Classifying infant cry patterns by the Genetic Selection of a Fuzzy Model

Alejandro Rosales-Pérez [a], Carlos A. Reyes-García [a,*], Jesus A. Gonzalez [a], Orion F. Reyes-Galaviz [b], Hugo Jair Escalante [a], Silvia Orlandi [c,d]

[a] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Computer Science Department, Luis Enrique Erro No. 1, Tonantzintla, Puebla 72840, Mexico
[b] University of Alberta, Department of Electrical and Computer Engineering, Edmonton, AB T6G 2V4, Canada,
[c] Department of Information Engineering, University of Firenze, Firenze, Italy
[d] Department of Electrical, Electronic and Information Engineering (DEI) "Guglielmo Marconi", University of Bologna, Bologna, Italy

ARTICLE INFO

ABSTRACT

Infant crying analysis is an important tool for identifying different pathologies at a very early stage of the life of a baby. Being able to perform this task with high accuracy is therefore important and required as a medical support system to assess a baby's health. In this research we propose an automatic classification model for infant crying for early disease detection. Our model mainly consists of two phases: (a) an acoustic features acquisition from the Mel Frequency Cepstral Coefficient and the Linear Predictive Coding from signal processing and (b) the selection/creation of an optimized fuzzy model through the Genetic Selection of a Fuzzy Model (GSFM) algorithm. GSFM searches for the best model by choosing a combination of a feature selection method, a type of fuzzy processing, a learning algorithm together with its associated parameters that best fit the input data. Our approach improves the predictive accuracy on the identification of the cause of crying and clearly helps to differentiate between normal and pathological cry. Experimental results show a significant accuracy improvement when using our optimized genetic selection method for most of the cases.

1. Introduction

Infant crying is at birth the only communication means of babies, which, while very restricted, takes the roll of adults' speech. Through crying, the baby shows his/her physical and psychological state. Several studies have been performed in order to distinguish between different kinds of cries. The crying wave carries useful information, as to detect possible physical pathologies from very early stages of life. Those studies bring the opportunity of helping babies by understanding their needs or by detecting specific diseases, in which case the appropriate treatment can be provided and future complications prevented. For example, world-wide statistics indicate that from one thousand new born, 1 or 2 present deep hearing loss or severe hearing loss. Nevertheless not all of them receive diagnosis and timely treatment. This fact generates a very serious problem, because the beginning of an opportune treatment is delayed. The earlier the deafness diagnosis the better guaranteed the possibilities of rehabilitation and acquisition of language.

In cases like those the application of non-invasive tools, like infant cry analysis, to produce early diagnosis could help to provide the needed treatment.

The first works with infant cry were initiated by Wasz–Hockert since the beginnings of the 60s [56]. In 1964 the research group of Wasz–Hockert showed that the four basic types of cry can be identified by listening: pain, hunger, pleasure and birth [56]. In a recent study Sheinkopf et al. examined differences in acoustic characteristics of infant cries in a sample of babies at risk for autism and a low-risk comparison group [49]. Besides, many other studies related to this line of research and to that of automatizing the recognition of cries have been reported.

Sergio D. Cano carried out and directed several works devoted to the extraction and automatic classification of acoustic characteristics of infant cry. In one of those studies, in 1999 Cano presented a work in which he demonstrates the utility of Kohonen's Self-Organizing Maps in the classification of Infant Cry Units [14]. In [12] a radial basis function (RBF) network is implemented for infant cry classification in order to find out relevant aspects concerned with the presence of Central Nervous System (CNS) diseases. First, an intelligent searching algorithm combined with a fast non-linear classification procedure is implemented, establishing the cry
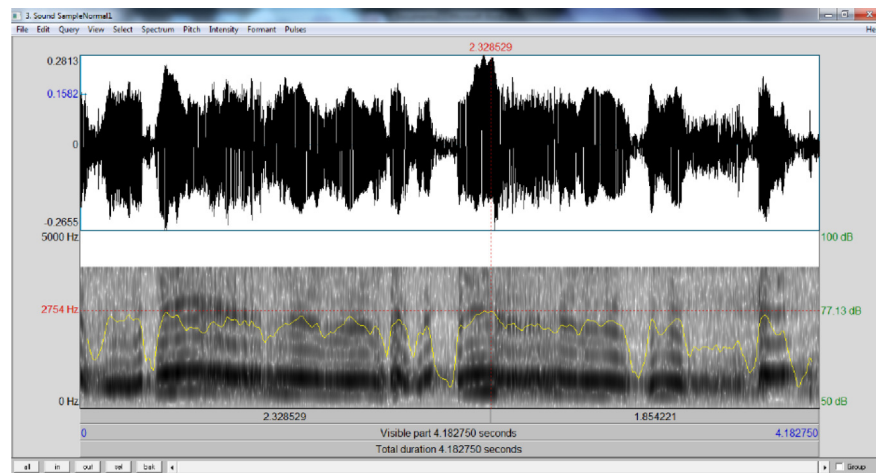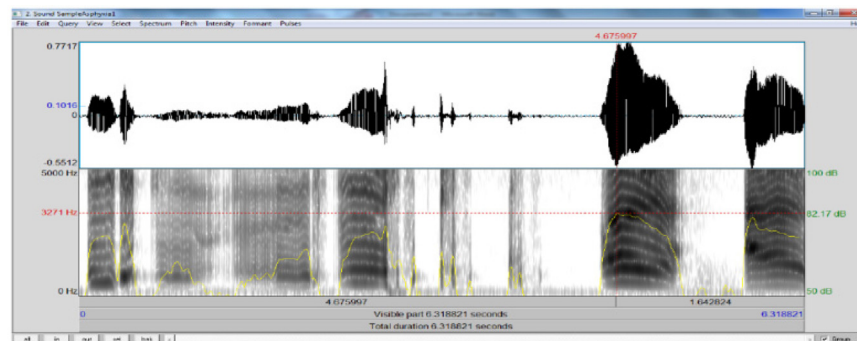

Fig. 1. Wave form and spectrogram of a normal cry.

parameters which better match the physiological status previously defined for the six control groups used as input data. Finally the optimal acoustic parameter set is chosen in order to implement a new non-linear classifier built on a RBF network, an ANN-based procedure which classifies the cry units into two categories, normal or abnormal class, as the ones shown in Figs. 1 and 2.

In 2004 a study is published in [55] which deals with classical and new methods of acoustic analysis of the infant cry, where the final goal is to detect hearing disorders according to the crying at the earliest possible moment. Manfredi et al. [38] developed, a new robust adaptive tool for new born infant cry analysis which is characterized by a high tracking capability, well suited for the signals under study. The tool performs F0 noise and resonance frequencies tracking, on signal frames of varying length. Moreover, voiced/unvoiced separation is implemented, allowing disregarding unvoiced parts of the signal where misleading results could be obtained. In a further study Manfredi et al. [37] presented a user-friendly software tool to deal with newborn infant cry signals, which allows robust tracking of main acoustic parameters on very short and time-varying signal frames.

