Penerbit Akademia Baru

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 I Vol. 4, No.1. Pages 12-28, 2016

Automatic Infant Cry Classification Using Radial Basis Function Network

N. S. A. Wahid [*,1,a], *P. Saad* [1,b] *and M. Hariharan* [2,c]

[1]School of Computer and Communication Engineering, Universiti Malaysia Perlis, 02600 Arau Perlis, Malaysia
[2]School of Mechatronic Engineering, Universiti Malaysia Perlis, 02600 Arau Perlis, Malaysia
[a,*]*nsyahirah91@gmail.com*, [b]*putehsaad@unimap.edu.my*, [c]*hari@unimap.edu.my*

Abstract – *This paper proposes the automatic infant cry classification to analyse infant cry signals. The cry classification system consists of three stages: (1) feature extraction, (2) feature selection, and (3) pattern classification. We extract features such as Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC), and dynamic features to represent the acoustic characteristics of the cry signals. Due to the high dimensionality of data resulting from the feature extraction stage, we perform feature selection in order to reduce the data dimensionality by selecting only the relevant features. In this stage, five different feature selection techniques are experimented. In pattern classification stage, two Artificial Neural Network (ANN) architectures: Multilayer Perceptron (MLP) and Radial Basis Function Network (RBFN) are used for classifying the cry signals into binary classes. Experimental results show that the best classification accuracy of 99.42% is obtained with RBFN.* ***Copyright © 2016 Penerbit Akademia Baru - All rights reserved.***

Keywords: Infant cry analysis, Feature selection, Feature extraction, Spectral features

1.0 INTRODUCTION

Crying is a type of communication for infants to express their physical and emotional conditions. Crying process involves several brain sections such as limbic and brainstem systems and is connected to the respiratory system. The cry characteristics show the integrity and progress of the central nervous system [1]. Therefore, automatic infant cry classification which is a non-invasive process is suitable to assess the physical and emotional states of infants.

In early studies, auditory analysis and sound spectrographic analysis are used to analyse the cry signals. Several types of cries and pathologies have been detected from the infant cry signals using the conventional analyses such as hunger, pain, pleasure, asphyxia, hydrocephalus, hypoglycaemia, brain damage, encephalitis, encephalitis, hypothyroidism, down syndrome, oropharyngeal abnormalities, and genetic defects [2], [3]. However, these analyses required subjective evaluation from medical experts and the evaluation process is time consuming. Besides, they are unsuitable for a large infant cry samples due to time constraint. Hence, automatic infant cry classification had been proposed to overcome the limitations of the conventional analyses. The automatic classification enables the cry signals to be automatically classified into different types of cries and pathologies using suitable techniques.

Significant progress has been obtained in the development of automatic cry classification system. The cry classification system has been applied to identify different types of cries and pathologies such as hunger and pain cries [4], [5], asphyxia [6]–[8], deaf [9]–[11], autism [12], and cleft palate [13]. However, the automatic infant cry classification which is a pattern recognition problem, often deals with a large input data that consists of redundant and irrelevant features. The redundant features do not provide any new information regarding the underlying structure of the data and irrelevant features do not have any effect on the underlying structure [14]. This situation may decrease the classifier predictive performance and simultaneously having high computational processing time [15]. The simplest way to solve this problem is by selecting the relevant features and eliminates the rest. This process is known as feature selection and can be categorized into two main techniques: filter techniques and wrapper techniques. Filter techniques are independent of a classifier, whereas wrapper techniques apply the classification algorithm as part of function evaluation to search for the relevant feature subsets. In this paper, due to the high dimensional of data, we only focus on the filter techniques for feature selection as they provide fast processing time during the selection of relevance subset of features.

Thus, in this study, we compare different types of filter techniques for feature selection process in automatic infant cry classification system. Features such as Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC), and dynamic features are extracted. 10-fold cross validation is used to evaluate the effectiveness of the features applied and the reliability of the classification results. The experimental results show that the classification system achieved highest classification accuracy up to 99.42%.

2.0 MATERIALS AND METHOD

2.1 Database

The database used is known as Baby Chillanto database which is a property of the Instituto Nacional de Astrofisica Optica y Electronica (INAOE) – CONACYT, Mexico. The database is described in reference [16]. The infant cry samples were recorded directly by specialized physicians from just born up to 6 month old infants. The samples were labelled with information about the cause of cry during the recording process.

Table 1: Data sets description

Data set	Total no. of samples	No. of samples from each category
(i) Asphyxia vs. normal and hungry	847	Asphyxia: 340
		Normal and hungry: 507
(ii) Deaf vs. normal and hungry	1386	Deaf: 879
		Normal and hungry: 507
(iii) Hungry vs. pain	542	Hungry: 350
		Pain: 192

Table 1 shows the description of infant cry data sets used in this study. All the samples in the database have 1 second (s) length and the sampling frequency used is 8000 Hertz (Hz). The database consists of 340 samples from asphyxia cries, 192 samples of pain cries, 350 samples of hungry cries, 879 samples from deaf cries, and 157 samples of normal cries. Pain and hungry cry samples are obtained from normal infants; hence they are also categorized under the normal cries category. In this study, three data sets are developed to perform binary classification of (i) asphyxia vs. normal and hungry, (ii) deaf vs. normal and hungry, and (iii) hungry vs. pain.

