Expert Systems with Applications 39 (2012) 9515–9523

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Infant cry classification to identify asphyxia using time-frequency analysis and radial basis neural networks

M. Hariharan [a,*], J. Saraswathy [a], R. Sindhu [b], Wan Khairunizam [a], 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

Keywords:
Infant cry
Feature extraction
Short-time Fourier transform
PNN and GRNN

ABSTRACT

A cry is the first verbal communication of infants and it is described as a loud, high-pitched sound made by infants in response to certain situations. 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. Two types of radial basis neural networks such as Probabilistic Neural Network (PNN) and General Regression Neural Network are employed as classifiers for discriminating infant cry signals. Two classes of infant cry signals are considered such as normal cry signals and pathological cry signals of infants with asphyxia. For comparison, the proposed features are also tested using two neural network models such as Multilayer Perceptron (MLP) and Time-Delay Neural Network (TDNN) trained by scaled conjugate gradient algorithm. The experimental results show that the PNN and GRNN give very promising classification accuracy compared to MLP and TDNN and the proposed methods can effectively classify normal and pathological infant cries of infants with asphyxia.

1. Introduction

Infant cry is defined as a loud, high-pitched sound made by infants in response to certain situations and it is the first verbal communication of infants. 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 such as pediatrician and pediatric nurse 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 (Barajas-Montiel & Reyes-García, 2006; Cano, Suaste-Rivas, Escobedo, Reyes-Garcia, & Ekkel, 2006; Cano et al., 1995; Escobedo, Cano, Coello, Regueiferos, & Capdevila, 2001; Garcia & Reyes García, 2003a, 2003b; Manfredi, Tocchioni, & Bocchi, 2006; Petroni, Malowany, Johnston, & Stevens, 1995; Reyes Galaviz & Reyes Garcia, 2004; Reyes-Galaviz, Cano-Ortiz, Reyes-Garcia, y Electronica, & Puebla, 2009; Reyes-Galaviz & Reyes-Garcia, 2004; Reyes-Galaviz, Verduzco, Arch-Tirado, & Reyes-García, 2005; Várallyay, 2007; Várallyay, Benyó, Illényi, Farkas, & Kovács, 2004; Wasz-Hockert et al., 1968). 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 (Barajas-Montiel & Reyes-García, 2006; Cano et al., 1995, 2006; Escobedo et al., 2001; Garcia & Reyes García, 2003a, 2003b; Manfredi et al., 2006; Petroni et al., 1995; Reyes Galaviz, Arch Tirado, & Reyes Garcia, 2004; Reyes Galaviz & Reyes Garcia, 2004; Reyes-Galaviz et al., 2005, 2009; Várallyay, 2007; Várallyay et al., 2004; Wasz-Hockert et al., 1968). Different classification algorithms and hybrid systems were used for infant cry classification (Barajas-Montiel & Reyes-García, 2006; Cano et al., 1995, 2006; Escobedo et al., 2001; Garcia & Reyes García, 2003a, 2003b; Manfredi et al., 2006; Petroni et al., 1995; Reyes Galaviz et al., 2004; Reyes-Galaviz & Reyes-Garcia, 2004; Reyes-Galaviz et al., 2005, 2009; Várallyay, 2007; Várallyay et al., 2004; Wasz-Hockert et al., 1968). 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

* Corresponding author. Tel.: +60 49798419; fax: +60 49885167.
E-mail address: hari@unimap.edu.my (M. Hariharan).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.102

proposed by researchers as it is a good tool for analyzing the infant cry signals both time and frequency scale simultaneously (Hariharan, Sindhu, & Yaacob, 2011c). Few works can be found from literature using time-frequency analysis for infant cry signal recognition (Cano et al., 1995; Escobedo et al., 2001; Manfredi et al., 2006; Petroni et al., 1995; Várallyay, 2007; Wasz-Hockert et al., 1968). But, the interpretation from time-frequency analysis is different. Many of them have used pitch, harmonic analysis, and noise analysis (Cano et al., 1995; Escobedo et al., 2001; Manfredi et al., 2006; Petroni et al., 1995; Várallyay, 2007; Wasz-Hockert et al., 1968). This paper presents the development of an intelligent learning system to classify normal and pathological cries using short-time Fourier transform and radial basis neural networks (PNN and GRNN). 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 pathological 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 radial basis neural networks are proposed for discriminating normal and pathological cries (Asphyxia). Two scheme of data validation methods are used (Conventional Validation-ConV and 10-fold cross validation-CrossV), 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 radial basis neural networks 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://www.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 reference (Reyes-Galaviz et al., 2009). All the samples of this database have the length of 1 s and we have taken the same samples for our analysis. It consists of 507 of normal cry signals and 340 of asphyxia (Asphyxia is defined as the failure to breathe well within one minute after delivery of the baby. This disease can cause damage to the brain, organs and tissues or even death if subjected to delayed or improper treatment) cry signals. In this experiment, we took the same number of samples for each class 340. The asphyxia 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 baby who is affected with asphyxia 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 PNN, GRNN, MLP and TDNN. This section briefly describes the feature extraction and classification methods.

4. Feature extraction using short-time Fourier transform (STFT)

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 (John, 2004; Rabiner & Juang, 1993). 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 ($N \geqslant L$). Usually, N is chosen to be a power of -2 for using an

Table 1
Some of the significant works on the classification of infant cry signals.

