Application of Pattern Recognition Techniques to the Classification of Full-Term and Preterm Infant Cry

*Silvia Orlandi, †Carlos Alberto Reyes Garcia, *,‡Andrea Bandini, §Gianpaolo Donzelli, and *Claudia Manfredi,
*§*Firenze*, ‡*Bologna, Italy*, †*Puebla, Mexico*

Summary: Objectives. Scientific and clinical advances in perinatology and neonatology have enhanced the chances of survival of preterm and very low weight neonates. Infant cry analysis is a suitable noninvasive complementary tool to assess the neurologic state of infants particularly important in the case of preterm neonates. This article aims at exploiting differences between full-term and preterm infant cry with robust automatic acoustical analysis and data mining techniques.

Study design. Twenty-two acoustical parameters are estimated in more than 3000 cry units from cry recordings of 28 full-term and 10 preterm newborns.

Methods. Feature extraction is performed through the *BioVoice* dedicated software tool, developed at the Biomedical Engineering Lab, University of Firenze, Italy. Classification and pattern recognition is based on genetic algorithms for the selection of the best attributes. Training is performed comparing four classifiers: Logistic Curve, Multilayer Perceptron, Support Vector Machine, and Random Forest and three different testing options: full training set, 10-fold cross-validation, and 66% split.

Results. Results show that the best feature set is made up by 10 parameters capable to assess differences between preterm and full-term newborns with about 87% of accuracy. Best results are obtained with the Random Forest method (receiver operating characteristic area, 0.94).

Conclusions. These 10 cry features might convey important additional information to assist the clinical specialist in the diagnosis and follow-up of possible delays or disorders in the neurologic development due to premature birth in this extremely vulnerable population of patients. The proposed approach is a first step toward an automatic infant cry recognition system for fast and proper identification of risk in preterm babies.

Key Words: Infant cry analysis–Preterm newborn–Automatic classification–Acoustical parameters–Feature selection.

INTRODUCTION

Scientific and clinical advances in perinatology and neonatology have enhanced the chances of survival of preterm and very low birth weight neonates. Clinical and ethical demands have emerged regarding the early assessment of these vulnerable children to detect markers of possible developmental deficits. The studies have shown that an early detection of the risks for vulnerable children would allow implementing prevention strategies and policies in childhood.[1]

The crying of newborns and infants is a functional expression of basic biological needs, and emotional or psychological conditions such as hunger, cold, pain, cramps, and even joy.[2] It requires a coordinated effort of several brain regions, mainly brainstem and limbic system and is linked to the breath and the lung mechanisms. Its characteristics reflect the development and possibly the integrity of the central nervous system. Thus, infant cry analysis is a suitable noninvasive complementary tool to assess the physical state of infants particularly important in the case of preterm neonates. Specifically, the distinction between a regular wailing and one with anomalies is of clinical interest.

Being cost-effective and contactless, the study of the newborn infant crying has had an outstanding growth in the last decades. Several studies concern both the subjective auditory analysis of voice and speech and the automatic acoustical analysis in adults. However, with respect to the newborn cry, few automated methods exist, some of them based on classical approaches such as Fourier transform and autocorrelation analysis[2–6] and other on parametric techniques.[7–9] Such methods allow estimating the main acoustical features such as the frequency of vibration of the vocal folds, the vocal tract resonance frequencies, the cry duration, and so forth. However, the high variability of the newborn cry signal has limited the development of methods as robust as those devoted to the analysis of adult voice for its automated analysis.

Preterm infants and infants with neurologic conditions may have different cry characteristics when compared to healthy full-term infants. Qualitative and quantitative research on possible neurophysiological differences between full-term and preterm infants has been carried on since the 1980s. Most of the studies investigated possible differences in infant gender, their neurophysiological maturity, and risks of brain damage for preterm infants caused by deoxygenation due to prolonged crying.[10–16]

In general, the automatic infant cry classification process is a pattern recognition problem. From the infant's cry (input pattern), the goal was to classify the kind of cry or pathology detected on the baby.

Accepted for publication August 7, 2015.

From the *Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy; †Computer Science Department, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico; ‡Department of Electrical, Electronic and Information Engineering (DEI) "Guglielmo Marconi", University of Bologna, Bologna, Italy; and the §Department of Neuroscience, Psychology, Drug Research and Child Health, Università degli Studi di Firenze, Firenze, Italy.

