2013 IEEE International Conference on Control System, Computing and Engineering, 29 Nov. - 1 Dec. 2013, Penang, Malaysia

Infant Cry Classification: Time Frequency Analysis

J. Saraswathy, M. Hariharan, Wan Khairunizam, Sazali Yaacob
School of Mechatronic Engineering
Universiti Malaysia Perlis (UniMAP)
Perlis, Malaysia

N. Thiyagar
Head of Department, Consultant Pediatrician & Adolescent Medicine Specialist, Department of Pediatrics
Hospital Sultanah Bahiyah
Alor Setar, Kedah

***Abstract*—Acoustic analysis of infant cry has been the subject of a number of researchers since half decades ago. This paper addresses a simple time-frequency analysis based signal processing technique using short-time Fourier transform (STFT) for the investigation and classification of infant cry signals. A cluster of statistical features are derived from the time-frequency plots of infant cry signals. The extracted feature vectors are used to model and train two types of radial basis neural network namely Probabilistic Neural Network (PNN) and General Regression Neural Network (GRNN) in classification phases. Three classes of infant cry signals are considered such as normal cry signals cry signals from deaf infants and infants with asphyxia. Promising classification results above 99% reveals that the proposed features and classification technique can effectively classify different infant cries.**

***Key words*—Acoustic analysis, infant cry, signal processing, classification.**

I. INTRODUCTION

Generally infants are incapable of expressing their physical, physiological and psychological status to their surrounding environment as they cannot communicate with words [1]. Cry is the most powerful and primary form of communication for babies. It has been proposed in the pediatric literature that the infant cry is a reflection of complex neurophysiologic functions and analysis of the cry itself can be used to identify the infant's physical and psychological status [2]. Detection of pathological status of babies using the conventional methods is strenuous, requires good expertise and intangible. Therefore computer based analytical tools can be very useful for in depth study and rapid classification of different cry signals. This non-invasive method has been widely used in infant cry analysis area and has shown very encouraging results [3-8].

This paper presents the development of an automatic classification system for classifying normal and pathological cries (asphyxia and deaf) using STFT. Two different types of radial basis neural networks namely PNN and GRNN are implemented in classification stage to evaluate the efficacy of proposed time-frequency based statistical features. In order to test the effectiveness of proposed features and reliability of the classification results, 10-fold cross validation and conventional validation are used. The empirical results emphasize that the STFT combined with statistical features and radial basis neural network classifiers can be used in early diagnosis of infant's pathological status and may avoid opportune treatment and therapies.

The paper is organized as follows: Section II presents the details of the database used in this analysis. Section III deals with the introduction to the STFT and feature extraction of the statistical features also includes the statistical analysis of the extracted feature vectors. Classifiers used are explained in Section IV. Section V presents the classification accuracy of the experiments with the interpretation of the results. The proposed study compared with other literature studies related to the automatic detection of pathological status from cry signals in Section VI. The paper concluded in Section VII with some future directions.

II. INFANT CRY 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 reference [8]. 340 of normal cry signals and 340 of deaf cry signals and 340 of asphyxia cry signals are used for our analysis. Asphyxia is a type of respiratory disorder which can cause damage to the brain, organs and tissues or even death if subjected to delayed or improper treatments [9]. According to World Health Organization (WHO) deaf or hypo-acoustic is defined as the insufficiency of hearing ability that may retard the performance of child's learning and development stages, especially in school life if not subjected to early treatments [10]. The length of cry signals is 1 sec. The sampling frequency of infant cry signals is set to 8000 Hz for our analysis.

III. SHORT TIME FOURIER TRANSFORM (STFT)

Feature extraction plays an important role in the area of automatic detection of pathological infant cry signals. Short-time Fourier transform is one of the simple and typical feature extraction method based on time-frequency analysis. It has been widely applied in different applications due to its simplest and good frequency resolution [11-13]. In STFT analysis, the interest signal is first multiplied with a window function with or without overlap. Then the standard Fourier transform is applied to the windowed segments and this is what termed the "short-term Fourier transform" as defined in Eq. (1) [14-15]:

978-1-4799-1508-8/13/$31.00 ©2013 IEEE

$$STFT(f,t)=\int_{-\infty}^{\infty} s(\tau)Y(t-\tau)e^{-j2\pi f\tau}d\tau \quad (1)$$

where *Y (t)* is a short-time windowing function of duration T_Y and size L.

