_f is alternatively the amplitude when it is declining. Correspondingly, D_f and D_r respectively measure the distance between the rising and falling parts of the F_0 contour. This feature set is described in detail in.[42]

For extracting the tilt features, the requirement is to find the accurate F_0 contour. Finding the F_0 in newborns' CAS is hindered by the high instability of the infants' CAS.[7] Among the popular software for extracting F_0, the most precise one is Praat software.[43] Therefore, we extracted the F_0 using Praat software. Table 3 shows an example of the result of F_0 extracting using Praat software.

The values of A_t, D_t and the F_0 of each episode of the CAS were computed. Finally, the statistical measures of the range, mean, standard deviation, median and interquartile range of these values were put in the feature vector.

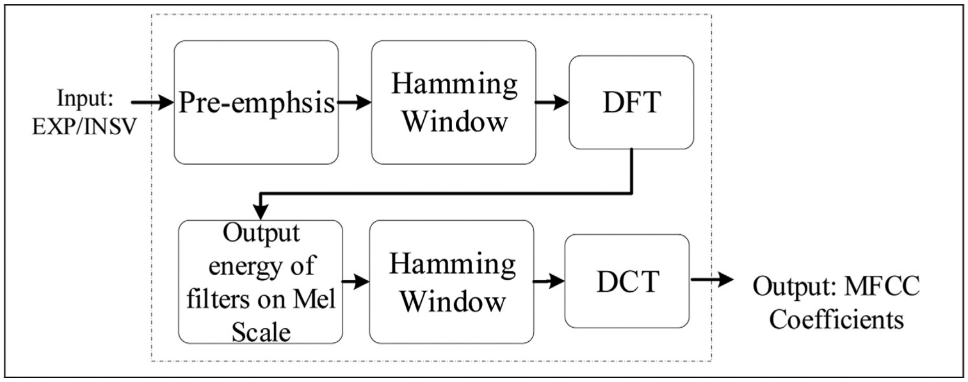

FIGURE 4. The block diagram of MFCC features extraction.

TABLE 3.
The Evaluated Values of F_0 and Intensity by Praat Software For a Portion of CAS Within the Period of 0.014 to 0.0.17 Seconds

Time	F_0	Intensity
0.0140	0	-undefined-
0.0280	373.3105	-undefined-
0.0420	376.4588	-undefined-
0.0560	379.6858	77.751
0.0700	370.2263	77.362
0.0840	361.8400	76.333
0.0980	362.1973	75.891
0.1120	367.0559	75.978
0.1260	364.5674	76.924
0.1400	363.7566	78.619
0.1540	365.9141	79.855
0.1680	369.2621	80.186

Rhythm Feature

In this study, we also investigated the usefulness of the duration feature, which is a subset of the rhythm feature. We calculated the duration of expiration and inspiration episodes within each CASs.

Intensity Feature

This feature was already used to automatically identify expiration and inspiration episodes of infants' CASs.[39] Intensity is the measure of the loudness of the signal. It measures the quantity of energy that the signal conveys per unit area. The intensity magnitude is calculated by Equation 4:

$$Intensity = 10loglog\ (\ A^2(n)w(n)) \qquad (4)$$

In this equation, w and A respectively refer to the window function and the amplitude of the CASs. We used Praat software to precisely estimate the intensity of infants' CAS. Table 3 shows the results of extracting this feature from a portion of CAS in our dataset. Like tilt feature extraction, the statistical measures of the range, mean, standard deviation, median and interquartile range of the values of intensity features were put in the feature vector.

Feature Reduction: Principal Component Analysis (PCA)

Feature selection is used for reducing the dimensionality size of measuring space by eliminating the low effect or useless features. The PCA method is one of the best methods for decreasing feature dimensionality linearly. It can identify key components and help the classifier analyze a set of features that are more valuable in conveying specific group information than just examining them all. This algorithm tries to represent the features in a way that highlights their similarities and differences. This technique defines new axes for the features, and these new axes display the features. The first axis should be placed in a direction, which maximizes the data variance. In other words, in a direction in which the distribution of features is highest. Then the second axis is perpendicular to the aforementioned axis. For more information on PCA, the authors suggest reading.[44]

Besides the favored method of PCA, we experimented with the statistical metrics as a feature reduction method. The statistical metrics include the range, mean, standard deviation,[45,46] median and interquartile range[15] for compressing the size of MFCCs vectors. In the evaluation section, we compare the results and the cost of processing time of each method of PCA and statistical measures.

Classifiers

The classification approaches taken in this study are classification by a single episode called SE experiment shown in Figure 2 and classification by the whole episodes in CAS called AE experiment shown in Figure 3. In the SE experiment, each episode of CAS, including expiration or inspiration (referred to "EXP" and "INSV" in Table 2) is considered a sample, and the model is trained to assign a label. While in the AE experiment, we used the majority voting technique to vote based on the number of the most predicted label in each CAS.

To develop a comparison, we investigated the performance of 14 classifiers from three families to investigate the most credible functional one in identifying the CASs of unhealthy infants suffering from sepsis from healthy ones. In the following, we describe the three families of classifiers.

