Automatic Infant Cry Pattern Classification for a Multiclass Problem

N.S.A.Wahid[1], P.Saad[1], M.Hariharan[2]
[1]School of Computer and Communication Engineering, Universiti Malaysia Perlis, 02600 Perlis, Malaysia.
[2]School of Mechatronic Engineering, Universiti Malaysia Perlis, 02600 Perlis, Malaysia.
nsyahirah91@gmail.com

***Abstract*—Crying is the only way of communication for infants to express their physical and emotional needs. Automatic infant cry analysis that provides fast and non-invasive process is suitable to assess the physical and emotional states of infants. The cry analysis provides an opportunity to understand infants' needs. It is also beneficial in clinical environment for identifying specific pathologies through infant cry. This paper presents an automatic infant cry classification system for a multiclass problem. The cry classification system consists of three stages: (1) feature extraction, (2) feature selection, and (3) pattern classification. We extracted spectral features, such as Mel Frequency Cepstral Coefficients (MFCC) and Linear Prediction Cepstral Coefficients (LPCC) to represent the acoustic characteristics of the cry signals. In addition, the combination of spectral and dynamic features was also investigated. Due to the high dimensionality of data resulting from the feature extraction stage, we selected relevant features to perform feature selection to reduce the data dimensionality. In this stage, five different feature selection techniques were experimented. In the pattern classification stage, two Artificial Neural Network (ANN) architectures: Multilayer Perceptron (MLP) and Radial Basis Function Network (RBFN) were used for classifying the cry signals into five categories: asphyxia, pain, hunger, deaf, and normal. Experimental results show that the best classification accuracy of 93.43% (Kappa value of 0.91) was obtained from MFCC + ∆MFCC + ∆∆MFCC feature set, when using CFS selection technique and RBFN.**

***Index Terms*—Artificial Neural Network; Dynamic Features; Feature Selection; Infant Cry Classification.**

I. INTRODUCTION

Crying is a type of communication for infants to express their physical and emotional condition. There are many reasons for infant to cry such as sadness, hunger, lonely, anger, and discomfort. Furthermore, vital information, such as the health status of the infant can be obtained from the cry itself. Thus, many researches have been conducted to analyze the characteristics of infant cry that give signals to different types of cries and pathologies. These researches allow for the understanding of various needs of the infants so that suitable treatment can be given, thus, helping to prevent any further complications of the infants.

Early researches have employed auditory and sound spectrographic analysis to analyse the signals of infant cry. 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, brain damage, encephalitis, hypothyroidism, down syndrome, oropharyngeal abnormalities, and genetic defects [1], [2]. However, these analyses require subjective evaluation from clinical specialists and consume time when performing the evaluation process. Moreover, they are unsuitable for a large database. Hence, automatic infant cry classification system has been proposed to overcome the limitations of the conventional analyses. The classification system provides fast and accurate diagnosis results. It is also suitable for large cry database and harmless to infants. This analysis has been applied widely and obtained promising results in the classification of different types of cries and pathologies, such as hunger and pain cries [3]–[5], asphyxia [6]–[8], deaf [9]–[11], autism [12], and cleft palate [13].