In [31], LaGasse et al. state that the typical infant cry analysis protocol involves 30 s of crying from a single application of the stimulus. The recorded cry is submitted to an automated computer analysis system that digitizes the cry and either presents a digital spectrogram of the cry or calculates measures of cry characteristics. Another study presented by Branco et al. [11] is focused in the extraction of the characteristics of pain vocal emission of newborns during venepuncture through acoustic analysis and relate them to the Neonatal Infant Pain Scale (NIPS) and some other variables of the newborns.

(a) Asphyxia Cry
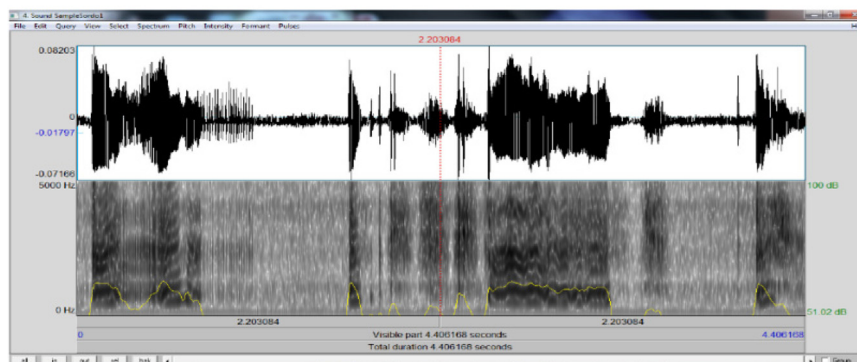
(b) Deaf Cry

Fig. 2. Wave forms and spectrograms of abnormal cry. (a) Asphyxia cry and (b) deaf.
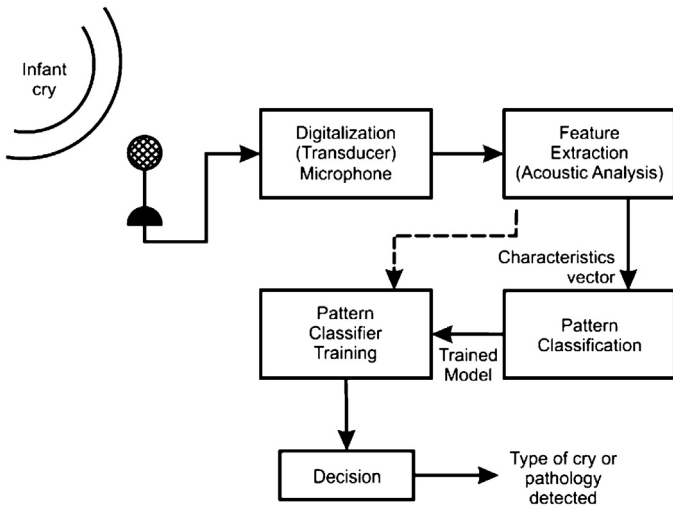
Fig. 3. Infant's cry automatic recognition process.

Fig. 4. Feature extraction process cry.

The infant cry automatic classification process is, in general, a pattern recognition problem, similar to Automatic Speech Recognition (ASR). The goal is to take the wave from the infant's cry as the input pattern, and at the end obtain the kind of cry or pathology detected on the baby [24,43]. Generally, the process of Automatic Cry Recognition is performed in two steps, see Fig. 3.

The first step is known as signal processing, or feature extraction, whereas the second is known as pattern classification. In the acoustical analysis phase, the cry signal is first normalized and cleaned, and then it is analyzed to extract the most important characteristics in function of time. The cleaning of the signal is applied in order to eliminate irrelevant and undesirable information, like background noise, channel distortion, and particular characteristics of the signal. Some of the more used techniques for the processing of the signals are those to extract: linear prediction coefficients, cepstral coefficients, pitch, intensity, spectral analysis, and Mel's filter bank. The set of obtained characteristics is represented by a vector, which, for the process purposes, represents a pattern. The set of all vectors is then used to train the classifier, which is later on evaluated on data that was not used for training the model. For the evaluation process, a set of unknown feature vectors is compared with the knowledge that the computer has to measure the classification output efficiency. When using supervised learning methods for classification, as in our case, the knowledge is utilized to measure the error between the actual and the desired output during training. This knowledge is represented by the labels attached to each sample training vector.

For our purposes, the study of the infant crying is performed through an automatic infant cry recognition process. In this work, besides applying some of the above mentioned acoustic extraction techniques, and forming the set of pattern vectors, we propose the use of a Genetic Selection method to generate a customized fuzzy model (GSFM) for the classification of infant cry. GSFM selects: the best combination of feature selection methods, the more adequate type of fuzzy processing, an appropriate learning algorithm, and its associated parameters that best fit the data. The viability of the proposed technique in the infant cry classification task is supported by the results obtained through experiments. In this paper we show the implementation stage as well as experiments and some results, in which up to 99.42% in recognition accuracy was obtained. According to our findings such accuracy rate can be obtained due to the right selection of a model totally adapted to the set of samples available for each experiment. Although the results were obtained for identifying cause of crying and differentiate between normal and pathological cry, it is our hypothesis that having available the
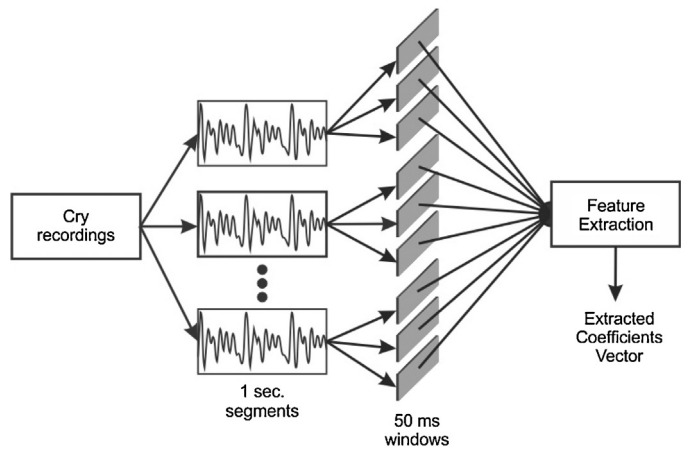
right sample data base for training, our proposed method will perform well in the early identification of other infant disorders as those belonging to the Autism Spectrum Disorders.

The remaining contents of our paper are organized as follows. In Section 2 the acoustic feature extraction methods used are described, namely Mel Frequency, Cepstral Coefficients and Linear Predictive Coding. Section 3 briefly describes the most used pattern recognition techniques and outlines some of the previous works conducting to the present model selection algorithms. It also explains the classification approach taken with particular detail on the presented Genetic Selection of a Fuzzy Model. The experiments designed and results obtained on each of them are detailed in Section 4, where comparisons with similar methods are also provided. Finally, to round up the advantages of the proposed method, Section 5 provides conclusions and outlines potential directions for future work.

