journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Normal and hypoacoustic infant cry signal classification using time–frequency analysis and general regression neural network

M. Hariharan[a,*], R. Sindhu[b], Sazali Yaacob[a]

[a] School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), 02600, Perlis, Malaysia
[b] School of Microelectronic Engineering, Universiti Malaysia Perlis (UniMAP), 02600, Perlis, Malaysia

ARTICLE INFO

Article history:
Received 12 October 2010
Received in revised form 2 June 2011
Accepted 19 July 2011

Keywords:
Acoustic analysis
Infant cry
Feature extraction
Short-time Fourier transform
Pattern classification
General Regression Neural Network

ABSTRACT

Crying is the most noticeable behavior of infancy. Infant cry signals can be used to identify physical or psychological status of an infant. Recently, acoustic analysis of infant cry signal has shown promising results and it has been proven to be an excellent tool to investigate the pathological status of an infant. This paper proposes short-time Fourier transform (STFT) based time–frequency analysis of infant cry signals. Few statistical features are derived from the time–frequency plot of infant cry signals and used as features to quantify infant cry signals. General Regression Neural Network (GRNN) is employed as a classifier for discriminating infant cry signals. Two classes of infant cry signals are considered such as normal cry signals and pathological cry signals from deaf infants. To prove the reliability of the proposed features, two neural network models such as Multilayer Perceptron (MLP) and Time-Delay Neural Network (TDNN) trained by scaled conjugate gradient algorithm are also used as classifiers. The experimental results show that the GRNN classifier gives very promising classification accuracy compared to MLP and TDNN and the proposed method can effectively classify normal and pathological infant cries.

1. Introduction

Cry is multimodal and dynamic in nature. Detection of pathological status of babies using the conventional methods takes several months or even years after the infant is born. It is necessary to detect the pathological status earlier to avoid unnecessary treatments and therapies. Infants cry is due to some possible reasons such as, hunger, pain, sleepiness, discomfort, feeling too hot or too cold, and too much noise or light. From the cry, a trained professional can understand the physical or psychological status of the baby. Acoustic analysis of infant cry signal is a non-invasive tool for the detection of certain pathological conditions [1–14]. In recent years, simple techniques have been proposed for analyzing the infant cry through linear prediction coding, Mel frequency cepstral coefficients, pitch information, harmonic analysis and noise analysis [1–14]. Different classification algorithms and hybrid systems were used for infant cry classification [1–14]. Infant cry is a highly non-stationary signal; Fourier transform is not a very useful tool for analyzing non-stationary signals as the time domain information is lost while performing the frequency transformation. When looking at a Fourier transform of a signal, it is impossible to tell when a particular event took place. In order to overcome the drawbacks of Fourier transform technique, time–frequency analysis has been proposed by researchers as it is a good tool for analyzing the infant cry signals both timeand frequency scale

[*] *Corresponding author*. Tel.: +6049885167.
E-mail address: hari@unimap.edu.my (M. Hariharan).

0169-2607/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2011.07.010

Table 1 – Some of the significant works on the classification of normal and deaf cry signals.

Author name	Feature extraction method	Classifier	Best accuracy
O.F. Reyes-Galaviz et al. [1]	Mel-frequency cepstral coefficients	Feed forward input delay neural network (normal, deaf cry, asphyxia cry, 3 class problem)	96.08–97.39%
J.O. Garcia [2]	Linear prediction technique	Scaled conjugate gradient neural networks (normal and deaf cry)	91.08% (314 samples) 86.20% (1036 samples)
J.O. Garcia [3]	Mel-frequency cepstral coefficients and linear prediction technique	Scaled conjugate gradient neural networks (normal and deaf cry)	97.43%
G. Várallyay [4]	Fundamental frequency detection using smoothed spectrum method	–	–
O.F. Reyes-Galaviz et al. [5]	Mel-frequency cepstral coefficients	Evolutionary–neural system (normal and pathological cry-deaf + asphyxia)	100% from some experiments
O.F. Reyes-Galaviz et al. [6]	Linear predictive coefficients, Mel-frequency cepstral coefficients	Evolutionary neural system and a neural network system (normal, deaf cry, asphyxia cry, 3 class problem)	96.49%

simultaneously. There are many works on infant cry signal recognition using the time–frequency analysis. But, the interpretation from time–frequency analysis is different. Many of them have used pitch, harmonic analysis, and noise analysis [7–12]. This paper presents the development of an intelligent learning system to classify normal and pathological cries using short-time Fourier transform and general regression neural network. Researchers have proposed approaches for problems of two class domain (normal or pathological) or more than two classes of infant cries (normal or 2 pathological cry signals). Table 1 presents some of the significant works on the classification of normal and deaf cry signals.

From the literature, it has been observed that the feature extraction plays an important role in the area of automatic detection of pathological cries. In this paper, a feature extraction method using STFT based time–frequency analysis for deriving features from infant cry signals and a GRNN are proposed for discriminating normal and pathological cries. Two schemes of data validation methods are used (10-fold cross validation and data independent validation scheme where the classifiers are trained with a selected set of samples and tested with samples which are not taken in count during training), in order to test the effectiveness of the proposed features and the reliability of the classification results. The experimental investigations elucidate that the STFT combined with statistical features and GRNN classifier can be used to detect certain pathological status of an infant from cry signals.

2. Database

The database of infant cry is downloaded from the website http://ingenieria.uatx.mx/orionfrg/cry/ called Baby Chillanto database and is a property of the Instituto Nacional de Astrofisica Optica y Electronica (INAOE) – CONACYT, Mexico. The database is described in Ref. [5]. All the samples of this database have the length of 1 second and we have taken the same samples for our analysis. It consists of 507 of normal cry signals and 879 of deaf cry signals. In this experiment, we took the same number of samples for each class 507. The deaf cry signals are recorded from 6 babies and normal cry signals are recorded from 5 babies. The sampling frequency of infant cry signals is set to 8000 Hz for our analysis. All the infant cry signals are subjected to feature extraction through STFT. The infant cry signal recorded from normal baby and deaf baby are plotted in Fig. 1.

3. Method

Classification of infant cries is a typical pattern recognition system and it consists of two blocks: short-time Fourier transform based signal processing and classification using general regression neural network, MLP and TDNN. This section briefly describes the feature extraction and classification methods.

3.1. *Short-time Fourier transform (STFT) based signal processing*

Infant cry is a dynamic or non-stationary signal. Fourier transform is not a very useful tool for analyzing non-stationary signals as the time domain information is lost while performing the frequency transformation. When looking at a Fourier transform of a signal, it is impossible to tell when a particular event took place. In order to overcome the drawbacks of Fourier transform approach, time–frequency analysis has been proposed by researchers as it is a good tool for analyzing the infant cry signals both time and frequency scale simultaneously. In order to produce good time–frequency spectrogram of infant cry signals, STFT is selected as feature extraction. The STFT based spectrogram is simple and fast technique compared to other time–frequency analysis. Short-time is a straightforward approach of slicing the waveform of interest into a number of short-segments and performing the analysis on each of the segments using standard Fourier transform [21,22]. A window function is applied to a segment of data, effectively isolating that segment from the overall waveform, and Fourier transform is applied to that segment. This is termed the spectrogram or “short-term Fourier transform”.