II. LITERATURE REVIEW

Significant progress has been achieved in the development of the automatic infant cry classification system. Lederman et al. [13] classified cries of infants with cleft palate using Mel Frequency Cepstral Coefficients (MFCC) and Linear Prediction Cepstral Coefficients (LPCC) features and Hidden Markov Model (HMM) as classifier. They obtained an average accuracy of 91% in the classification of two categories of cries: cleft palate with plate and without plate. In [14], Orlandi et al. applied Random Forest classifier to identify cries from pre-term and full-term infants with the accuracy up to 87%. In [15], an optimal wavelet feature was used with Artificial Neural Network (ANN) to determine the normal and pathological infant cries. The highest classification accuracy achieved was more than 99%. In [16], MFCC feature and feed forward neural network were used in automatic classification of normal and deaf cry with 97.43% classification accuracy. In [9], a General Regression Neural Network (GRNN) was implemented for infant cry classification with two class problems: normal cry and deaf cry. Short-time Fourier transform (STFT) analysis was used to extract the features, and a method was proposed. This has resulted in a maximum correct classification of 99%. Furthermore, in [17], Multilayer Perceptron (MLP) trained with scaled conjugate gradient was applied to classify the normal and deaf cry with 93.2% classification accuracy. In [18], several features such as LPCC, MFCC, and Bark Frequency Cepstral Coefficients (BFCC) were applied to classify three types of infant cries with an accuracy up to 80%. Moreover, in [19], an embedded infant cry classification system using Support Vector Machine (SVM) was proposed to identify four types of cries with the classification rate around 56%. Hariharan et al. [20] proposed an automatic infant cry classification with three classes: normal, deaf, and asphyxia. Weighted Linear Prediction Cepstral Coefficients (WLPCC) and PNN were applied to classify the cry signals with above 98% classification accuracy. Saraswathy et al. [6] successfully developed an automatic infant cry classification using PNN and GRNN. The cry signals were classified into three categories: asphyxia, deaf and normal cry with the highest classification accuracy above 99%. Finally, in [21], three categories of cries were considered: pain, wet diaper, and hunger. Gaussian mixture models (GMM) were used as a classifier in the cry classification systems developed with 78.96% overall classification accuracy.

These studies proved that the infant cry consists of specific patterns that enable the classification of different types of pathologies and cries using automated techniques. However, from the literature reviewed, while two-class classification, also known as binary classification have been widely studied [9], [11], [13]–[17], [22]–[25], only a small number of studies focuses on multiclass classification [6], [18]–[21]. Therefore, in this paper, we propose an automatic infant cry classification system for a multiclass problem. We studied different features, selection techniques, and classifiers to determine which of these techniques best performs in the classification system. Five different categories of cries namely, asphyxia, pain, hunger, deaf, and normal were identified. 10-fold cross validation was used to evaluate the effectiveness of the features applied and the reliability of the classification results. The experimental results showed that the classification system achieved the highest classification accuracy up to 93.43% (Kappa value of 0.91).

III. RESEARCH METHODOLOGY

This section explains the methodology of this research. The database information used in the research is also provided. There are three main stages involved in this research namely, the feature extraction, feature selection, and pattern classification

A. Database

The database used is known as the 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 [5]. The infant cry samples ranging from just born to 6-month old infants were recorded directly by specialized physicians. The samples were labeled with information about the cause of crying during the recording process.

Table 1 shows the description of the infant cry database used in this study. All the samples in the database have 1 second (s) length and the sampling frequency used for our study is 8000 Hertz (Hz). The database consists of 340 samples from asphyxia cries, 192 samples of pain cries, 350 samples of hunger cries, 879 samples from deaf cries, and 157 samples of normal cries. In this study, all samples from the five categories of cries were used in the classification process.

Table 1
Database Description

	Infant Cry Category	Total no. of samples used	No. of samples from each category
1	Asphyxia		340
2	Pain	1918	192
3	Hunger		350
4	Deaf		879
5	Normal		157

B. Feature Extraction

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

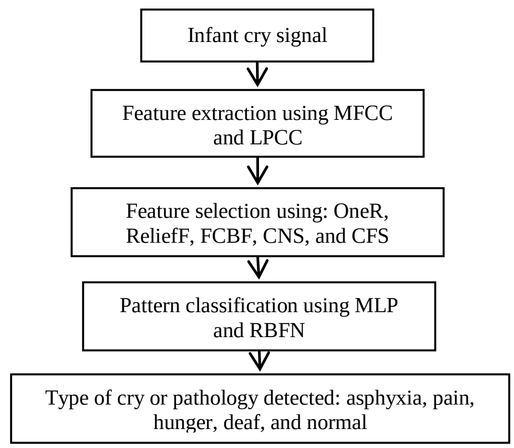

Figure 1: Block diagram of the automatic infant cry classification system

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 have been widely applied in the field of automatic speech recognition (ASR) since the mid-eighties. In addition, MFCC and LPCC have been proven to be the appropriate representations of infant cry signals [5], [26].