Support Vector Machine (SVM) Algorithm: Five Classifiers

The SVM learning algorithm is known as one of the best classifications and outlier detection methods. The basis of the SVM classifier is the linear classification of data. The SVM approach selects the decision boundary to maximize the minimum distance between the particular classes. This selection mechanism makes the classifiers' decisions in practice well tolerable to noise conditions. The border selection in SVM is based on support vector points.[47] In this study the linear, cubic, quadratic, fine Gaussian, and medium Gaussian SVM classifiers are included.

Decision Tree Algorithm: Six Classifiers

This algorithm develops a set of conditions in tree construction to predict the class of a feature. The tree algorithm is based on minimizing the diversity at nodes. The lack of uniformity in the nodes is measurable using the criteria of impurity measure. The difference between tree classifiers is due to the impurity measure, splitting method, and pruning tree nodes.[48] This paper evaluated the performance of six tree classifiers, including simple, medium, complex, bagged, boosted, and RUSBoost trees.

Discriminant Analysis Algorithm: Three Classifiers

In this algorithm, the assumption is that different classes generate data based on different Gaussian distributions. In other words, every class is assumed to be a normally distributed cluster of data points. In this survey, we constructed the linear, quadratic, and subspace discriminant analysis algorithm.

After performing the SE and AE experiments using the explained method, we put together the most competent classifiers for each feature set. These classifiers' predicted labels were then fed to a majority voting block to predict the CAS class as healthy or septic. This idea is based on the assumption that the classifiers perform in a complementary way to enhance the predictive result. We also concatenated all feature sets together and fed them to all classifiers to compare the results.

MODEL EVALUATION AND RESULTS

All the feature extraction, classification, and evaluation stages were performed using Matlab. We utilized features from several domains and different classifiers with several kernels to capture the best result. For measuring each frameworks' ability to identify the CAS of infants with sepsis disease from healthy ones, we used the standard metrics in the pathology diagnostic field, including specificity, recall, and F-measure.[49] The followings are the definitions of our evaluation measures.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{5}$$

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{6}$$

$$Fscore = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

In our case, a "True Positive" would be correctly identifying an infant with the septic pathology. The performance of the classifiers was measured with 5-fold cross-validation. To ensure the validity of our model, we designed the distribution of CASs between the folds to guarantee the independence of the folds. In other words, there are no samples of the same infants in more than one fold. Accordingly, in each iteration, the models learn on four folds (called trained folds) on the CASs of some infants and then test the one fold (called test fold) which does not include any sample of the infants in the training folds. In each iteration, one fold becomes the test fold. We used two datasets in our research. The dataset includes the expiration and inspiration episodes of CAS. These episodes are called EXP and INSV in Table 2. Table 4 presents the number of samples in each fold for each dataset of expiration and inspiration. There are more expiration samples than inspiration samples because there were some inspiration segments that we could not determine its $F0$ values or they were very short intervals.

TABLE 4.
The Number of Samples From Healthy and Septic Newborns in Each Fold

	Class Healthy (EXP)	Class Sepsis (EXP)	Class Healthy (INSV)	Class Sepsis (INSV)
Fold One	507	507	140	140
Fold Two	517	517	141	141
Fold Three	524	524	139	139
Fold Four	523	523	132	132
Fold Five	453	453	109	109

TABLE 5.
The Classification Results of Discriminant Analysis Family Classifiers Using Statistical Measure and PCA for MFCC Features

	EXP Dataset		INSV Dataset	
	PCA	Statistical Measures	PCA	Statistical Measures
Linear Discriminant (SE experiment)	54.60%	65.50%	54.70%	65.20%
Linear Discriminant (AE experiment)	70.20%	77.30%	74.60%	80.85%
Quadratic Discriminant (SE experiment)	59.20%	65.10%	65.60% *	64.30%
Quadratic Discriminant (AE experiment)	72.40%	78.10%	**83.00% ***	79.20%
Subspace Discriminant (SE experiment)	54.50%	68.10%	54.20%	57.10%
Subspace Discriminant (AE experiment)	70.40%	**81.00%**	72.80%	73.20%

Percentages refer to F-score. The results of the best frameworks are bolded. The * sign indicates the results of the classification frameworks in which the PCA method resulted in a better recognition power than the statistical measure reduction method.

Evaluation of MFCC Features

In this study, we decided to investigate further the MFCC features from the previous study that we presented in.[15] We evaluated the results of dimension reduction techniques for MFCC features, including statistical measures and PCA. These results are presented in Tables 5, 6, and 7, respectively, showing the classification results by families of discriminant analysis, decision tree, and SVM models. We also compared the results of SE and AE experiments. As explained in the previous section, in the AE experiment, a majority voting technique was used to label the CAS based on the labels of its episodes resulting from the SE experiment.