2.2 Feature Extraction

The aim of feature extraction process is to extract the important characteristics from the cry signal and eliminates irrelevant information such as channel distortion, particular characteristics of the signal, and background noise. Thus, due to this reason, feature extraction was applied as a first stage in the cry classification system. Figure 1 shows the block diagram of the automatic infant cry classification. The input of this process is the cry signals and the output is the type of cry or pathology identified at the infant.

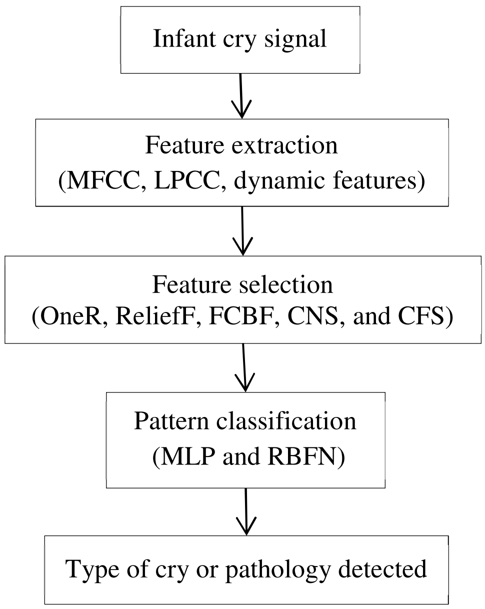

Figure 1: Block diagram of the automatic infant cry classification system

In this study, MFCC and LPCC features were extracted to represent the acoustic characteristics of the cry signals. The MFCC and LPCC which are the spectral features are widely applied in automatic speech recognition (ASR) field since the mid-eighties. In addition, MFCC and LPCC have been proven to be the appropriate representations of infant cry signals [16], [17].

Figure 2 illustrates the extraction process of MFCC and LPCC features. The first step in feature extraction is to pre-process the signal with a pre-emphasis filter. The purpose of this step is to flatten the spectrum of the signal and reduce the effect of finite precision in the signal processing steps later [18]. The infant cry signal is a non-stationary signal as it is constantly changing. Therefore, a short term analysis must be applied by blocking the signal into short frames usually with a duration of 10ms to 50ms [19]. Then, each frame was windowed by Hamming window to minimize the signal discontinuities. This process was done by tapering the signal to zero at the beginning and end part of each frame.

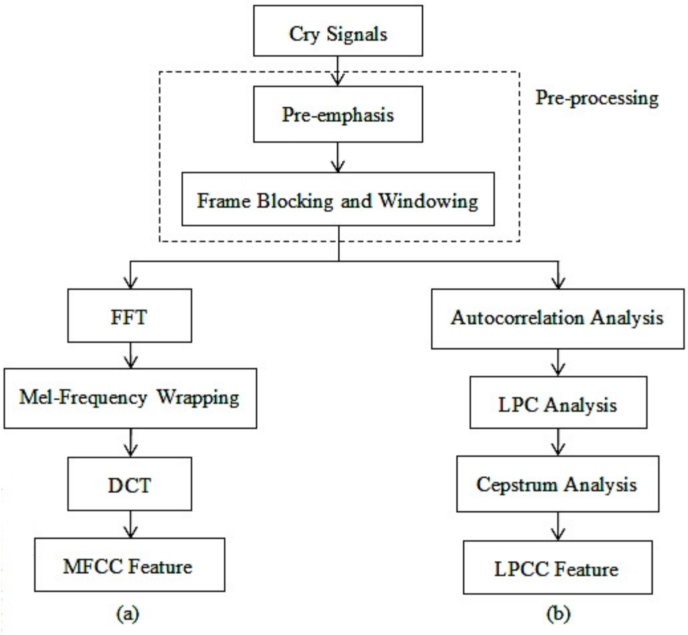

Figure 2: Block diagram of feature extraction process: (a) MFCC feature, (b) LPCC feature

Next, MFCC and LPCC features were extracted. The process for extracting the MFCC feature is illustrated in Figure 2(a). After the pre-processing step, the Fast Fourier Transform (FFT) was applied to the windowed signal. The aim of FFT is to convert the signal from time domain to frequency domain. The obtained values from the FFT step were then grouped and weighted by a set of triangular filters known as mel-spaced filterbanks. The first filter is very narrow and acts an indicator to calculate energy that exists near 0 Hertz (Hz). As the frequency increases, the following filters become wider and less concern about variations. This process is similar to human auditory system as it can detect the frequencies which are below than 1 kHz in linear scale and frequencies above 1 kHz in logarithmic scale. The formula for computing *mels* for a given frequency (f) in Hz is shown in equation (1).

$$mel(f) = 2952\ log_{10}(1 + f/700) \tag{1}$$

The last step is to convert the log mel spectrum back into time domain by using Discrete Cosine Transform (DCT). The cepstral representation of the cry spectrum gives a good representation of the local spectral characteristics of the signal for the given frame analysis. The output of this step is called MFCC which is an acoustic vector.

The process for extracting the LPCC feature is illustrated in Figure 2(b). After the pre-processing step, each windowed frame was auto correlated using equation (2) [20]:

$$s[n] = \sum_{k=1}^{p} a[k]s[n-k] \tag{2}$$

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 | Vol. 4, No.1. Pages 12-28, 2016

where $s[n]$ is the signal samples, $a[k]$ denotes as the linear predictor coefficients, and p is the order of the linear predictor. Next, the aim of Linear Prediction Coefficients (LPC) analysis is to convert the autocorrelation coefficients into LPC. This analysis was performed by using Levinson-Durbin recursive algorithm [21]. Finally, LPCC feature was derived from the LPC using a recursion technique [22].

