Identification of Reasons Behind Infant Crying Using Acoustic Signal Processing and Deep Neural Network for Neonatal Intensive Care Unit

Nagaraj V. Dharwadkar, Computer Science and Engineering, Rajarambapu Institute of Technology, Sakhrale, India
https://orcid.org/0000-0003-3017-0011

Amulya A. Dixit, Rajarambapu Institute of Technology, Sakhrale, India

Anil K. Kannur, Nagarjuna College of Engineering and Technology, India

Mohammad Ali Bandusab Kadampur, College of Engineering, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia

Santosh Joshi, Florida International University, USA

ABSTRACT

The infants admitted in the neonatal intensive care unit (NICU) always need a hygienic environment and round the clock observations. Infants or the just born babies always express their physical and emotional needs through cry. Thus, the detection of the reasons behind the infant cry plays a vital role in monitoring the health of the babies in the NICU. In this paper, the authors have proposed a novel approach for detecting the reasons for Infant's cry. In the proposed approach the cry signal of the infant is captured, and from this signal, the unique set of features are extracted using MFCCs, LPCCs, and pitch. This set of features is used to differentiate the patters signals to recognize the reasons for the cry. The reasons for cry such as hunger, pain, sleep, and discomfort are used to represent different classes. The neural network multilayer classifier is designed to recognize the reasons for the cry using the standard dataset of infant cry. The proposed classifier can achieve accuracy of 93.24% from the combined features of MFCCs, LPCCs, and pitch.

KEYWORDS

Acoustic Characteristics, Feature Extraction, Infant Cry Signal, Neural Networks

1. INTRODUCTION

The just-born babies or infants are not able to orally communicate with parents since they are not able to speak. They use to cry as their communication medium to express their physical and emotional needs. When they required more attention or care they cry. Thus, the parents are not able to understand the crying of the baby every time completely. The only solution to this problem is to study the acoustic

DOI: 10.4018/IJIRR.289576

speech pattern of the infant cry and determine the reason behind the cry. The acoustic speech pattern produced by infant cry is always different for different reasons. Thus, the different situations in the baby of hungry or sleepy or in pain generate different acoustic speech patterns. The infant crying pattern can be used as a biological alarm system to alert the parents. To differentiate the acoustic speech patterns using signal processing algorithms. Various types of features can be extracted to identify and analyze the cry signal patterns (R.P. Balandong, 2013; S. Bhattacharya, 2016). At the earliest stage the reason for infant cry helps the parents to take appropriate steps and for pediatricians guides them properly in treatment. The acoustic characteristics of the cry patterns are directly influenced by the infant's physical and psychological state (D. Lederman, 2002). The acoustic signal of infant cry contains valuable information such as gender, health, identity, and emotions (R. Cohen, 2012). By using these properties (features) and analyzing them, we can detect the infant's reason behind crying. The basic objective is to find unique features. As the infants are not able to speak their cry defines it all. We have to analyze their cry signal patterns to find similar and discriminating features. It will be a great help for pediatricians and parents if they know the reason behind cry. Proper treatment can be given to that baby.

There are many attempts made to detect the reason behind infant cry. Orozco in (J. Orozco, 2003) has used Neural networks primarily for this purpose. As a large amount of database is used, neural networks work best. They achieved good accuracy results (J. Orozco, 2003). Another research work related to Baby cry detection did the comparison between classical and new methods of acoustic analysis of Infant cry in (G.J. Varallyay, 2004). They used fundamental frequency detection and dominant frequency detection. In the cry of infants with normal hearing and hard of hearing, the ratio seems to be different between fundamental frequency and dominant frequency (G.J. Varallyay, 2004). When it comes to the processing of cry signal the main focus is in the extraction and analysis of the fundamental frequency (F0) and the first three formats F1, F2, and F3 of infant cry signal as implemented in [P. Pal, 2006]. These parameters contain important information regarding the emotional state of the infant. The Harmonic Product Spectrum (HPS) method has been used to obtain the values of the fundamental frequency (F0) as for infant cries, the fundamental frequency varies widely and rapidly (P. Pal, 2006). The detection is done in five categories such as pain, hunger, fear, sadness, anger.

The features such as pitch frequency, short-time energy, MFCCs, Harmonicity factor are implemented in (R. Cohen, 2012). They have developed a cry detection algorithm. To achieve a low false-negative error rate, the algorithm analyses the signal at various time-scales (segments of several seconds, sections of about 1 second, and frames of several tens of msec). The proposed algorithm is composed of three main stages which are the Voice Activity Detector (VAD), Classification Using k-nearest neighbors (k-NN) algorithm, Post-processing for validating the classification stage to reduce false-negative errors. The results of this research work were based purely on the algorithm mentioned earlier (R. Cohen, 2012). In a paper published in the year 2013 (J. Saraswathy, 2013), Baby Chillanto Database was used for their research. The database contained 340 of normal signals, 340 of deaf cry signals, and 340 of asphyxia cry signals. In their results, authors can classify the baby cry signals into normal, deaf, and asphyxia. The Short-Time Fourier Transform (STFT) features were extracted from the baby cry signals. The performance of the proposed using PNN and GRNN Radial Basis Neural Network classifiers scheme was analyzed. Their results show that they have achieved an accuracy of 99.22% (J. Saraswathy, 2013). The reasons behind infant cry what we are intending to classify such as hunger, pain, sleep, and discomfort differ from the classification categories implemented in (J. Saraswathy, 2013).

Furthermore, to address the issues related to the reason for infant cry considering factors such as hunger, pain, discomfort, and illness (R. P. Balandong, 2013). This author used features such as MFCCs, Pitch, Linear Prediction Coefficients (LPCCs) and Multilayer Perceptron (MLP), and Support Vector Machine(SVM) classifiers for binary classification of normal cry and different disease cry [Amulya A. Dixit, 2018]. In this paper, the classification of reasons such as hunger, pain, discomfort, and illness, they have used different Neural networks are applied as classifiers (R. P. Balandong, 2013) and accuracy achieved was 71.4%. An accuracy of 90.4% was achieved for the classification of anger, fear, and pain through the infant's cry. The limitation of this paper was the classification accuracy. The combinations of different types of features and classifier to increase the accuracy of the classifier using deep neural networks. To increase the accuracy using the multiclass classification of reasons, efforts have been made in (N. Wahid, 2016). They have used Baby Chillanto Dataset. The MFCCs and LPCCs features were extracted from cry signal patterns. They have shown that the Radial Basis Function Network (RBFN) classifier obtained an accuracy of 93.43% (N. Wahid, 2016).