Regarding Tables 5–7, the AE experiment consistently outperformed the SE experiment in all evaluations, except in the case of using fine Gaussian SVM classifier using PCA reduction method. We highlighted this result using ⊕ sign in Table 7. Meanwhile, the statistical measure resulted in better recognition power in all cases except in cases of using fine Gaussian SVM for the SE experiment, cubic SVM in the SE experiment, and quadratic discriminant analysis for both AE and SE experiments. We highlighted these results using * sign in Tables 5 and 7. Notably, these mentioned exceptional cases are related to the inspiration dataset.

In discriminant analysis family classifiers, as it shows in Table 5, the best method for feature reduction for MFCC in the expiration dataset is the use of statistical measures, which resulted in 81% F-score using subspace discriminant analysis classifier. However, in inspiration datasets, the best result is 83% F-score which belongs to using PCA techniques using the quadratic discriminant analysis classifier. Tables 6 and 7 illustrate the results obtained from decision tree and SVM classifiers. In Table 6, for the expiration

TABLE 6.
The Classification Results of Decision Tree Family Using Statistical Measure and PCA For MFCC Features

	EXP Dataset		INSV Dataset	
	PCA	Statistical Measures	PCA	Statistical Measures
Simple Tree				
SE experiment	55.30%	65.60%	55.50%	55.60%
AE experiment	74.10%	85.00%	70.60%	74.60%
Medium Tree				
SE experiment	54.90%	62.80%	55.60%	60.60%
AE experiment	72.50%	82.30%	71.90%	**81.80%**
Complex Tree				
SE experiment	50.10%	50.10%	55.40%	60.40%
AE experiment	63.00%	63.00%	71.20%	**81.80%**
Bagged Tree				
SE experiment	56.60%	66.60%	62.60%	62.80%
AE experiment	80.10%	**85.50%**	77.30%	80.60%
Boosted Tree				
SE experiment	58.10%	67.10%	56.40%	59.30%
AE experiment	80.60%	81.30%	68.10%	77.20%
RUSBoost Tree				
SE experiment	55.20%	63.40%	56.90%	59.60%
AE experiment	74.60%	82.30%	76.20%	81.10%

Percentages refer to F-score. The results of the best frameworks are bolded.

TABLE 7.
The Classification Results of SVM Classifiers Using Statistical Measure and PCA Method for MFCC Features

	EXP Dataset		INSV Dataset	
	PCA	Statistical Measures	PCA	Statistical Measures
Linear SVM				
SE experiment	54.20%	68.70%	56.50%	62.00%
AE experiment	73.60%	85.30%	74.20%	79.80%
Cubic SVM				
SE experiment	56.80%	66.30%	59.50% *	56.90%
AE experiment	77.30%	**85.70%**	77.10%	78.20%
Quadratic SVM				
SE experiment	55.60%	67.40%	57.00%	59.30%
AE experiment	77.60%	82.60%	75.70%	78.00%
Fine Gaussian SVM				
SE experiment	56.60%	64.00%	61.60% ⊕ *	56.60%
AE experiment	80.10%	81.50%	57.10%	63.60%
Medium Gaussian SVM				
SE experiment	53.90%	68.90%	59.60%	63.20%
AE experiment	76.40%	84.20%	71.30%	**81.10%**

Percentages refer to F-score. The results of the best frameworks are bolded. The ⊕ sign indicates the classification framework in which the SE experiment outperformed the AE experiment. The * sign indicates the results of the classification frameworks in which the PCA method resulted in a better recognition power than the statistical measure reduction method.

dataset and inspiration dataset, the best F-score results are 85.50% for the bagged tree and 81.80% for both complex tree and medium tree classifiers. For SVM classifiers, as shown in Table 7, we see that the cubic SVM and medium Gaussian SVM outperformed others respectively in the expiration dataset with 85.70% F-score and the inspiration dataset with 81.10% F-score.

Evaluation of Prosodic Features

Regarding the results obtained using tilt and intensity feature sets shown in Tables 8 and 9. In every case, the method of the AE experiment resulted better than the SE experiment. Among the classifiers for the tilt feature set, the expiration dataset and inspiration dataset, boosted tree with 79% F-score and quadratic discriminant analysis with 83.9% F-score defeated other classifiers.

In intensity feature set investigation, as shown in Table 9 we observed that cubic SVM is the best classifier for both expiration dataset and inspiration dataset with the F-score of 70.9% and 74.60%.

Table 10 shows the efficacy of the rhythm feature using different classifiers. This durational feature only was

TABLE 8.
The Classification Results of Different Classifiers Using Tilt Features

	EXP Dataset		INSV Dataset	
	SE experiment	AE experiment	SE experiment	AE experiment
Linear Discriminant	54.60%	67.00%	65.00%	76.00%
Quadratic Discriminant	47.00%	49.40%	66.90%	**83.90%**
Subspace Discriminant	54.90%	70.00%	58.70%	73.30%
Simple Tree	38.80%	45.80%	60.20%	69.20%
Medium Tree	53.70%	68.60%	52.40%	70.30%
Complex Tree	54.90%	74.70%	54.00%	68.70%
Bagged Tree	59.30%	78.70%	60.80%	76.50%
Boosted Tree	57.70%	**79.00%**	58.90%	74.30%
RUSBoost Tree	56.30%	74.80%	56.20%	70.20%
Linear SVM	55.50%	69.00%	61.30%	72.40%
Cubic SVM	55.90%	78.50%	60.80%	75.90%
Quadratic SVM	54.50%	74.20%	61.20%	74.10%
Fine Gaussian SVM	56.00%	75.60%	61.20%	71.60%
Medium Gaussian SVM	55.10%	70.10%	63.70%	71.70%

Percentages refer to F-score. The results of the best frameworks are bolded.