STFT is represented in the discrete domain given by Eq. (1):

$$X(m, k) = \sum_{n=1}^{N} x(n)[W(n-k)e^{-jnm/N}] \quad (1)$$

where $W[n]$ is a short-time windowing function of size L, centered at time location m, and N is the number of discrete frequencies ($N \geq L$). Usually, N is chosen to be a power of -2 for using an efficient fast Fourier transform (FFT). Since the

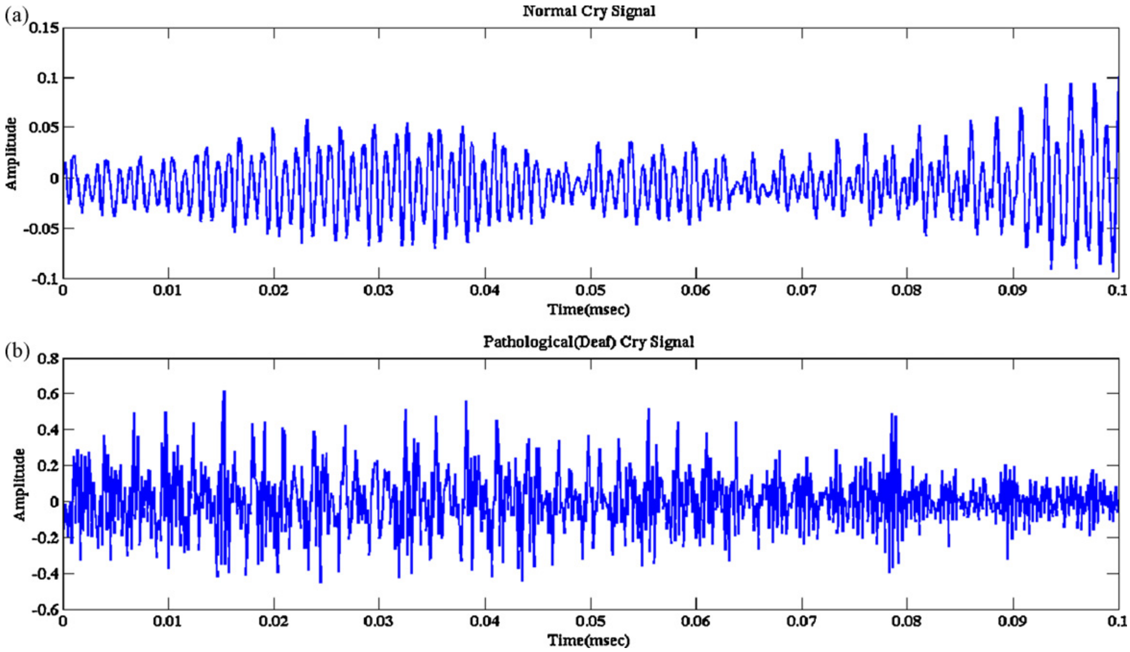

Fig. 1 – Infant cry signals (normal and deaf baby).

Fourier transform is a complex function, the power spectrum density (PSD) is used and is given by Eq. (2):

$$P_s[m,k] = \frac{1}{N}\left|X[m,k]\right|^2 \tag{2}$$

The spectrogram can be used for observing the temporal and spectral characteristics at any point in the infant cry signals. Generally the frame length is chosen between 10 ms and 50 ms in the area of speech signal analysis [21] and hence in this work, the infant cry signals are segmented into different frame length of 20 ms, 30 ms, 40 ms, and 50 ms with 50% overlap between the frames. The effect of different frame length has been studied and its results are presented in this work. The output of the STFT is a matrix whose rows pertain to frequency and columns to time. From the STFT–PSD of the cry signals (Figs. 3(a), 3(b), 6(a), and 6(b)), time–frequency, time–amplitude, and frequency–amplitude plots can be generated and which can clearly display the discrimination among the different types of cry signals. The block diagram of the feature extraction and classification is shown in Fig. 2.

Fig. 3(a) and (b) illustrates the time–frequency plot of pathological cry signal (deaf, segment 6) and normal cry signal (segment 4). Fig. 6(a) and (b) illustrates the time–frequency plot of pathological cry signal (deaf, segment 300) and normal cry signal (segment 200). Figs. 4(a), 5(a), 7(a), and 8(a) depict the time–maximum amplitude plot, which is maximum amplitude versus time by finding columns of time frequency plot. Figs. 4(b), 5(b), 7(b), and 8(b) illustrate the frequency–maximum amplitude plot, which is maximum amplitude versus frequency by finding rows of time–frequency plot at every frequency. Figs. 4(c), 5(c), 7(c), and 8(c) depict the frequency–standard deviation plot, which shows the standard deviation versus normalized frequency by finding rows of time–frequency plot at every frequency. Feature extraction plays a vital role in the area of classification of infant cry signals. Using Figs. 3(a), 3(b), 6(a) and 6(b), one can differentiate the normal and pathological cry through visual inspection. However, there is a possibility of wrong interpretation from the time–frequency plots and also the results depend on the expertise of the medical professionals. Hence in this paper, a simple feature extraction method is proposed by applying standard statistical techniques to the time–frequency plots of infant cry signals, time–maximum amplitude plots of infant cry signals, frequency maximum amplitude plots of infant cry signals, and frequency–standard deviation amplitude plots of infant cry signals. The standard statistical features are found to be useful for quantification and classification of infant cry signals.

Set 1. Feature extraction from time–frequency plots

Mean and standard deviation of amplitude of time–frequency plots (2 features, Feature 1 and Feature 2).

Set 2. Feature extraction from time–maximum amplitude plots, frequency–maximum of amplitude plots and frequency–standard deviation plots.

Maximum, minimum, mean, standard deviation, skewness, and kurtosis of time–maximum amplitude plots, frequency–maximum of amplitude plots and frequency–standard deviation plots (Features 3–8, Features 9–14, and Features 15–20, totaling 18 features). Twenty features are extracted from each frame of an infant cry signal and finally the average of the features is used as input for the classifiers to distinguish the cry signals between normal and deaf cries.

Fig. 9(a)–(d) shows the scatter plots between features. From the scatter plots, it is observed that the features extracted from normal and pathological cry signals are almost distinguishable.

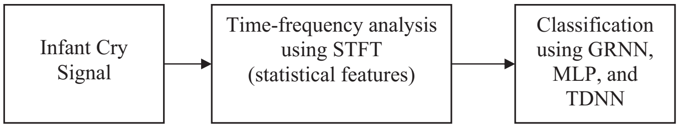
Infant Cry Signal
Time-frequency analysis using STFT (statistical features)
Classification using GRNN, MLP, and TDNN

Fig. 2 – Block diagram of the feature extraction and classification phase.