2. Acoustic processing

The acoustic analysis implies the application and selection of filter techniques, feature extraction, signal segmentation, and normalization. With the application of these techniques we try to describe the signal in terms of its fundamental components. One cry signal is complex and codifies more information than the one needed to be analyzed and processed in real time applications. For this reason, in our cry recognition system we use a feature extraction function as a first plane processor. As illustrated in Fig. 4, the input is a cry signal, and its output is a vector of features that characterizes key elements of the cry's sound wave.

The acoustic features are extracted as follows (see Fig. 4): the original infant cry recordings are divided into 1-s segments, for each segment we extract 16 coefficients to every 50 ms window, generating a total of 304 features from each 1-s sample. One should note that we do not use any noise-removal algorithms prior to the initial segmentation; since infant cry is very different from voice samples, it is normally mistaken as noise and most algorithms degrade the signal to an extent that the information contained in the sound wave is eliminated.

In our study, we have been experimenting with diverse types of acoustic features, emphasizing, by their utility, Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) [32].

2.1. Mel Frequency Cepstral Coefficient

In the literature, one can find a plethora of methods and techniques used for Automatic Speech Recognition (ASR)
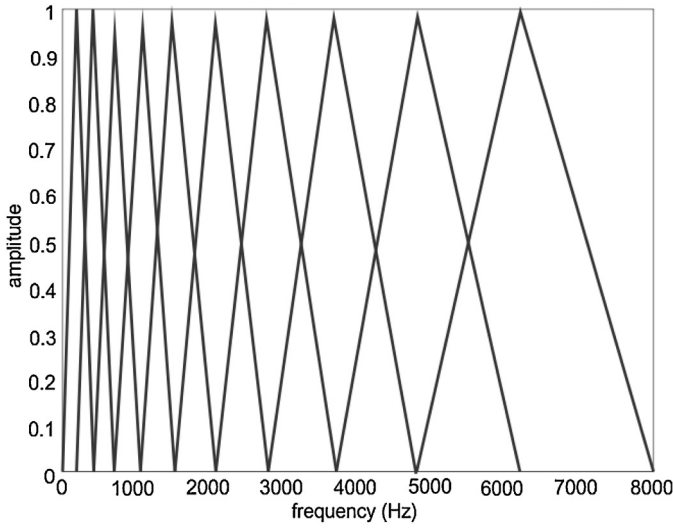
Fig. 5. The mel filter bank.

(a) Normal Cry

(b) Hipo-acoustical Cry

Fig. 6. MFCC pattern of (a) normal cry and (b) hipo-acoustical cry.

applications. Many of these works employ Mel Frequency Cepstral Coefficients (MFCC) to process the recorded sound waves [1,20,24,30,35,40,43,54]. In point of fact, since the mid-eighties MFCC has been the most widely used feature extraction method in the field of ASR [19]. The MFCC use a spectrum represented by a Mel filter bank (refer to Fig. 5) which has a frequency resolution similar to that of a human auditory system, and the hair-cell distribution inside, which has high resolution at low frequencies; the Mel Frequency scale corresponds to a linear scale below 1 kHz, and a logarithmic scale above 1 kHz (see Eq. (1)), making them useful for speech synthesis, recognition, and a better representation of sound [22].

$$F_{Mel} = \frac{100}{\log(2)} \times \left[1 + \frac{F_{Hz}}{1000} \right] \tag{1}$$

where F_{Mel} is the resulting frequency on the mel-scale measured in mels, and F_{Hz} is the normal frequency measured in Hertz (Hz). Commonly, the speech is divided into overlapping frames of 25 or 30 ms, with 10 ms overlap for the consecutive frames. Each one of these frames is processed by a Hamming window, to avoid distortion, and then a Fast Fourier Transform (FFT) is applied to each windowed excerpt. Next, the powers obtained in the last step are mapped onto the Mel filter bank by using triangular overlapping windows (see Fig. 5). A log energy output of each of the Mel frequencies is evaluated by applying the N-point discrete Fourier Transform of the discrete input signal $x(n)$:

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(\frac{-j\pi 2nk}{N} \right) \tag{2}$$

Next, an equal area filter-bank $H_i(k)$ is employed in the computation of the low-energy output:
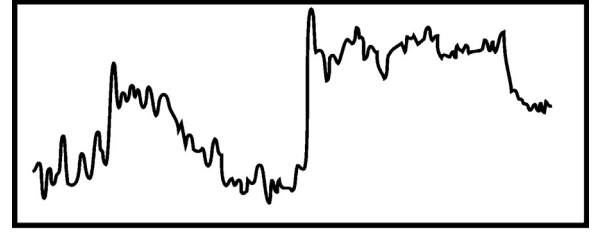
$$X_i = \log_{10} \left(\sum_{k=0}^{N-1} |X(k)| \times H_i(k) \right) \tag{3}$$

where $i = 1, 2, \ldots, M$ and M is the number of filters in the filter-bank. Finally the discrete cosine transform of the list of Mel log powers is used as a signal (refer to Eq. (4)). This last step results in the MFCC features, which are the amplitudes of the resulting spectrum.

$$C_j = \sum_{i=1}^{M} X_i \cos \left(j \times \left(i - \frac{1}{2} \right) \times \frac{\pi}{M} \right) \tag{4}$$
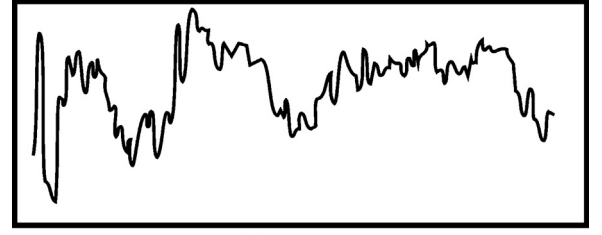
where $j = 0, 1, \ldots, J - 1$, and J is the number of MFCC that are needed. This technique is applied to the infant cry to generate the patterns that will help us recognize the type of pathology or message that the infant is trying to express. It is important to highlight that the frames used to divide the speech are different to the frames used for infant cry. Since there is no articulation of phonemes in the sound a child produces, we have more flexibility in the choice of time frames used to divide the sound waves. Moreover, it is expected that after processing the infant cry, each resulting vector represents a pattern which contains distinguishing MFCC features, as shown in Fig. 6 [43].