Figure 2 illustrates the extraction process of MFCC and LPCC features. The first step in the 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 [27]. 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 within a duration of 10ms to 50ms [28]. Then, each frame was windowed by a 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.

Next, the 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.

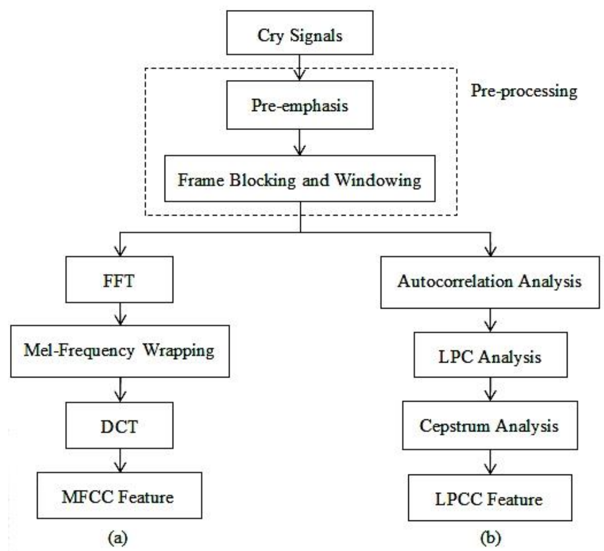

Figure 2: Block diagram of features extraction process: MFCC feature (a) and LPCC feature (b)

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 as an indicator to calculate energy that exists near 0 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 below 1 kHz in linear scale and frequencies above 1 kHz in logarithmic scale. The formula for computing the *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 the 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) [29]:

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

where $s[n]$ is the signal samples, $a[k]$ denotes 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 [30]. Finally, the LPCC feature was derived from the LPC using a recursion technique [31].

In addition to the spectral features extracted, we also investigated the combination of spectral features with dynamic features. Dynamic features is the time derivatives of the spectrum-based features [32]. These features contain the dynamic characteristics of the spectral features. The first order derivatives, also known as Delta (Δ) features [28], can be calculated using Equation (3) as follows [28]:

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

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 (ΔΔ) features [33] 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

C. *Feature Selection*

Feature extraction from high dimensional data 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, a large number of features will possibly cause the classifier to over fit the training data as the redundant or irrelevant features may negatively influence the learning algorithm [34]. 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 [35].

In general, the feature selection can be categorised into two techniques: filter techniques and wrapper techniques [36]. Filter techniques are independent of a classifier, whereas wrapper techniques apply the classification algorithm as part of the 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. The following are the filter techniques applied in our work:

OneR [37] calculates the weight or value of each feature individually. The OneR algorithm constructs one rule for each feature in the data by determining the most frequent class for each feature value. In other words, the most frequent class is the class that occurs most often for that particular feature value. It then calculates the error rate for each rule constructed from each feature. Finally, it selects the features with the smallest error rate.

ReliefF [38] 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 used to update the individual relevance scores of each feature. The theory is that a relevance feature should have the ability to discriminate between instances from other classes and have the same value for instances within the same class.

Fast Correlation-Based Filter (FCBF) [39] applies Symmetrical Uncertainty (SU) [40] 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.

Consistency-Based Subset Evaluation (CNS) [41] searches for subsets of features which contain a strong single class majority. In general, the algorithm searching process prefers small feature subsets with a high-class consistency. Thus, a search strategy is 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 a simple genetic algorithm (GA) [42].

Correlation-Based Feature Selection (CFS) [43] 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 [43]. 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 [42].

D. Pattern Classification

ANN has been 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 [44]. In this work, we applied MLP and RBFN to compare the effectiveness of feature selection techniques used.

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 [45]. 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 was determined experimentally and we set it to 10. The learning rate and momentum factor were set to 0.3 and 0.2 respectively.

Radial Basis Function Network (RBFN) consists of three-layer feed forward type neutral network. The input is converted using the basis functions in the hidden layer and the output layer contains the 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 centers 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 centers and the spreads) were determined using *K*-means clustering algorithm [46] with a predetermined cluster number. The number of clusters was determined experimentally and we set it to 10. Finally, the connection weights were updated using backpropagation method.