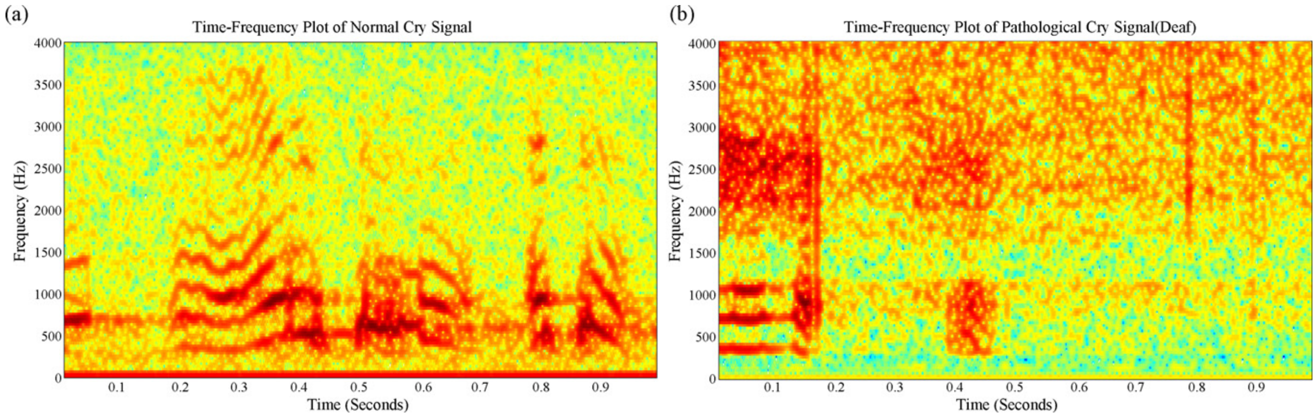

Fig. 3 – (a) Time frequency plot of normal cry signal (segment 4) and (b) time frequency plot of pathological cry signal (deaf, segment 6).

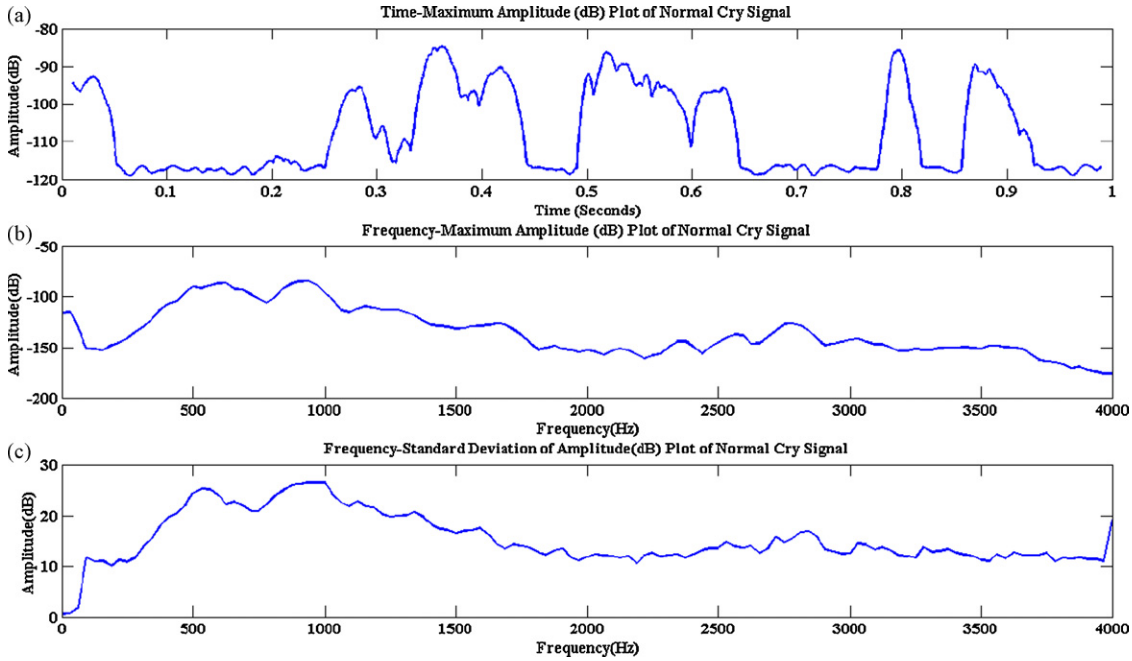

Fig. 4 – (a) Time–maximum amplitude (dB) plot of normal cry signal (segment 4), (b) frequency–maximum amplitude (dB) plot, and (c) frequency–standard deviation of amplitude (dB) plot.

4. Classifiers

Artificial neural networks are widely used in pattern recognition and classification problems by learning from examples. Different neural network models are available for classifying the patterns. In this work, a general regression neural network is used for the classification of normal and pathological cries since it was successfully applied in different pattern recognition applications [15–20]. To prove the reliability of the proposed features, two neural network models such as Multilayer Perceptron and Time-Delay Neural Network trained by scaled conjugate gradient algorithm are also used as classifiers.

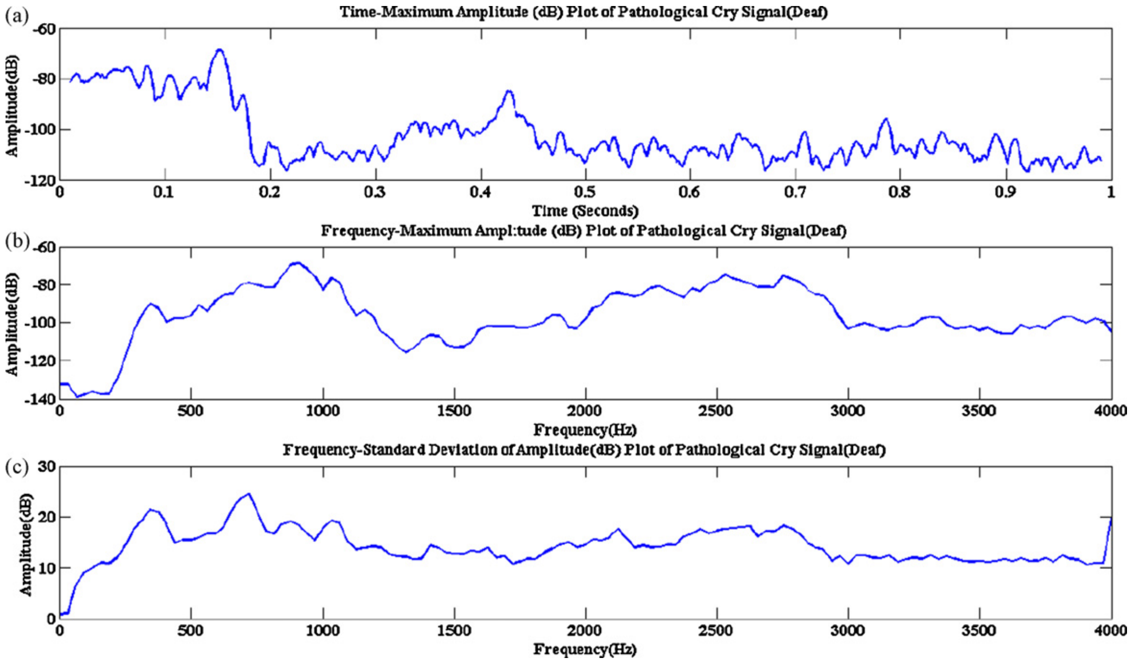

Fig. 5 – (a) Time–maximum amplitude (dB) plot of pathological cry signal (deaf, segment 6), (b) frequency–maximum amplitude (dB) plot, and (c) frequency–standard deviation of amplitude (dB) plot.