TABLE 9.
The classification results of different classifiers using intensity features.

	EXP Dataset		INSV Dataset	
	SE Experiment	AE Experiment	SE Experiment	AE Experiment
Linear Discriminant	51.00%	62.90%	44.10%	59.00%
Quadratic Discriminant	45.80%	50.60%	49.30%	71.40%
Subspace Discriminant	52.40%	61.50%	48.90%	59.50%
Simple Tree	47.10%	57.10%	38.80%	60.90%
Medium Tree	50.60%	60.50%	47.60%	66.00%
Complex Tree	53.80%	68.00%	45.00%	58.30%
Bagged Tree	53.10%	65.70%	45.80%	62.20%
Boosted Tree	48.80%	60.10%	48.30%	65.20%
RUSBoost Tree	48.30%	58.20%	43.90%	61.70%
Linear SVM	50.30%	58.90%	52.30%	66.30%
Cubic SVM	58.70%	**70.90%**	53.20%	**74.60%**
Quadratic SVM	46.60%	56.70%	53.40%	69.50%
Fine Gaussian SVM	48.00%	58.70%	44.50%	58.00%
Medium Gaussian SVM	49.10%	60.20%	43.90%	52.60%

Percentages refer to F-score. The results of the best frameworks are bolded.

measured for the AE experiment as it requires a longer length of CAS. In the expiration dataset and the inspiration dataset, the cubic SVM with 75.60% F-score and quadratic SVM with 77.70% F-score were the best classifiers.

TABLE 10.
The Classification Results of Different Classifiers Using Rhythm Features

	EXP Dataset (AE experiment)	INSV Dataset (AE experiment)
Linear Discriminant	41.80%	62.80%
Quadratic Discriminant	20.40%	77.00%
Subspace Discriminant	41.80%	62.80%
Simple Tree	64.40%	58.30%
Medium Tree	62.40%	64.30%
Complex Tree	70.70%	65.50%
Bagged Tree	65.50%	62.70%
Boosted Tree	62.40%	61.30%
RUSBoost Tree	61.70%	63.10%
Linear SVM	55.30%	44.40%
Cubic SVM	**75.60%**	15.50%
Quadratic SVM	55.20%	**77.70%**
Fine Gaussian SVM	37.20%	48.40%
Medium Gaussian SVM	17.30%	55.40%

Percentages refer to F-score. The results of the best frameworks are bolded.

Evaluation of Feature Set Manipulation and Use of Multiple Classifiers

After acquiring the results of different classifiers using MFCC features and the prosodic features of intensity, tilt, and rhythm, we inspected the performance of two other frameworks. The first approach was to concatenate all features and see the best result was obtained by which classifier. The second approach was the use majority voting technique which inputs were the results of the most capable classifiers for each feature set that outperformed in the AE experiment. Table 11 and 12, respectively show the results of mentioned frameworks for the expiration dataset and the inspiration dataset. We only included the feature sets of MFCC, tilt, and rhythm in the majority voting framework as they consistently outperformed intensity features in both datasets.

In Table 11, it is shown that the feature set concatenation framework using quadratic SVM classifier outperformed the majority voting model, which respectively resulted in the F-scores of 86% and 83.30%. However, there was no significant difference between these two methods in the inspiration dataset as both methods resulted in about 82% F-score.

DISCUSSION

In this research, we examined machine learning techniques to develop an NCDS for investigating the potential of newborns' CASs for diagnosing septic infants from healthy ones. The sepsis pathology has not been studied while it is ranked as the 6th cause leading to the death among newborns in Canada on 2019.[35] Several evaluations were carried out to develop a comparison between the performance of each framework. In total, four feature sets of MFCC, tilt,

TABLE 11.
Best Classifiers for the Expiration Dataset

Feature Set	Classifier	Recall	Precision	F-score
MFCCs	Cubic SVM	85%	86.44%	85.70%
Tilt	Boosted Tree	78.30%	79.70%	79.00%
Intensity	Cubic SVM	71.50%	70.30%	70.90%
Rhythm	Cubic SVM	68.70%	83.90%	75.60%
All feature Concatenation	Quadratic SVM	83.90%	88.10%	**86.00%**
All feature Majority Voting	best classifiers in the AE experiment	71.80%	**99.10%**	83.30%

Percentages refer to F-score. The results of the best frameworks are bolded.

rhythm, and intensity were supplied to three families of classifiers, including SVM, discriminant analysis, and decision tree. We also assessed the performance of the concatenation method of all feature sets together and the method of collecting the votes of the most accurate classifiers for each feature set, and then labeled the test sample using the majority voting method. The input data of our proposed NCDS were two datasets of expiration and inspiration of infants' CASs.