IV. RESULTS AND DISCUSSIONS

Both feature selection and pattern classification have been performed in WEKA environment [47]. 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 the same size. A classifier was built and tested 10 times, the testing was done on one of the folds and the training process was done on the remaining folds. The process was repeated until all folds were used for testing and training the classifier. For each fold, the dimensionality was reduced by each feature selection technique, before passing it 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 [48]. Feature selection was performed only on the training data and the classifier was tested using the selected features on the test data.

In this work, two performance metrics that have been widely considered to evaluate the binary and multiclass problems were used: accuracy and Kappa statistic [49]. In the multiclass problem, considering only the accuracy may not be an appropriate evaluation particularly when the class has imbalance proportions. Thus, we included Kappa statistic to evaluate the agreement between the predicted and the observed classifications of dataset, while correcting the agreement that occurs by chance [50]. Kappa statistic can be calculated using Equation (4) [51] as follows:

$$K = \frac{n\sum_{i=1}^{m} h_{ii} - \sum_{i=1}^{m} T_{ri}T_{ci}}{n^2 - \sum_{i=1}^{m} T_{ri}T_{ci}} \tag{4}$$

where h_{ii} is the number of true positives of each class, n is the total number of samples, m is the number of classes, and T_{ri} and T_{ci} are the rows' and columns' counts respectively. Kappa statistic is a simple and useful metric for evaluating the classification rate in a multiclass problem while compensating for random successes. The difference between classification accuracy and Kappa statistic is the scoring of the correct classifications. Accuracy scores all the correct classifications over all classes, whereas Kappa statistic scores the correct classifications independently for each class and aggregates them [51]. Kappa statistic is less sensitive to randomness caused by a different number of examples in each class. It ranges from -1 (total disagreement) to 1(total agreement), and K=0 shows a random classification.

The accuracy (%) and Kappa value, averaged over 10-fold cross validation were 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 of feature selection with MLP and RBFN respectively. From Table 2, it can be seen that no feature selection techniques performed significantly better than MLP without feature selection (Unselect). However, ReliefF and CFS performed better than the other feature as they do not significantly degrade the performance of the classifier. OneR degraded the classifier performance on two feature sets and both FCBF and CNS showed degradations for all four feature sets used. From Table 3, the best result is from OneR, which improved RBFN on one feature set but also degraded it on one. ReliefF and CFS both are in the second place as they only degraded on one feature set. FCBF and CNS showed the worst performance as they degraded on all four feature sets used.

Table 2
Results of Feature Selection with MLP

Feature set	Performance metrics	MLP (Unselect)	OneR	ReliefF	FCBF	CNS	CFS
MFCC	Accuracy	88.22	85.61•	85.87	83.63•	84.20•	87.96
	Kappa statistic	0.83	0.80•	0.80	0.77•	0.78•	0.83
MFCC+ΔMFCC +ΔΔMFCC	Accuracy	89.47	90.25	88.90	84.46•	75.75•	90.04
	Kappa statistic	0.85	0.86	0.84	0.78•	0.66•	0.86
LPCC	Accuracy	88.79	86.08•	88.95	84.36•	85.45•	87.49
	Kappa statistic	0.84	0.80•	0.84	0.78•	0.79•	0.82
LPCC+ΔLPCC +ΔΔLPCC	Accuracy	89.16	89.05	88.48	84.93•	67.47•	89.73
	Kappa statistic	0.85	0.85	0.84	0.79•	0.53•	0.85

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.

Table 3
Results of Feature Selection with RBFN

Feature set	Performance metrics	RBFN (Unselect)	OneR	ReliefF	FCBF	CNS	CFS
MFCC	Accuracy	91.61	88.27•	88.16•	88.22•	88.32•	90.67
	Kappa statistic	0.88	0.83•	0.83•	0.83•	0.83•	0.87
MFCC+ΔMFCC +ΔΔMFCC	Accuracy	92.54	93.06	92.91	89.26•	80.55•	93.43
	Kappa statistic	0.89	0.90	0.90	0.85•	0.72•	0.91
LPCC	Accuracy	88.63	90.25◦	89.99	85.87•	86.08•	89.26
	Kappa statistic	0.84	0.86◦	0.86	0.80•	0.80•	0.85
LPCC+ΔLPCC +ΔΔLPCC	Accuracy	91.97	91.55	92.54	87.48•	73.36•	88.89•
	Kappa statistic	0.89	0.88	0.89	0.82•	0.62•	0.84•

RBFN (UnSelect), OneR, ReliefF, FCBF, CNS, and CFS denote the RBFN 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.