STFT spectrogram is the squared magnitude of STFT (as shown in Eq. 2) and it can be used for observing the temporal and spectral characteristics at any point of interest signals.

$$Spectrogram=\left[STFT(f,t)\right]^2 \quad (2)$$

Since the Fourier transform is a complex function, the power spectrum density (PSD) is used and is given by Eq. (3):

$$P_s[f,t]=\frac{1}{N}\left|STFT[f,t]\right|^2 \quad (3)$$

where N is the number of discrete frequencies (N≥L).

In this paper, the infant cry signals are segmented into different frame length of 20ms, 30ms, 40ms, and 50ms with 50% overlap using STFT technique [14]. The hamming window is chosen due to its superior ability to avoid the discontinuities during segmentation and taper the signal at the beginning and the end of each frame [15]. From the STFT–PSD of the cry signals (Fig.1), different standard time-frequency plots (TF) which can clearly displays the discrimination among the different types of cry signals namely time-frequency (TF), time-maximum amplitude (TMA), frequency-maximum amplitude (FMA) and frequency-standard deviation of amplitude (FSDA) plots are generated.

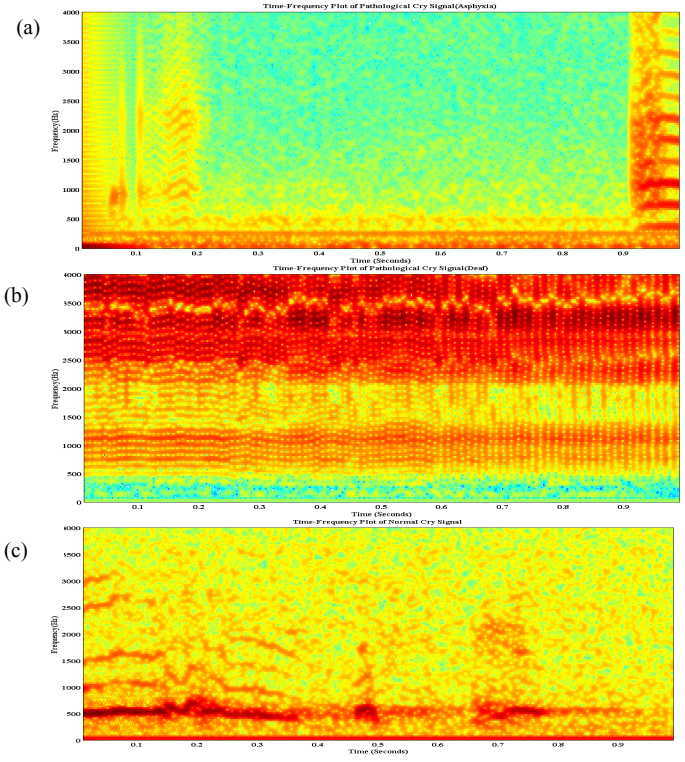

Fig. 1. Time frequency plots of infant cry signals (a) asphyxia cry, (b) deaf cry and (c) normal cry

By applying standard statistical techniques different sets of significant features (20 features) are extracted from the TF-plots as shown in Fig 2. The discriminatory ability of the statistical features is illustrated in scatter plots (Fig. 3). From the Fig. 3 it is observed that the features extracted from normal and pathological cries are almost distinguishable. In addition, the discriminatory ability of extracted features is further analyzed by performing one-way ANOVA test using SPSS software and found all of the features are statistically significant (p<0.001) as shown in Table 1.