Author name	Feature extraction method	Classifier	Best accuracy %
Garcia and Reyes García (2003a, 2003b)	Linear prediction technique	Scaled conjugate gradient neural networks (normal and deaf cry)	91.08% (314 samples) 86.20% (1036 samples)
Garcia and Reyes García (2003a, 2003b)	Mel-frequency cepstral coefficients and linear prediction technique	Scaled conjugate gradient neural networks (normal and deaf cry)	97.43%
Várallyay et al. (2004)	Fundamental frequency detection using smoothed spectrum method	–	–
Reyes-Galaviz and Reyes-Garcia (2004)	Mel-frequency cepstral coefficients	Adaptive neuro fuzzy inference system (normal, deaf cry, asphyxia cry:3 class problem)	93–96.67%
Reyes-Galaviz et al. (2005)	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%
Reyes-Galaviz et al. (2005)	Mel-frequency cepstral coefficients	Feed forward input delay neural network (normal, deaf cry, asphyxia cry, 3 class problem)	96.08–97.39%
Hariharan et al. (2010)	Weighted Linear Prediction Coefficients	Probabilistic Neural Network (normal, deaf cry, asphyxia cry, 3 class problem)	99%
Hariharan et al. (2011a, 2011b, 2011c)	Wavelet Packet Transform	Probabilistic Neural Network (normal and asphyxia cry)	99%
Hariharan et al. (2011a, 2011b, 2011c)	Time-frequency analysis based statistical features	General Regression Neural Network (normal and deaf cry)	99%

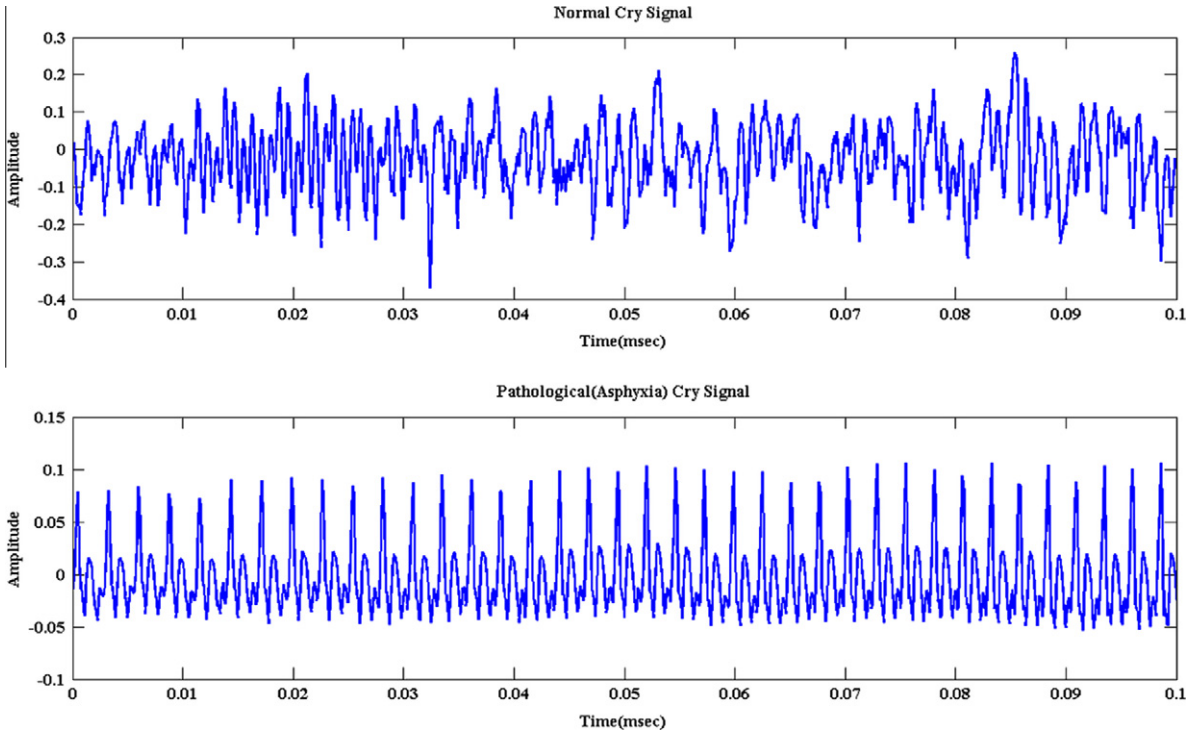

Fig. 1. Infant cry signals (normal and asphyxia).

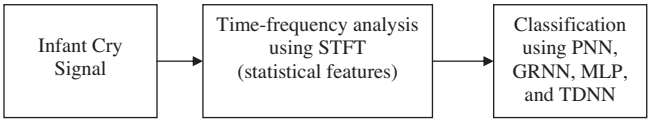

Fig. 2. Block diagram of the classification of normal and asphyxia infant cry.

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

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

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 (Rabiner & Juang, 1993) 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. 3a and 3b), 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.

Figs. 3a and 3b illustrate the time-frequency plot of normal cry signal (segment 350) and pathological cry signal (Asphyxia, segment 250). Figs. 4 and 5(a) depicts the time-maximum amplitude plot, which is maximum amplitude versus time by finding columns of time frequency plot. Figs. 4 and 5(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 and 5(c) depicts 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 the Figs. 3a and 3b, one can differentiate the normal and pathological cry through visual inspection.

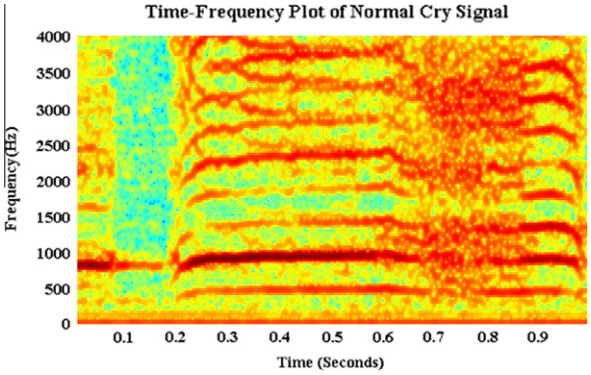

Fig. 3a. Time frequency plot of normal cry signal (segment 350).

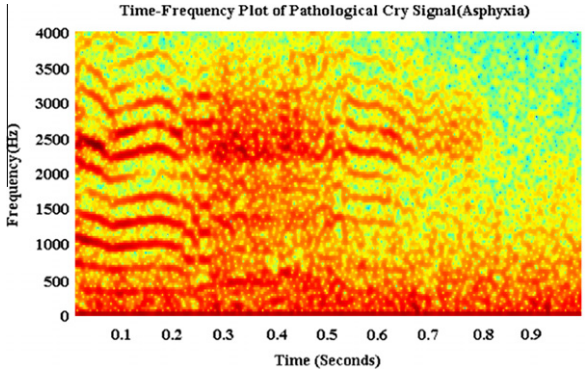

Fig. 3b. Time frequency plot of pathological cry signal (asphyxia, segment 250).