In addition to the spectral features, we also extracted dynamic features. Dynamic features is the time derivatives of the spectrum-based features [23]. These features contain the dynamic characteristics of the spectral features. The first order derivatives, also known as Delta (Δ) features [19], can be calculated using equation (3) as follows [19]:

$$\Delta F(m) = \frac{\sum_{k=-K}^{K} kF_{l-k}(m)}{\sum_{k=-K}^{K} k^2}\ , \quad 1 \leq m \leq Q \tag{3}$$

where F defines the spectral feature, l is the number of frames, and Q is the feature order. Also, the time derivatives of the Delta (Δ) features are often calculated to yield Delta-Delta ($\Delta\Delta$) features [24] using equation (3).

In this work, each 1s cry sample was divided into short frames with 50ms duration and from each frame 16 coefficients were extracted to produce vectors with 304 coefficients from each sample. The feature sets generated for our experiments are:

a) 304 MFCC
b) 304 MFCC + 304 ΔMFCC + 304 ΔΔMFCC
c) 304 LPCC
d) 304 LPCC + 304 ΔLPCC + 304 ΔΔLPCC

2.3 Feature selection

Feature extraction resulted in high dimensional data which often contains redundant and irrelevant features. Theoretically, large number of features should offer better discriminating ability. However, in practice, given a limited amount of training data, large number of features possibly will cause the classifier to over fit the training data as the redundant or irrelevant features may negatively influence the learning algorithm [25]. Moreover, excessive features will significantly increase the computational time.

Hence, in this study we incorporate feature selection before the classification task. Feature selection extracts the important information from the data and reduces the dimensionality so that the most significant parts of the data are represented by the selected features. The goals of feature selection are to simplify the classifier by selecting only the relevant features; reduce the data dimensionality; and improve or not significantly reduce the classification performance [26]. The techniques applied in this study are further explained in the following sections.

2.3.1 OneR

OneR [27] calculates the weight or value of each feature individually. The OneR technique constructs one rule for each feature in the data by determining the most frequent class for each feature value. In other word, the most frequent class is the class that occurs most often for that

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 | Vol. 4, No.1. Pages 12-28, 2016

particular feature value. It then calculates the error rate for each rule constructed from each feature. Finally, it selects the features with smallest error rate.

2.3.2 ReliefF

ReliefF [28] randomly selects an instance from the data and calculates its nearest neighbours from the same and different class. The values of the features of the nearest neighbours are compared to the sampled instance and are used to update the individual relevance scores of each feature. The theory is that a relevance feature should has the ability to discriminate between instances from other classes and have the same value for instances within the same class.

2.3.3 Fast Correlation-Based Filter (FCBF)

Fast Correlation-Based Filter (FCBF) [29] applies Symmetrical Uncertainty (SU) [30] to measure the correlation between features. FCBF consists of two stages: (1) choosing a subset of relevant features and (2) choosing predominant features from the relevant features. FCBF searches for the best feature subset using backward selection technique with sequential search strategy. The searching process stops when there is no more feature to be discarded.

2.3.4 Consistency-Based Subset Evaluation (CNS)

Consistency-Based Subset Evaluation (CNS) [31] searches for subsets of features which contain a strong single class majority. In general, the algorithm searching process preferred small feature subsets with high class consistency. Thus, a search strategy was applied in conjunction with CNS in order to select the smallest feature subset with consistency similar to that of full set of features. In this work, the search strategy applied in CNS algorithm is simple genetic algorithm (GA) [32].

2.3.5 Correlation-Based Feature Selection (CFS)

Correlation-Based Feature Selection (CFS) [33] evaluates the relevance subsets of features instead of the individual features. The algorithm consists of a heuristic merit of subset evaluation that measures the relevance of individual feature for class prediction and also the inter-correlation level among features. The main hypothesis of CFS is that a good feature subset consists of features that are highly correlated with the class, yet poorly correlated with each other [33]. CFS consists of two main stages. It first calculates the matrix of feature-class and feature-feature correlations. In the second stage, CFS searches the feature subset space in order to select the best feature subset. In this work, the search strategy applied in CNS algorithm is simple GA [32].

2.4 Pattern classification

Artificial neural network (ANN) is widely applied in many areas due to its characteristics such as high learning accuracy, robustness, and strong ability for non-linear mapping. Among various architectures of ANN, RBFN and MLP have the ability to avoid local minima as these networks follow the supervised learning process by using the information from input and output for training the network weights [34]. In this work, we applied MLP and RBFN to compare the effectiveness of feature selection techniques used.

2.4.1 Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is a feed forward neural network that consists of several layers of neurons with unidirectional connections between them and usually trained with back-propagation algorithms [35]. The MLP architecture used in our work consisted of three layers: one input layer, a hidden layer, and an output layer. The hidden layer processed and transmitted the information in the input pattern to the output layer. A sigmoid activation was used in the hidden layer. The number of hidden neurons in MLP was varied with 10 to 30 with increment of 5.

2.4.2 Radial Basis Function Network (RBFN)

Radial Basis Function Network (RBFN) consists of three-layer feed forward type ANN. The input is converted using the basis functions in the hidden layer and the output layer contains weighted sum of linear combinations of the hidden nodes responses. The basis functions applied in this work is the normalized Gaussian radial basis function. RBFN training phase was executed in two steps. In the first step, the centres and the spreads of the radial basis function were obtained from the input variable. In the second step, the weights were adjusted in order to reduce the error function. In this work, the parameters of the radial basis function (the centres and the spreads) were determined using *K*-means clustering algorithm [36] with a predetermined cluster number. The number of clusters was varied with 10 to 30 with increment of 5. Finally, the connection weights were updated using backpropagation method.