We have summarized the overall performance of all schemes in Table 1. This Table describes the overall comparison of the existing authors who have worked to address the issues related to address the identification of the reason for infant cry. This study is considered the approaches used by the researchers and the comparative study is done considering the similarities and differences in classification accuracy. Since in NICU, it requires detection of the accurate reason for infant cry for immediate medical follows up.

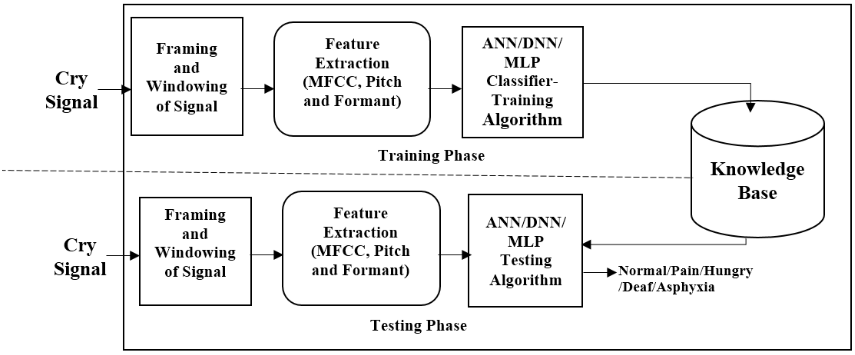
Furthermore, the study of different approaches shows that the most commonly used features by the researchers were MFCCs, pitch, LPCCs, formants (R. P. Balandong, 2013; R. Cohen, 2012), and few researchers used zero-crossing rate and short term energy as their features. The Deep neural network models were used by the researchers for many applications for classification tasks (R. P. Balandong, 2013; J. Orozco, 2003; N. Wahid, 2016; Y. Lavner, 2016). Generally, the multilayer perceptrons are popular when deep neural networks are considered for the classification and recognition of multiple tasks. In literature, the three-layered feed-forward neural network known as Radial Basis Function Network (RBFN) was used for increasing accuracy of classifier (N. Wahid, 2016). As shown in Table 1, the higher accuracy in classification was achieved by many researchers. But, most of the researcher has used binary classification of infant cries (R. P. Balandong, 2013; R. Cohen & Y. Lavner, 2012; S. E. Barajas-Montiel 2005), very few researchers have used an approach of multiclass classification (J. Saraswathy, 2013; N. Wahid, 2016). Hence, there is a lot of scope in the multiclass classification of reasons behind infant cry. Accuracy results can also be improved as compared with previous accuracy results achieved. The Baby Chillanto database is used by most researchers. However, this database cannot be available easily as it is a property of the Instituto Nacional de Astrofisica Optica y Electronica (INAOE) CONACYT, Mexico (N. Wahid, 2016).

The multiclass classification has a scope in increasing its accuracy as compared to existing schemes shown in Table 1. By proposing this approach, different reasons can be detected and analyzed which will be useful for the treatment of the infants. Accuracy results will matter most in this approach. Efforts will be made to achieve good accuracy through a multiclass classification approach. The more

Table 1. Literature study of the previous research papers

Database used	Features Used	Classifiers	Classification	Accuracy Results
Collected By them (S.E. Barajas-Montiel 2005)	MFCCs	SVM, Feed Forward Neural Networks	Pain & no Pain cry Hunger & no Hunger Cry	96.41% 87.61
Baby Chillato Database (J. Saraswathy, 2013)	Short Time Fourier Transform (STFT)	PNN GRNN	Asphyxia, Deaf and Normal Cry	98.22% 98.02%
Baby Chillato Database (N. Wahid, 2016)	MFCC LPCC	MLP, Radial Basis Function Network	Asphyxia, pain, hunger and deaf	93.43%

Figure 1. Block diagram of the proposed approach


Cry Signal
Framing and Windowing of Signal
Feature Extraction (MFCC, Pitch and Formant)
ANN/DNN/ MLP Classifier-Training Algorithm
Training Phase
Knowledge Base
Cry Signal
Framing and Windowing of Signal
Feature Extraction (MFCC, Pitch and Formant)
ANN/DNN/ MLP Testing Algorithm
Normal/Pain/Hungry /Deaf/Asphyxia
Testing Phase

reasons we can detect behind infant cry, it will benefit in all aspects. The crying of an infant is not every time completely understood. Parents cannot figure out why their kid is crying. There can be several reasons. Maybe the baby is hungry or feeling sleepy or baby is in pain. But, we cannot detect the exact reason for crying. If this detection is done, then it will be very useful for parents and also for pediatricians in NICU. Because if parents know why their baby is crying and if the baby is crying because he /she is hungry or feeling sleepy then they can feed him/her. But, in another case, if the baby is crying because of pain then they can go to a pediatrician for treatment. Detection of reasons behind Infant's cry is an important aspect related to the health and treatment of infants. A lot of work has been done in this area and implemented various new approaches to achieve good performance. Still, there are many challenges to overcome.

The remaining part of the paper is organized as follows. In Section 2, we have described the proposed method for the identification of Reasons behind Infant Cry. In Section 3, we have explained the experimental setup used in the proposed method. Section 4 describes the results and discussions of the experiments conducted. The paper is concluded and the future scope of the research is discussed in section 5.

2. RESEARCH METHODOLOGY

The research methodology plays an important part in any research work. Based on the literature, a method is proposed to implement the task of detection of reasons behind infant cry. The proposed framework of our research work is shown in Figure 1.