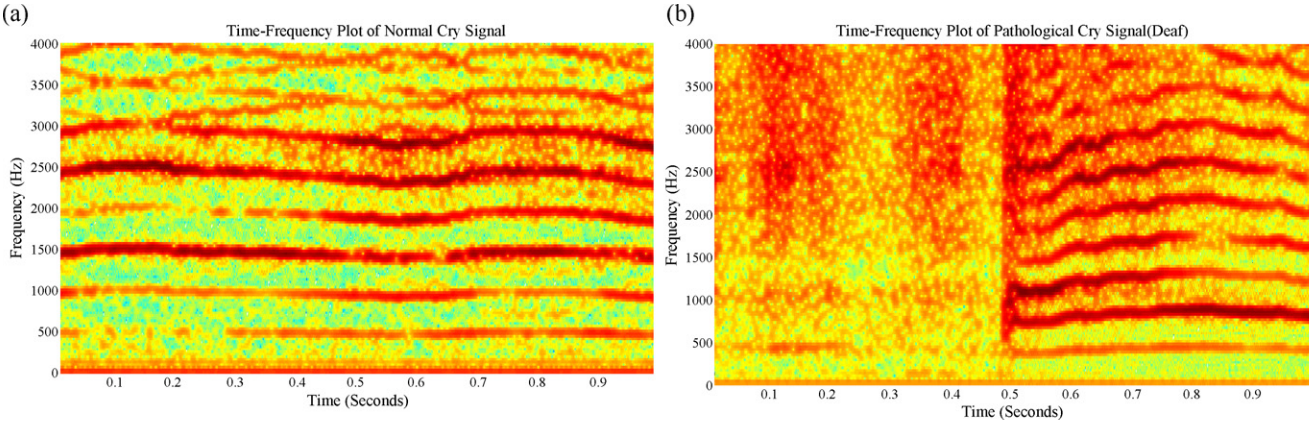

Fig. 6 – (a) Time frequency plot of normal cry signal (segment 300) and time frequency plot of pathological cry signal (deaf, segment 200).

4.1. *General regression neural network*

GRNN is a kind of radial basis networks and the training is conducted using one pass learning. This network does not require an iterative training procedure; it presents much faster learning than Multilayer Perceptron, it is more accurate than MLP and relatively insensitive to outliers [15–20]. For the GRNN the target variable is continuous. Radial basis function networks compute activations using an exponential of a distance measure (usually the Euclidean distance or a weighted norm) between the input vector and a prototype vector that characterizes the signal function at a hidden neuron rather than employing an inner product between the input vector and the weight vector [23]. D.F. Specht has proposed the model of GRNN to perform general (linear or nonlinear) regressions [24]. GRNN is based on the theory of probability regression analysis. It usually uses Parzen window estimates to set up the probability density function (PDF) from the observed data samples. Supposing x is a random vector variable, y is a random scalar variable, X and Y are measured values, $f(x, y)$ is the known continuous joint PDF. The expected value of y (the regression value on X) is given by Eq. (3) [24].

$$E(y|X) = \frac{\int_{-\infty}^{\infty} y\, f(X, y) dy}{\int_{-\infty}^{\infty} f(X, y) dy} \tag{3}$$

where y is the output predicted by GRNN. X the input vector $(x_1, x_2, \ldots, x_n)$ which consists of n predictor variables, $E(y|X)$ the expected value of the output y given an input vector X, and $f(X, y)$ the joint probability density function of X and y.

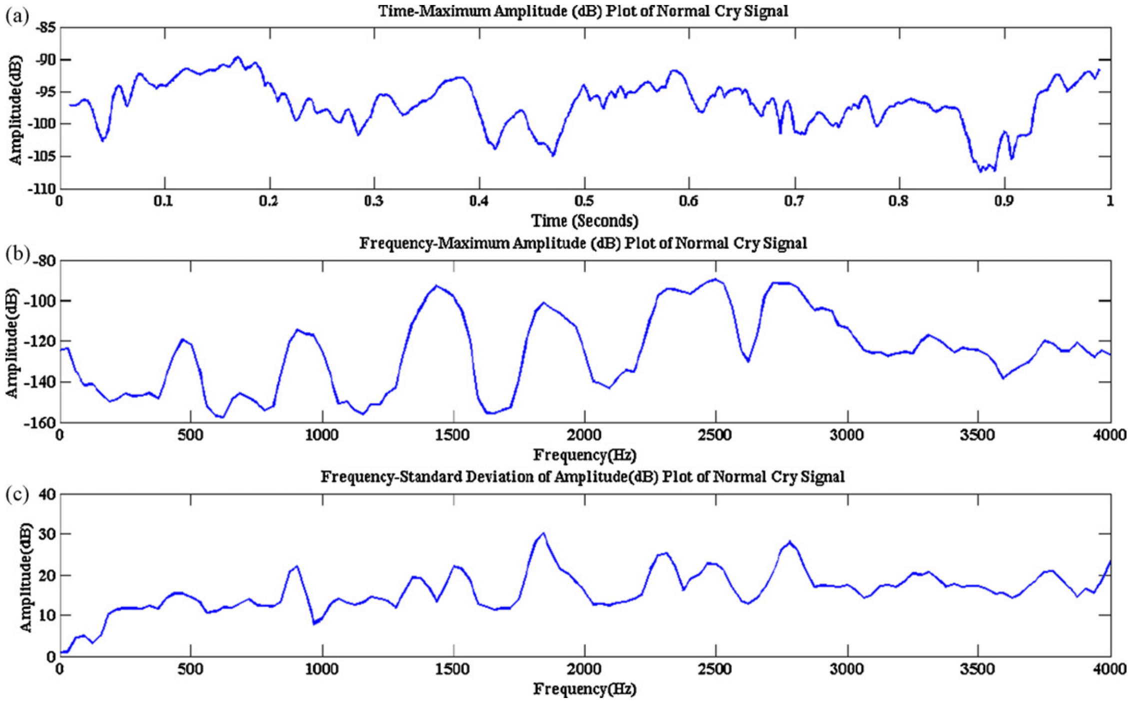

Fig. 7 – (a) Time–maximum amplitude (dB) plot of normal cry signal (segment 300), (b) frequency–maximum amplitude (dB) plot, and (c) frequency–standard deviation of amplitude (dB) plot.