3.0 RESULTS AND DISCUSSION

Both feature selection and pattern classification are performed in WEKA environment [37]. In this study, we applied 10-fold cross validation scheme to prove the reliability of the classification results obtained. This process randomly separates the data into 10 subsets or folds of approximately same size. A classifier is built and tested 10 times and the testing is done on one of the folds and the training process is done on the remaining folds. The process was repeated until all folds are used for testing and training the classifier. For each fold, the dimensionality was reduced by each feature selection technique before being passed to the classifiers. Dimensionality reduction was performed by cross validating the feature rankings generated by each selection technique with respect to the current classifier. Features with the best cross validated performance was selected as the best subset [38]. Feature selection was performed only on the training data and the classifier was tested using the selected features on the test data.

The classification accuracy (%), averaged over 10-fold cross validation was calculated for each feature set before and after feature selection. To determine whether the difference is statistically significant or not, we performed Wilcoxon Signed-Rank Test with 95% of confidence using each result obtained before and after feature selection. Table 2 and Table 3 present the results based on classification accuracy for MLP and RBFN respectively. The tables (Table 2 and Table 3) present how often each technique performs significantly better (denoted by "◦") or worse (denoted by "•") than without feature selection (column 3).

From Table 2, it can be seen that the best result is from ReliefF which improved the MLP performance on one feature set and degraded it on two. OneR and CFS showed degradations on two and three feature sets respectively. Meanwhile, CNS and FCBF performed worst as

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 I Vol. 4, No.1. Pages 12-28, 2016

they degraded the classifier performance on five and seven feature sets respectively. From Table 3, no feature selection techniques were able to improve the RBFN performance. However, ReliefF obtained the best result as it managed to maintain the classifier performance after the feature selection process for all cases. OneR is second best as it only showed degradations on two feature sets and followed by CFS which degraded RBFN performance on three. CNS and FCBF obtained the worst results as they showed degradations for most cases.

Table 2: Results of feature selection using MLP

Data set	Feature set	MLP (Unselect)	OneR	ReliefF	FCBF	CNS	CFS
Asphyxia vs. normal	MFCC	95.15	**96.58**	95.99	93.50	92.80	95.28
	MFCC + Δ + ΔΔ	96.34	**96.58**	95.87	91.14•	87.84•	94.45•
	LPCC	94.92	94.10	95.15	93.03	93.85	95.28
	LPCC + Δ + ΔΔ	96.70	94.80	95.28	93.63•	86.78•	93.75•
Deaf vs. normal	MFCC	97.33	98.05	97.91	95.81•	96.54	96.68
	MFCC + Δ + ΔΔ	97.91	97.69	96.47•	95.03•	95.17•	97.26
	LPCC	98.85	97.18•	98.27	96.90•	97.69•	98.63
	LPCC + Δ + ΔΔ	99.49	**99.21**	98.63•	96.68•	95.89•	99.06
Hungry vs. pain	MFCC	72.36	70.32	72.90	64.41•	75.65	72.35
	MFCC + Δ + ΔΔ	75.47	71.42•	**79.90**	71.23	73.45	72.14•
	LPCC	72.34	70.88	71.23	66.99	73.64	72.14
	LPCC + Δ + ΔΔ	69.78	69.22	74.39◦	67.55	69.39	71.22

MLP (UnSelect), OneR, ReliefF, FCBF, CNS, and CFS denote the MLP classifier without feature selection or using five different selection techniques respectively. The table presents how often each technique performs significantly better (denoted by "◦") or worse (denoted by "•") than without feature selection. The bold values are the highest accuracy for each data set.

In addition to classification accuracy, we also recorded the number of features selected and time taken (in seconds) to select the features and train the classifier. Table 4 shows the number of selected features and time taken to select features and train the classifier in seconds (s). We find that the feature selection techniques were able to greatly reduce the feature space. From Table 4, OneR, ReliefF, and CFS retained around 29% of the original features on average. CNS retained 18% of the features on average and it can be seen that FCBF selected the least number of features compared to the other techniques with only 7% of the features on average. In addition, RBFN showed faster performance than MLP in selecting features and train the classifier. For example, in Table 4, the highest classification accuracy for RBFN is 99.42% with 36.88 seconds and MLP is 96.58% with 356.30 seconds.

Table 3: Results of feature selection using RBFN

Data set	Feature set	RBFN (Unselect)	OneR	ReliefF	FCBF	CNS	CFS
Asphyxia vs.	MFCC	98.46	98.46	97.52	95.04•	96.81	98.58
normal	MFCC + Δ + ΔΔ	98.82	**99.29**	98.82	95.87•	90.31•	98.82
	LPCC	96.81	97.87	97.75	93.75•	94.81	97.40
	LPCC + Δ + ΔΔ	97.28	98.11	98.23	94.57•	92.10•	96.11
Deaf vs. normal	MFCC	98.63	98.12	98.20	96.82•	97.69•	98.19
	MFCC + Δ + ΔΔ	98.92	98.99	98.77	97.18•	98.20	97.90
	LPCC	99.57	99.06	99.13	98.12•	98.12•	99.13
	LPCC + Δ + ΔΔ	99.49	**99.42**	99.13	98.41•	94.95•	**99.42**
Hungry vs. pain	MFCC	81.19	76.56•	76.41	69.19•	76.57•	74.19•
	MFCC + Δ + ΔΔ	86.55	82.47•	85.61	75.11•	81.21•	82.10•
	LPCC	72.52	75.09	72.70	66.97•	72.34	75.12
	LPCC + Δ + ΔΔ	83.21	80.63	**86.54**	72.53•	77.67•	77.14•

RBFN (UnSelect), OneR, ReliefF, FCBF, CNS, and CFS denote the MLP classifier without feature selection or using five different selection techniques respectively. The table presents how often each technique performs significantly better (denoted by “◦”) or worse (denoted by “•”) than without feature selection. The bold values are the highest accuracy for each data set.