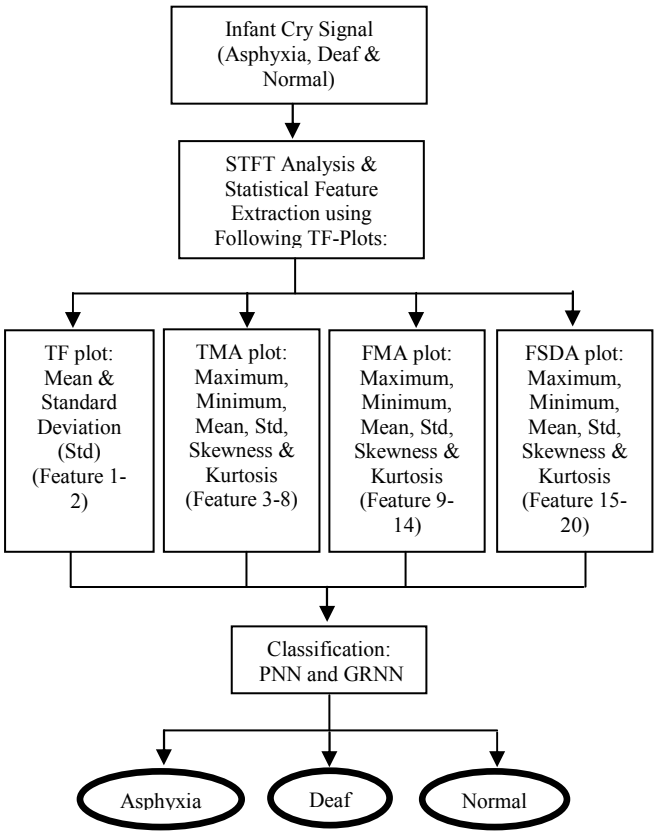

Fig. 2. Block diagram of the proposed study

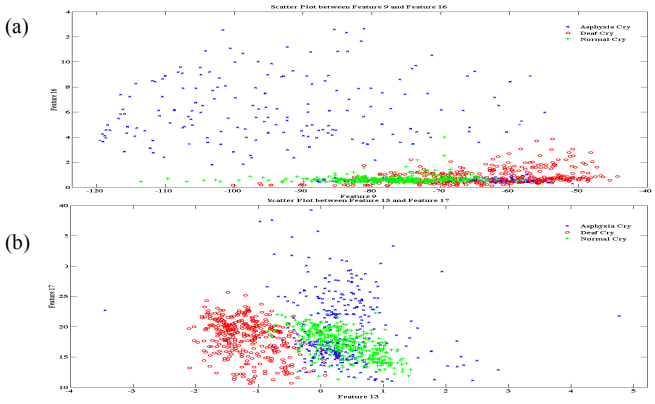

Fig. 3. Scatter plots (a) between feature 9 and feature 16 and (b) between feature 13 and feature 17