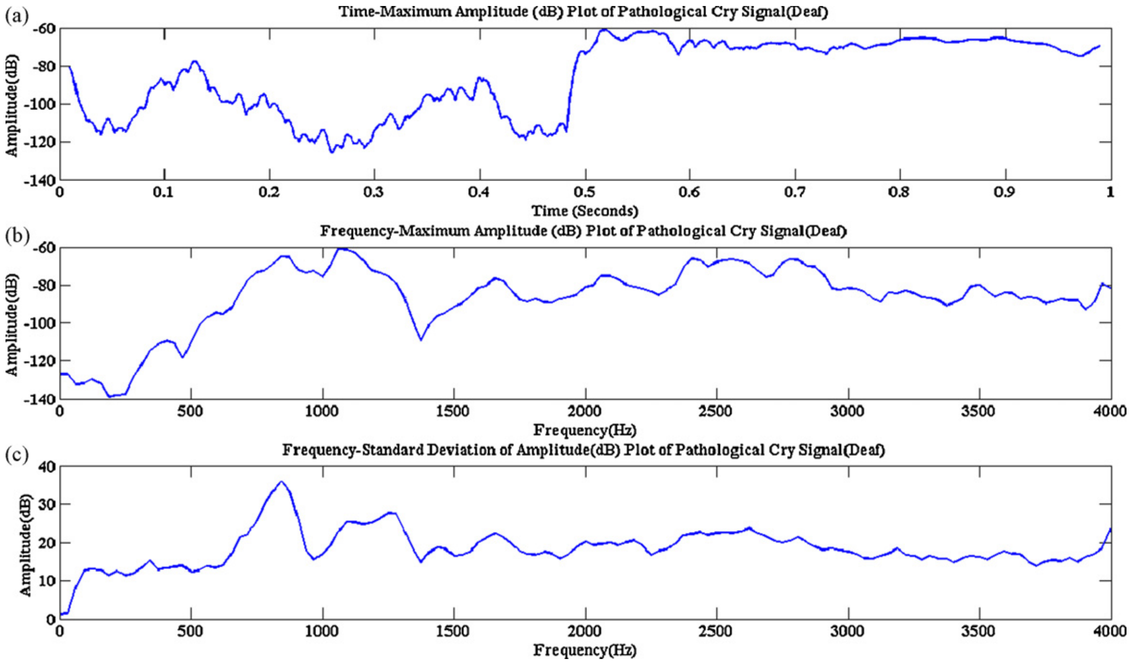

Fig. 8 – (a) Time–maximum amplitude (dB) plot of pathological cry signal (deaf, segment 200), (b) frequency–maximum amplitude (dB) plot, and (c) frequency–standard deviation of amplitude (dB) plot.

The estimated value Y is an exponentially weighted average value of all observed values Y^i given as in Eq. (4) [24]:

$$\hat{Y}(x) = \frac{\sum_{i=1}^{n} Y^i \exp(-(D_i^2/2\sigma^2))}{\sum_{i=1}^{n} \exp(-(D_i^2/2\sigma^2))} \tag{4}$$

where D_i is defined as in Eq. (5)

$$D_i^2 = (X - X^i)^T * (X - X^i) \tag{5}$$

The variable σ is a smoothing parameter that can be made large to smooth out noisy data or small to allow the estimated

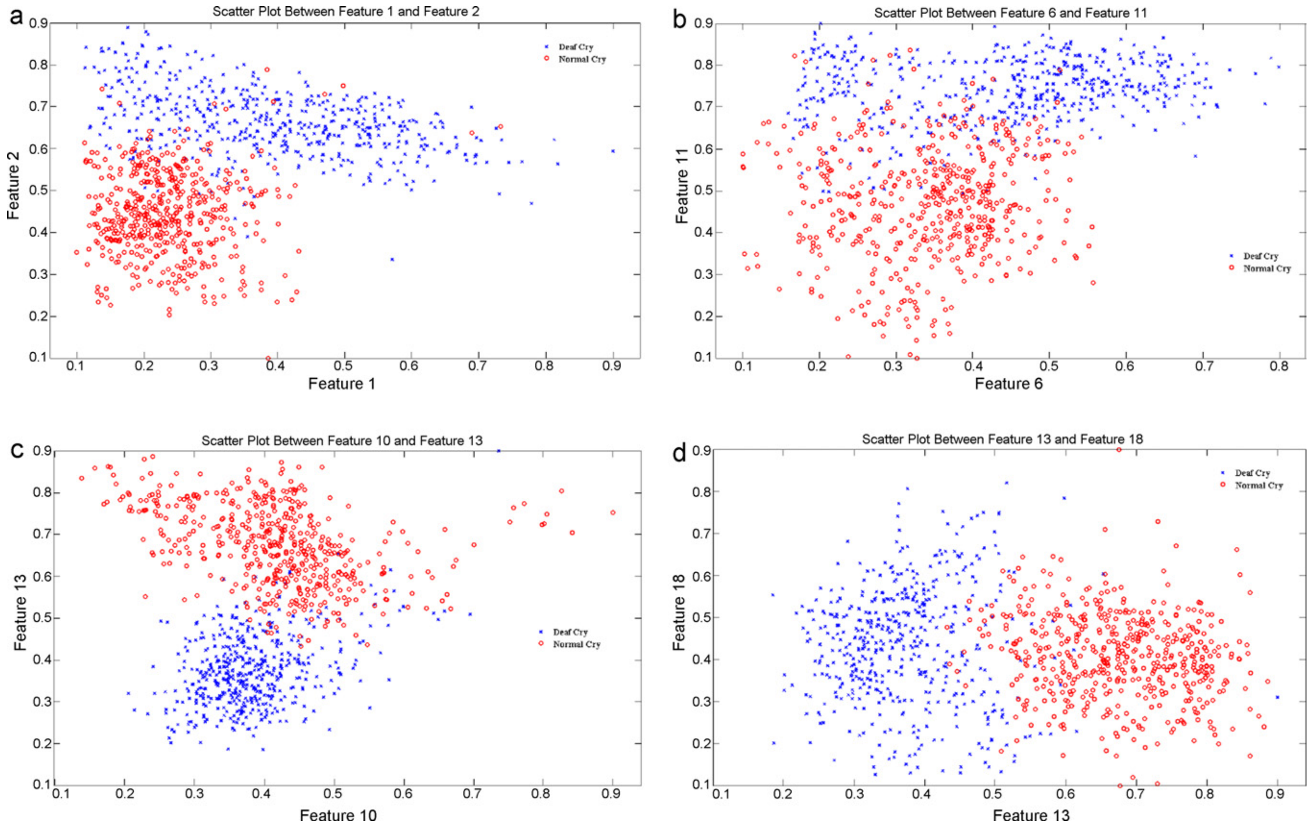

Fig. 9 – (a) Scatter plot between Feature 1 and Feature 2, (b) scatter plot between Feature 6 and Feature 11, (c) scatter plot between Feature 10 and Feature 13 and (d) scatter plot between Feature 13 and Feature 18.

regression surface to be as nonlinear as it is required to approximate closely the actual observed values of Y^i. The GRNN has 4 different layers: input layer, pattern layer, summation layer and output layer. In this work, GRNN architecture is constructed using *newgrnn()* in MATLAB function [25]. The detailed information about the GRNN architecture and mathematical equations can be found in Specht's paper [24]. The performance of the GRNN classifier highly depends upon the smoothing parameter or spread factor (σ). Based on the experimental investigations, the σ value is varied between 0.03 and 0.12 in steps of 0.01.

4.2. *Multilayer Perceptron classifier*

A three layer neural network model is developed with 20 input neurons, the hidden neurons which are varied between 10 and 20 in steps of 2 and 1 output neuron. The performance goal, learning rate, momentum factor are chosen as 0.001, 0.1, and 0.9 respectively. Scaled conjugate algorithm is chosen for training the neural network model [2,3]. The hidden and output neurons are activated by binary sigmoidal activation function. In this work, MLP architecture is constructed using *newff()* in MATLAB function [25]. The performance of the MLP classifier highly depends upon the different learning parameters, such as number of hidden neurons, learning rate, momentum factor, stopping criteria and activation functions. Based on the several experimental investigations, the best learning parameters are found and used during the training and testing of the MLP classifier.