For feature selection techniques, ReliefF, OneR, and CFS achieved excellent performance. The success of ReliefF, OneR, and CFS are due to their ability to determine the dependencies between features. Although they were not able to determine the strongly interacting features in a reduced feature subset, they managed to maintain the performance of classifiers on most cases by selecting the relevant features under moderate interaction levels [33]. FCBF and CNS conversely were not able to determine dependencies between features. One reason why FCBF performed poorly among others could be accounted for its search strategy. In FCBF, a predominant feature was used to eliminate features that were redundant to it. However, in a situation where the features were highly correlated, FCBF may eliminate a large number of features as they were considered to be redundant [39]. This has been proved from the experiments in Table 4 that FCBF retained the lowest number of features compared to the other feature selection techniques. For CNS, the reason it performed poorly among others could be because CNS focuses on finding the smallest feature subset with consistency similar to that of full set of features. Since a feature subset is considered consistent if there are no two instances with similar feature values have different class labels, the searching algorithm may select a small feature subset that has a complicated pattern or information, while ignoring larger feature sets admitting simple information [26].

Penerbit Akademia Baru
Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 | Vol. 4, No.1. Pages 12-28, 2016

Table 4: Number of features selected and time taken (s) to select features and train the classifiers

Data set	Feature set			OneR	ReliefF	FCBF	CNS	CFS
Asphyxia vs. normal	MFCC	Selected features		80 (26%)	70 (23%)	30 (10%)	41 (13%)	84.8 (28%)
		Time (s)	MLP	58.65	81.43	29.84	37.98	79.58
			RBFN	9.00	31.22	2.55	3.09	7.66
	MFCC + Δ + ΔΔ	Selected features		300 (33%)	300 (33%)	50.5 (6%)	52.8 (6%)	289.9 (32%)
		Time (s)	MLP	356.30	357.49	50.63	54.59	294.46
			RBFN	28.89	89.67	4.90	3.03	61.20
	LPCC	Selected features		100 (33%)	100 (33%)	67.5 (22%)	41 (13%)	121.6 (40%)
		Time (s)	MLP	108.38	142.42	80.19	48.36	132.82
			RBFN	10.54	26.98	4.18	1.97	7.14
	LPCC + Δ + ΔΔ	Selected features		150 (16%)	170 (19%)	84.7 (9%)	90.6 (10%)	152.8 (17%)
		Time (s)	MLP	134.48	176.08	55.16	56.85	131.67
			RBFN	21.15	73.11	5.11	3.00	46.16
Deaf vs. normal	MFCC	Selected features		120 (39%)	100 (33%)	22.9 (8%)	41 (13%)	98.8 (33%)
		Time (s)	MLP	158.15	227.54	37.68	58.61	142.52
			RBFN	16.23	68.12	4.09	4.27	12.27

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 | Vol. 4, No.1. Pages 12-28, 2016

Table 4: Number of features selected and time taken (s) to select features and train the classifiers (continued)

Data set	Feature set			OneR	ReliefF	FCBF	CNS	CFS
Deaf vs. normal	MFCC + Δ + ΔΔ	Selected features		250 (27%)	250 (27%)	38.5 (4%)	148 (16%)	342.1 (38%)
		Time (s)	MLP	191.07	362.34	33.42	102.24	305.57
			RBFN	48.06	217.27	8.79	12.10	111.03
	LPCC	Selected features		100 (33%)	100 (33%)	22.8 (8%)	42.3 (14%)	95.4 (31%)
		Time (s)	MLP	221.99	268.03	51.89	85.56	182.16
			RBFN	17.89	97.37	4.05	4.65	14.78
	LPCC + Δ + ΔΔ	Selected features		300 (33%)	300 (33%)	37.3 (4%)	152.9 (17%)	299.7 (33%)
		Time (s)	MLP	244.36	417.30	31.18	126.25	359.62
			RBFN	36.88	235.52	5.61	8.60	93.85
Hungry vs. pain	MFCC	Selected features		90 (30%)	90 (30%)	17 (6%)	91.8 (30%)	70.7 (23%)
		Time (s)	MLP	44.03	51.97	9.73	41.90	35.71
			RBFN	6.01	10.59	0.54	1.59	3.02
	MFCC + Δ + ΔΔ	Selected features		250 (27%)	250 (27%)	29.5 (3%)	257.9 (28%)	190.9 (21%)
		Time (s)	MLP	238.09	279.40	38.47	277.76	229.44
			RBFN	18.16	33.57	1.61	4.33	29.18

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 | Vol. 4, No.1. Pages 12-28, 2016

Table 4: Number of features selected and time taken (s) to select features and train the classifiers (continued)

Data set	Feature set			OneR	ReliefF	FCBF	CNS	CFS
Hungry vs. pain	LPCC	Selected features		80 (26%)	80 (26%)	11 (4%)	89.5 (29%)	67 (22%)
		Time (s)	MLP	43.32	49.83	7.72	42.91	34.99
			RBFN	5.80	11.57	0.48	1.83	3.01
	LPCC + Δ + ΔΔ	Selected features		250 (27%)	250 (27%)	19.7 (2%)	268.1 (29%)	209.4 (23%)
		Time (s)	MLP	197.78	221.29	19.70	203.57	189.60
			RBFN	16.89	32.76	1.07	5.03	30.94

Information in brackets show the percentage of the original features retained.