Address correspondence and reprint requests to Claudia Manfredi, Department of Information Engineering, Università degli Studi di Firenze, Via di S. Marta 3, 50139 Firenze, Italy. E-mail: claudia.manfredi@unifi.it

Journal of Voice, Vol. 30, No. 6, pp. 656-663
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.08.007

Several authors have proposed classification methods for a wide range of pathologies. Reyes et al[17–19] have investigated normal, deaf, and asphyxiating newborns using classification methods such as neural networks, genetic selection, and fuzzy logic. Some of these powerful techniques are successfully applied also in the present work and will be shortly described in the next section. Poel and Ekkel[20] present results concerning the classification of newborn cry into normal and hypoxia-related disorder using Radial Basis Function Neural Networks with 85% overall classification performance. Lederman et al[21] propose the classification of infants with cleft palate on the basis of parallel Hidden Markov Models with an average of 91% correct classification rate in a subject- and age-dependent experiment. Mijovic et al[22] propose Empirical Mode Decomposition techniques to assess the existence and extent of decoupling in term neonates and its possible relation to clinical pain expression. Sahak et al[23] applied Combined Support Vector Machine (SVM) and Principal Component Analysis to recognize the infant cries with asphyxia with a classification accuracy of 95.86%. Zabidi et al[24] applied a new algorithm to optimize Mel frequency cepstrum coefficients to extract an optimal feature set for the diagnosis of hypothyroidism in infants using a Multilayer Perceptron (MLP) neural network. Nonaka et al[25] used a Hidden Markov Model architecture. The algorithm yields up to 95% classification precision (86% average) to identify expiratory and inspiratory phases from the baby cries. In the study by Hariharan et al,[26,27] a General Regression Neural Network is used as a classifier for discriminating normal cry signals and pathologic cry signals from deaf infants and babies with asphyxia. Etz et al[28] propose a decision tree to classify infant cries to find differences between infants with normal development, hearing impairment, and unilateral cleft lip and palate, whereas Alaie et al[29] apply Gaussian mixture models to distinguish between healthy full-term and premature infants and those with specific medical problems with a true positive (TP) rate of 80.77% and a true negative rate of 86.96%. Finally, in the study by Singh et al,[30] three different types of infant cries are considered: hunger, pain, and wet diaper. Gaussian mixture models are used to classify the previously mentioned cries.

This nonexhaustive list of studies shows that the newborn cry contains specific features that enable the classification of various diseases and conditions by automated techniques.

This article aims at highlighting and differentiating the features of newborn cry in the two groups of healthy term and preterm infants. To this aim, a robust tool (*BioVoice*) specifically developed for the acoustical analysis of newborn cry is applied[16] that provides 22 acoustical parameters both in time and frequency domain. The proposed classification method allows pointing out the relevant cry features capable to assess differences between preterm and full-term newborns with about 87% of accuracy. This result may be a valuable aid to the diagnosis of possible delays or disorders in the neurologic development due to premature birth in this extremely vulnerable population of patients.

MATERIALS AND METHODS

Recording protocol

Newborn cry signals were recorded in a quiet room of neonatology unit (S. Giovanni di Dio Hospital, Firenze, Italy) and neonatal intensive care unit (Children Hospital A. Meyer, Firenze, Italy), respectively, for term and preterm infants. All parents of the infants were native Italian speakers, and they signed informed consent to participate in this study.

A unidirectional microphone (Shure SM58; Shure Inc. Chicago, IL) was positioned at a fixed distance (25 cm) from the baby's mouth and equipped with Tascam US-144 (TEAC Corp. Montebello, CA) portable audio/musical instrument digital interface (96 kHz/24-bit recording). Recordings were stored on a multimedia laptop in a single channel audio track. The sampling rate was $Fs = 44$ kHz with 16-bit resolution. Each recording lasts at least 15 minutes to include several cry sequences. A cry sequence is defined here as a set of multiple cries, the so-called cry units (CUs). Cry sequences are spaced one from the other by a suitable amount of time, lasting more than 30 seconds.[31,32]