4.3. *Time delay neural network*

Time delay neural network has been used in speech recognition applications [26,27] as well as in the infant cry classification [1,28]. It was proposed to use in infant cry classification since the cry data are not static and are time dependent on crying patterns [1,28]. The detailed information about the TDNN can be found in [1,26–28]. A TDNN model is developed and trained by scaled conjugate gradient algorithm. It consists of 20 neurons and the input delay specified by user, in this case the delay [28] is (0, 1), the hidden neurons which are varied between 10 and 20 in steps of 2 and 1 output neuron. The performance goal, learning rate, momentum factor are chosen as 0.001, 0.1, and 0.9 respectively. Scaled conjugate algorithm is chosen for training the TDNN model [2,3]. The hidden and output neurons are activated by binary sigmoidal activation function. In this work, TDNN architecture is constructed using *newfftd()* in MATLAB function [25]. The performance of the TDNN classifier highly depends upon the different learning parameters, such as number of hidden neurons, number of input delay, learning rate, momentum factor, stopping criteria and activation functions. Based on the several experimental investigations, the best learning parameters are found and used during the training and testing of the TDNN classifier.

Table 2 – Results of MLP classifier trained by Scaled conjugate gradient algorithm for the frame length 20 ms, 30 ms, 40 ms and 50 ms (10-fold cross validation).

Hidden neurons	Frame length (20 ms)			Frame length (30 ms)			Frame length (40 ms)			Frame length (50 ms)		
	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC
10	97.04	96.85	96.94	97.63	97.44	97.53	97.02	96.28	96.65	97.06	97.62	97.34
12	**97.81**	**97.06**	**97.44**	97.22	96.67	96.94	**97.62**	**97.25**	**97.44**	97.05	97.23	97.14
14	97.02	96.47	96.75	97.24	97.24	97.24	96.65	96.84	96.75	96.67	97.22	96.94
16	97.60	96.49	97.04	97.44	97.44	97.44	96.48	97.41	96.94	**97.64**	**98.02**	**97.83**
18	96.81	95.71	96.25	**98.22**	**97.84**	**98.03**	98.00	96.69	97.34	96.28	97.02	96.65
20	97.42	96.67	97.04	96.84	96.65	96.75	97.02	96.28	96.65	98.00	96.88	97.44

Table 3 – Results of TDNN classifier trained by scaled conjugate gradient algorithm for the frame length 20 ms, 30 ms, 40 ms and 50 ms (10-fold cross validation).

Spread factor	Frame length (20 ms)			Frame length (30 ms)			Frame length (40 ms)			Frame length (50 ms)		
	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC
10	**97.82**	**97.45**	**97.63**	97.44	97.44	97.44	96.87	97.61	97.24	97.06	97.62	97.34
12	97.43	97.05	97.24	96.30	97.60	96.94	97.81	97.06	97.44	96.86	97.42	97.14
14	97.80	96.31	97.04	97.61	96.87	97.24	97.61	96.68	97.14	97.04	96.85	96.94
16	96.46	96.64	96.55	97.45	97.82	97.63	97.24	97.43	97.34	**97.83**	**97.83**	**97.83**
18	96.84	96.84	96.84	**97.64**	**98.02**	**97.83**	97.04	97.04	97.04	97.01	95.91	96.45
20	97.60	96.49	97.04	98.01	97.07	97.53	**97.63**	**97.44**	**97.53**	97.83	97.83	97.83

5. Results and discussion

In this work, two validation schemes (10-fold cross validation [29] and data independent validation) are used to prove the reliability of the classification results. In 10-fold cross validation scheme, the proposed feature vectors are divided randomly into 10 sets and training is repeated for 10 times. For each run of cross validation the number of normal and pathological cases is equal. In data independent validation scheme, the classifiers are trained with a selected set of samples and are tested with different samples which are not taken in count for the training stage and also the training and testing dataset are prepared as follows:

670 segments are used for training (335 segments from 3 deaf babies + 335 segments from 2 normal babies) and remaining 344 segments are used for testing (172 segments from remaining 3 deaf babies + 172 segments from remaining 3 normal babies). In order to test the classifier performance, three measures namely, sensitivity (SE), specificity (SP), and the overall accuracy (AUC) are considered. These measures are calculated from the measures true positive (TP, number of correctly classified pathological samples), true negative (TN, number of correctly classified normal samples), false positive (FP, number of incorrectly classified pathological samples), and false negative (FN, number of incorrectly classified normal samples).

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$$

$$\text{Overall accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}$$

The GRNN is trained with different spread factor or smoothing factor between 0.03 and 0.12 and its effects on the classification performance are analyzed. The MLP and TDNN are trained with different number of hidden neurons between 10 and 20 and its effects on the classification performance are analyzed. The results for the MLP classifier, TDNN, and GRNN classifier using 10-fold cross validation scheme are tabulated in Tables 2–4. The maximum classification accuracy was highlighted in Tables 2–4 for every frame length. From Table 2, the best overall accuracy of 97.44% (20 ms and 12 hidden neurons), 98.03% (30 ms and 18 hidden neurons), 97.44% (40 ms and 12 hidden neurons), 97.83% (50 ms and 16 hidden neurons) are obtained using MLP classifier. From Table 3, the best overall accuracy of 97.63% (20 ms and 10 hidden neurons), 97.83% (30 ms and 18 hidden neurons), 97.53% (40 ms and 20 hidden neurons), and 97.83% (50 ms and 16 hidden neurons) are obtained using TDNN classifier. From Table 4, it is observed that the GRNN classifier gives maximum overall accuracy of 99.01% (20 ms, 0.06-spread factor), 99.01% (30 ms, 0.05-spread factor), 99.21% (40 ms, 0.05-spread factor), and 99.31% (50 ms, 0.08-spread factor). In all the classifiers, there are no specific changes in the classification accuracies due to the different frame length. From the results of Tables 2 and 3, the best number of hidden neurons can lie between 10 and 20 to obtain maximum classification accuracy using MLP and TDNN classifier. From the results of Table 4, the best spread factor can lie between 0.06 and 0.10 to obtain maximum classification accuracy using GRNN classifier.

The results for the MLP, TDNN, and GRNN classifier using the data independent validation scheme (the classifiers are trained with a selected set of samples and are tested with different samples which are not taken in count for the training stage) are tabulated in Tables 5–7. The maximum classification accuracy was highlighted in Tables 5–7 for every frame length. From Table 5, the best overall classification accuracy of 89.10% (20 ms and 12 hidden neurons), 89.39% (30 ms and

Table 4 – Results of GRNN classifier for the frame length 20 ms, 30 ms, 40 ms and 50 ms (10-fold cross validation).