As the results of the experiments show from Tables 5 to 10, the technique of majority voting in the AE experiment enhanced the performance of the model in all cases by far, except in the case of classification of inspiration episode dataset using the framework of MFCC and fine Gaussian SVM classifier with the PCA feature reduction technique as shown in Table 7 (highlighted by ⊕ sign). In Figure 5, we brought the minimum, maximum, and mean of the increase using the majority voting technique in the AE experiment in datasets of expiration and inspiration.

Consequently, the successive classification of episodes in the CAS and then use of majority voting to predict the CAS (AE experiment) resulted quite assuring in enhancing the performance of NCDS rather than classifying the single episode (SE experiment). This idea was inspired by[34] which was also successful in the domain of environmental sound classification.

Regarding the MFCC features, we analyzed the comparison of the use of two methods of feature reduction, including PCA and statistical measures. These results are presented in Tables 5 to 7. The results consistently show the superiority of using statistical measures over the PCA method in feature reduction in all classifiers in both datasets

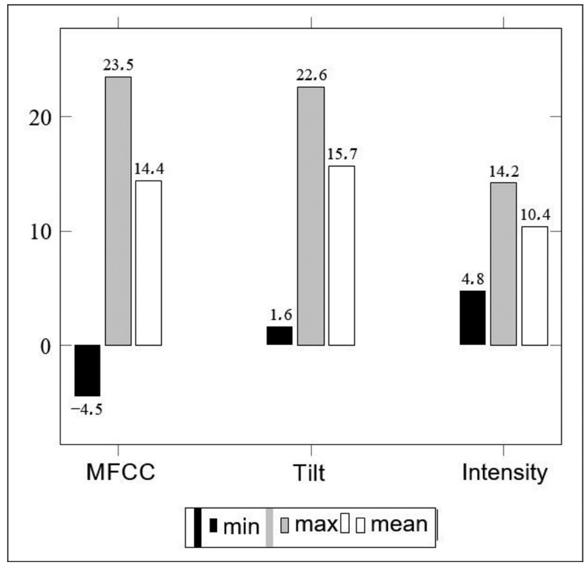

FIGURE 5. The minimum, maximum and mean of the improvement using the majority voting technique in the AE experiment in datasets of expiration and inspiration.

except in some cases for classification of the inspiration dataset. These cases include quadratic discriminant for both experiments of SE and AE (Table 5) and cubic SVM and fine Gaussian SVM in the SE experiment (Table 7). These cases are marked using * in mentioned tables.

The importance of feature selection is based on the problem, dataset properties and number, the interconnection condition among samples in the dataset, the desirable running time, and the considered classifier scheme. Through

TABLE 12.
Best Classifiers for Inspiration Dataset

Feature Set	Classifier	Recall	Precision	F-score
MFCCs	Quadratic Discriminant	**78.80%**	87.60%	83.00%
Tilt	Quadratic Discriminant	74.10%	**96.60%**	**83.90%**
Intensity	Cubic SVM	65.80%	82.00%	74.60%
Rhythm	Quadratic Discriminant	69.40%	86.50%	77.70%
All feature Concatenation	Quadratic Discriminant	76.20%	89.90%	82.80%
All feature Majority Voting	best classifiers in the AE experiment	71.70%	**96.60%**	82.30%

Percentages refer to F-score. The results of the best frameworks are bolded.

TABLE 13.
Elapsed Running Time for Extracting Each Feature Set

Feature Set	Elapsed Time (Minutes)
MFCC + PCA	23.20
MFCC + stats	15.80
Tilt	10.30
Intensity	10.60
Rhythm	0.08

TABLE 14.
List of Acronyms Used in the Manuscript

Abbreviation	Full Name
CAS	Cry Audio Signal
NCDS	Newborn Cry Diagnostic System
MFCC	Mel Frequency Cepstral Coefficient
LPC	Linear Predictive Coefficient
SVM	Support Vector Machine
SE	Single Episode
AE	All Episodes
PCA	Principal Component Analysis

these examinations, we found out that in all experiments for expiration dataset, and most cases for inspiration dataset, the statistical measures are more powerful in terms of their discriminatory properties to represent the features that are most relevant to the classifiers experimented within this work, including classifiers of discriminant, decision tree, and SVM, compared to the use PCA algorithm. Moreover, as a feature reduction method, we noticed that statistical measures are a more low-cost approach in terms of computational resources compared to PCA. Table 13 shows the running time of feature extraction for each feature set. Thus, the statistical measures feature reduction method saves the execution time and, in the majority of cases, it elevates the model's predictive power. The statistical method was applied successfully in[15, 45, 46] in the domain of automatic emotion recognition in speech, and developing NCDSs for infants with deafness, asphyxia, and respiratory distress.