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 I Vol. 4, No.1. Pages 12-28, 2016

In comparing the classifiers, RBFN obtained better classification performance than MLP on all feature sets. RBFN showed better performance due to its proper consideration of data distribution by prior clustering [40]. Moreover, RBFN required significantly less time to select features and train the classifier. The MLP is computationally time intensive as it is trained in fully supervised manner and requires more number of iterations during the network training process in order to obtain the best classification result. In contrast, the RBFN performed faster than MLP due to unsupervised training process in the hidden layer.

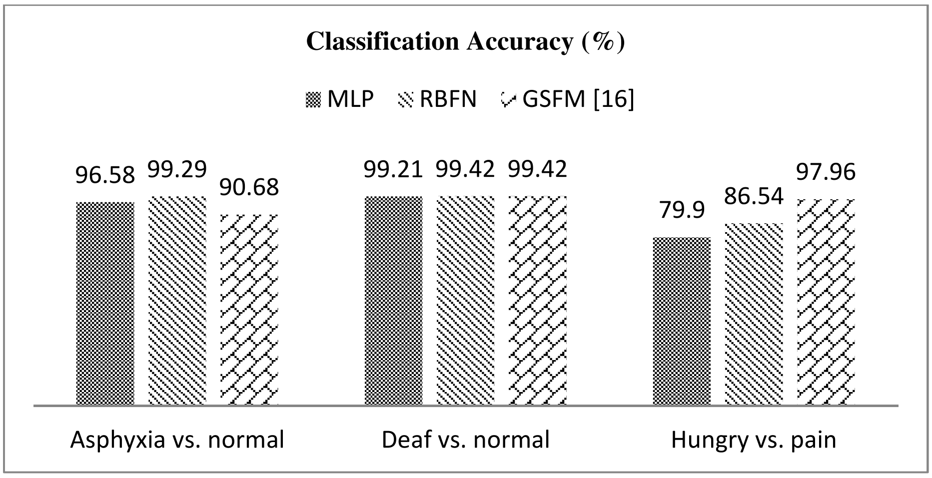

Figure 3: Results of binary classification accuracy

Figure 3 shows the results of the proposed study and results obtained from method in [16]. The classification accuracy for MLP and RBFN are obtained from the best results in Table 2 and Table 3 respectively. From Figure 3, MLP and RBFN were able to generate very competitive classification results for asphyxia vs. normal and deaf vs. normal data sets. However, the performance for hungry vs. pain data set is low for both classifiers. One reason could be due to the distribution of hungry vs. pain data set which is more complex than the other two as shown in Figure 4. The MLP trained with backpropagation performs well on simple training problems. However, as the problem complexity increases (in this case due to higher complexity of the data), the performance of backpropagation decreases rapidly [41]. The method reported in [16] managed to performed better in solving the complex pattern classification tasks such as the hungry vs. pain data set. However, MLP and RBFN outperformed the method reported in [16] for asphyxia vs. normal data set and obtained almost the same results with [16] for deaf vs. normal data set.

4.0 CONCLUSION

In this paper, we compare five different feature selection techniques in the automatic infant cry classification system. OneR, ReliefF, and CFS achieved excellent performance on most cases. FCBF and CNS on the other hand showed worst performance as they reduced the system performance after feature selection for all cases. For classifiers, RBFN obtained faster processing time and better classification accuracy than MLP. Thus, we suggest that OneR, ReliefF, and CFS can be applied in feature selection process and RBFN is a suitable classifier for the automatic infant cry classification system. In future, we would like to explore other feature selection techniques to select features in complex data distribution such as in hungry vs. pain dataset.

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 I Vol. 4, No.1. Pages 12-28, 2016