Table 4
Number of features selected for MLP and RBFN and time taken (s) to select features and train the classifiers.

Feature set		OneR	ReliefF	FCBF	CNS	CFS
MFCC	No of features	90 (30%)	90 (30%)	28.2 (9%)	41 (13%)	92.8 (31%)
	MLP	140.13	307.59	52.17	65.64	139.10
	RBFN	72.86	177.26	476.42	449.69	71.33
MFCC+ΔMFCC +ΔΔMFCC	No of features	320 (35%)	320 (35%)	40 (4%)	54.4 (6%)	329.7 (36%)
	MLP	431.36	754.23	54.19	66.30	447.25
	RBFN	63.96	457.49	480.63	1859.61	165.14
LPCC	No of features	125 (41%)	125 (41%)	39.5 (13%)	41 (13%)	125.9 (41%)
	MLP	139.18	275.71	52.97	49.51	133.93
	RBFN	69.48	201.31	569.53	474.34	57.40
LPCC+ΔLPCC +ΔΔLPCC	No of features	200 (22%)	200 (22%)	54.7 (6%)	52.8 (6%)	207.2 (23%)
	MLP	224.54	629.18	83.45	81.18	365.47
	RBFN	81.90	431.27	805.66	1489.68	192.64

In addition to accuracy and Kappa statistic, 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 approximately an average of 32% of the features. Meanwhile, FCBF and CNS retained an average of only 8% and 10% of the features, respectively. In comparison to the time taken to select the features and train the MLP classifier, FCBF and CNS showed the fastest time. OneR and CFS required about the same amount of time, while ReliefF has the slowest time. For RBFN, OneR and CFS were faster than the other techniques. On the other hand, FCBF and CNS showed the slowest speed.

The success of ReliefF and CFS could be due to their ability to determine the dependencies between features. Although they were not able to determine the strong interacting features in a reduced feature subset, they managed to maintain the performance of classifiers on most feature sets by selecting the relevant features under moderate interaction levels [43]. FCBF and CNS conversely were not able to determine the 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. From Table 4, FCBF retained the lowest number of features among others with an average of only 8% of the original features. All the selection techniques applied were not able to significantly improve the performance of the classifiers on most feature sets. Techniques, such as OneR and ReliefF only calculated the individual score of each feature and ignored the feature dependencies. Although FCBF, CNS, and CFS searched for the best subset by calculating both the individual score of each feature and the dependencies between features, they did not consider selecting strong correlated features because these features were considered to be redundant. However, two completely irrelevant features by themselves can become relevant when combined together, and the combination of two strongly correlated features is better than any independent features [52]. Therefore, ignoring strongly correlated features may significantly reduce the classification performance.

To demonstrate the performance of selection techniques depending on feature sets visually, we produced Figure 3 and Figure 4 using Table 2 and Table 3 respectively. Figure 3 and Figure 4 were produced based on kappa statistic results.

From Figure 3, all selection techniques except ReliefF and CNS showed improvement when using the combination of spectral and dynamic features. ReliefF showed no performance improvement for LPCC features as it obtained similar results for LPCC and LPCC+ΔLPCC +ΔΔLPCC with Kappa value of 0.84. For CNS, the performance was significantly reduced when using the combination of spectral and dynamic features.

From Figure 4, all techniques except CFS and CNS showed improvement in performance when using the combination of spectral and dynamic features. CFS showed a slightly reduced performance from Kappa value of 0.85 (LPCC) to Kappa value of 0.84 (LPCC + ΔLPCC + ΔΔLPCC). Meanwhile, CNS performance reduced significantly when using MFCC + ΔMFCC + ΔΔMFCC and LPCC + ΔLPCC + ΔΔLPCC feature sets.