Regarding the prosodic features, by the present study we learned that the inspiration dataset (labeled as INSV in Table 2) as a strong predictor for sepsis compared to healthy infants which is consistent with our previous study.[15] From computation point of view the assessment of tilt and intensity features took nearly the same amount of time. However, tilt features showed better distinctive properties. The rhythm feature had the lowest computational cost. Rhythm feature was effortless and fast to extract, while it had better F-score results than intensity features. According to[50] an authoritative classifier has an error rate lower than the random guessing on an untrained dataset, therefore the present study shows that septic infants of less than two months cry differently than healthy ones in terms of spectral features, duration feature, the pattern of changes of the F_0 and the energy of their CAS, which makes this method promising as a possible diagnostic tool. For further analysis, we concatenated all feature sets together and fed them to each classifier. Unlike the promising results in our previous study in which we concatenated tilt, rhythm, and MFCC,[15] the results of the concatenation of MFCC with tilt, rhythm, and intensity in both episodes were not improving in the present study. In a previous study, the control group was infants with respiratory distress. Thus, the idea of feature manipulation for diagnosing septic infants from healthy infants did not reproduce the good results of training based on the individual feature set. We also examined the idea of aggregating the results of the best classifiers for each feature set extracted from the same dataset and voting for the most recurred label. The intuition was to generate a framework in which the classifiers would complement their errors, thus would enhance the diagnostic power of the NCDS. Accordingly, the predicted labels achieved from the most competent classifiers for each feature set shown in Tables 11 and 12 were collected and aggregated to predict the final result. In practice, this framework could not enhance the performance of the NCDS and had a more computational cost; however, in the expiration dataset, it could improve the precision measure up to 99% (Table 11).

The unexpected performance of the multiple classifiers scheme might be explained by the fact that the integration of best classifiers was chosen globally. We generalized the model to predict for all test samples. However, in the case of noise existence around some test samples in the feature space, this scheme probably would not guarantee the best prediction for those test samples. Thus, we have to employ an approach that selects the outperforming classifiers locally. In every region of feature space, the competency of classifiers is estimated based on local information. This approach is called the dynamic selection of the classifier. We hope to address the shortcoming of our proposed multiple classifier scheme in future work by experimenting with the scheme of dynamic selection of classifiers, and the stacked classifier scheme. The method should handle the feature sets that do not degrade the feature space or the system performance. In the future extraction phase, we also expect to examine the performance of other feature sets, such as the auditory inspiration modulated feature set in the NCDS.

In our study, we generalized that the CASs are initiated by any reason, which in practice makes the task of diagnosing is difficult as newborns cry rhythmically different for their different needs.[51] Moreover, the CASs in our dataset belong to infants from different geographical regions. Infants in a linguistic group was proven to have a similar pattern of F_0 contour.[52] Thus, the state of a more uniform database in terms of rhythmicity and melody by experience would probably help the overall performance of the NCDS. However, the motivation was to develop an NCDS to make a precise decision under different situations, be unbiased by reason for crying, the surrounding noise, and be flexible with the length of the sample.

As a final point, it is worth explaining why we did not use pervasive deep learning techniques in our study. While the use of deep learning techniques is becoming rapidly prevalent, there are yet classification problems that have the limitation of dataset shortage which massively hinders the use of such techniques.[49] Notably, there are fewer applications of deep learning in the infant diagnostic task based on CASs due to the absence of enough CASs dataset. The number of infants and their CASs for each disease is often inadequate. Thus, in the case of enough dataset, it is worth attempting deep learning techniques; however, there is no certainty that they work better than other classifiers,[53] as the choice of a classifier is dataset-based.

CONCLUSION

The experiments presented here evaluate the functionality of our proposed NCDS for the unstudied disease of sepsis which is one of the most common leading to death factors in infant mortality. In our suggested NCDS, we used the well-known MFCC features and the prosodic features of tilt, rhythm, and intensity in a configuration with different families of classifiers, including SVM, decision tree, and discriminant analysis. These parameters were applied on CASs of groups of healthy and septic newborns. The obtained results show the vital contributions of the proposed features and classifiers to distinguish the septic infants from healthy ones, only based on their CASs. The best accomplished F-score results are for the framework of the concatenation of all feature sets using quadratic SVM with 86%, and the framework of tilt feature set with quadratic discriminant analysis with 83.90% respectively for the two datasets of expiration and inspiration episodes of newborns' CAS. Hence, we conclude that septic infants cry differently than healthy infants from the spectral and temporal views. The scheme proposed in this study is promising to be used as a tool to assist pediatricians and address the lack of pediatricians in deprived areas.

Acknowledgement

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2016-05067] and has been made possible through the funding provided by the Bill and Melinda Gates Foundation [OPP1025091]. We would also like to thank Prasun Lala from the ETS library for his feedback on the manuscript.