Figure 1 shows the proposed approach. Cry signal is provided as the input from database files. This input cry signal is then divided into short frames. These short frames are further processed. A windowing function is applied to these frames. Feature extraction is performed. The features to be used are MFCC, LPCC, pitch, formants. After the extraction of features, a feature vector file is prepared. This file is then given to classifiers in the classification step. Neural networks are selected to carry out the classification task. Different types of neural networks such as Artificial Neural Networks (ANN), Deep Neural Networks (DNN), Multilayer Perceptrons (MLP) will be implemented as classifiers. The classification process will give the output which will be the reasons behind infant cry. These reasons are represented at the end of Figure 1. The output classes which we are expected after classification are Pain cry, Hunger cry, Asphyxia cry, Normal Cry, Deaf cry (Amulya A. Dixit, 2018).

Our proposed approach distinguishes earlier researches concerning using multi-layered combinations of features and classifiers. It is expected that this approach will result in higher accuracy. A detailed explanation of the steps involved in this proposed approach is given below.

2.1 Pre-Processing

This step is necessary if the database files containing cry sounds of infants have many unwanted regions. Unwanted regions refer to some unwanted silences and noises in the audio file. These regions can make the cry signal processing task difficult. Therefore, it is better to remove these unwanted regions from the cry signals if present. If the audio file does not have any noise then there is no need to implement this pre-processing step.

2.2 Features Extraction

The feature extraction process extracts some important characteristics from the cry signal pattern. Features will be selected on the basis that will discriminate between different reasons such as hunger, pain, discomfort, sleep behind infant cry effectively. After completion of the features extraction step, a feature vector file will be generated, which will be given to classifiers. The features that are planned to use in this detection of reasons behind Infant cry are shown below.

2.2.1 Mel Frequency Cepstral Coefficients (MFCC)

MFCCs are extracted to represent the acoustic characteristics of the signal. Short term power spectrum of a signal is represented by MFCCs (R. Cohen, 2012). MFCCs are mostly used for the speaker recognition task. Most of the previous researches has suggested that MFCCs are efficient features to be used in this task (R. P. Balandong, 2013; S. Bhattacharya, 2016; S. Yamamoto, 2013). MFCCs when combined with suitable classifier, gives better accuracy results.

MFCC features are obtained by processing several steps. The cry signal is non-stationary as it is changing constantly. Therefore, the signal is divided into short frames. Each frame is then windowed by the Hamming window to minimize the discontinuities present in the signal. The windowed signal is then applied with Fast Fourier Transform (FFT) to convert the time domain into the frequency domain. The values obtained from the FFT step will be then given to a set of triangular filters called melspaced filter banks. The basic formula for calculating mels for frequency(F) in Hz is shown below in Equation (1) (N. Wahid, 2016; Amulya A. Dixit, 2018):

$$mel\left(F\right) = 2952\log_{10}\left(1 + \frac{F}{700}\right) \tag{1}$$

Finally, the log mel spectrum is converted into the time domain using a Discrete Cosine Transform (DCT). After this step, the output obtained is MFCCs (N. Wahid, 2016). These extracted MFCC coefficients are used further for the classification task. A total of 13 MFCC coefficients are used for our research work and a feature vector file is generated in MATLAB software after extraction of the MFCC feature.

2.2.2 Delta MFCC and Delta Delta MFCC

They are also known as differential and acceleration coefficients. The addition of delta MFCC features to the static 13-dimensional MFCC features strongly improves the accuracy, and a further (smaller) improvement is provided by the addition of double-delta coefficients.

2.2.3 Pitch

The pitch feature which is often known as the fundamental frequency is by far the most important feature reported in cry research publications as a differentiating feature among cries of infants who suffer from different health problems. When a baby is crying, the tone, volume, and pitch change (R. P. Balandong, 2013; S. Bhattacharya, 2016; D. Lederman, 2002). It is observed that, when a baby is crying because of pain, the pitch at that time is higher. As compared to adults, newborn babies have a vocal tract with higher fundamental frequency and resonances (D. Lederman, 2002). Therefore, the detection of the pitch in the cry signal will help to discriminate against the cry signals to give different reasons for the cry.

The most commonly used method for pitch estimation is based on the Autocorrelation function. The highest value of this function is detected in the region of interest (D. Lederman, 2002; K. Rakesh, 2011). The signal is divided into short segments/frames, consisting of N samples. The short-time autocorrelation function is given by Equation (2):

$$R_x\left(\tau\right)=\frac{1}{L}\sum_{n=0}^{L-1-m}x\left(n\right)\cdot x\left(n+\tau\right),\quad 0\leq\tau\leq T \tag{2}$$

where L is the length of the frame, T is the number of autocorrelation points to be computed. T is called the delay or lag. The value of lag τ equals the value of pitch. In this, 4 features include minimum, maximum, mean, and standard deviation of the fundamental frequency (Amulya A. Dixit, 2018).

2.2.4 Delta Pitch

Pitch derivatives are also calculated too. These are nothing but the difference in the values of pitch per frame. These features can also contribute to improving accuracy when combined with the basic pitch feature (Amulya A. Dixit, 2018).

2.3 Formants

The relation between a glottal airflow velocity input and vocal track airflow velocity output can be approximated by a linear filter with resonances called formants. Formants change with different vocal tract configurations corresponding to different resonant cavities (Holmes, 1997). Estimation of location and bandwidths of formants is done in formant feature extraction (Snell, 1993). The total 4 number of formants were extracted for our classification purpose.

2.4 Classification

The classification step involves various classifiers. Neural networks have been giving impressive accuracy results as the previous research papers suggest (R. P. Balandong, 2013; N. Wahid, 2016; Y. Lavner, 2016). The classifiers considered for implementing the classification task are shown below.

2.4.1 Deep Neural Networks

DNN is a classifier based on feedforward artificial neural networks (R. Cohen, 2012). DNN consists of multiple layers of nodes. At least 3 hidden layers of nodes are present in a DNN with nonlinearly activating functions making it a deep neural network (G.J. Varallyay, 2004; S. Yamamoto, 2013). Each node uses a nonlinear activation function. Activation functions that are used for classification tasks are relu, tanh, sigmoid, and softmax. Combinations of these activation functions are implemented using the Deep Neural Network Algorithm.