2.2. Linear Predictive Coding

Similar to MFCC, Linear Predictive Coding (LPC) is another method widely used in the literature of ASR [15,18,28,29,39,42,58]. Numerous researchers select LPC as their feature extraction method because it is one of the most powerful speech analysis techniques, and one of the most advantageous methods applied to encode good quality speech at low bit rates, while providing at the same time accurate estimates of the speech signal. LPC was proposed for the first time as a method for encoding human speech by the United States Department of Defense in the federal standard 1015, published in 1984. The proposed algorithm enabled the user to encode understandable speech, with a very unnatural and synthetic quality, and with a file size 20 times smaller than a MP3 file [58]. In digital signal processing, LPC can be viewed as a subset of filter theory [32].

In principle, LPC tries to imitate the resonant structure of the human vocal tract when a sound is produced [19]. In a nutshell, LPC starts with the assumption that a signal is produced by a buzzer at the end of a tube (vocal tract), with occasional hissing and popping sounds (sibilants and plosive sounds); this is a crude but a very close description of speech production. To extract the features from the sound previously produced, the input signal is divided into segments; again a Hamming window is used. Next, the LPC analysis is carried out by estimating the autocorrelation coefficients, as shown in the following expression [19,58]:

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n - k) \tag{5}$$

Table 1
Studies on infant cry classification. We differentiate the existing studies in terms of the type of cry to be recognized, the type of features extracted from the cries, and the classifier used for this aim.

Reference	Type of cry	Features	Classifier	Performance
Lederman et al. [33]	Healthy infants and no healthy	20-MFCC	Hidden Markov Models	63.00%
Orozco-Garcia and Reyes-Garcia [41]	Normal and pathological cries	LPC	Neural Networks	94.30%
Cano Ortiz et al. [13]	Normal and abnormal	Statistical	Neural Networks	85.00%
Reyes-Galaviz et al. [44]	Normal, deaf, and asphyxia	MFCC	ANFIS	96.67%
Barajas and Reyes [5]	Normal, deaf, and asphyxia	MFCC/LPC	FRNN	91.08% (overall)
Lederman et al. [34]	Clef palate	20-MFCC	Hidden Markov Model	91%
Sahak et al. [48]	Normal and asphyxia	MFCC	SVM	95.86%
Hariharan et al. [26]	Normal and pathological	Wavelet	PNN	99.00%

where $\hat{s}(n)$ is the predicted signal value, $s(n-k)$ is the previous observed values, a_k is the predictor coefficients, and p is the linear predictor order, or a linear combination of p past samples. Finally, the autocorrelation between the frames is evaluated, and the LPC analysis is performed on the resulting autocorrelation coefficients. When processing speech, usually the segments have a similar size as in the previous method in order to detect the production of phonemes. But with infant cry, we again have more flexibility when setting the amount of time used to divide the cry. At the end, we use LPC to extract features and help represent the crying signals.

2.3. Feature extraction process remarks

All the infant cry recordings are processed with the freeware program Praat v4.0.8 [10]. A major advantage of this publicly available software is that it has a scripting language, which allows the user to prepare scripts for the automatic processing of the recordings. In this study we used three scripts to perform the needed tasks: (1) a script to automatically divide each file into a desired length (in seconds) and save each segment into a raw wave file; (2) another script to open each wav segment, extract the required coefficients (e.g., MFCC or LPC) into a short text file and save each text file; and (3) since each text file comes with a header, which is not needed for the training process, we used one final script to remove the header from each text file.

Once all the text files are cleaned, we can use them for training or testing the models. The time to process a set of recordings heavily relies on the number and length of the sound files available to process. Nevertheless, while this part of the process is necessary for the whole processing of the infant cry recordings during the training and testing phases of the model, it is not necessary to repeat the same steps when the model is potentially used as a diagnostic tool. In a hypothetical real-world environment, the users will have access to the trained model, and a tool which receives a sound file (pre-recorded or acquired in-site) and returns a quick diagnostic, since the segmentation, processing, and testing of one single wave file is performed virtually in real-time.

3. Pattern recognition techniques

Several pattern recognition techniques have been used for the classification of infant cry. Among these techniques, artificial neural networks have been one of the most widely used [13,26,41,45]. With the same purpose, the use of support vector machines (SVMs) [7,48], hidden Markov models [33,34], as well as several hybrid approaches that combine fuzzy logic with neural networks [44,50–52], fuzzy logic with support vector machines [6] or evolutionary strategies with neural networks [23] have also been explored. Table 1 summarizes the characteristics of previous studies around the infant cry recognition.

Most of the previous works have reported favorable results in infant cry recognition. However, many efforts had to be devoted to the manual design of the classifiers with the intend to determine the set of adequate parameters for each technique to get the right classification of infant cry. In order to avoid useless hard work and many drawbacks, alternative approaches that combine genetic algorithms with fuzzy logic and neural networks have been proposed [5,46]. Nonetheless, most of the works only determine the parameters for a specific learning algorithm.

Choosing an algorithm to classify instances from a specific domain with high accuracy is not an easy task, the task is even harder if we want to obtain the best possible result with the available resources. There are some domains for which obtaining the best possible accuracy might not be relevant but there are others in which this is a critical issue; this is the case of some medical systems.

Finding the "best" classification method for a particular data set is the goal of a model selection algorithm. In order to effectively perform this task, the search space for a model considers different variables such as: selecting preprocessing method, features and instances that lead to better performance, as well as the classification method to use, together with its parameters.

Previous work in this area has been done as in the case of Ayat et al. [2], in which the authors optimize the kernel parameters of a support vector machine (SVM) while reducing the number of support vectors in order to decrease the generalization error. Bao et al. [4] use a particle swarm optimization algorithm to optimize the parameters of an SVM too. Chapelle et al. [16] choose the parameters of an SVM by minimizing estimates of their generalization error using a gradient descent algorithm over the set of parameters.

For the selection of models there are also multi-objective optimization approaches. In [3], the authors use Artificial Immune Systems (AIS) to optimize the parameters of the SVM using a multi-objective strategy (optimizing the kernel and penalizing the parameters of the SVM). Chatelain et al. [17] created a multi-objective and multi-model selection approach for model selection when the misclassification costs are unknown or might change over time. Suttorp and Igel [53] explore different multi-objective evolutionary approaches for SVM model selection. Li et al. [36] proposed a multi-objective uniform design search method for SVM model selection in order to reduce the computational cost of the model selection task while improving the classification ability. This model was applied to the face recognition problem.

There are some other methods that try to reduce the intensive computation time consumed by model selection algorithms. This is the case of Gorissen et al. [25] that uses a surrogate environment with adaptive sampling guided through evolution.