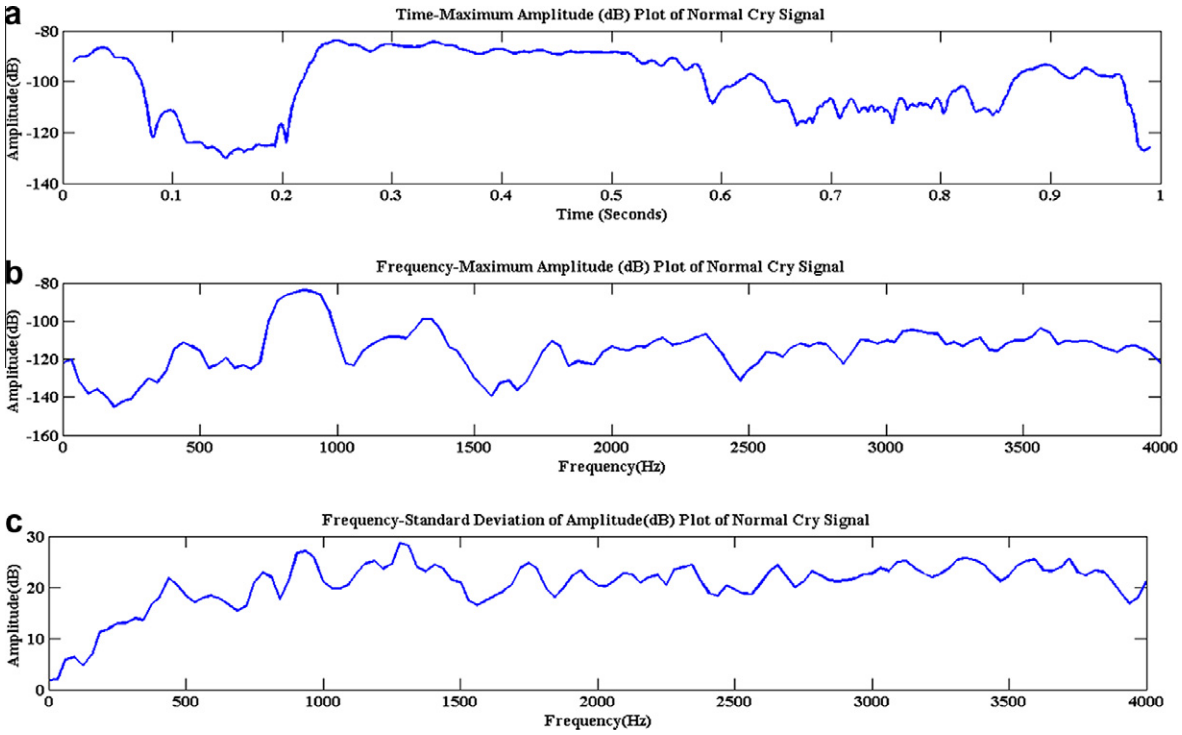

Fig. 4. (a) Time-maximum amplitude (dB) plot of normal cry signal (segment 350), (b). Frequency-maximum amplitude (dB) plot, and (c). Frequency – standard deviation of amplitude (dB) plot.

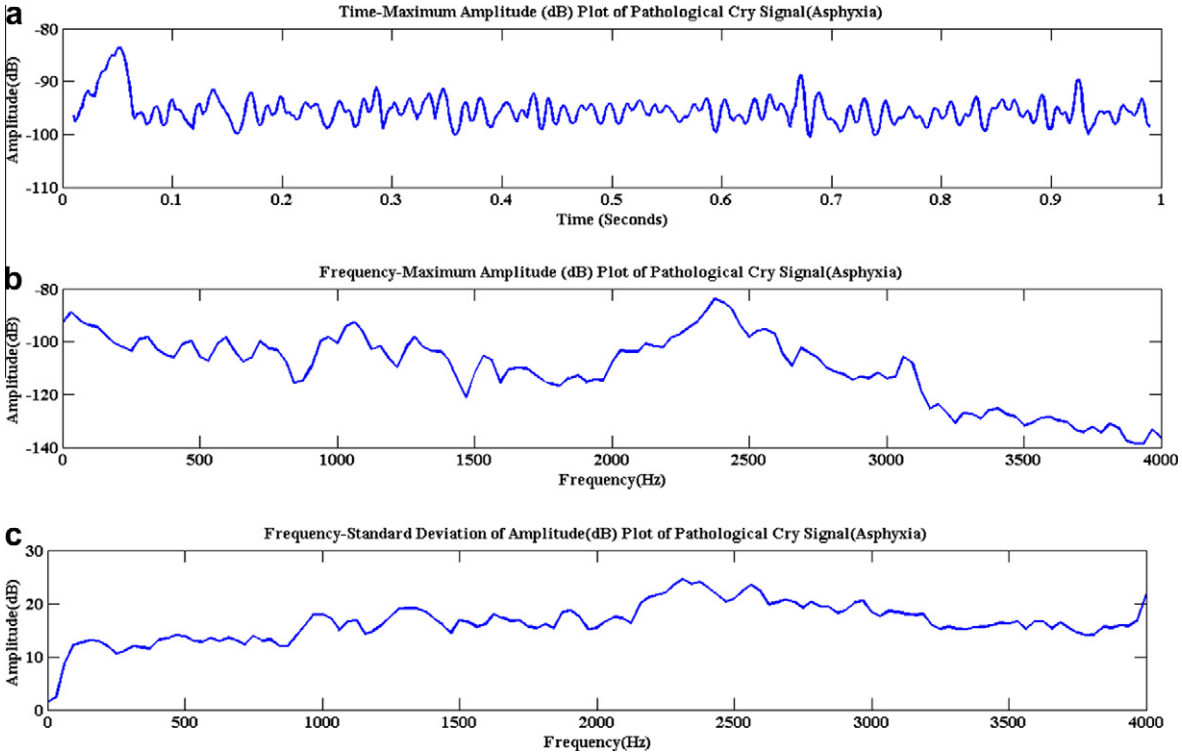

Fig. 5. (a) Time-maximum amplitude (dB) plot of pathological cry signal (asphyxia, segment 250), (b). Frequency-maximum amplitude (dB) plot, and (c). Frequency – standard deviation of amplitude (dB) plot.

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.

Table 2
Statistics of extracted parameters (mean ± std) for the different frame length.