2.4.2 Multilayer Perceptrons (MLPs)

The field of artificial neural networks is often just called neural networks or multi-layer perceptrons after perhaps the most useful type of neural network. Multi-Layer perceptron is a feedforward neural network with one or more layers between the input and output layers. Feedforward means that data flows in one direction from the input to the output layer (forward).

2.4.3 Recurrent Neural Networks

A recurrent neural network (RNN) is a class of artificial neural networks where connections between units form a directed graph along a sequence. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs.

2.4.4 Support Vector Machines

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. SVM is a supervised learning technique. When we have a dataset with features & class labels both then we can use Support Vector Machine. There are two types of SVM classifiers such as Linear SVM classifier and Non-linear SVM classifier. When there is more number of classes and the dataset is dispersed up to some extent, then a non-linear SVM classifier is used (Online, https://dataaspirant.com/2017/01/13/support-vectormachine-algorithm).

3. EXPERIMENTAL SETUP

The databases for Infant cry detection are rarely available commercially. As this work is concerned about the baby's health and treatment, it is a quite sensitive domain for research. The primary basis of this research work depends on the database. Therefore, the database files should be authenticated to work on. Many researchers have used the databases which are created by themselves. Therefore, these databases are not accessible to others. The collection of a database is the basic requirement for our research. As mentioned in the literature survey, some researchers have used Baby Chillanto Database. However, the database restricted to users. The Baby Chillanto database is the most suitable database for our research work. For getting this database, we have to contact the authorities who own this database. By contacting them, some formalities have to be done. The database is then made accessible to us for use.

3.1 Database Description

The Baby Chillanto Database is the property of the Instituto Nacional de Astrofisica Optica y Electronica (INAOE) – CONACYT, Mexico. The database contains 5 classes. The classes are Pain, Asphyxia, Hunger, Normal, Deaf. The cry signals of infants are present in these 5 categories. The description of database files is given in Table 2

The duration of these database files is 1 second. They have also provided the full length of these cry audio files. However, for training and testing purpose 1-second files are most suitable.

3.2 Analysis of Data

The analysis of data is a very important and basic step. As the data is in the form of speech i.e. cry signals, the analysis of data is performed by extracting some speech signal properties present in that cry signal files. As we are intending to do the categorical classification of these 5 classes which are the reasons behind the infant cry, the properties we extract the information should represent their difference between these classes. The system represents speech content by attributed relational features of objects and relationships between them. This representation relies on the assumption that a fixed

Table 2. Database description

Infant Cry Category	Pain	Hunger	Asphyxia	Normal	Deaf
No. of Samples	192	350	340	507	885
Total			2274		

number of objects are common in many sound contents. All these common objects are "labeled." For this application, the feature extraction function considers for each labeled object in the database. These features are sufficient for medical purposes. Currently, the classification system is the main method of organizing database collections. The method typically employs the classification system, developed by the authors. Information retrieval has several unique characteristics. First, examiners search for sound by primitive features. Second, registries hold large collections of sound contents in electronic format. And finally, in the field, successful retrieval criteria are well-defined. The feature extraction from the database of speech contents makes an ideal pattern of information retrieval.

3.2.1 Spectrograms

A spectrogram is a visual representation of the spectrum of frequencies of sound or another signal as they vary with time. Spectrograms are used extensively in the fields of music, sonar, and speech processing. A common format is a graph with two geometric dimensions. One axis represents time and the other axis is frequency; a third dimension indicating the amplitude of a particular frequency at a particular time is represented by the intensity or color of each point in the image. The spectrogram is computed as a sequence of FFTs of windowed data segments.

As seen from the Figure 2, Figure 3, Figure 4, Figure 5 and, Figure 6, the pain cry signal spectrogram has the highest frequency (energy) and asphyxia cry signal has the lowest frequency in its signal. These spectrograms indicate the discrimination of these classes from each other.

4. RESULTS AND DISCUSSIONS

4.1 Results From WEKA

These results are achieved by using the WEKA software. In this, the feature vector file containing all the data extracted from features and is given input to the classifier. The classifiers used are Multilayer Perceptron and Support Vector Machine. The highest results are achieved from the MLP classifier and the accuracy is 85.817%. SVM classifier has given accuracy up to 78.0655%. SVM classifier