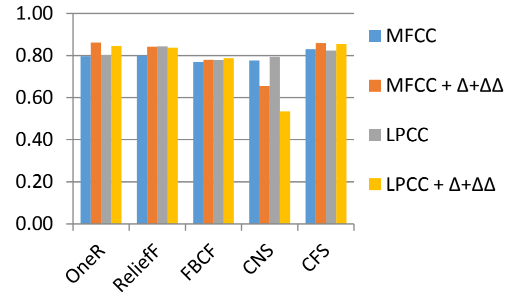

Figure 3: Kappa statistic results using MLP

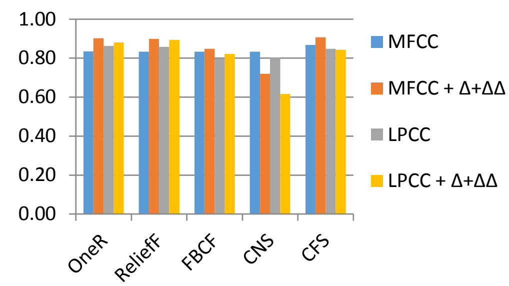

Figure 4: Kappa statistic results using RBFN

Figure 3 and Figure 4 indicate that most of the selection techniques showed improvement in performance when using the combination of spectral and dynamic features. The spectral features are the representation of short-term stationary signals, thus time domain information is not presented. However, the infant cry is a non-stationary signal as the articulators constantly change their position at a certain rate during the production of cry sounds. The dynamic information of the spectral features can be obtained by extracting the dynamic features. Yet, CNS showed significant reduction in performance when using the combination of spectral and dynamic features. The technique was not able to select the best feature subset when the feature dimension increased significantly. One reason 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 that have different class labels, the searching algorithm may select a small feature subset that has a complicated function, while ignoring larger feature sets admitting simple rules [35].

In comparing the classifiers, RBFN obtained better classification performance than MLP on all feature sets. Moreover, RBFN required significantly less time to select features and train the classifier, except when using FCBF and CNS. The MLP was computationally time intensive as it was trained in fully supervised manner and required more number of iterations during the network training process in order to obtain the best classification result. In contrast, the RBFN performed faster due to unsupervised training process in the hidden layer.

V. CONCLUSION

In this paper, we studied different features, selection techniques and classifiers to perform multiclass classification of infant cry. We found that the combination of spectral and dynamic features was able to improve the performance of the classification system for all selection techniques except CNS. For selection techniques, OneR, ReliefF, and CFS achieved good performance on most cases. FCBF and CNS, on the other hand showed the worst performance as they reduced the system performance after the feature selection for all cases. For classifiers, RBFN obtained better performance in terms of accuracy and Kappa statistic than MLP. Moreover, RBFN required less time to select features and train the classifier when applied with OneR, ReliefF, and CFS selection techniques. The best classification accuracy of 93.43% (Kappa value of 0.91) was obtained from MFCC + ΔMFCC + ΔΔMFCC feature set when using CFS selection technique and RBFN classifier. Although CFS was not able to significantly improve the classifier performance, it was able to achieve the goal of feature selection by maintaining the performance of the classifier with a reduced feature subset in most cases.

ACKNOWLEDGMENT

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].

REFERENCES

[1] M. J. Carwin, B. M. Lester, and H. L. Golub, "The infant cry: what can it tell us?," *Current Problems in Pediatrics*, vol. 26, no. 9, pp. 324–334, 1996.

[2] O. Wasz-Höckert, T. J. Partanen, V. Vuorenkoski, and K. Michelsson, "The identification of some specific meanings in infant vocalization," *Experientia*, vol. 20, no. 3, pp. 154–154, 1964.

[3] S. E. Barajas-Montiel and C. A. Reyes-Garcia, "Identifying pain and hunger in infant cry with classifiers ensembles," in *2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)*, 2005, vol. 2, pp. 770–775.