Features	20 ms		30 ms		40 ms		50 ms	
	Normal	Asphyxia	Normal	Asphyxia	Normal	Asphyxia	Normal	Asphyxia
1	2.64 ± 0.68	3.04 ± 1.06	2.40 ± 0.67	2.83 ± 1.02	2.27 ± 0.66	2.73 ± 1.00	2.22 ± 0.65	2.69 ± 0.98
2	−156.90 ± 8.33	−169.66 ± 25.77	−157.94 ± 8.38	−170.37 ± 25.66	−158.24 ± 8.46	−170.50 ± 25.56	−158.36 ± 8.54	−170.43 ± 25.46
3	−80.85 ± 7.91	−82.75 ± 17.88	−78.71 ± 8.01	−81.68 ± 18.59	−77.43 ± 8.05	−80.97 ± 18.87	−76.76 ± 8.10	−80.66 ± 19.08
4	−125.34 ± 7.41	−124.66 ± 27.64	−121.60 ± 7.21	−118.24 ± 26.05	−118.98 ± 6.83	−114.36 ± 25.15	−116.93 ± 6.38	−111.44 ± 24.32
5	−101.51 ± 8.68	−103.08 ± 20.14	−99.08 ± 8.71	−100.56 ± 20.30	−97.61 ± 8.57	−98.77 ± 20.27	−96.73 ± 0.68	−97.49 ± 20.20
6	13.17 ± 4.28	9.97 ± 4.78	13.06 ± 4.28	9.39 ± 4.86	12.80 ± 4.13	8.90 ± 4.90	12.51 ± 4.03	8.40 ± 4.85
7	−0.31 ± 0.92	0.02 ± 0.71	−0.28 ± 0.94	0.18 ± 0.86	−0.23 ± 0.91	0.26 ± 0.89	−0.17 ± 0.92	0.35 ± 0.97
8	2.79 ± 2.53	2.95 ± 1.46	2.81 ± 2.78	3.17 ± 2.31	2.73 ± 2.40	3.19 ± 2.24	2.73 ± 2.62	3.37 ± 2.96
9	−80.85 ± 7.91	−82.75 ± 17.88	−78.71 ± 8.01	−81.68 ± 18.59	−77.43 ± 8.05	−80.97 ± 18.87	−76.76 ± 8.10	−80.66 ± 19.08
10	−153.47 ± 9.68	−162.28 ± 18.23	−157.04 ± 9.03	−165.15 ± 18.46	−158.82 ± 8.98	−166.96 ± 18.70	−160.70 ± 8.65	−168.55 ± 18.70
11	−119.99 ± 11.23	−125.63 ± 20.97	−121.35 ± 11.27	−127.20 ± 21.25	−122.45 ± 11.21	−128.42 ± 21.32	−123.61 ± 11.13	−129.68 ± 21.39
12	16.84 ± 3.49	19.32 ± 4.65	17.60 ± 3.38	19.79 ± 4.59	17.98 ± 3.32	20.01 ± 4.52	18.28 ± 3.24	20.14 ± 4.52
13	0.33 ± 0.46	0.21 ± 0.68	0.34 ± 0.47	0.25 ± 0.65	0.37 ± 0.47	0.27 ± 0.62	0.39 ± 0.47	0.31 ± 0.63
14	2.87 ± 0.71	2.84 ± 2.22	2.99 ± 0.74	2.88 ± 2.11	3.04 ± 0.76	2.89 ± 1.97	3.09 ± 0.76	2.96 ± 2.18
15	28.08 ± 3.58	32.03 ± 7.36	28.78 ± 3.70	32.23 ± 7.34	29.03 ± 3.57	32.55 ± 7.33	29.32 ± 3.47	32.72 ± 7.18
16	1.26 ± 0.52	5.55 ± 4.11	0.96 ± 0.54	4.70 ± 4.00	0.79 ± 0.43	4.05 ± 3.80	0.65 ± 0.30	3.46 ± 3.36
17	17.25 ± 2.41	19.57 ± 5.51	17.20 ± 2.33	19.37 ± 5.35	17.13 ± 2.31	19.31 ± 5.33	17.10 ± 2.30	19.26 ± 5.35
18	4.42 ± 0.83	4.94 ± 1.72	4.39 ± 0.79	4.95 ± 1.66	4.41 ± 0.75	4.99 ± 1.61	4.41 ± 0.73	5.03 ± 1.58
19	−0.70 ± 0.77	−0.09 ± 0.87	−0.51 ± 0.74	−0.09 ± 0.84	−0.42 ± 0.69	−0.07 ± 0.78	−0.27 ± 0.63	−0.03 ± 0.80
20	6.10 ± 2.06	4.58 ± 2.19	5.90 ± 1.86	4.65 ± 2.01	5.55 ± 1.55	4.47 ± 1.68	5.08 ± 1.30	4.48 ± 1.99

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 (Feature 3–8, Feature 9–14, and Feature 15–20, totaling 18 features). Twenty features are extracted from every infant cry signals and used as input for the classifiers to distinguish the cry signals between normal and asphyxia cries.

Table 2 shows the statistics of extracted features (mean ± std) for the different frame length. From the Table 2, it is observed that the features extracted from normal and pathological cry signals are almost distinguishable.

5. Dimensionality reduction using Principal Component Analysis (PCA)

Principal Component Analysis is a type of projection method and it is used to reduce the redundant features in the feature matrix. PCA maximizes the variance of the projected vectors (Jolliffe, 1986; Lindsay, 2002). Feature extraction or feature selection is a key step of any pattern recognition process. High dimensional features affect the performance of the classifier and also increase the computational complexity. Feature selection is a process whereby a high dimensional data space is transformed into a feature space, which has exactly the same dimension as the original data. From the transformed feature space, non-redundant features are selected based on the Eigen value criterion, where the number of features corresponding to Eigen values more than 1.0 (Jolliffe, 1986; Lindsay, 2002). In this work, 13 features are identified from the transformed feature space, which are corresponding to Eigen values more than 1.0.