TABLE I. ANALYSIS OF VARIANCE (ANOVA) OF EXTRACTED STATISTICAL FEATURES USING SPSS

		Sum of Squares	df	Mean Square	F	*p*-value
Mean of TF	Between Groups	345.440	2	172.720	150.642	.000
	Within Groups	1166.056	1017	1.147		
	Total	1511.496	1019			
Std of TF	Between Groups	200120.561	2	100060.280	383.682	.000
	Within Groups	265223.081	1017	260.790		
	Total	465343.641	1019			
Maximum of TMA	Between Groups	52691.056	2	26345.528	146.026	.000
	Within Groups	183483.658	1017	180.417		
	Total	236174.714	1019			
Minimum of TMA	Between Groups	16061.095	2	8030.547	24.673	.000
	Within Groups	331012.358	1017	325.479		
	Total	347073.452	1019			
Mean of TMA	Between Groups	9452.573	2	4726.287	20.193	.000
	Within Groups	238037.843	1017	234.059		
	Total	247490.416	1019			
Std of TMA	Between Groups	13468.551	2	6734.275	216.123	.000
	Within Groups	31689.167	1017	31.159		
	Total	45157.717	1019			
Skewness of TMA	Between Groups	55.190	2	27.595	33.028	.000
	Within Groups	849.715	1017	.836		
	Total	904.905	1019			
Kurtosis of TMA	Between Groups	98.048	2	49.024	7.963	.000
	Within Groups	6261.009	1017	6.156		
	Total	6359.057	1019			
Maximum of FMA	Between Groups	52691.056	2	26345.528	146.026	.000
	Within Groups	183483.658	1017	180.417		
	Total	236174.714	1019			
Minimum of FMA	Between Groups	10532.755	2	5266.378	32.988	.000
	Within Groups	162357.079	1017	159.643		
	Total	172889.835	1019			
Mean of FMA	Between Groups	172960.177	2	86480.088	415.591	.000
	Within Groups	211626.776	1017	208.089		
	Total	384586.953	1019			
Std of FMA	Between Groups	2413.043	2	1206.521	73.179	.000
	Within Groups	16767.469	1017	16.487		
	Total	19180.511	1019			
Skewness of FMA	Between Groups	530.731	2	265.365	981.432	.000
	Within Groups	274.983	1017	.270		
	Total	805.714	1019			
Kurtosis of FMA	Between Groups	431.493	2	215.746	93.410	.000
	Within Groups	2348.940	1017	2.310		
	Total	2780.432	1019			
Maximum of FSDA	Between Groups	2257.275	2	1128.637	35.873	.000
	Within Groups	31996.766	1017	31.462		
	Total	34254.041	1019			
Minimum of FSDA	Between Groups	1652.263	2	826.131	210.641	.000
	Within Groups	3988.655	1017	3.922		
	Total	5640.918	1019			
Mean of FSDA	Between Groups	799.804	2	399.902	27.499	.000
	Within Groups	14789.766	1017	14.543		
	Total	15589.569	1019			
Std of FSDA	Between Groups	68.634	2	34.317	23.797	.000
	Within Groups	1466.607	1017	1.442		
	Total	1535.241	1019			
Skewness of FSDA	Between Groups	57.821	2	28.910	62.290	.000
	Within Groups	472.018	1017	.464		
	Total	529.839	1019			
Kurtosis of FSDA	Between Groups	162.323	2	81.162	28.896	.000
	Within Groups	2856.506	1017	2.809		
	Total	3018.829	1019			

TABLE II. CONFUSION MATRIX FOR THE CLASSIFICATION OF INFANT CRIES USING PNN

Class	Asphyxia	Deaf	Normal
Asphyxia	338/340	1	2
Deaf	1	337/340	1
Normal	1	2	337/340
Overall Accuracy: 1012/1020 = 99.22%			

TABLE III. CONFUSION MATRIX FOR THE CLASSIFICATION OF INFANT CRIES USING GRNN

Class	Asphyxia	Deaf	Normal
Asphyxia	338/340	2	4
Deaf	1	336/340	0
Normal	1	2	336/340
Overall Accuracy: 1010/1020 = 99.02%			

IV. CLASSIFICATION

In this work, two types of radial basis neural network structures namely PNN and GRNN are selected and used for the pattern identification due to their some advantages namely relatively insensitive to outliers, easy to train than MLP and TDNN and does not require an iterative training procedure [16-19]. Based on the experimental investigations, the smoothing parameter for PNN and GRNN is varied between 0.03 and 0.12 in steps of 0.01. The detailed information about these radial basis neural networks can be found in some literature papers [16-21].

In order to develop robust classifiers that are adequately generalized to perform well on extracted vectors with unknown class labels, two different validation schemes (10-fold cross validation and conventional validation) which can be used for estimating the classifier performance are performed [22-23]. In 10-fold cross validation scheme, the proposed feature vectors are divided randomly into 10 sets and training is repeated for 10 times. 1020 segments (340 asphyxia + 340 deaf cry + 340 normal cry) of cry signals are used. In conventional validation, 60% of data are used for training and remaining 40% are used for testing. From 1020 segments (340 asphyxia + 340 deaf cry + 340 normal cry) 612 segments are used for training and remaining 408 data are used for testing. In this work, the feature extraction and classification algorithms are developed in MATLAB function [24].

V. EMPIRICAL RESULTS

The extracted statistical features from normal and pathological cries were evaluated by two different types of neural networks (PNN & GRNN) which were trained with different smoothing parameters, through 10-fold cross validation and conventional validation schemes. This section presents the performance of the proposed features for different frame lengths and spread factors of classifiers.