Spread factor	Frame length (20 ms)			Frame length (30 ms)			Frame length (40 ms)			Frame length (50 ms)		
	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC
0.03	95.95	99.18	96.94	95.76	**99.39**	96.84	96.35	**99.59**	97.63	95.44	**99.38**	97.14
0.04	98.82	**99.21**	98.82	98.62	99.01	98.72	98.82	99.40	98.92	98.81	99.01	98.72
0.05	98.62	99.21	98.82	**99.21**	99.02	**99.01**	99.21	99.41	**99.21**	99.40	98.82	99.01
0.06	**99.40**	98.82	**99.01**	98.81	98.62	98.62	**99.60**	99.02	99.21	**99.80**	98.64	99.11
0.07	99.40	98.82	99.01	99.01	98.82	98.92	98.81	98.82	98.72	99.80	98.83	99.21
0.08	99.21	98.82	99.01	99.01	98.62	98.82	98.82	99.01	98.92	99.80	98.83	**99.31**
0.09	99.01	98.82	98.92	98.81	98.24	98.52	98.81	98.62	98.72	99.60	98.83	99.21
0.10	98.81	98.04	98.42	98.81	98.04	98.42	98.80	97.85	98.32	99.60	98.44	99.01
0.11	98.61	97.66	98.13	98.22	98.03	98.13	98.41	97.65	98.03	99.40	98.05	98.72
0.12	98.19	96.14	97.14	98.21	97.46	97.83	97.60	96.49	97.04	99.00	97.29	98.13

Table 5 – Results MLP classifier trained by scaled conjugate gradient algorithm for the frame length 20 ms, 30 ms, 40 ms and 50 ms (670 segments used for training and 344 segments for testing).

Hidden neurons	Frame length (20 ms)			Frame length (30 ms)			Frame length (40 ms)			Frame length (50 ms)		
	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC
10	84.82	95.76	89.53	81.40	97.30	87.65	80.88	96.05	86.92	82.23	96.79	88.08
12	**84.53**	**95.10**	**89.10**	**83.66**	**97.69**	**89.39**	80.76	96.40	86.92	81.81	96.96	87.88
14	83.75	95.12	88.52	82.83	97.86	88.92	81.99	96.70	87.94	82.41	97.30	88.43
16	81.73	95.21	87.24	81.91	97.03	87.94	81.23	96.72	87.38	82.92	97.81	88.95
18	83.32	94.86	88.14	82.75	97.43	88.69	81.76	96.75	87.79	**83.85**	**97.74**	**89.56**
20	82.52	94.87	87.65	83.02	97.53	88.92	**82.27**	**96.66**	**88.11**	81.80	96.82	87.85

Table 6 – Results of TDNN classifier trained by Scaled conjugate gradient algorithm for the frame length 20 ms, 30 ms, 40 ms and 50 ms (670 segments used for training and 344 segments for testing).

Spread factor	Frame length (20 ms)			Frame length (30 ms)			Frame length (40 ms)			Frame length (50 ms)		
	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC
10	84.76	94.55	89.04	82.56	97.91	88.72	81.77	96.92	87.82	83.48	96.24	88.81
12	84.91	95.20	89.39	82.33	97.54	88.43	82.10	96.85	88.08	80.74	97.05	87.15
14	82.82	95.15	87.94	82.93	98.28	89.13	82.11	96.57	87.97	84.18	96.78	89.45
16	84.08	94.78	88.55	82.18	97.47	88.31	**83.26**	**96.90**	**88.90**	84.26	96.50	89.42
18	**84.98**	**95.27**	**89.45**	**83.65**	**97.62**	**89.42**	82.36	96.85	88.23	83.61	96.81	89.10
20	84.74	94.87	89.13	82.21	97.54	88.37	80.98	96.50	87.15	**84.55**	**97.65**	**90.03**

Table 7 – Results of GRNN classifier for the frame length 20 ms, 30 ms, 40 ms and 50 ms (670 segments used for training and 344 segments for testing).

Spread factor	Frame length (20 ms)			Frame length (30 ms)			Frame length (40 ms)			Frame length (50 ms)		
	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC	SE	SP	AUC
0.03	67.33	98.90	75.29	70.12	99.01	78.20	71.31	99.05	79.36	72.46	100.00	80.81
0.04	82.44	99.27	88.66	82.84	98.56	88.95	82.52	99.27	88.95	83.33	99.28	89.53
0.05	85.86	99.31	91.28	85.71	97.96	90.70	86.73	99.32	91.86	85.86	99.31	91.28
0.06	85.43	99.31	90.99	85.71	97.96	90.70	86.29	99.32	91.57	87.56	98.67	92.15
0.07	**86.22**	**98.64**	**91.28**	**87.11**	**98.00**	**91.86**	86.73	99.32	91.86	87.56	98.67	92.15
0.08	86.22	97.97	91.28	87.11	98.00	91.86	**87.63**	**98.67**	**92.44**	88.48	98.68	92.73
0.09	85.35	97.95	90.70	86.15	97.32	90.99	87.63	98.67	92.44	**89.95**	**98.71**	**93.90**
0.10	84.62	95.30	89.24	85.28	97.28	90.41	87.11	98.00	91.86	88.54	98.68	93.02
0.11	83.51	93.33	87.79	83.58	97.20	89.24	86.15	97.32	90.99	87.50	97.37	91.86
0.12	83.33	92.11	87.21	82.59	95.80	88.08	85.71	97.30	90.70	87.37	96.10	91.28

12 hidden neurons), 88.11% (40 ms and 20 hidden neurons) and 89.56% (50 ms and 18 hidden neurons) are obtained using MLP classifier. From Table 6, best overall classification accuracy of 89.40% (20 ms and 18 hidden neurons), 89.42% (30 ms and 18 hidden neurons), 88.90% (40 ms and 16 hidden neurons) and 90.03% (50 ms and 20 hidden neurons) are obtained using TDNN classifier. From Table 7, it is observed that the GRNN classifier gives maximum overall accuracy of 91.28% (20 ms, 0.07-spread factor), 91.86% (30 ms, 0.07-spread factor), 92.44% (40 ms, 0.08-spread factor), and 93.90% (50 ms, 0.09-spread factor). In all the classifiers, there are no specific changes in the classification accuracies due to the different frame lengths.

From the results of Tables 5 and 6, the best number of hidden neurons can lie between 10 and 20 to get maximum classification accuracy using MLP and TDNN classifier. From the results of Table 7, the best spread factor can lie between 0.06 and 0.10 to get maximum classification accuracy using GRNN classifier.

From the above discussion, it has been observed that the suggested time–frequency analysis based statistical features can be used to provide the most discriminating representation of normal and deaf cry signals. In this paper, twenty simple and efficient statistical features are derived through STFT based time–frequency analysis to provide robust representation of infant cry signals. In Table 1, some of significant works are reported and the maximum classification accuracy of 100% was obtained [5]. The number of features used in the works reported in Table 1 is different and also different classification algorithms and hybrid systems were used for infant cry classification. In [5], the authors have proposed evolutionary approach using two different set of Mexican and Cuban babies. They classified the infant cry signals either into normal or pathological (deaf babies and asphyxiating babies) infant cry signals, but the pathological signals were not further classified into asphyxia cry signals or deaf cry signals. They have used 30 features or more than that to obtain 100% accuracy for the infant cry signals recognition recorded from Mexican babies. But we have obtained the classification accuracy of above 99% with only twenty times–frequency analysis based statistical features and GRNN classifier. It shows that the suggested features and GRNN classifier provides closer results with the earlier works. Using the data independent validation scheme, the maximum classification accuracy of 93% (GRNN), 89% (TDNN), and 89% (MLP) are obtained. Finally, the experimental result indicates the strength of the suggested method and has the potential in detecting pathological problem of an infant from cry signals.