REFERENCES

1. Soltis J. The signal functions of early infant crying. *Behav Brain Sci*. 2004;27:443–458.
2. Rutledge J. Fundamentals of speech recognition, by lawrence rabiner and bing-hwang juang. *Ann Biomed Eng*. 1995;23:526.. -526.
3. Boukydis CZ, Lester BM. *Infant crying: Theoretical and research perspectives*. Springer Science & Business Media; 2012.
4. Mukhopadhyay J, Saha B, Majumdar B, Majumdar AK, Gorain S, Arya B, et al. 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)*. IEEE; 2013:64–67.
5. Michelsson K, Michelsson O. Phonation in the newborn, infant cry. *Int J Pediatr Otorhinolaryngol*. 1999;49:S297–S301.
6. Abdulaziz Y, Ahmad SMS. Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients. *2010 International Conference on Information Retrieval & Knowledge Management (CAMP)*. IEEE; 2010:260–263.
7. Manfredi C, Bandini A, Melino DViellevoye R, Kalenga M, and Orlandi S, et al. Automated detection and classification of basic shapes of newborn cry melody. *Biomed Signal Process Control*. 2018;45:174–181.
8. Abou-Abbas L, Tadj C, Gargour C, Montazeri L, et al. Expiratory and inspiratory cries detection using different signals' decomposition techniques. *J Voice*. 2017;31. pp. 259. e13-259. e28.
9. Kim MJ, Kim Y, Hong S, Kim H, et al. 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)*. IEEE; 2013:1–4.
10. Aucouturier J-J, Nonaka Y, Katahira K, Okanoya K, et al. Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models. *J Acoust Soc Am*. 2011;130:2969–2977.
11. Wahid N, Saad P, Hariharan M. Automatic infant cry classification using radial basis function network. *J Adv Res Appl Sci Eng Technol*. 2016;4:12–28.
12. Saha B, Purkait PK, Mukherjee J, Majumdar AK, Majumdar B, and Singh AK, et al. An embedded system for automatic classification of neonatal cry. *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*. IEEE; 2013:248–251.
13. Orozco-García J, Reyes-García CA. A study on the recognition of patterns of infant cry for the identification of deafness in just born babies with neural networks. *Iberoamerican Congress on Pattern Recognition*. Springer; 2003:342–349.
14. Alaie HF, Abou-Abbas L, Tadj C. Cry-based infant pathology classification using GMMs. *Speech Commun*. 2016;77:28–52.
15. Matikolaie FS, Tadj C. On the use of long-term features in a newborn cry diagnostic system. *Biomed Signal Process Control*. 2020;59: 101889.
16. Rosales-Pérez A, Reyes-García CA, Gonzalez JA, Reyes-alaviz OF, Escalante HJ, Orlandi S, et al. Classifying infant cry patterns by the Genetic Selection of a Fuzzy Model. *Biomed Signal Process Control*. 2015;17:38–46.
17. Kheddache Y, Tadj C. Identification of diseases in newborns using advanced acoustic features of cry signals. *Biomed Signal Process Control*. 2019;50:35–44.
18. Martinez-Cañete Y, Cano-Ortiz SD, Lombardía-Legrá L, Rodríguez-ernández E, Veranes-icet L, et al. Data Mining Techniques in Normal or Pathological Infant Cry. *International Workshop on Artificial Intelligence and Pattern Recognition*. Springer; 2018:141–148.
19. Hariharan M, et al. Improved binary dragonfly optimization algorithm and wavelet packet based non-linear features for infant cry classification. *Comput Methods Programs Biomed*. 2018;155:39–51.
20. Matikolaie FS, Kheddache Y, Tadj C. Automated newborn cry diagnostic system using machine learning approach. *Biomed Signal Process Control*. 2022;73: 103434.
21. Sahak R, Mansor W, Khuan LY, Zabidi A, Yassin A, et al. Detection of asphyxia from infant cry using support vector machine and multilayer perceptron integrated with Orthogonal Least Square. In: *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics*. IEEE; 2012:906–909.
22. Hariharan M, Chee LS, Yaacob S. Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network. *J Med Syst*. 2012;36:1309–1315.
23. Lederman D, Cohen A, Zmora E, Wermke K, Hauschildt S, Stellzig-isenhauer A, et al. On the use of hidden Markov models in infants' cry