The best individual accuracy of each class was presented in the form of confusion matrix as shown in Table I and Table II. From the Table I and II, it is perceived that the overall best accuracy was 99.22% using PNN and 99.02% using GRNN. From the Table III, the following results were observed. The best accuracy for frame length of 20ms was 98.73% (PNN, CrossV, 0.10), 98.77% (PNN, ConV, 0.07), 98.82% (GRNN, CrossV, 0.12) and 98.53% (GRNN, ConV, 0.10). For frame length 30ms, the best accuracy was 98.53% (PNN, CrossV, 0.09), 98.60% (PNN, ConV, 0.11), 98.73% (GRNN, CrossV, 0.09) and 98.77% (GRNN, ConV, 0.10).The best accuracy for frame length of 40ms was 98.73% (PNN, CrossV, 0.06), 98.38% (PNN, ConV, 0.04), 98.82% (GRNN, CrossV, 0.06) and 98.58% (GRNN, ConV, 0.04). Furthermore, the best accuracy for frame length of 50ms was 99.02% (PNN, CrossV, 0.10), 99.22% (PNN, ConV, 0.12), 98.82% (GRNN, CrossV, 0.12) and 99.02% (GRNN, ConV, 0.07).

TABLE IV. CLASSIFICATION ACCURACY OF PATHOLOGICAL CRY SIGNALS (ASPHYXIA AND DEAF) AND NORMAL CRY SIGNALS USING PNN AND GRNN FOR DIFFERENT FRAME LENGTHS.

Frame length	**Architecture**	**Validation Type**	**Spread factor**									
			0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12
20	PNN	CrossV	98.63	98.33	98.43	98.53	98.43	98.53	98.63	**98.73**	98.63	98.53
		ConV	98.18	98.31	98.38	98.14	**98.77**	97.73	98.55	98.50	98.09	98.23
	GRNN	CrossV	98.63	98.73	98.63	98.43	98.63	98.53	98.53	98.63	98.73	**98.82**
		ConV	94.02	97.45	97.94	97.79	97.97	98.16	98.31	**98.53**	98.06	98.24
30	PNN	CrossV	98.24	98.53	98.33	98.43	98.24	98.24	**98.53**	98.43	98.43	98.33
		ConV	98.41	98.26	98.09	98.26	98.28	98.33	98.50	98.50	**98.60**	98.38
	GRNN	CrossV	98.14	98.63	98.53	98.43	98.33	98.24	**98.73**	98.33	98.33	98.53
		ConV	92.65	96.59	97.50	97.87	98.41	98.24	98.50	**98.77**	98.16	98.48
40	PNN	CrossV	98.24	98.24	98.33	**98.73**	98.73	98.43	98.63	98.73	98.53	98.33
		ConV	98.04	**98.38**	98.36	98.21	98.31	98.09	98.19	98.36	98.16	98.26
	GRNN	CrossV	98.43	98.33	98.43	**98.82**	98.53	98.63	98.53	98.63	98.63	98.43
		ConV	92.92	96.79	97.94	97.67	98.41	**98.58**	98.30	98.31	98.36	97.94
50	PNN	CrossV	98.63	98.73	98.92	98.63	98.92	98.82	98.92	**99.02**	98.73	98.53
		ConV	98.71	98.58	98.50	98.50	98.46	98.68	98.70	98.97	98.60	**99.22**
	GRNN	CrossV	98.63	98.73	98.63	98.43	98.63	98.53	98.53	98.63	98.73	**98.82**
		ConV	94.85	97.45	98.26	98.36	**99.02**	98.53	98.85	98.63	98.80	98.55

From the empirical results, it were inferred that the best spread factor can lie between 0.03 and 0.12 to obtained maximum classification accuracy using PNN and GRNN and the 50ms can be the best frame length for infant cry signal analysis using STFT technique since the maximum accuracy up to 99% was yielded from that.