[4] C. Y. Chang, Y. C. Hsiao, and S. T. Chen, "Application of Incremental SVM Learning for Infant Cries Recognition," in *2015 18th International Conference on Network-Based Information Systems*, 2015, pp. 607–610.

[5] A. Rosales-Pérez, C. A. Reyes-García, J. A. Gonzalez, O. F. Reyes-Galaviz, H. J. Escalante, and S. Orlandi, "Classifying infant cry patterns by the Genetic Selection of a Fuzzy Model," *Biomedical Signal Processing and Control*, vol. 17, pp. 38–46, 2015.

[6] J. Saraswathy, M. Hariharan, W. Khairunizam, S. Yaacob, and N. Thiyagar, "Infant cry classification: time frequency analysis," in *2013 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2013, pp. 499–504.

[7] A. Zabidi, 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*, vol. 37, Á. Jobbágy, Ed. Springer Berlin Heidelberg, 2012, pp. 61–65.

[8] A. Rosales-Pérez, C. A. Reyes-García, and P. Gómez-Gil, "Genetic fuzzy relational neural network for infant cry classification," in *Pattern Recognition*, vol. 6718, J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, C. B.-Y. Brants, and E. R. Hancock, Eds. Springer Berlin Heidelberg, 2011, pp. 288–296.

[9] M. Hariharan, R. Sindhu, and S. 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, no. 2, pp. 559–569, Nov. 2012.

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

[11] J. Varallyay, Z. Benyo, A. Illenyi, Z. Farkas, and L. Kovacs, "Acoustic analysis of the infant cry: classical and new methods," in *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2004 (IEMBS'04)*, 2004, pp. 313–316.

[12] S. Orlandi, 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 (EMBC)*, 2012, pp. 2953–2956.

[13] D. Lederman, E. Zmora, S. Hauschildt, A. Stellzig-Eisenhauer, and K. Wermke, "Classification of cries of infants with cleft-palate using parallel hidden markov models," *Medical and Biological Engineering and Computing*, vol. 46, no. 10, pp. 965–975, 2008.

[14] S. Orlandi, C. A. R. Garcia, A. Bandini, G. Donzelli, and C. Manfredi, "Application of pattern recognition techniques to the classification of full-term and preterm infant cry," *Journal of Voice*, 2015.

[15] J. Saraswathy, M. Hariharan, T. Nadarajaw, W. Khairunizam, and S. Yaacob, "Optimal selection of mother wavelet for accurate infant cry classification," *Australasian Physical and Engineering Sciences in Medicine*, vol. 37, pp. 439–456, 2014.

[16] J. O. Garcia and C. A. R. Garcia, "Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks," in *International Joint Conference on Neural Networks, 2003*, 2003, pp. 3140 – 3145.

[17] M. M. Jam and H. Sadjedi, "Wavelet-based automatic cry recognition system for detecting infants with hearing-loss from normal infants," *The Journal of Engineering*, vol. 1, no. 1, pp. 1–2, 2013.

[18] L. Liu, K. Kuo, and S. M. Kuo, "Infant cry classification integrated ANC system for infant incubators," in *2013 10th IEEE International Conference on Networking, Sensing and Control (ICNSC)*, 2013, pp. 383–387.

[19] B. Saha, P. K. Purkait, J. Mukherjee, A. K. Majumdar, B. Majumdar, and A. K. Singh, "An embedded system for automatic classification of neonatal cry," in *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*, 2013, pp. 248–251.

[20] M. Hariharan, L. S. Chee, and S. Yaacob, "Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network," *Journal of Medical Systems*, vol. 36, no. 3, pp. 1309–1315, Jun. 2012.

[21] A. K. Singh, J. Mukhopadhyay, and K. S. Rao, "Classification of infant cries using epoch and spectral features," in *2013 National Conference on Communications (NCC)*, 2013, pp. 1–5.

[22] O. F. Reyes-Galaviz, S. D. Cano-Ortiz, and C. A. Reyes-García, "Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies," in *Seventh Mexican International Conference on Artificial Intelligence, 2008 (MICAI '08)*, 2008, pp. 330–335.