6. 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, two types of radial basis neural network structures are selected such as Probabilistic Neural Network and General Regression Neural Network and they are used for the classification of normal and pathological cries since they were successfully applied in different applications (Bowden, Bixon, Dandy, Maier, & Holmes, 2006; Erkmen & Yildirim, 2008; Feng, Chu, & Song, 2004; Firat & Gungor, 2009; Hariharan, Paulraj, & Yaacob, 2011b; Hariharan, Sin Chee, & Yaacob, 2010; Hariharan, Yaacob, & Awang, 2011a; Hariharan et al., 2011c; Leung, Chen, & Daouk, 2000; Polat & Yildirim, 2008). PNN and GRNN have similar architectures. The target variable is categorical for PNN classifier whereas for the GRNN network the target variable is continuous. Radial basis function networks that 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 (Kumar, 2004). 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.

6.1. Probabilistic Neural Network

Specht has proposed the probabilistic neural net based on Bayesian classification and classical estimators for probability density function (Fausett, 1994; Specht, 1990). PNN comprises of four units, such as input units, pattern units, summation units and output units. All the units are fully interconnected and the pattern units are activated by exponential function, instead of sigmoidal activation function. The pattern unit computes distances from the input vector to the training input vectors, when an input is presented, and produces a vector whose elements indicate how close the input is to a training input. The summation unit sums these contributions for each class of inputs and produces a net output which is a vector of probabilities. From the maximum of these probabilities, output units produce a 1 for that class and a 0 for the other classes using compete transfer function. Consider the two class problem, namely Class *A* and Class *B*. PNN uses the following estimator for the probability density function as given by Eq. (3)

$$f_A(x) = \frac{1}{(2\pi)^{n/2}\sigma^n}\frac{1}{m_a}\sum_{i=1}^{m_A}\exp\left[-\frac{(x-x_{Ai})^T(x-x_{Ai})}{2\sigma^2}\right] \quad (3)$$

where x_{Ai} is the ith training pattern from Class A, n is the dimension of the input vectors, m_A is the number of training patterns in Class A, and σ is a smoothing parameter corresponding to the standard deviation of the Gaussian distribution.

The net can be used for classification as soon as an example of a pattern from each of the two classes has been presented to it. However, PNN generalizes well as it is trained with more examples. Varying smoothing parameter (σ) gives control over the degree of nonlinearity of the decision boundaries for the net. A decision boundary approaches a hyperplane for large values of σ and approximates the highly nonlinear decision surface of the nearest neighbour classifier for small values of σ that are close to zero. In this paper, PNN architecture is constructed using *newpnn*() in MATLAB function (Matlab, version 7.0, 2004). The detailed information about the PNN architecture and mathematical equations can be found in the Specht's paper (Specht, 1990). The performance of the PNN classifier highly depends upon the smoothing parameter or spread factor (σ). Based on the experimental investigations, the σ value is varied between 0.04 and 0.085 in steps of 0.005.

6.2. General Regression Neural Network

Specht has proposed the model of GRNN to perform general (linear or nonlinear) regressions (Specht, 1991). GRNN is based on the theory of probability regression analysis. It usually uses Parzen window estimates to set up the 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 the Eq. (4) (Specht, 1991)

$$E(y|X) = \frac{\int_{-\infty}^{\infty} yf(X,y)dy}{\infty\int_{-\infty}^{\infty} f(X,y)dy} \quad (4)$$

where y = 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.

The estimated value Y is an exponentially weighted average value of all observed values Y^i given as in Eq. (5) (Specht, 1991):

$$\widehat{Y}(x) = \frac{\sum_{i=1}^{n} Y^i exp(-\frac{D_i^2}{2\sigma^2})}{\sum_{i=1}^{n} exp(-\frac{D_i^2}{2\sigma^2})} \quad (5)$$

where D_i is defined as in Eq. (6)

$$D_i^2 = (X - X^i)^T * (X - X^i) \quad (6)$$

The variable σ is a smoothing parameter that can be made large to smooth out noisy data or small to allow the estimated 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 (Matlab, version 7.0, 2004). The detailed information about the GRNN architecture and mathematical equations can be found in the Specht's paper (Specht, 1991). Based on the experimental investigations, the σ value is varied between 0.04 and 0.085 in steps of 0.005.

6.3. 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 (Garcia & Reyes García, 2003a, 2003b). The hidden and output neurons are activated by binary sigmoidal activation function. In this work, MLP architecture is constructed using *newff*() in MATLAB function (Matlab, version 7.0, 2004). 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.

6.4. Time delay neural network

Time delay neural network has been used in speech recognition applications (Hampshire & Waibel, 1990; Waibel, Hanazawa, Hinton, Shikano, & Lang, 1989) as well as in the infant cry classification (Reyes Galaviz & Reyes Garcia, 2004; Reyes-Galaviz et al., 2005). It was proposed to use in infant cry classification since the cry data are not static and are time dependent on crying patterns (Reyes Galaviz & Reyes Garcia, 2004; Reyes-Galaviz et al., 2005). The detailed information about the TDNN can be found in Hampshire and Waibel (1990), Reyes Galaviz and Reyes Garcia (2004) Reyes-Galaviz et al. (2005) and Waibel et al. (1989). A TDNN model is developed and training by scaled conjugate gradient algorithm. It consists of 20 neurons and the input delay specified by user, in this case the delay (Reyes Galaviz & Reyes Garcia, 2004) 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 (Garcia & Reyes García, 2003a, 2003b). The hidden and output neurons are activated by binary sigmoidal activation function. In this work, TDNN architecture is constructed using *newfftd*() in MATLAB function (Matlab, version 7.0, 2004). 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.

7. Results and discussion

In this work, two validation schemes (Conventional Validation-ConV and 10-fold cross validation (Kohavi, 1995)) 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 Conventional Validation scheme (one training set and one testing set), 680 segments (340 asphyxia + 340 normal) are used. Out of the features extracted from 680 segments, the features extracted from randomly selected 408 segments (60%) are used for training and the features extracted from remaining 272 segments (40%) are used for testing. All the networks are trained and tested with 20 original features and also 13 reduced features which are found by using PCA analysis. Results for the MLP, TDNN, PNN and GRNN classifier using 10-fold cross validation scheme and

Conventional Validation scheme with 20 original features and also with reduced features (13) found by using PCA analysis are tabulated in Tables 3 and 4. Average and standard deviation of the classification accuracies of normal and pathological infant cry signals (asphyxia) are tabulated. The standard deviation of the classification clearly reveals the consistency of the classifier results. If the standard deviation is higher, the classification results are inconsistent and also it reveals that the learning parameters of the classifiers affects the performance of the classifiers.