VI. DISCUSSIONS

In this section, we present a summary of the techniques used in the three class classification of infant cry signals in literature as shown in Table V. As seen in Table V, different signal processing and classification algorithms were proposed by authors for classifying three different cries (asphyxia, deaf and normal). A method based on Mel frequency cepstral coefficient (MFCC) and fuzzy-relational neural network (FRNN) developed with accuracy of 96.49% [25]. The MFCC features and Time delay neural network (TDNN) which trained with adaptive back propagation function used for classifying normal and pathological cries. The obtained accuracy was varied from 96.08% to 97.39% [3]. Fuzzy support vector machine (FSVM) and MFCC features applied and reported highest correct recognition rate of 94.98% [26]. A highest accuracy of 99% reported through conventional validation by using weighted linear prediction cepstral coefficient (WLPCC) and PNN [18].

In this present study, simple statistical features were derived through STFT based time-frequency analysis to provide robust representation of infant cry signals. The extracted features were classified using radial basis neural networks (PNN and GRNN) through cross validation and conventional validation schemes. The maximum classification accuracy of above 99% was obtained and it shows that the suggested features and classification algorithm provides greater results with the literature works as tabulated in Table V. From the above discussion, it has been inferred that the suggested time-frequency analysis based statistical features can be the most discriminating representation of normal, deaf and asphyxia cry signals. Finally, the experimental results demonstrate the proposed methods' superiority towards clinical implications.

TABLE V. A PERFORMANCE COMPARISON OF THE PROPOSED METHODOLOGY AND OTHER SIGNIFICANT INFANT CRY CLASSIFICATION RELATED STUDIES

Literature studies	Feature Extraction	Classifier	Accuracy (%)
[25] (2004)	MFCC	FRNN	98
[3] (2005)	MFCC	TDNN	96.08 - 97.39
[26] (2006)	MFCC	FSVM	94.98
[18] (2012)	WLPCC	PNN	99
Proposed methodology	Time frequency features	PNN and GRNN	99.22

VII. CONCLUSION

Infant cry being the first verbal communication tool also can be used as a reliable indicator of infant's early pathological status. This paper presents the analysis of newborn infant cry signals based on time frequency analysis (STFT) and radial basis classifiers (PNN and GRNN). A total of 20 statistical features were extracted from different time frequency plots and their discriminatory ability was validated through one way ANOVA test by fixing $p<0.001$. The frame length, smoothing parameters of PNN and GRNN were varied and their effects on the performance of classification results were presented. The maximum classification accuracy of above 99% was obtained by using PNN and GRNN classifiers. In future work, the reduced feature set with predominant features will be proposed with implemented different feature reduction techniques. We attempt to validate our results using larger databases and different pathological cry signals. We intended to improve the accuracy of this analysis by exploring other time frequency based analysis namely, Wigner Ville, Choi William and extra.

ACKNOWLEDGMENT

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.

REFERENCES

[1] O.F. Reyes Galaviz, E. Arch Tirado, and C.A. Reyes Garcia, "Classification of infant crying to identify pathologies in recently born babies with ANFIS", LNCS, vol. 3118, pp. 408-415, 2004.

[2] Hemant A. Patil, "Cry Baby: Using Spectrographic Analysis to Assess Neonatal Health Status", Chapter 14, Advances in Speech Recognition, Mobile Environments, 323 Call Centers and Clinics, 2010.

[3] O.F. Reyes-Galaviz, A.Verduzco, E.Arch-Tirado, and C.A Reyes-Garcia, "Analysis of an infant cry recognizer for the early identification of pathologies", Nonlinear Speech Modeling and Applications, vol. 3445, pp. 404-409, 2005.

[4] J. Orozco, and C. Garcia, "Detecting pathologies from infant cry applying scaled conjugate gradient neural networks", European Symposium on Artificial Neural Networks, Bruges (Belgium), 2003, pp. 349–354.

[5] A. Rosales-Perez, C.A. Reyes-Garcia, and P. Gomez-Gil, "Genetic fuzzy relational neural network for infant cry classification", LNCS, vol. 6718, pp. 288-296, 2011.