6. Conclusions

This paper presents a simple feature extraction method based on time–frequency analysis using STFT for the investigation of infant cry signals. Simple statistical features are derived from time–frequency plots, time–maximum amplitude plots, frequency–maximum amplitude plots, and frequency–standard deviation plots. A GRNN classifier is employed to classify the cry signals into normal or pathological. To prove the reliability of the proposed features, two neural network models such as Multilayer Perceptron and Time-Delay Neural Network trained by scaled conjugate gradient algorithm are also used as classifiers. 10-fold cross validation and data independent validation scheme are performed, in order to test the generalizability and reliability of the GRNN, MLP and TDNN classifier. The suggested method provides maximum classification accuracy of 99% (GRNN), 97% (TDNN), and 97% (MLP) using 10-fold cross validation scheme. Using the data independent validation scheme, the maximum classification accuracy of 93% (GRNN), 88% (TDNN), and 88% (MLP) are obtained. From the results, it can be inferred that the GRNN gives higher accuracy compared to MLP and TDNN. The classification results indicate that the suggested method could be used as a valuable tool for classifying the infant cry signals into normal and pathological. In the future work, the suggested method will be used to classify more than one pathological cry signal from normal cry signal. Feature reduction techniques will be implemented to propose the reduced feature set with predominant features. The proposed method will be validated with larger samples.

Conflict of interest statement

None.

Acknowledgements

The Baby Chillanto Data Base is a property of the Instituto Nacional de Astrofisica Optica y Electronica – CONACYT, Mexico. We like to thank Dr. Carlos A. Reyes-Garcia, Dr. Emilio Arch-Tirado and his INR-Mexico group, and Dr. Edgar M. Garcia-Tamayo for their dedication of the collection of the Infant Cry data base. The authors would like to thank Dr. Carlos Alberto Reyes-Garcia, Researcher, CCC-Inaoep, Mexico for providing infant cry database. The authors are grateful to the anonymous reviewers for the valuable comments.

REFERENCES

[1] O.F. Reyes-Galaviz, A. Verduzco, E. Arch-Tirado, C.A. Reyes-García, Analysis of an infant cry recognizer for the early identification of pathologies, Nonlinear Speech Modeling and Applications 3445 (2005) 404–409.

[2] J.O. Garcia, C.A. Reyes García, Detecting pathologies from infant cry applying scaled conjugate gradient neural networks, in: European Symposium on Artificial Neural Networks, Bruges (Belgium), 2003, pp. 349–354.

[3] J.O. Garcia, C.A. Reyes García, Acoustic features analysis for recognition of normal and hypoacoustic infant cry based on neural networks, Lecture Notes in Computer Science, Artificial Neural Nets Problem Solving Methods 2687 (2003) 615–622, doi: 10.1007/3-540-44869-1_78.

[4] G. Várallyay Jr., Z. Benyó, A. Illényi, Z. Farkas, L. Kovács, Acoustic analysis of the infant cry: classical and new methods, in: Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA, 2004, pp. 313–316.

[5] O.F. Reyes-Galaviz, S. Cano-Ortiz, C. Reyes-Garca, Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies, in: Proceedings of the 8th Mexican International Conference on Artificial Intelligence, MICAI 2009, Guanajuato, Mexico, 2009, pp. 330–335.

[6] O.F. Reyes-Galaviz, C. Reyes-Garcia, A system for the processing of infant cry to recognize pathologies in recently born babies with neural networks, in: Proceedings of the 9th Conference Speech and Computer (SPECOM'2004), St. Petersburg, Russia, 2004.

[7] D. Escobedo, S. Cano, E. Coello, L. Regueiferos, L. Capdevila, Rising shift of pitch frequency in the infant cry pf some pathologic cases, in: Proceedings of the 2nd International Conference MAVEBA 2001, Firenze, Italy, 2001.

[8] S. Cano, et al., The spectral analysis of infant cry: an initial approximation, in: Proceedings of the EUROSPEECH'95 (sponsored by ESCA & IEEE), Madrid, 1995.

[9] C. Manfredi, V. Tocchioni, L. Bocchi, A robust tool for newborn infant cry analysis, in: Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, USA, Aug 30–Sept 3, 2006, pp. 509–512.

[10] G. Várallyay Jr., The melody of crying, International Journal of Pediatric Otorhinolaryngology 71 (11) (2007) 1699–1708.

[11] Wasz-Hockert, et al., The Infant Cry: A Spectrographic and Auditory Analysis, William Heinemann Medical Books Ltd., 1968.

[12] M. Petroni, A. Malowany, C. Johnston, B. Stevens, International Infant Cry Research Group, Identification of pain from infant cry vocalizations using artificial neural networks (ANNs). Applications and science of artificial neural networks, The International Society for Optical Engineering 2492 (1995) 729–738.

[13] S. Cano, I. Suaste-Rivas, D. Escobedo, C.A. Reyes-Garcia, T. Ekkel, A combined classifier of cry units with new acoustic attributes, Lecture Notes in Computer Sciences (LNCS) 4225 (2006) 416–425.

[14] S.E. Barajas-Montiel, C.A. Reyes-García, Fuzzy Support vector machines for automatic infant cry recognition Lecture Notes in Control and Information Sciences (LNCIS), vol. 345, Springer, 2006, pp. 876–881.

[15] Z. Feng, F. Chu, X. Song, Application of general regression neural network to vibration trend prediction of rotating machinery, Lecture Notes in Computer Sciences (LNCS) 3174 (2004) 767–772.

[16] B. Erkmen, T. Yildirim, Improving classification performance of sonar targets by applying general regression neural network with PCA, Expert Systems with Applications 35 (2008) 472–475.

[17] O. Polat, T. Yildirim, Hand geometry identification without feature extraction by general regression neural network, Expert Systems with Applications 34 (2008) 845–849.

[18] G.J. Bowden, J.B. Bixon, G.C. Dandy, H.R. Maier, M. Holmes, Forecasting chlorine residuals in a water distribution system using a general regression neural network, Mathematical and Computer Modelling 44 (2006) 469–484.

[19] M.T. Leung, A.S. Chen, H. Daouk, Forecasting exchange rates using general regression neural network, Computers and Operation Research 27 (2000) 1093–1110.

[20] M. Firat, M. Gungor, Generalized regression neural networks and feed forward neural networks for prediction of scour depth around bridge piers, Advances in Engineering Software 40 (2009) 731–737.

[21] L. Rabiner, B. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993.

[22] John L. Semmlow, Biosignal and Biomedical Image Processing, Marcel Dekker Inc., 2004.

[23] S. Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill, New Delhi, 2004.

[24] D.F. Specht, A general regression neural network, IEEE Transactions on Neural Networks 2 (6) (1991) 568–576.

[25] Matlab® Documentation, Version 7.0, Release 14, The Math-Works, Inc., 2004.

[26] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, K.J. Lang, Phoneme recognition using time delay neural network, IEEE Transactions on Acoustics, Speech, and Signal Processing 37 (3) (1989) 328–339.

[27] J.B. Hampshire, A.H. Waibel, A novel objective function for improved phoneme recognition using Time-Delay Neural Network, IEEE Transactions on Neural Network 1 (2) (1990) 216–228.

[28] O.F. Reyes Galaviz, C.A. Reyes Garcia, Infant Cry classification to identify hypoacoustics and asphyxia with neural networks, MICAI 2004, LNAI 2972 (2004) 69–78.

[29] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the14th International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada, 1995.