PNN and GRNN are trained with different spread factor or smoothing factor between 0.04 and 0.085 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 maximum classification accuracy was highlighted in the Tables 3 and 4 for every frame length. From the Table 3, the best overall accuracy of 96.11 ± 1.50% (20 ms and 16 hidden neurons), 96.47 ± 1.18% (30 ms and 14 hidden neurons), 96.54 ± 1.32% (40 ms and 12 hidden neurons), 96.99 ± 1.12% (50 ms and 12 hidden neurons) are obtained using MLP classifier with 20 original features (Conventional Validation). The best overall accuracy of 95.81 ± 0.50% (20 ms and 10 hidden neurons), 95.04±3.02% (30 ms and 10 hidden neurons), 94.93 ± 1.65% (40 ms and 10 hidden neurons), 96.47 ± 1.26% (50 ms and 10 hidden neurons) are obtained using MLP classifier with 13 reduced features (Conventional Validation).

From the Table 3, the best overall accuracy of 96.88 ± 0.51% (20 ms and 10 hidden neurons), 97.03 ± 0.32% (30 ms and 10 hidden neurons), 96.85 ± 0.55% (40 ms and 12 hidden neurons), 97.47 ± 0.79% (50 ms and 10 hidden neurons) are obtained using MLP classifier with 20 original features (10-fold cross validation). The best overall accuracy of 95.85 ± 0.70% (20 ms and 10 hidden neurons), 95.96 ± 0.45% (30 ms and 10 hidden neurons), 96.12 ±

Table 3
Results of MLP and TDNN classifier trained by Scaled conjugate gradient algorithm for the frame length 20 ms, 30 ms, 40 ms and 50 ms (Conventional Validation and 10−fold cross validation).

Frame length	Validation method	Type of classifier	Hidden neurons					
			10	12	14	16	18	20
20 ms	ConV	MLP	95.40 ± 1.39	96.03 ± 1.37	95.22 ± 1.11	**96.11 ± 1.50**	94.49 ± 2.21	95.37 ± 1.58
		MLP + PCA13	**95.81 ± 0.50**	94.45 ± 1.53	93.49 ± 1.18	93.42 ± 2.64	93.24 ± 1.79	94.52 ± 2.10
		TDNN	**96.40 ± 0.69**	95.44 ± 0.85	96.07 ± 1.27	95.11 ± 1.48	95.55 ± 1.65	96.32 ± 0.99
		TDNN + PCA13	93.97 ± 2.10	93.71 ± 1.81	93.49 ± 1.44	94.89 ± 2.01	**95.15 ± 0.96**	94.34 ± 1.79
	CrossV	MLP	**96.88 ± 0.51**	96.65 ± 0.61	96.37 ± 0.78	96.24 ± 0.39	96.35 ± 0.29	96.18 ± 0.48
		MLP + PCA13	**95.85 ± 0.70**	95.51 ± 0.51	95.75 ± 0.53	95.72 ± 0.57	95.35 ± 0.71	94.94 ± 0.52
		TDNN	96.50 ± 0.65	96.20 ± 0.22	96.56 ± 0.30	96.47 ± 0.10	**96.56 ± 0.25**	96.47 ± 0.44
		TDNN + PCA13	**96.35 ± 0.49**	96.29 ± 0.48	95.88 ± 0.78	95.82 ± 0.68	95.50 ± 0.47	95.71 ± 0.75
30 ms	ConV	MLP	96.14 ± 1.26	96.07 ± 1.30	**96.47 ± 1.18**	95.74 ± 1.26	96.29 ± 1.23	95.00 ± 1.19
		MLP + PCA13	**95.04 ± 3.02**	94.45 ± 1.39	94.63 ± 1.68	93.53 ± 2.50	93.64 ± 1.83	93.93 ± 1.63
		TDNN	96.40 ± 0.96	**96.54 ± 1.44**	96.29 ± 0.94	96.43 ± 1.04	95.74 ± 1.14	95.70 ± 1.14
		TDNN + PCA13	95.18 ± 1.59	94.93 ± 1.45	93.90 ± 1.63	**95.59 ± 1.51**	94.38 ± 1.23	94.23 ± 1.46
	CrossV	MLP	**97.03 ± 0.32**	96.47 ± 0.50	96.62 ± 0.45	96.47 ± 0.63	96.59 ± 0.43	96.62 ± 0.28
		MLP + PCA13	**95.96 ± 0.45**	95.76 ± 0.59	95.78 ± 0.67	95.38 ± 0.51	95.72 ± 0.50	95.38 ± 0.57
		TDNN	96.82 ± 0.47	96.59 ± 0.22	**96.88 ± 0.39**	96.44 ± 0.28	96.73 ± 0.38	96.32 ± 0.28
		TDNN + PCA13	95.71 ± 0.69	**96.26 ± 0.71**	95.88 ± 0.36	96.03 ± 0.74	95.59 ± 0.89	95.94 ± 0.66
40 ms	ConV	MLP	96.18 ± 1.11	**96.54 ± 1.32**	95.74 ± 1.29	94.82 ± 1.36	95.44 ± 1.37	95.99 ± 1.57
		MLP + PCA13	**94.93 ± 1.65**	94.63 ± 1.92	94.34 ± 1.65	94.15 ± 1.47	94.56 ± 2.37	93.38 ± 1.20
		TDNN	95.77 ± 1.38	96.10 ± 1.33	**96.99 ± 1.37**	94.38 ± 1.34	95.51 ± 1.53	95.77 ± 1.29
		TDNN + PCA13	95.00 ± 1.56	94.82 ± 1.63	**95.37 ± 0.72**	94.82 ± 1.54	94.15 ± 2.90	94.23 ± 1.35
	CrossV	MLP	96.53 ± 0.44	**96.85 ± 0.55**	96.65 ± 0.39	96.71 ± 0.30	96.76 ± 0.42	96.62 ± 0.21
		MLP + PCA13	96.60 ± 0.55	95.87 ± 0.44	95.94 ± 0.63	**96.12 ± 0.55**	96.12 ± 0.89	95.94 ± 0.33
		TDNN	96.47 ± 0.50	96.65 ± 0.28	**97.03 ± 0.45**	96.79 ± 0.42	96.68 ± 0.34	96.50 ± 0.38
		TDNN + PCA13	**96.53 ± 0.68**	96.50 ± 0.55	96.41 ± 0.54	96.26 ± 0.25	95.79 ± 0.52	96.24 ± 0.42
50 ms	ConV	MLP	95.99 ± 0.74	**96.99 ± 1.12**	96.84 ± 1.43	96.43 ± 1.18	96.58 ± 1.87	96.25 ± 1.04
		MLP + PCA13	**96.47 ± 1.26**	95.99 ± 1.42	95.96 ± 1.92	95.66 ± 1.57	95.81 ± 0.63	94.63 ± 2.57
		TDNN	96.76 ± 1.11	96.21 ± 1.50	**97.06 ± 1.46**	96.69 ± 1.40	96.77 ± 1.21	96.95 ± 1.39
		TDNN + PCA13	96.03 ± 1.80	96.21 ± 1.48	96.32 ± 1.72	96.07 ± 0.95	**97.24 ± 1.11**	95.85 ± 1.21
	CrossV	MLP	**97.47 ± 0.79**	97.26 ± 0.30	97.41 ± 0.56	97.47 ± 0.24	97.20 ± 0.56	97.32 ± 0.66
		MLP + PCA13	97.04 ± 0.37	**97.14 ± 0.44**	96.84 ± 0.36	96.79 ± 0.39	96.84 ± 0.47	96.47 ± 0.42
		TDNN	97.35 ± 0.36	97.38 ± 0.30	97.44 ± 0.45	97.05 ± 0.33	**97.76 ± 0.28**	97.41 ± 0.22
		TDNN + PCA13	97.18 ± 0.38	97.15 ± 0.73	97.26 ± 0.20	**97.35 ± 0.21**	97.15 ± 0.52	96.85 ± 0.38