classification. *The 22nd Convention on Electrical and Electronics Engineers in Israel, 2002*. IEEE; 2002:350–352.
24. Lederman D, Zmora E, Hauschildt S, Stellzig-isenhauer A, Wermke K, et al. Classification of cries of infants with cleft-palate using parallel hidden Markov models. *Med Biol Eng Comput*. 2008;46:965–975.
25. Hariharan M, Sindhu R, Yaacob S. Normal and hypoacoustic infant cry signal classification using time−frequency analysis and general regression neural network. *Comput Methods Programs Biomed*. 2012;108:559–569.
26. Santiago-Sánchez K, Reyes-García CA, Gómez-Gil P. Type-2 fuzzy sets applied to pattern matching for the classification of cries of infants under neurological risk. *International Conference on Intelligent Computing*. Springer; 2009:201–210.
27. Orlandi S, Manfredi C, Bocchi L, Scattoni ML, et al. 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*. IEEE; 2012:2953–2956.
28. Reyes-Galaviz OF, Verduzco A, Arch-Tirado E, Reyes-arcía CA, et al. Analysis of an infant cry recognizer for the early identification of pathologies. *International School on Neural Networks, Initiated by IIASS and EMFCSC*. Springer; 2004:404–409.
29. Sahak R, Mansor W, Lee Y, Yassin AM, Zabidi A, et al. Orthogonal least square based support vector machine for the classification of infant cry with asphyxia. *2010 3rd International Conference on Biomedical Engineering and Informatics*. 3, IEEE; 2010:986–990.
30. Zabidi A, Mansor W, Lee KY. Optimal feature selection technique for mel frequency cepstral coefficient feature extraction in classifying infant cry with asphyxia. *Indonesian J Electr Eng Comput Sci*. 2017;6 (3):646–655.
31. Zabidi A, Mansor W, Khuan LY, Yassin IM, Sahak R, et al. Classification of infant cries with hypothyroidism using multilayer perceptron neural network. *2009 IEEE International Conference on Signal and Image Processing Applications*. IEEE; 2009:246–251.
32. Grau SM, Robb MP, Cacace AT. Acoustic correlates of inspiratory phonation during infant cry. *J Speech Lang Hear Res*. 1995;38(2):373–381.
33. Robb MP, Goberman AM. Application of an acoustic cry template to evaluate at-risk newborns: preliminary findings. *Neonatology*. 1997;71 (2):131–136.
34. Abdoli S, Cardinal P, Koerich AL. End-to-end environmental sound classification using a 1D convolutional neural network. *Expert Syst Appl*. 2019;136:252–263.
35. S. C. Government of Canada, "Leading causes of death, infants," (in eng), 2020/11/26/2020.
36. Prathibha A, Putta R, Srinivas H, Satish S. An eclectic approach for detection of infant cry and wireless monitoring of swinging device as an alternative warning system for physically impaired and better neonatal growth. *World J Sci Technol*. 2012;2(5):62–65.
37. Ruvolo P, Movellan J. Automatic cry detection in early childhood education settings. *2008 7th IEEE International Conference on Development and Learning*. IEEE; 2008:204–208.
38. Yamamoto S, Yoshitomi Y, Tabuse M, Kushida K, Asada T, et al. Recognition of a baby's emotional cry towards robotics baby caregiver. *Int J Adv Rob Syst*. 2013;10(2):86.
39. Abou-Abbas L, Tadj C, Fersaie HA. A fully automated approach for baby cry signal segmentation and boundary detection of expiratory and inspiratory episodes. *J Acoust Soc Am*. 2017;142(3):1318–1331.
40. Manfredi C, Viellevoye R, Orlandi S, Torres-arcía A, Pieraccini G, Reyes-arcía C, et al. Automated analysis of newborn cry: relationships between melodic shapes and native language. *Biomed Signal Process Control*. 2019;53: 101561.
41. Jurafsky D, Martin JH. *Speech and language processing*. Pearson London; 2014.
42. Mary L. Automatic extraction of prosody for speaker, language and speech recognition. *Extraction and Representation of Prosody for Speaker, Speech and Language Recognition*. Springer; 2012:19–33.
43. Orlandi S, Bandini A, Fiaschi F, Manfredi C, et al. Testing software tools for newborn cry analysis using synthetic signals. *Biomed Signal Process Control*. 2017;37:16–22.
44. Ayesha S, Hanif MK, Talib R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Inf Fusion*. 2020;59:44–58.
45. M. Bhargava and T. Polzehl, "Improving automatic emotion recognition from speech using rhythm and temporal feature,"*arXiv preprint arXiv:1303.1761,*2013.
46. Amaro-Camargo E, Reyes-García CA. Applying statistical vectors of acoustic characteristics for the automatic classification of infant cry. *International Conference on Intelligent Computing*. Springer; 2007:1078–1085.
47. Wang L. *Support vector machines: theory and applications*. Springer Science & Business Media; 2005.
48. Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern*. 1991;21:660–674.
49. Wimalarathna H, Ankmnal-eeranna S, Allan C, Agrawal SK, Allen P, Samarabandu J, et al. Comparison of machine learning models to classify Auditory Brainstem Responses recorded from children with Auditory Processing Disorder. *Comput Methods Programs Biomed*. 2021;200: 105942.
50. Dietterich TG. Ensemble methods in machine learning. *International workshop on multiple classifier systems*. Springer; 2000:1–15.
51. Michelsson K, Christensson K, Rothgänger H, Winberg J, et al. Crying in separated and non-separated newborns: sound spectrographic analysis. *Acta Paediatr*. 1996;85:471–475.
52. Manfredi C, Pieraccini G, Viellevoye R, Reyes-arcía A. T.-. C., et al. "Relationships between newborns'cry melody shapes and native languagE,". *Models and Analysis of Vocal Emissions for Biomedical Applications: 10th International Workshop, December 13-15, 2017 2017 Nov 30*. 117, 201747.
53. Fernández-Delgado M, Cernadas E, Barro S, Amorim D, et al. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res*. 2014;15:3133–3181.