Figure 2. Spectrogram of Pain cry signal

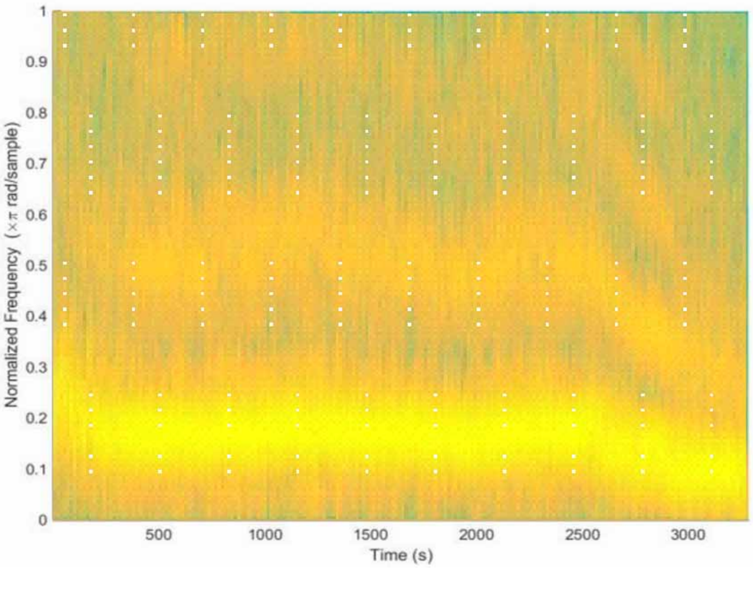


Figure 3. Spectrogram of Asphyxia cry signal

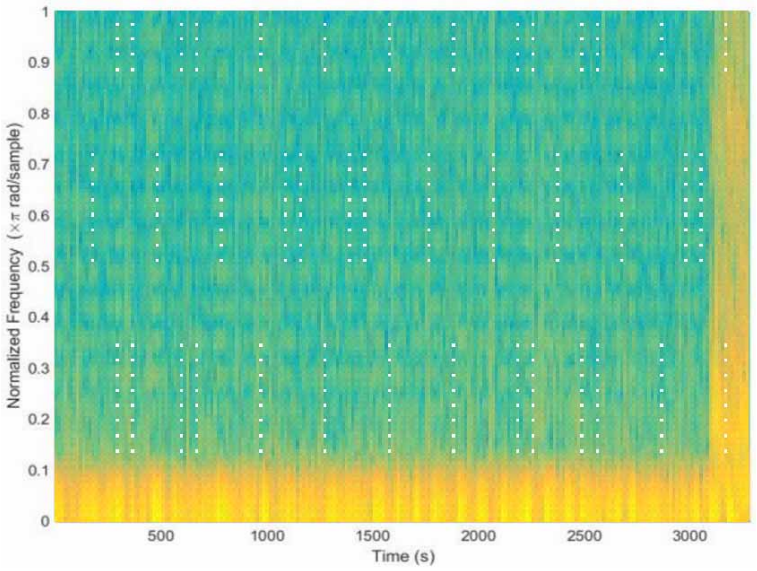
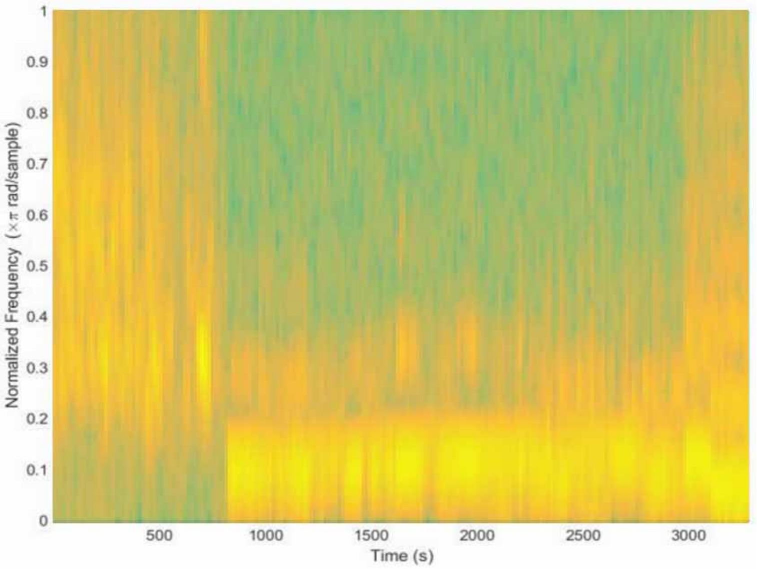


Figure 4. Spectrogram of Hunger cry signal



works best for the linear data. Therefore, as our data is a non-linear MLP classifier has given good accuracy results.

4.2 Results From Python

These results are achieved by using Anaconda software. The language used is Python. Python codes for classification purposes are used. The TensorFlow libraries are used. The python codes represent the Deep neural classifier codes. As we are using DNN, there are variations in the hidden layers,

Figure 5. Spectrogram of Deaf cry signal

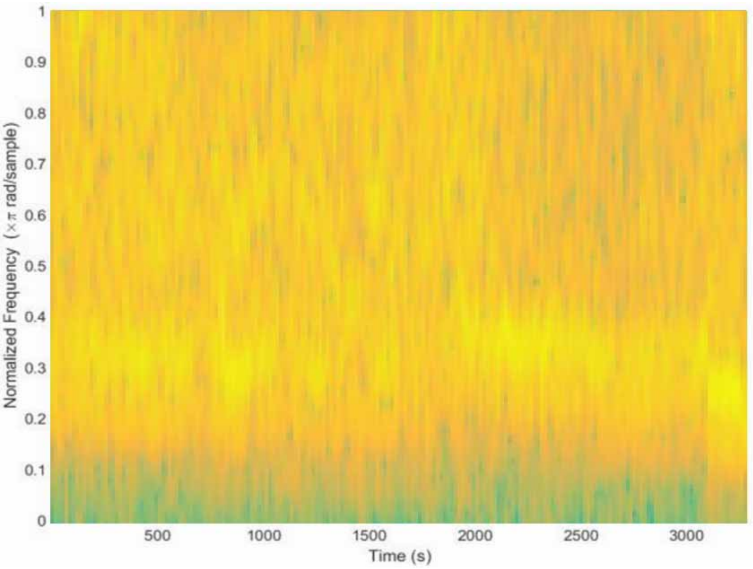
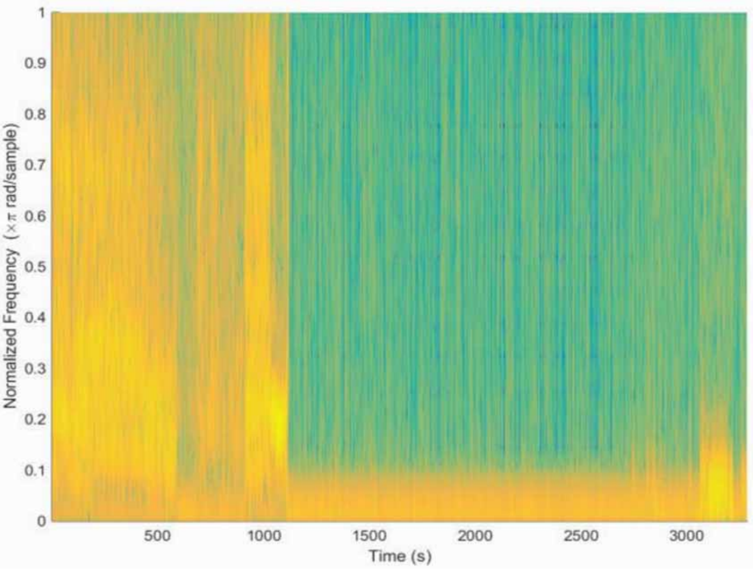


Figure 6. Spectrogram of Normal cry signal



epochs, and batch size. Accuracies are achieved by trial and error basis by changing these parameters for obtaining higher accuracy results. The highest accuracy achieved is 93.24%.