Table 4
Results of PNN and GRNN Classifiers (Conventional Validation + 10-fold Cross Validation).

Validation method	Type of classifier	Frame length			
		20 ms	30 ms	40 ms	50 ms
ConV	PNN	**98.87 ± 0.15**	98.72 ± 0.11	98.32 ± 0.20	98.41 ± 0.15
	PNN + PCA13	**98.88 ± 0.19**	98.73 ± 0.24	98.18 ± 0.18	98.46 ± 0.23
	GRNN	97.78 ± 0.25	97.79 ± 0.42	98.17 ± 0.27	**98.57 ± 0.28**
	GRNN + PCA13	**97.15 ± 1.10**	96.81 ± 0.84	96.25 ± 0.92	95.81 ± 1.13
CrossV	PNN	98.79 ± 0.12	98.53 ± 0.14	99.16 ± 0.07	**99.18 ± 0.14**
	PNN + PCA13	**99.19 ± 0.17**	99.01 ± 0.25	98.57 ± 0.14	98.72 ± 0.12
	GRNN	98.13 ± 0.35	98.07 ± 0.26	**98.28 ± 0.31**	98.74 ± 0.29
	GRNN + PCA13	**97.75 ± 1.19**	97.25 ± 0.73	96.56 ± 1.28	96.23 ± 1.32

0.55% (40 ms and 16 hidden neurons), 97.14 ± 0.44% (50 ms and 12 hidden neurons) are obtained using MLP classifier with 13 reduced features (10-fold cross validation).

From the Table 3, the best overall accuracy of 96.40 ± 0.69% (20 ms and 10 hidden neurons), 96.54 ± 1.44% (30 ms and 12 hidden neurons), 96.99 ± 1.37% (40 ms and 14 hidden neurons), 97.06 ± 1.46% (50 ms and 14 hidden neurons) are obtained using TDNN classifier with 20 original features (Conventional Validation). The best overall accuracy of 95.15 ± 0.96% (20 ms and 18 hidden neurons), 95.59 ± 1.51% (30 ms and 16 hidden neurons), 95.37 ± 0.72% (40 ms and 14 hidden neurons), 97.24 ± 1.11% (50 ms and 10 hidden neurons) are obtained using TDNN classifier with 13 reduced features (Conventional Validation). From the Table 3, the best overall accuracy of 96.56 ± 0.25% (20 ms and 18 hidden neurons), 96.88 ± 0.39% (30 ms and 14 hidden neurons), 97.03 ± 0.45% (40 ms and 14 hidden neurons), 97.76 ± 0.28% (50 ms and 18 hidden neurons) are obtained using TDNN classifier with 20 original features (10-fold cross validation). The best overall accuracy of 96.35 ± 0.49% (20 ms and 10 hidden neurons), 96.26 ± 0.71% (30 ms and 12 hidden neurons), 96.53 ± 0.68% (40 ms and 10 hidden neurons), 97.35 ± 0.21% (50 ms and 12 hidden neurons) are obtained using TDNN classifier with 13 reduced features (10-fold cross validation).

From the Table 4, the best overall accuracy of 98.87 ± 0.15% (20 ms) is obtained using PNN classifier with 20 original features and an overall accuracy of 98.88 ± 0.19 (20 ms) is obtained with reduced (13) features (Conventional Validation). GRNN classifier gives an overall accuracy of 98.57 ± 0.28% with 20 original features and an overall accuracy of 97.15 ± 1.10% with reduced features (Conventional Validation). Using the 10-fold cross validation scheme, the best overall accuracy of 99.18 ± 0.14% and 99.19 ± 0.17% are obtained using PNN classifier with 20 original features and reduced features respectively. Similarly, GRNN gives an overall accuracy of 98.28 ± 0.31% and 97.75 ± 1.19% using 20 original features and reduced features respectively. In all the classifiers, there are no specific changes in the classification accuracies due to the different frame length.

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 asphyxia 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. We have obtained the classification accuracy of above 98% with only twenty time-frequency analysis based statistical features and radial basis neural networks (PNN and GRNN). It shows that the suggested features and PNN and GRNN classifier provides closer results with the earlier works. 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.