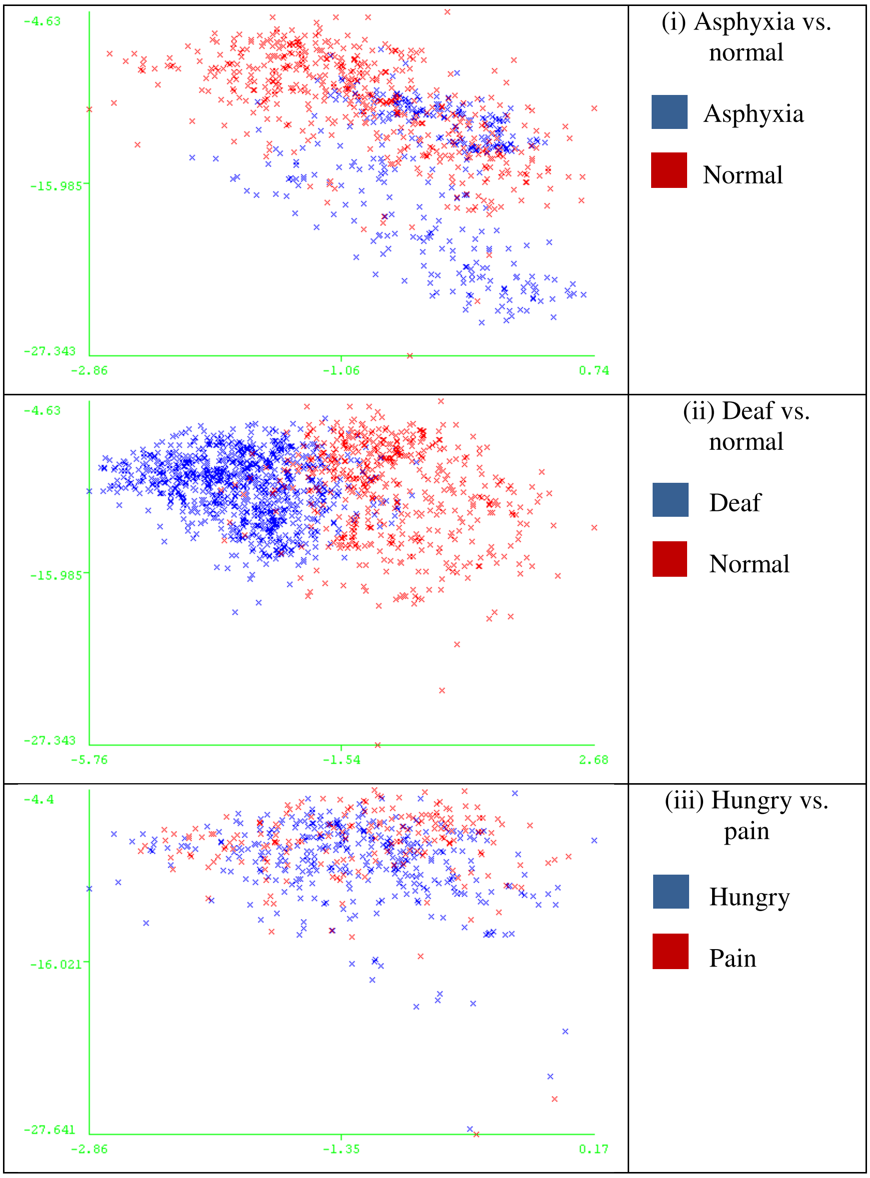

Figure 4: Data set distribution plot

ACKNOWLEDGEMENT

The Baby Chillanto database 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 database. The authors would like to thank Dr. Carlos Alberto Reyes-Garcia, Researcher, CCC-Inaoep, Mexico for providing the infant cry database. In addition, the authors appreciate the support for this research received under Fundamental Research Grant Scheme (FRGS) from Ministry of Education, Malaysia. [Grant No: 9003-00485].

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 I Vol. 4, No.1. Pages 12-28, 2016

REFERENCES

[1] Orlandi, Silvia, Carlos Alberto Reyes Garcia, Andrea Bandini, Gianpaolo Donzelli, and Claudia Manfredi. "Application of Pattern Recognition Techniques to the Classification of Full-Term and Preterm Infant Cry." Journal of Voice (2015).

[2] Corwin, Michael J., Barry M. Lester, and Howard L. Golub. "The infant cry: what can it tell us?." Current Problems in Pediatrics 26, no. 9 (1996): 313-334.

[3] Wasz-Höckert, O., T. J. Partanen, V. Vuorenkoski, K. Michelsson, and E. Valanne. "The identification of some specific meanings in infant vocalization."Experientia 20, no. 3 (1964): 154-154.

[4] Barajas-Montiel, Sandra E., and Carlos A. Reyes-Garcia. "Identifying pain and hunger in infant cry with classifiers ensembles." In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), vol. 2, pp. 770-775. IEEE, 2005.

[5] Petroni, Marco, Alfred S. Malowany, C. Celeste Johnston, and Bonnie J. Stevens. "Classification of infant cry vocalizations using artificial neural networks (ANNs)." In Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on, vol. 5, pp. 3475-3478. IEEE, 1995.

[6] Saraswathy, Jeyaraman, Muthusamy Hariharan, Wan Khairunizam, Sazali Yaacob, and N. Thiyagar. "Infant cry classification: time frequency analysis." In Control System, Computing and Engineering (ICCSCE), 2013 IEEE International Conference on, pp. 499-504. IEEE, 2013.

[7] Zabidi, A., W. Mansor, L. Y. Khuan, I. M. Yassin, and R. Sahak. "Binary Particle Swarm Optimization and F-Ratio for Selection of Features in the Recognition of Asphyxiated Infant Cry." In 5th European Conference of the International Federation for Medical and Biological Engineering, pp. 61-65. Springer Berlin Heidelberg, 2011.

[8] Rosales-Pérez, Alejandro, Carlos A. Reyes-García, and Pilar Gómez-Gil. "Genetic fuzzy relational neural network for infant cry classification." In Mexican Conference on Pattern Recognition, pp. 288-296. Springer Berlin Heidelberg, 2011.

[9] Hariharan, Muthusamy, 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 108, no. 2 (2012): 559-569.

[10] Orozco, José, and Carlos A. Reyes García. "Detecting pathologies from infant cry applying scaled conjugate gradient neural networks." In European Symposium on Artificial Neural Networks, Bruges (Belgium), pp. 349-354. 2003.

[11] Varallyay, G. Jr, Z. Benyó, A. Illényi, Zs Farkas, and L. Kovács. "Acoustic analysis of the infant cry: classical and new methods." In Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, vol. 1, pp. 313-316. IEEE, 2004.

[12] Orlandi, S., C. Manfredi, L. Bocchi, and M. L. Scattoni. "Automatic newborn cry analysis: a non-invasive tool to help autism early diagnosis." In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2953-2956. IEEE, 2012.