[6] M.A. Ruiz Diaz, C.A Reyes Garcia, L.C. Altamirano Robles, and J.E. Xaltena Altamirano, "Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis", Biomedical Signal Processing and Control, vol. 7, pp. 43-49, 2012.

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

[8] O.F. Reyes-Galaviz, S. Cano-Ortiz, and C. Reyes-Garca, "Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies", Proceedings of the 8th Mexican International Conference on Artificial

Intelligence, MICAI 2009, Guanajuato, Mexico, 2009, pp. 330–335.

[9] C.F.P. Office of Health and Nutrition, Health and Nutrition, Bureau for Global Programs Field Support and Research, U.S, "Detecting and treating newborn asphyxia, maternal neonatal & health", pp. 1-2, 2004. Accessed 18 June 2012. http://pdf.usaid.gov/pdf_docs/Pnacy993.pdf.

[10] M. Cunningham, and O. Edward, "Hearing assessment in infants and children: recommendations beyond neonatal screening", American Academy of Pediatrics, Clinical Report, vol.111, pp.436-440, 2003.

[11] J.C. Braun, D.F. Hanley, and N.V. Thakor, "Detection of neurological injury using time-frequency analysis of the somatosensory evoked potential", Electroencephalography and clinical Neurophysiology", vol.100, pp.310-318, 1996.

[12] Y. Hu, K.D. Luk, W.W. Lu, and J.C. Leong, "Comparison of time–frequency analysis techniques in intraoperative somatosensory evoked potential (SEP) monitoring", Computers in Biology and Medicine, vol. 32, pp. 13–23, 2002.

[13] L. Lu, and L. Jianhui, "The comparison of time-frequency analysis on the feature extraction of maglev' track long-wave irregularity", Proceedings of the Eighth International Conference on Electronic Measurement and Instruments, 2007, pp. 315-319.

[14] L. Rabiner, and B. Juang, "Fundamentals of speech recognition", Prentice hall, 1993.

[15] John L. Semmlow, "Biosignal and biomedical image processing", Marcel Dekker Inc, 2004.

[16] M. Hariharan, R. Sindhu, and Sazali Yaacob, "Normal and hypoacoustic infant cry signal classification using time-frequency analysis and general regression neural network", Computer Methods and Programs in Biomedicine, vol. 108, pp. 559-569, 2012.

[17] M. Hariharan, J. Saraswathy, R. Sindhu, Wan Khairunizam, and Sazali Yaacob, "Infant cry classification to identify asphyxia using time-frequency analysis and radial basis neural networks" Expert Systems with Applications, vol. 39, pp. 9515-9523, 2012.

[18] M. Hariharan, Lim Sin Chee, and Sazali Yaacob, "Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network", Journal of Medical Systems, vol. 36, pp. 1309-1315, 2012.

[19] M. Hariharan, Sazali Yaacob, and Saidatul Ardeenaawatie Awang, "Pathological infant cry analysis using wavelet packet transform and probabilistic neural network", Expert Systems with Applications, vol. 38, pp. 15377-15382, 2011.

[20] D.F. Specht, "Probabilistic neural networks. Neural Networks", vol. 3, pp. 109-118, 1990.

[21] D.F. Specht, "A general regression neural network. Neural Networks", vol. 2, pp. 568-576, 1991.

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

[23] T. Gandhi, B.K. Panigrahi, and S. Anand, "A comparative study of wavelet families for EEG signal classification", Neurocomputing, vol.74, pp. 3051-3057, 2011.

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

[25] Israel Suaste-Rivas, Orion F. Reyes-Galviz, Alejandro Diaz-Mendez, and Carlos A. Reyes Garcia, "Implementation of a linguistic fuzzy relational neural network for detecting pathologies by infant cry recognition" IBERAMIA 2004, LNAI, vol. 3315, pp. 953–962, 2004.

[26] Sandra E. Barajas-Montiel, and Carlos A. Reyes-Garcia, "Fuzzy support vector machines for automatic infant cry recognition", LNCIS, vol. 345, pp. 876 – 881, 2006. Springer-Verlag Berlin Heidelberg 2006.