Results from Python Code without cross-validation: In this python code, training and testing data are separated by 70% and 30% respectively instead of K-fold cross-validation. In Machine Learning, the general rule of thumb is to partition the data set into the ratio of 3:1:1 (60:20:20) for training, validation and testing respectively. When a learning system is trained with some data samples, you might not know to which extent it can predict unseen samples correctly. The concept of cross

Table 3. Accuracy Results from WEKA software

Classifiers	Hidden Layers	Feature Combinations		
		MFCC+Pitch	MFCC+Pitch+Delta Pitch	MFCC + Pitch + Delta Pitch + Formant
MLP	13	83.16%	84.02%	85.44%
MLP	17,14,20	85.82%	84.25%	86.66%
SVM	-	78.55%	75.05%	76.23%

Table 4. Accuracy Results from Python Code with stratified K fold Cross-Validation

Classifiers	Hidden Layers				Epoch	Batch Size	Feature Combinations	Accuracy
MLP	17	20	25	5	50	10	MFCC+ Pitch	88.25%
MLP	47	40	45	5	50	10	MFCC+ pitch+ delta MFCC+ Delta Delta MFCC+ Delta Pitch	85.00%
RNN	47	40	43	5	60	10	MFCC+ pitch+ delta MFCC+ Delta Delta MFCC+ Delta Pitch	93.19%
RNN	47	36	40	5	50	10	MFCC+ pitch+ delta MFCC+ Delta Delta MFCC+ Delta Pitch	93.24%
DNN							MFCC+ pitch+ delta MFCC+ Delta Delta MFCC+ Delta Pitch	97.00%

validation is done to tweak the parameters used for training in order to optimize its accuracy and to nullify the effect of over-fitting on the training data. This shouldn't be done on the test set itself and hence the separation between testing set and cross validation set. In cases where cross validation is not applicable, it is common to separate the data in the ratio of 7:3 (70:30) for training and testing respectively.

The highest accuracy is achieved when we used the MFCC+pitch+delta MFCC+ Delta Delta MFCC+ Delta Pitch+Formant feature combination. The highest accuracy is 97%. The results achieved are as shown in Table 5.

4.3 Analysis

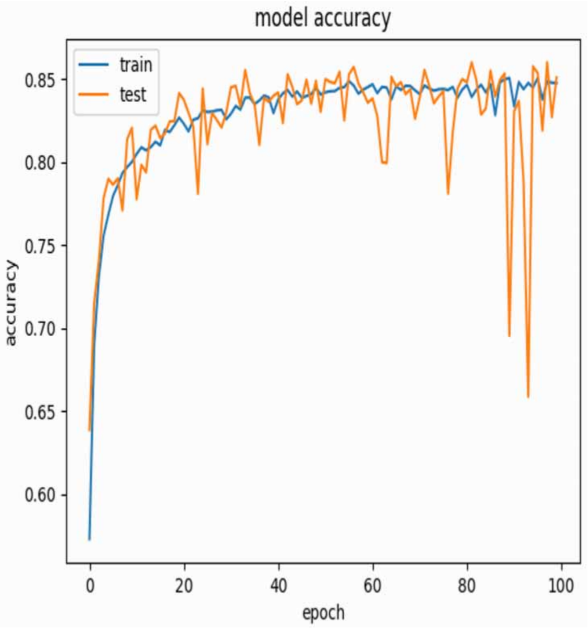
As we detailed the results achieved by using different tools and codes, the accuracy graphs obtained are also given below. The graphs are of 2 types. One is the model accuracy graph and the other is the model loss graph. The model accuracy graph represents the accuracy achieved versus epochs. The accuracy of both training and testing data is shown. The model loss graph represents the loss of training data and testing data concerning epochs set.

As from the above graphs, it is clear that the model is a good fit model. A good fit is a case where the performance of the model is good on both the train and validation sets. This can be diagnosed from a plot where the train and validation loss decrease and stabilize around the same point.

Table 5. Accuracy Results from Python Code without K-fold Cross-Validation

Classifiers	Hidden Layers				Epoch	Feature Combinations	Accuracy
MLP	85	80	90	5	80	MFCC+pitch+delta MFCC+ Delta Delta MFCC+ Delta Pitch	86.31%
MLP	90	85	92	5	100	MFCC+pitch+delta MFCC+ Delta Delta MFCC+ Delta Pitch	86.01%
MLP	47	36	60	5	100	MFCC+pitch+delta MFCC+ Delta Delta MFCC+ Delta Pitch	85.92%
MLP	47	36	60	5	100	MFCC+pitch+delta MFCC+ Delta Delta MFCC+ Delta Pitch	86.58%
MLP	85	72	92	5	150	MFCC+pitch+delta MFCC+ Delta Delta MFCC+ Delta Pitch	86.50%
MLP	90	95	100	5	80	MFCC+pitch+delta MFCC+ Delta Delta MFCC+ Delta Pitch+Formant	87.00%
DNN						MFCC+pitch+delta MFCC+ Delta Delta MFCC+ Delta Pitch+Formant	97%

Figure 7. Training and testing accuracy model with 100 epochs



4.4 Discussion

The accuracy results achieved from the python code are quite satisfactory. The features used in this process are MFCC, Delta MFCC, Delta Delta MFCC, Pitch, Delta Pitch, and formants. The results obtained by using only the pitch feature were not good. The accuracy achieved by using only the pitch feature was only 46.95%. Therefore, for improving the accuracy there was a need to increase the feature and combine them. The MFCC and Pitch features were then combined to improve accuracy.