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 I Vol. 4, No.1. Pages 12-28, 2016

[13] Lederman, Dror, Ehud Zmora, Stephanie Hauschildt, Angelika Stellzig-Eisenhauer, and Kathleen Wermke. "Classification of cries of infants with cleft-palate using parallel hidden Markov models." Medical & biological engineering & computing 46, no. 10 (2008): 965-975.

[14] Dash, Manoranjan, Huan Liu, and Hiroshi Motoda. "Consistency based feature selection." In Pacific-Asia conference on knowledge discovery and data mining, pp. 98-109. Springer Berlin Heidelberg, 2000.

[15] Arauzo-Azofra, Antonio, Jose Manuel Benitez, and Juan Luis Castro. "Consistency measures for feature selection." Journal of Intelligent Information Systems 30, no. 3 (2008): 273-292.

[16] Rosales-Pérez, Alejandro, Carlos A. Reyes-García, Jesus A. Gonzalez, Orion F. Reyes-Galaviz, Hugo Jair Escalante, and Silvia Orlandi. "Classifying infant cry patterns by the Genetic Selection of a Fuzzy Model." Biomedical Signal Processing and Control 17 (2015): 38-46.

[17] Cohen, Rami, and Yizhar Lavner. "Infant cry analysis and detection." InElectrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of, pp. 1-5. IEEE, 2012.

[18] Liu, Li, Jialong He, and Günther Palm. "Signal modeling for speaker identification." In Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, vol. 2, pp. 665-668. IEEE, 1996.

[19] Rabiner, Lawrence, and Biing-Hwang Juang. "Fundamentals of speech recognition." (1993).

[20] Kinnunen, Tomi, Ville Hautamäki, and Pasi Fränti. "Fusion of spectral feature sets for accurate speaker identification." In 9th Conference Speech and Computer. 2004.

[21] Harrington, Jonathan, and Steve Cassidy. Techniques in speech acoustics. Vol. 8. Springer Science & Business Media, 1999.

[22] Atal, Bishnu S. "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification." the Journal of the Acoustical Society of America 55, no. 6 (1974): 1304-1312.

[23] Mason, J. S., and X. Zhang. "Velocity and acceleration features in speaker recognition." In Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on, pp. 3673-3676. IEEE, 1991.

[24] Kinnunen, Tomi. "Spectral features for automatic text-independent speaker recognition." Licentiate's Thesis, University of Joensuu.–2003 (2003).

[25] Yu, Lei, and Huan Liu. "Efficient feature selection via analysis of relevance and redundancy." Journal of machine learning research 5, no. Oct (2004): 1205-1224.

[26] Dash, Manoranjan, and Huan Liu. "Consistency-based search in feature selection." Artificial intelligence 151, no. 1 (2003): 155-176.

[27] Holte, Robert C. "Very simple classification rules perform well on most commonly used datasets." Machine learning 11, no. 1 (1993): 63-90.

[28] Kononenko, Igor. "Estimating attributes: analysis and extensions of RELIEF." In European conference on machine learning, pp. 171-182. Springer Berlin Heidelberg, 1994.

Journal of Advanced Research in Applied Sciences and Engineering Technology
ISSN (online): 2462-1943 I Vol. 4, No.1. Pages 12-28, 2016

[29] Yu, Lei, and Huan Liu. "Feature selection for high-dimensional data: A fast correlation-based filter solution." In ICML, vol. 3, pp. 856-863. 2003.

[30] Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. Numerical recipes in C. Vol. 2. Cambridge: Cambridge university press, 1996.

[31] Liu, Huan, and Rudy Setiono. "A probabilistic approach to feature selection-a filter solution." In ICML, vol. 96, pp. 319-327. 1996.

[32] Golberg, David E. "Genetic algorithms in search, optimization, and machine learning." Addion wesley 1989 (1989): 102.

[33] Hall, Mark A. "Correlation-based feature selection for machine learning." PhD diss., The University of Waikato, 1999.

[34] Chiddarwar, Shital S., and N. Ramesh Babu. "Comparison of RBF and MLP neural networks to solve inverse kinematic problem for 6R serial robot by a fusion approach." Engineering Applications of Artificial Intelligence 23, no. 7 (2010): 1083-1092.

[35] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. Learning internal representations by error propagation. No. ICS-8506. CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR COGNITIVE SCIENCE, 1985.

[36] Moody, John, and Christian J. Darken. "Fast learning in networks of locally-tuned processing units." Neural computation 1, no. 2 (1989): 281-294.

[37] Witten, Ian H., and Eibe Frank. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.

[38] Hall, Mark A., and Geoffrey Holmes. "Benchmarking attribute selection techniques for data mining." (2000).

[39] Senliol, Baris, Gokhan Gulgezen, Lei Yu, and Zehra Cataltepe. "Fast Correlation Based Filter (FCBF) with a different search strategy." InInternational Symposium on Computer and Information Sciences (ISCIS 2008), pp. 27-29. Istanbul Technical University, Suleyman Demirel Cultural Center, Istanbul, Turkey, 2008.

[40] Yeung, Daniel S., Defeng Wang, Wing WY Ng, Eric CC Tsang, and Xizhao Wang. "Structured large margin machines: sensitive to data distributions."Machine Learning 68, no. 2 (2007): 171-200.

[41] Montana, David J., and Lawrence Davis. "Training Feedforward Neural Networks Using Genetic Algorithms." In IJCAI, vol. 89, pp. 762-767. 1989.