8. 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. Radial basis neural networks (PNN and GRNN) are employed to classify the cry signals into normal or pathological (Asphyxia). 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. Conventional Validation and 10-fold cross validation are performed, in order to test the generalizability and reliability of the PNN, GRNN, MLP and TDNN classifier. The suggested method provides maximum classification accuracy of 99% (PNN), 97% (TDNN), and 97% (MLP) using 10-fold cross validation scheme. Using the Conventional Validation scheme, the maximum classification accuracy of 98% (PNN and GRNN), 97% (TDNN), and 96% (MLP) are obtained. From the results, it can be inferred that the PNN and 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, different feature reduction techniques will be implemented to propose the reduced feature set with predominant features. The proposed method will be validated with larger samples.

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

Barajas-Montiel, S. E., & Reyes-García, C. A. (2006). Fuzzy support vector machines for automatic infant cry recognition. *Springer, Lecture Notes In Control And Information Sciences, 345*, 876–881.

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

Cano, S., Suaste-Rivas, I., Escobedo, D., Reyes-Garcia, C. A., & Ekkel, T. (2006). A combined classifier of cry units with new acoustic attributes. *Lecture notes in computer sciences, 4225*, 416–425.

Cano. S. et al. (1995). The spectral analysis of infant cry: An initial approximation. In *Paper presented at conference of EUROSPEECH'95 (sponsored by ESCA & IEEE), Madrid*.

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

Escobedo, D., Cano, S., Coello, E., Regueiferos, L., & Capdevila, L. (2001). Rising shift of pitch frequency in the infant cry pf some pathologic cases. In *Paper presented at the 2nd International Conference, MAVEBA 2001, Firenze, Italy, 2001*.

Fausett, L. (1994). *Fundamentals of neural network*. New Jersey, USA: Prentice Hall.

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

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

Garcia, J. O., & Reyes García, C.A. (2003a). Detecting pathologies from infant cry applying scaled conjugate gradient neural networks. In *Paper presented at the European Symposium on Artificial Neural Networks, Bruges (Belgium), 23–25 April 2003*.

Garcia, J. O., & Reyes García, C. A. (2003b). 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*, 615–622.

Hampshire, J. B., & Waibel, A. H. (1990). A novel objective function for improved phoneme recognition using time-delay neural network. *Neural Network, 1*(2), 216–228.

Hariharan, M., Sin Chee, L., & S., Yaacob (2010). Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network. *Journal of Medical Systems*, 1–7.

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

Hariharan, M., Paulraj, M. P., & Yaacob, S. (2011b). Detection of vocal fold paralysis and oedema using time-domain features and probabilistic neural network. *International Journal of Biomedical Engineering and Technology, 6*(1), 46–57.

Hariharan, M., Sindhu, R., & Yaacob, S. (2011c). Normal and hypoacoustic infant cry signal using time-frequency analysis and general regression neural network. *Computer Method and Programs in Biomedicine*, 1–11.

Jolliffe, I. T. (1986). *Principal component analysis*. Springer-Verlag.
Semmlow, J. L. (2004). *Biosignal and biomedical image processing*. Marcel Dekker Inc.
Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Paper presented at the14th International Joint Conference on Artificial Intelligence, Montreal, Quebec, Canada, 1995*.
Kumar, S. (2004). *Neural networks: A classroom approach*. New Delhi: Tata McGraw-Hill.
Leung, M. T., Chen, A. S., & Daouk, H. (2000). Forecasting exchange rates using general regression neural network. *Computers and Operation Research, 27*, 1093–1110.
Lindsay I. S. (2002). A tutorial on principal component analysis.
Manfredi, C., Tocchioni, V., & Bocchi, L. (2006). A robust tool for newborn infant cry analysis. In *Paper presented at the international Conference of the IEEE Engineering in Medicine and Biology Society, 2006*.
Matlab® Documentation., Version 7.0, Release 14, The Math-Works, Inc., 2004.
Petroni, M., Malowany, A., Johnston, C., & Stevens, B. (1995). 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*, 729–738.
Polat, O., & Yildirim, T. (2008). Hand geometry identification without feature extraction by general regression neural network. *Expert Systems with Applications, 34*, 845–849.
Rabiner, L., & Juang, B. (1993). *Fundamentals of speech recognition*. Prentice hall.
Reyes Galaviz, O. F., Arch Tirado, E., & Reyes Garcia, C. A. (2004). Classification of infant crying to identify pathologies in recently born babies with ANFIS. *LNCS, 3118*, 408–415.
Reyes Galaviz, O. F., & Reyes Garcia, C. A. (2004). Infant cry classification to identify hypoacoustics and asphyxia with neural networks. MICAI 2004. *LNAI, 2972*, 69–78.
Reyes-Galaviz, O. F., & Reyes-Garcia, C. (2004). A system for the processing of infant cry to recognize pathologies in recently born babies with neural networks. In *Paper presented at 9th Conference Speech and Computer (SPECOM'2004), St. Petersburg, Russia, September 20–22, 2004*.
Reyes-Galaviz, O. F., Verduzco, A., Arch-Tirado, E., & Reyes-García, C. A. (2005). Analysis of an infant cry recognizer for the early identification of pathologies. *Nonlinear Speech Modeling and Applications, 3445*, 404–409.
Reyes-Galaviz, O. F., Cano-Ortiz, S., Reyes-Garcia, C., y Electronica, O., & Puebla, M. (2009). Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies. In *Paper presented at the 8th Mexican international conference on artificial intelligent, MICAI 2009, Guanajuato, Mexico*.
Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks, 3*(1), 109–118.
Specht, D. F. (1991). A general regression neural network. *Neural Networks, 2*(6), 568–576.
Várallyay, G., Jr., Benyó, Z., Illényi, A., Farkas, Z., & Kovács, L. (2004). Acoustic analysis of the infant cry: classical and new methods. In *Paper presented at the 26th annual international conference of the IEEE EMBS, San Francisco, CA, USA*.
Várallyay, G., Jr. (2007). The melody of crying. In *Paper presented at the international journal of pediatric otorhinolaryngology, 2007*.
Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. J. (1989). Phoneme recognition using time delay neural network. *IEE Trasactions on Acoustics, Speech, and Signal Processing, 37*(3), 328–339.
Wasz-Hockert, O. et al. (1968). *The infant cry: A spectrographic and auditory analysis*. William Heinemann Medical Books Ltd.