Figure 8. Training and testing loss model with 100 epochs

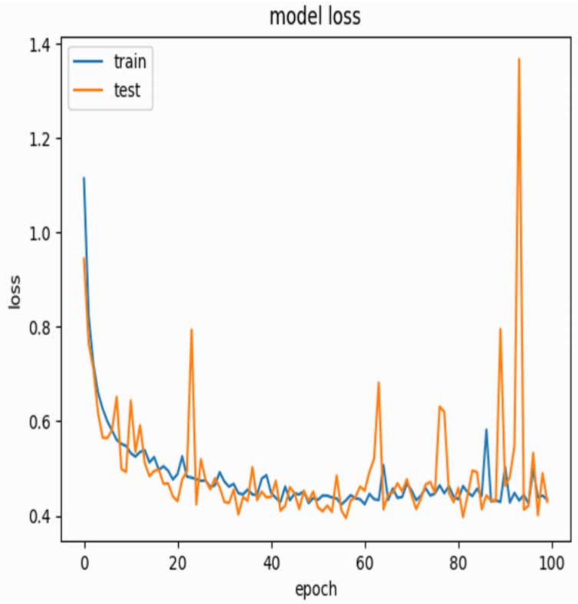
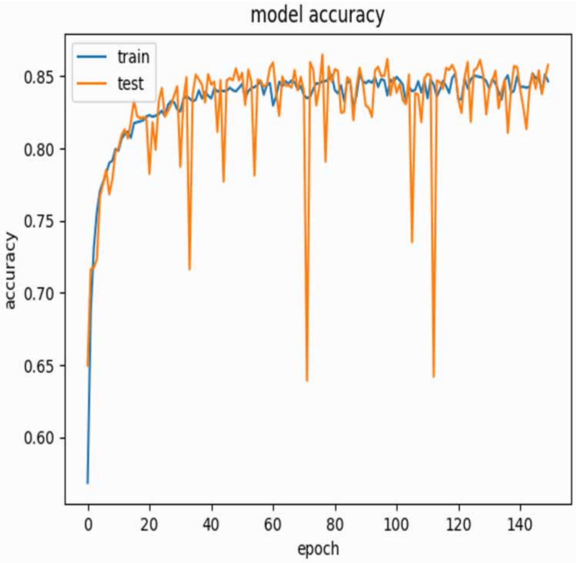
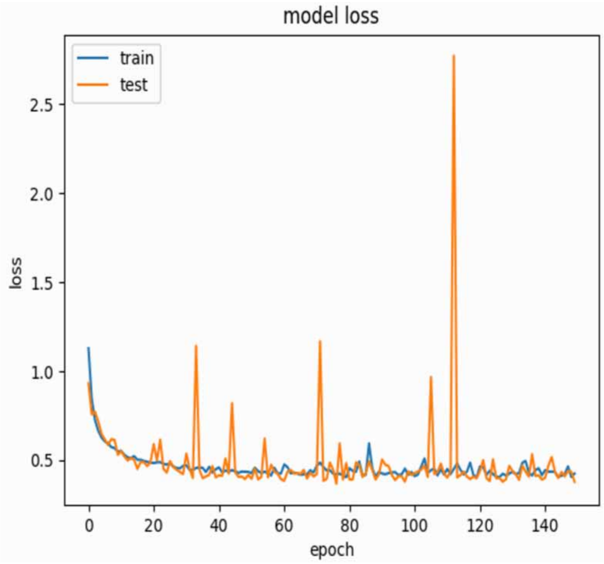


Figure 9. Training and testing accuracy model with 150 epochs



The accuracy was improved up to 88.25%. From this, it was clear that the MFCC feature performs well on this data and when combined with pitch it achieved good accuracy results. However, the accuracy improvement is the main objective of this research, the combination of MFCC, Delta MFCC, Delta Delta MFCC, Pitch, and Delta Pitch has been implemented and it achieved significant improvement in the accuracy up to 93.24% and DNN with an accuracy of 97%. By combining more features to this feature vector set, the accuracy can be improved more.

Figure 10. Training and testing loss model with 150 epochs



4.5 Difference Between WEKA Software and Python

On a very high level, the biggest difference that between Weka and the others is flexibility. Weka is very much a plug and play Machine Learning solution – it's packaged nicely into a .jar file, and it comes with a GUI that you can run most simple analysis and model building through. Compared to the others, which are languages that can be run though an interactive shell, there's more guidance in Weka, and running ML via Weka seems quite magical. The downside to this ease of use is that Weka is far less flexible than the others for statistical analysis and data exploration. This really boils down to the fact that the others are programming langauges with ML packages and libraries that you can import, whereas Weka itself is a ML package. As it probably goes without saying, this means that the others provide a much greater degree of freedom to clean, explore and transform your data sets, as well as a much greater freedom to tune and tweak the underlying algorithms. In my limited experience, I've found Weka to be an easy introduction to machine learning on toy data sets, as things just work out of the box (although that's not to say that tools that you can use from R or python don't work out of the box). However, in practice, Matlab, python and R and their respective packages and libraries are much more flexible and practical for data science.

5. CONCLUSION

The proposed novel methodology comprises of utilizing multi-layered combinations of classifiers and features. Such combination will be chosen considering different parameters of newborns cry. This methodology had three stages incorporated with it. The initial step comprises of the pre-processing of the cry signal. The unwanted noise and silences were removed if present in the cry signal. From that point onward, the cry signal was partitioned into edges of short length and windowing capacity is applied. The windowed signal was utilized for additional processing. The subsequent step was the features extraction. Extracted features given to the classifier for recognition. In the last advance, the characterization of purposes for newborn child cry was performed. Deep neural networks, artificial neural networks, and multilayer perceptron will be the primarily selected classifiers to carry out this task. Baby cry recognition is a very difficult and challenging task as it is identified with the wellbeing and treatment of children. Subsequently, legitimate consideration must be taken while executing this exploration work. In light of the past research works, another methodology is proposed to identify the purposes for newborn child cry. Features which are mostly used by other researchers will be selected for the feature extraction. The features planned for the extraction purpose are MFCCs, Pitch, Formants, etc. As long as classifiers are concerned, Neural networks are the first choice to classify the reasons of the infant cry.

The highest accuracy results achieved were 93.24% from the combined features MFCC, Delta MFCC, Delta Delta MFCC, Pitch, and Delta Pitch. These combined features generated a feature vector file containing 47 features as columns. The highest accuracy is achieved using DNN Classifier is 97%. DNN classifier worked best for this data because of its great ability to perform on non-linear data. Furthermore, as the data was large in amount, DNN was the most suitable classifier for implementing this multiclass classification task. In this DNN classifier implementation, we implemented some of its types such as MLP and RNN. From these types, RNN achieved good accuracy as compared to the MLP classifier. The efforts can be made to improve the accuracy results achieved until now. For this purpose, different methods and approaches can be used. There is a scope to implement a CNN classifier. The spectrograms obtained from the audio cry signal files can be given as an input to the CNN classifier, it will then extract the features from these spectrogram images and will put the labels of classes also. This can result in a great accuracy of classification. Another approach is using time-domain features such as Shimmer and Jitter. The addition of these features to the present feature combination can help in improving accuracy.