In this work we propose to explore the use of the Genetic Selection of a Fuzzy Model (GSFM) approach for infant cry classification. GSFM was recently proposed and it was applied to acute leukemia subtypes classification [47]. A genetic algorithm is used in GSFM for selecting the right combination of a feature selection method, the type of fuzzy processing, a learning algorithm, and their associated parameters that better fit to a data set. The goal of GSFM is to choose a combination of methods that minimizes the average error rate of the classes for a given data set. The rest of this section
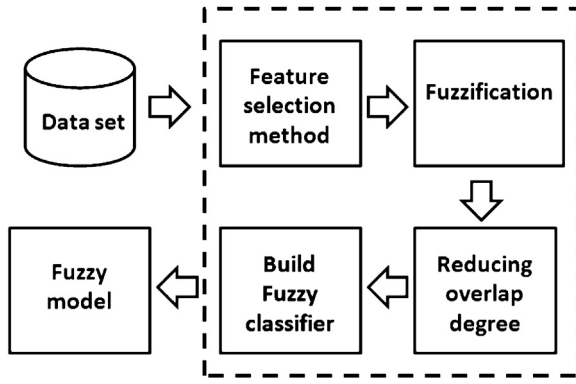
Fig. 7. Process for building a fuzzy model.

Table 2
Methods considered in GSFM. The combination of feature selection method, type of fuzzy processing, and fuzzy classifier is done with a genetic algorithm.

Method	Number of parameters	Description
Feature selection		
ReliefF	3	Ranking of features based on the ReliefF algorithm.
χ^2	1	Ranking of features based on the χ^2 statistical test.
InfoGain	1	Features are ranked according to their information gain.
Correlation	2	A subset of features is selected according to the correlation among themselves.
Fuzzy processing		
NLP	1	Defines the number of linguistic properties. It can take the values: 3 (Low, Medium, and High), 5 (Very Low, Low, Medium, High, and Very High), and 7 (Very Low, Low, More or Less Low, Medium, More or Less High, High, and Very High).
TMF	1	The type of fuzzy membership function. It can take the values of Trapezoid, Triangle, Gaussian, and Bell.
Fuzzy classifier		
FDT	1	A fuzzy decision tree.
FDF	2	A fuzzy decision forest.
FKNN	2	A fuzzy version of the K nearest neighbor algorithm.
FRNN	5	A fuzzy relational neural network classifier.

describes classification approach and in Section 4 we show experimental results on the application of GSFM to the infant-cry data set.

3.1. Classification approach

Genetic algorithms, proposed by Holland [27] in the 70s, are heuristic search techniques, which are inspired in Darwin's theory of evolution to solve problems using computational models. The genetic algorithms are based on the idea of survival of the fittest and reproduction strategies where stronger individuals have higher chances to create offsprings, and consequently are considered in the evolution process. Generally, genetic algorithms have five basic components: an encoding scheme, in a form of chromosomes or individuals, that represents the potential solutions to the problem, a form to create potential initial solutions, a fitness function to measure how close a chromosome is to the desired solution, selection operations and reproduction operators [21].

In this work we use a hybrid algorithm called Genetic Selection of a Fuzzy Model (GSFM) to perform infant cry classification. The goal of GSFM is to find an effective classification model to be used for infant cry classification. We use GSFM to find a classification model for each of the considered data sets (see Section 4.1). The obtained model was trained with a training data set, and this model was used to predict the testing set. In the following we detail the GSFM technique.

3.1.1. Genetic Selection of a Fuzzy Model

The process of the construction of a fuzzy model is shown in Fig. 7. A labeled data set is the input. Given that each sample is described by a set of N features, and that N is usually large, the first step is to reduce the dimensionality of the data set. This task is done by applying a feature selection method. Then, the subset of selected features is converted into fuzzy values, which is the fuzzification step. Next, the parameters of fuzzy membership are fitted to reduce the overlapping degree. Finally, with the fuzzy features a fuzzy classifier is built. Given a pool of feature selection methods, fuzzy processing and learning algorithms, GSFM selects the combination of them that minimizes the error.

GSFM has the advantage of considering different methods, Table 2 describes each of the methods considered by GSFM and the corresponding hyper-parameters. We considered four feature selection methods (reliefF, χ^2, information gain (InfoGain), and one based on correlation), the fuzzy pre-processing, which consists on choosing the number of linguistic properties (NLP), and the type of membership function (TMF), and four fuzzy classifiers (a fuzzy decision tree (FDT), A fuzzy decision forest (FDF), the fuzzy K nearest neighbors (FKNN), and the fuzzy relational neural network (FRNN)). Therefore, GSFM explores the models space and it picks up the

best one according to a given criterion. It is worthnoting that each model is composed by a combination of a feature selection step, a pre-processing step and the learning algorithm together with its hyper-parameters. This results in a vast model space, which could be computationally intractable. Stochastic search techniques are well-suited to deal with these cases. In particular, GSFM uses a genetic algorithm to this aim.

3.1.2. Representation

In GSFM, each model (a combination of feature selection method, pre-processing method and learning algorithm together with its hyper-parameters) is represented by a chromosome. This chromosome, also known as individual, is a 15-dimensional binary vector, as follows:

$$\mathbf{x}^{(i)} = [fs_1, fs_2, hp_{1_{fs}}, \dots, hp_{4_{fs}}, hp_{1_{fp}}, \dots,$$
$$hp_{4_{fp}}, fc_1, fc_2, hp_{1_{Fc}}, \dots, hp_{3_{fc}}] \tag{6}$$

where fs_1, fs_2 are two bits that control the feature selection method to be used; $hp_{1_{fs}}, \dots, hp_{4_{fs}}$ are the hyper-parameters for the selected feature selection method; $hp_{1_{fp}}, \dots, hp_{4_{fp}}$ are the fuzzy pre-processing's hyper-parameters, and $fc_1, fc_2, hp_{1_{Fc}}, \dots, hp_{3_{fc}}$ are the learning algorithm and its hyper-parameters.

3.1.3. Avoiding extinction

The goal of genetic operations is to create new individuals from existing ones. The selection operation, which is one of the genetic operations, selects the parents to be crossed, giving preference to the fittest individuals. However, individuals who represent a type of learning algorithm may be less selected, due to its low fitness value,

than others causing the number of these individuals to become smaller and disappear in future generations.

When the size of the population of individuals that represent a type of learning algorithm is lower than a threshold α, the surviving individuals for this type are copied, mutated and subsequently added to the population. This step is performed in order to avoid the extinction of these individuals, as well as to maintain the diversity of the different models. In order to ensure that the population size remains constant, the individuals (i.e., classification models) with a bad performance are removed, starting with the one with the worst performance, following by the second one and so on until the number of individuals is equal to the population size.