[23] R. Sahak, W. Mansor, Y. K. Lee, A. I. M. Yassin, and A. Zabidi, "Orthogonal least square based support vector machine for the classification of infant cry with asphyxia," in *2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI)*, 2010, vol. 3, pp. 986–990.

[24] A. Zabidi, W. Mansor, L. Y. Khuan, I. M. Yassin, and R. Sahak, "The effect of f-ratio in the classification of asphyxiated infant cries using

multilayer perceptron neural network," in *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 2010, pp. 126–129.

[25] M. Hariharan, S. Yaacob, and S. A. Awang, "Pathological infant cry analysis using wavelet packet transform and probabilistic neural network," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15377–15382, Nov. 2011.

[26] R. Cohen and Y. Lavner, "Infant cry analysis and detection," in *2012 IEEE 27th Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, 2012, pp. 1–5.

[27] L. Liu, J. He, and G. Palm, "Signal modeling for speaker identification," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996. ICASSP-96*, 1996, vol. 2, no. 3, pp. 665–668.

[28] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. New Jersey: Prentice Hall, 1993.

[29] T. Kinnunen, V. Hautamäki, and P. Fränti, "Fusion of spectral feature sets for accurate speaker identification," in *9th Conference Speech and Computer (SPECOM-2004)*, 2004, pp. 361–365.

[30] J. Harrington and S. Cassidy, *Techniques in Speech Acoustics*. Dordrecht: Kluwer Academic Publishers, 1999.

[31] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *Journal of the Acoustical Society of America*, vol. 55, no. 6, pp. 1304–1312, 1974.

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

[33] T. Kinnunen, "Spectral features for automatic text-independent speaker recognition (Licentiate's thesis)," University of Joensuu, 2003.

[34] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy," *Journal of Machine Learning Research*, vol. 5, pp. 1205–1224, 2004.

[35] M. Dash and H. Liu, "Consistency-based search in feature selection," *Artificial Intelligence*, vol. 151, no. 1, pp. 155–176, 2003.

[36] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1–2, pp. 245–271, 1997.

[37] R. C. Holte, "Very simple classification rules perform well on most commonly used datasets," *Machine Learning*, vol. 11, no. 1, pp. 63–91, 1993.

[38] I. Kononenko, "Estimating Attributes: Analysis and Extensions of Relief," in *Machine Learning: ECML-94*, Springer Berlin Heidelberg, 1994.

[39] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *International Conference on Machine Learning (ICML)*, 2003, pp. 856–863.

[40] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes in C*, 1st ed. Cambridge: Cambridge University Press, 1988.

[41] H. Liu and R. Setiono, "A probabilistic approach to feature selection - a filter solution," in *Proc 13th International Conference on Machine Learning*, 1996, vol. 96, pp. 319–327.

[42] D. E. Golberg, *Genetic algorithms in search, optimization, and machine learning*. Addion wesley, 1989, 1989.

[43] M. A. Hall, "Correlation-based feature selection for machine learning," Waikato University, New Zealand, 1999.

[44] S. S. Chiddarwar and N. R. 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*, vol. 23, no. 7, pp. 1083–1092, 2010.

[45] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, D. E. Rumelhart and J. L. McClelland, Eds. Cambridge MA: MIT Press, 1986, pp. 318–362.

[46] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, no. 2, pp. 281–294, 1998.

[47] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[48] M. Hall and G. Holmes, "Benchmarking attribute selection techniques for data mining," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 15, no. 6, pp. 1437–1447, 2003.

[49] J. Cohen, "A coefficient of agreement for nominal scales," *Educat. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[50] A. Ozcift and A. Gulten, "A robust multi-class feature selection strategy based on rotation forest ensemble algorithm for diagnosis of erythemato-squamous diseases," *Journal of Medical Systems*, vol. 36, no. 2, pp. 941–949, 2012.

[51] M. Galar, A. Fernandez, E. Barrenechea, and F. Herrera, "DRCW-OVO: Distance-based relative competence weighting combination for One-vs-One strategy in multi-class problems," *Pattern Recognition*, vol. 48, no. 1, pp. 28–42, 2015.

[52] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.