REFERENCES

Balandong. (2013). *Acoustic analysis of baby cry.* Department of Bıomedıcal Engıneerıng, Faculty of Engıneerıng, Unıversıty of Malaya.

Barajas-Montiel, S. E., & Reyes-Garcia, C. A. (2005). Identifying pain and hunger in infant cry with classifiers ensembles. In *Computational Intelligence for Modelling, Control, and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference*. IEEE. doi:10.1109/CIMCA.2005.1631561

Bhattacharya, S. (2016). *Infant cry detection* (thesis). University of Miami.

Cohen, R., & Lavner, Y. (2012). Infant cry analysis and detection. *Electrical & Electronics Engineers in Israel (IEEE), 2012 IEEE 27th Convention of. IEEE*, 1–5. doi:10.1109/EEEI.2012.6376996

Dixit & Dharwadkar. (2018). A Survey on Detection of Reasons Behind Infant Cry Using Speech Processing. Proceedings of International Conference on Communication and Signal Processing (ICCSP).

Holmes, Holmes, & Garner. (1997). Using formant frequencies in speech recognition. *Eurospeech Book*, *97*, 2083-2087.

Lavner, Y., Cohen, R., Ruinskiy, D., & Ijzerman, H. (2016). Baby cry detection in the domestic environment using deep learning. *Science of Electrical Engineering (ICSEE), IEEE International Conference on the*, 1–5. doi:10.1109/ICSEE.2016.7806117

Lederman. (2002). *Automatic classification of infants cry* (Thesis). Ben-Gurion University of the Negev.

Orozco, J., & Garc'ıa, C. A. R. (2003). Detecting pathologies from infant cry applying scaled conjugate gradient neural networks. *European Symposium on Artificial Neural Networks*, 349–354.

Pal, P., Iyer, A. N., & Yantorno, R. E. (2006). Emotion detection from infant facial expressions and cries. *Acoustics, Speech and Signal Processing, ICASSP Proceedings 2006 IEEE International Conference*, *2*. doi:10.1109/ICASSP.2006.1660444

Rakesh, K., Dutta, S., & Shama, K. (2011). Gender recognition using speech processing techniques in LabVIEW. *International Journal of Advances in Engineering and Technology*, *1*(2), 51–63.

Reyes-Galaviz, O. F., Cano-Ortiz, S. D., & Reyes-García, C. A. (2008). Evolutionary-Neural System to Classify Infant Cry Units for Pathologies Identification in Recently Born Babies. *Proceedings of the Special Session MICAI*, 330-335. doi:10.1109/MICAI.2008.73

Saraswathy, Hariharan, Khairunizam, Yaacob, & Thiyagar. (2013). Infant cry classification: time-frequency analysis. *Control Systems, Computing, and Engineering (ICCSCE), 2013 IEEE International Conference on*, 499–504.

Snell & Milinazzo. (1993). Formant location from LPC analysis data. *IEEE Transactions on Speech and Audio Processing Journal*, *1*, 129-134.

Support Vector Machine. (n.d.). https://dataaspirant.com/2017/01/13/support-vectormachine-algorithm

Varallyay, G. J., Benyo´, Z., Ille'nyi, A., Farkas, Z., & Kovac's, L. (2004). Acoustic analysis of the infant cry: classical and new methods. *Engineering in Medicine and Biology Society, IEMBS'04. 26th Annual International Conference of the IEEE*, *1*, 313–316. doi:10.1109/IEMBS.2004.1403155

Wahid, N., Saad, P., & Hariharan, M. (2016). Automatic infant cry pattern classification for a multiclass problem. *Journal of Telecommunication, Electronic and Computer Engineering*, *8*(9), 45–52.

Yamamoto, S., Yoshitomi, Y., Tabuse, M., Kushida, K., & Asada, T. (2013). Recognition of a baby's emotional cry towards robotics baby caregiver. *International Journal of Advanced Robotic Systems*, *10*(2), 86. doi:10.5772/55406

Anil Kannur, Assistant Professor, Department of Computer Science & Engineering, Rajarambapu Institute of Technology, Islampur, India. He has obtained his B E in Computer Science & Engineering in 2001, M.Tech. (Computer Science & Engineering) in 2006 and presently pursuing Ph.D. in Computer Science and Engineering (Registered under Visvesvaraya Technological University, Belagavi, India). He has published 14 research papers in peer reviewed International Journals and conferences. His research area of interest is Image Processing, Pattern Recognition and Computational Forensics.

Mohammad Ali Kadampur is an Assistant Professor in the College of Engineering, Al-Imam Muhammad Bin Saud Islamic University Riyadh, Saudi Arabia. His research areas are data mining and knowledge discovery, deep learning applications, Community detection in social networks, Artificial Intelligence in web & education. He has over 30 publications in national and international journals of repute. He has obtained his Master of Technology from the National Institute of Technology, Surathkal, India, and Ph.D. from the National Institute of Technology, Warangal India, respectively both in the fields of Computer Science and Engineering. Earlier, he has a Bachelor of Engineering degree from Basaveshwar Engineering College Bagalkot, India, in Electronics & Communication Engineering. He received national merit scholarship and Institute fellowships in recognition of his merit and supported his education. He is an active member of ACM and reviewer on many leading journals including Elsevier journals. He has successfully executed several sponsored projects and currently active in guiding research and teaching Computer Science and allied courses.

Santosh Joshi is a Database Architect at Applied Research Center (ARC) at Florida International University (FIU). He has a B.S. in Computer Science Engineering from Karnatak University India and Master's in Engineering Management from Florida International University. He has 14 years' experience in IT industry in implementation of technology solutions designed to address complex problems. Expertise in Application development, Maintenance and Technical Project Management. He worked on multiple IT and Research projects mainly into the Database implementations, Data warehousing, Data Analysis, Machine Learning and Big Data Analytics in Cyber Security domain.