3.1.4. Fitness function

In each generation in the genetic algorithm, a population of individuals is evaluated with the fitness function, and the fitness value is used to choose the individuals for reproduction, and this process is repeated until a number of generations is reached. In GSFM, individuals are classification models, hence we need a way to evaluate the *goodness* (fitness) of a classification model, in this case we considered the balanced error rate (BER). The BER is the average of the errors incurred by a classifier on each class i and it is computed as follows:

$$\text{BER} = \frac{1}{j} \sum_{i=1}^{j} e_i \tag{7}$$

where j is the number of classes in the data set and e_i is the error rate of the classification model in the ith class, that is, the fraction of instances of class i that were misclassified by the model. Since all we have for estimating the performance of models is training data, we may compute the performance of models in these data. However, as the models were trained using such data, estimating the BER on training data directly would lead to optimistic estimates. Instead we adopted a 2-fold cross validation procedure over the training set to estimate it. In this way, training data is split into two disjoint subsets, one is used for training the model and in the other we assess the model's performance, next we interchange the subsets and repeat the process. The fitness of the model is the average BER over the two subsets. In this manner, the performance of the model is estimated on data that was not used to train the model.

The BER is in the range [0, 1], where a value of 0 indicates that a classification model correctly predicts all samples, whereas a value of 1 indicates that a classifier has incorrectly predicted all of the samples. Hence, the goal of GSFM is to minimize the BER.

4. Experiments and results

This section describes the experimental study we performed to assess the effectiveness of the GSFM method for infant cry classification. First we describe the data set considered for experimentation and next we report experimental results of GSFM.

4.1. Data set description

For our experiments we used the Baby Chillanto® infant cry database property of INAOE-CONACyT, Mexico. The infant cries were collected by recordings done directly by medical doctors, the cry samples were carefully labeled at the time of the recording with references like infant age and the reason for the cry. Then each signal wave was divided in segments of 1 s, each segment represents a sample. Then, acoustic features were obtained by means of Frequencies in the Mel scale (MFCC), by the use of the freeware program Praat v4.0.8 [9]. Every 1 s sample is divided in frames of 50 ms and from each frame (Hamming window) we

Table 3
Description of infant cry data sets.

Data set	No. samples	Samples by class
Asphyxia vs. normal and hungry	847	Asphyxia: 340 Normal and hungry: 507
Deaf vs. normal and hungry	1386	Deaf: 879 Normal and hungry: 507
Hungry vs. pain	542	Hungry: 350 Pain: 192

extract 16 coefficients, this procedure generates vectors with 304 coefficients by sample.

The infant cry corpus has 340 samples of cries of asphyxia, 192 for pain, 350 for hunger, 879 cries of babies who are deaf and 157 of normal cries. Pain and hunger cries come from normal babies, so they are also part of the normal cries collection. This corpus was used to build different binary data sets: asphyxia vs. normal and hunger, deaf vs. normal and hunger and hunger vs. pain. Table 3 shows the different data sets and the number of samples of each case. These data sets were used in our experiments.

4.2. Experimental study

We performed several experiments considering binary classification problems. First, we considered the binary problems of discriminating between asphyxia and normal[1] cries; deaf and normal; and hungry and pain cries. For each experiment, we used GSFM to determine the best model.

The evaluation was done using 10-fold cross validation. This technique divides the data set into 10 disjoint subsets, and in each fold a subset is left apart for testing and the remaining subsets for training. This process is repeated until all subsets have been used for testing and training. As evaluation metrics we used accuracy (ACC), true positive rate (TPR), true negative rate (TNR) and area under the ROC[2] curve (AUC) [8].

Table 4 shows the best results obtained in our experiments and those reported by Rosales-Pérez et al. [46]. Please note that even though other works have tackled the infant cry classification problem, a direct comparison against those methods was not performed; this is because they do not apply 10-fold cross validation. Nonetheless, accuracy percentages of 99% and of 95% are reported in [26] and [48], respectively, for asphyxia vs. normal cries; this values can be used as reference, although we emphasize that results are not comparable at all.

From Table 4 we can see that GSFM is able to generate very competitive classification models (according to the different measures). In problems 2 and 3 (deaf vs. normal and hungry vs. pain, respectively) the performance of our method is close to perfect, which evidences the usefulness of GSFM to provide models that could be used as support tools for medical diagnosis. The performance in the other problem is slightly worse, although still it is quite competitive.

Also, it can be seen from Table 4 that our approach clearly outperforms the method reported in [46]. We performed a Wilcoxon signed-rank test [57] with 95% of confidence to determine whether the differences are statistically significant or not. This test was done across the obtained results in each fold. According to this test, GSFM outperforms significantly the method in [46] in asphyxia vs. normal and deaf vs. normal data sets.

Table 5 describes the obtained models for each classification problem. For each model, the feature selection method (FSM), type of membership function (TMF), number of linguistic properties

---

[1] We considered both, the hungry and pain cries as normal ones.
[2] Receiver Operating Characteristic.

Table 4
Classification results for each experiment in percentages. Results are the average of using 10-fold cross validation. Reported results by Rosales-Pérez et al. [46] are also shown. The best result is shown in bold font for each case.

ID	Data set	Accuracy		TPR		TNR		AUC	
		GSFM	[46]	GSFM	[46]	GSFM	[46]	GSFM	[46]
1	Asphyxia vs. normal	**90.68**	88.67	85.29	**90.00**	**94.29**	87.78	**95.79**	92.85
2	Deaf vs. normal	**99.42**	97.55	**100.00**	98.75	**98.42**	95.47	**100.00**	99.75
3	Hungry vs. pain	**97.96**	96.03	**99.43**	95.59	95.26	**96.67**	**98.89**	98.35

Table 5
Selected models for infant cry data sets using GSFM.

ID	FSM	TMF	NLP	LA	Parameters
1	Correlation	Trapezoid	3	FKNN	nn = 7
					sm = correlation
2	InfoGain	Bell	7	FDT	cv = 0.87
3	InfoGain	Gauss	3	FKNN	nn = 5
					sm = chord

(NLP), the learning algorithm (LA), and its associated parameters are shown. It can be seen that different configurations were selected for the different problems, hence showing the main benefit of GSFM: the capability of selecting highly effective customized models for each approached problem.

5. Conclusion and future work

In this work we proposed a method that can be used as a medical diagnosis support tool for babies at early stages of life. This is a non-invasive technique. We described the design and implementation of GSFM which allows selecting an adequate model to differentiate between different types of cries with precise results. Among the main advantages of the adopted approach is the fact that our system releases the user from having to determine the right combination of model components, which is a huge search space since it considers the feature, preprocessing, and fuzzy method (considering its parameters) selection steps. As for the knowledge needed to obtain the classification accuracy reported, belonging to the supervised learning kind, this knowledge is provided to the generated models as the label included in each training vector. This label identifies the class to which this vector belongs to. As long as the sample vectors are correctly labeled, the set of labels can be taken as a reliable source of knowledge on which to support precise training and classification processes. Our experimental results show that our approach outperforms results reported in the literature from methods that consider a single learning algorithm.

Since customizing classification models requires to perform a search using a evolutionary algorithm, this process can be computationally expensive. However, some advantages can be remarked from this approach; for instance, it is not required a depth knowledge on machine learning algorithms to generate a reliable model. Moreover, once the model is trained, the prediction time (e.g., for the diagnosis of new cases) can be considered negligible.

As a future work we would like to explore the use of other techniques to extract features from the audio signals, as well as to test strategies such as ensembles of different models. We also want to extend the proposal to a multi-objective problem, taking into account the model complexity, and to explore alternatives to reduce the computational cost. We will also try infant cry classification as a multi-class problem.

Acknowledgment

References

[1] H. Arsikere, S. Lulich, A. Alwan, Estimating speaker height and subglottal resonances using MFCCs and GMMs, IEEE Signal Process. Lett. 21 (2) (2014) 159–162.
[2] N. Ayat, M. Cheriet, C. Suen, Automatic model selection for the optimization of SVM kernels, Pattern Recogn. 38 (10) (2005) 1733–1745.
[3] I. Aydin, M. Karakose, E. Akin, A multi-objective artificial immune algorithm for parameter optimization in support vector machine, Appl. Soft Comput. 11 (1) (2011) 120–129.
[4] Y. Bao, Z. Hu, T. Xiong, A PSO and pattern search based memetic algorithm for SVMs parameters optimization, Neurocomputing 117 (0) (2013) 98–106.
[5] S. Barajas, C. Reyes, Your fuzzy relational neural network parameters optimization with a genetic algorithm, in: The 14th IEEE International Conference on Fuzzy Systems, 2005 (FUZZ'05), IEEE, 2005, pp. 684–689.
[6] S. Barajas-Montiel, C. Reyes-García, Fuzzy support vector machines for automatic infant cry recognition, in: D.-S. Huang, K. Li, G. Irwin (Eds.), Intelligent Computing in Signal Processing and Pattern Recognition. Vol. 345 of Lecture Notes in Control and Information Sciences, Springer, Berlin, Heidelberg, 2006, pp. 876–881.
[7] S. Barajas-Montiel, C. Reyes-García, E. Arch-Tirado, M. Mandujano, Improving baby caring with automatic infant cry recognition, in: K. Miesenberger, J. Klaus, W. Zagler, A. Karshmer (Eds.), Computers Helping People with Special Needs. Vol. 4061 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006, pp. 691–698.
[8] J. Beck, E. Shultz, et al., The use of relative operating characteristic (ROC) curves in test performance evaluation, Archiv. Pathol. Lab. Med. 110 (1) (1986) 13.
[9] P. Boersma, D. Weenink, Praat v. 4.0.8: A System For Doing Phonetics by Computer, Institute of Phonetic Sciences of the University of Amsterdam, 2002.
[10] P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer [Computer program]. Version 5.3.82, 2010, Retrieved from http://www.praat.org/
[11] A. Branco, S.M. Fekete, L.M. Rugolo, M.I. Rehder, The newborn pain cry: descriptive acoustic spectrographic analysis, Int. J. Pediatr. Otorhinolaryngol. 71 (4) (2007) 539–546.
[12] S. Cano Ortiz, D. Escobedo Beceiro, T. Ekkel, A radial basis function network oriented for infant cry classification. Progress in Pattern Recognition, Image Anal. Appl. (2004) 15–36.
[13] S. Cano Ortiz, D. Escobedo Beceiro, T. Ekkel, A radial basis function network oriented for infant cry classification, in: A. Sanfeliu, J. Martínez Trinidad, J. Carrasco Ochoa (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. Vol. 3287 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2004, pp. 15–36.
[14] S. Cano-Ortiz, D. Escobedo-Becerro, Clasificación de unidades de llanto infantil mediante el mapa auto-organizado de koheen. I Taller AIRENE sobre Reconocimiento de Patrones con Redes Neuronales, Universidad Católica del Norte, Chile, 1999, pp. 24–29.
[15] C.-F. Chan, S.-P. Chui, Efficient codebook search procedure for vector-sum excited linear predictive coding of speech, Electron. Lett. 30 (22) (1994) 1830–1831.
[16] O. Chapelle, V. Vapnik, O. Bousquet, S. Mukherjee, Choosing multiple parameters for support vector machines, Mach. Learn. 46 (1–3) (2002) 131–159.
[17] C. Chatelain, S. Adam, Y. Lecourtier, L. Heutte, T. Paquet, A multi-model selection framework for unknown and/or evolutive misclassification cost problems, Pattern Recogn. 43 (3) (2010) 815–823.
[18] R. Christiansen, C. Rushforth, Detecting and locating key words in continuous speech using linear predictive coding, IEEE Trans. Acoust. Speech Signal Process. 25 (5) (1977) 361–367.
[19] M. Cutajar, E. Gatt, I. Grech, O. Casha, J. Micallef, Comparative study of automatic speech recognition techniques, IET Signal Process. 7 (1) (2013) 25–46.
[20] A. De la Torre, A. Peinado, J. Segura, J. Perez-Cordoba, M. Benitez, A. Rubio, Histogram equalization of speech representation for robust speech recognition, IEEE Trans. Speech Audio Process. 13 (3) (2005) 355–366.
[21] A. Engelbrecht, Computational Intelligence: An Introduction, Wiley, 2007.
[22] T. Fukada, K. Tokuda, T. Kobayashi, S. Imai, An adaptive algorithm for mel-cepstral analysis of speech, in: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-92), vol. 1, 1992, pp. 137–140.
[23] O. Galaviz, C. García, Infant cry classification to identify hypo acoustics and asphyxia comparing an evolutionary-neural system with a neural network system, in: A. Gelbukh, A. de Albornoz, H. Terashima-Marín (Eds.), MICAI 2005: Advances in Artificial Intelligence. Vol. 3789 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2005, pp. 949–958.
[24] J. Garcia, C. Reyes Garcia, Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward

neural networks, in: Proceedings of the International Joint Conference on Neural Networks, vol. 4, 2003, pp. 3140–3145.

[25] D. Gorissen, T. Dhaene, F.D. Turck, Evolutionary model type selection for global surrogate modeling, J. Mach. Learn. Res. 10 (December) (2009) 2039–2078.

[26] M. Hariharan, S. Yaacob, S.A. Awang, Pathological infant cry analysis using wavelet packet transform and probabilistic neural network, Expert Syst. Appl. 38 (12) (2011) 15377–15382.

[27] J. Holland, Adaptation in Artificial and Natural Systems, The University of Michigan Press, Ann Arbor, 1975.

[28] R.A. Kazi, V.M. Prasad, J. Kanagalingam, C.M. Nutting, P. Clarke, P. Rhys-Evans, K.J. Harrington, Assessment of the formant frequencies in normal and laryngectomized individuals using linear predictive coding, J. Voice 21 (6) (2007) 661–668.

[29] W. Kleijn, D. Krasinski, R. Ketchum, An efficient stochastically excited linear predictive coding algorithm for high quality low bit rate transmission of speech, Speech Commun. 7 (3) (1988) 305–316.

[30] S.G. Koolagudi, D. Rastogi, K.S. Rao, Identification of language using mel-frequency cepstral coefficients (MFCC), Procedia Eng. 38 (0) (2012) 3391–3398.

[31] L.L. LaGasse, A.R. Neal, B.M. Lester, Assessment of infant cry: acoustic cry analysis and parental perception, Ment. Retard. Dev. Disabil. Res. Rev. 11 (1) (2005) 83–93.

[32] D. Lederman, Automatic classification of infants' cry (M.Sc. thesis), Ben Gorion University, 2002, pp. 1–11.

[33] D. Lederman, A. Cohen, E. Zmora, K. Wermke, S. Hauschildt, A. Stellzig-Eisenhauer, On the use of hidden markov models in infants' cry classification, in: The 22nd Convention of Electrical and Electronics Engineers in Israel, IEEE, 2002, pp. 350–352.

[34] D. Lederman, E. Zmora, S. Hauschildt, A. Stellzig-Eisenhauer, K. Wermke, Classification of cries of infants with cleft-palate using parallel hidden Markov models, Med. Biol. Eng. Comput. 46 (2008) 965–975.

[35] V. Leutnant, A. Krueger, R. Haeb-Umbach, A new observation model in the logarithmic mel power spectral domain for the automatic recognition of noisy reverberant speech, IEEE/ACM Trans. Audio Speech Lang. Process. 22 (1) (2014) 95–109.

[36] W. Li, L. Liu, W. Gong, Multi-objective uniform design as a SVM model selection tool for face recognition, Expert Syst. Appl. 38 (6) (2011) 6689–6695.

[37] C. Manfredi, L. Bocchi, S. Orlandi, L. Spaccaterra, G. Donzelli, High resolution cry analysis in preterm newborn infants, Med. Eng. Phys. 31 (5) (2009) 528–532.

[38] C. Manfredi, V. Tocchioni, L. Bocchi, A robust tool for newborn infant cry analysis, in: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006 (EMBS'06), 2006, pp. 509–512.

[39] D. Manolakis, G. Carayannis, A lattice–ladder structure for multipulse linear predictive coding of speech, IEEE Trans. Acoust. Speech Signal Process. 35 (2) (1987) 228–231.

[40] K. Murty, B. Yegnanarayana, Combining evidence from residual phase and MFCC features for speaker recognition, IEEE Signal Process. Lett. 13 (1) (2006) 52–55.

[41] J. Orozco-García, C. Reyes-García, A study on the recognition of patterns of infant cry for the identification of deafness in just born babies with neural networks, in: A. Sanfeliu, J. Ruiz-Shulcloper (Eds.), Progress in Pattern Recognition, Speech and Image Analysis. Vol. 2905 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2003, pp. 342–349.

[42] D. O'Shaughnessy, Linear predictive coding, IEEE Potentials 7 (1) (1988) 29–32.

[43] O. Reyes-Galaviz, S. Cano-Ortiz, C. Reyes-Garcia, 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.

[44] O. Reyes-Galaviz, E. Tirado, C. Reyes-Garcia, Classification of infant crying to identify pathologies in recently born babies with ANFIS, in: K. Miesenberger, J. Klaus, W. Zagler, D. Burger (Eds.), Computers Helping People with Special Needs. Vol. 3118 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2004, p. 625.

[45] O. Reyes-Galaviz, A. Verduzco, E. Arch-Tirado, C. Reyes-García, Analysis of an infant cry recognizer for the early identification of pathologies, in: G. Chollet, A. Esposito, M. Faundez-Zanuy, M. Marinaro (Eds.), Nonlinear Speech Modeling and Applications. Vol. 3445 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2005, pp. 20–22.

[46] A. Rosales-Pérez, C. Reyes-García, P. Gómez-Gil, Genetic fuzzy relational neural network for infant cry classification, in: J. Martínez-Trinidad, J. Carrasco-Ochoa, C. Ben-Youssef Brants, E. Hancock (Eds.), Pattern Recognition. Vol. 6718 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2011, pp. 288–296.

[47] A. Rosales-Pérez, C. Reyes-García, P. Gómez-Gil, J. Gonzalez, L. Altamirano, Genetic selection of fuzzy model for acute leukemia classification, in: I. Batyrshin, G. Sidorov (Eds.), Advances in Artificial Intelligence. Vol. 7094 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2011, pp. 537–548.

[48] R. Sahak, W. Mansor, Y. Lee, A. Yassin, A. Zabidi, Performance of combined support vector machine and principal component analysis in recognizing infant cry with asphyxia, in: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2010, pp. 6292–6295.

[49] S.J. Sheinkopf, J.M. Iverson, M.L. Rinaldi, B.M. Lester, Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder, Autism Res. 5 (5) (2012) 331–339.

[50] I. Suaste-Rivas, A. Daz-Méndez, C. Reyes-García, O. Reyes-Galaviz, Hybrid neural network design and implementation on FPGA for infant cry recognition, in: P. Sojka, I. Kopecek, K. Pala (Eds.), Text, Speech and Dialogue. Vol. 4188 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2006, pp. 703–709.

[51] I. Suaste-Rivas, O. Reyes-Galaviz, A. Diaz-Mendez, C. Reyes-Garcia, A fuzzy relational neural network for pattern classification, in: A. Sanfeliu, J. Martínez Trinidad, J. Carrasco Ochoa (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. Vol. 3287 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2004, pp. 275–299.

[52] I. Suaste-Rivas, O. Reyes-Galviz, A. Diaz-Mendez, C. Reyes-Garcia, Implementation of a linguistic fuzzy relational neural network for detecting pathologies by infant cry recognition, in: C. Lematre, C. Reyes, J. González (Eds.), Advances in Artificial Intelligence – IBERAMIA 2004. Vol. 3315 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2004, pp. 953–962.

[53] T. Suttorp, C. Igel, Multi-objective optimization of support vector machines, in: Multi-Objective Machine Learning, Springer, 2006, pp. 199–220.

[54] C. Valentini-Botinhao, J. Yamagishi, S. King, R. Maia, Intelligibility enhancement of hmm-generated speech in additive noise by modifying mel cepstral coefficients to increase the glimpse proportion, Comput. Speech Lang. 28 (2) (2014) 665–686.

[55] G. Varallyay, Z. Benyo, A. Illenyi, Z. Farkas, 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), vol. 1, 2004, pp. 313–316.

[56] O. Wasz-Höckert, T. Partanen, V. Vuorenkoski, K. Michelsson, E. Valanne, The identification of some specific meanings in infant vocalization, Cell. Mol. Life Sci. 20 (3) (1964) 154.

[57] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bull. 1 (6) (1945) 80–83.

[58] J.-D. Wu, B.-F. Lin, Speaker identification based on the frame linear predictive coding spectrum technique, Expert Syst. Appl. 36 (4) (2009) 8056–8063.