A CU is defined here as a high-energy voiced frame lasting at least 260 ms. This choice comes from literature where different time lengths are considered for CUs, ranging from 60 to 500 ms.[21,33–36] In fact, CUs of very short duration do not allow the assessment of some relevant features such as their melodic shape. Moreover, inspiratory sounds that have duration less than 200 ms must be disregarded.[33]

Database

We recorded 28 healthy term newborns (TN, 17 boys and 11 girls) and 10 preterm infants (PN, 5 boys and 5 girls). Gestational age of TN at birth was between 37 weeks and 2 days and 42 weeks; the weight was between 2400 and 4250 g. Gestational age of PN at birth was between 23 weeks and 5 days and 34 weeks. The weight at birth was between 590 g and 2700 g. At the recording time (20–30 days after birth), the PN gestational age was between 35 weeks and 1 day and 43 weeks and 1 day; the weight ranged between 1380 and 2430 g.

The TN infants were recorded within the first 2 days of life, whereas PN newborns could be recorded only about 20–30 days after birth, because of their long staying in the incubator. Specifically, the PN infants were recorded within the first 45 days after the normal end of pregnancy (37 weeks). PN infants were recorded in a quiet room in the Neonatal Intensive Care Unit at the Children Hospital A. Meyer, Firenze, Italy. The newborns were hospitalized because of prematurity and other diseases: respiratory distress, obstructive sleep apnea, hypoglycemia, bowel obstruction, bleeding, and anemia. TN infants were recorded in a quiet room of the neonatology clinic at S. Giovanni di Dio Hospital, Firenze, Italy. They did not suffer from any disease.

We collected an audio recording for each infant of at least 1 hour of duration consisting of at least 10% of crying. Recordings were performed before the afternoon feeding. Full-term infants usually cried without stimulation, whereas for preterm infants, sometimes a little solicitation was made on the foot sole to start crying. From the whole recording, we manually selected 2 or 3 minutes of crying. Audio analysis was performed with *BioVoice* on all CUs.

Signal processing

Acoustic analysis of infant's cry is often performed through commercial[37] or free software.[38] However, such tools are not specifically developed for the analysis of high-pitched quasi-stationary signals as newborn cries are. Thus, their use may fail in correctly estimated parameters such as fundamental frequency F_0 and vocal tract resonance frequencies F_1, F_2, and F_3 unless specific ranges are manually set.[39]

Thanks to its high-resolution characteristics cry analysis is carried out here through the *BioVoice* tool.[16] *BioVoice* is applicable to the analysis of a wide range of voice signals also lasting several minutes. It has been successfully compared with most commonly used software tools on synthesized signals.[7,8,32,40] *BioVoice* provides a user-friendly interface for uploading the audio file(s), selecting the category (adult male or female, infant, etc) and the type of analysis (adult, singer, newborn, etc). Several signals can be uploaded and analyzed sequentially with a considerable time saving. *BioVoice* does not require any manual setting thus being well suited also for nonexpert users.

BioVoice automatically detects CUs using a robust voiced/unvoiced (V/UV) selection procedure.[32] This method has proved successful in preventing improper splitting of a single voiced frame into several parts. This frequently occurs in the case of irregular and quasi-stationary signals as newborn infant cries are. On each CU, F_0 is estimated by means of a two-step procedure that was shown to outperform other methods.[7,8] Its strength comes from the adaptive procedure implemented for the local definition of the length of each signal frame on which the acoustic parameters are estimated: the higher the F_0, the shorter the length of the frame. On each CU, *BioVoice* computes the number of estimated F_0 values. This number is variable both due to the varying frame length and because the outliers (the values of F_0 outside the range 200–1050 Hz) are removed. The first three resonance frequencies of the vocal tract (F_1, F_2, and F_3) are estimated by peak picking in the power spectral density obtained by a parametric approach that was found more robust and with higher resolution capability than the traditional fast Fourier transform–based technique.[7,8] Moreover, *BioVoice* computes the number of CUs, the vocalic percentage, and the number and length of the voice breaks in each recording.

Features extraction

BioVoice provided 5182 CUs for TN and 1662 CUs for PN babies. Thus, for classification, the same number of 1662 CUs was used for both groups, randomly selected in the TN group. On each CU, the following 22 attributes (parameters) that gained great scientific interest in the last years[31,33,41] were estimated (CU length is given in seconds, values concerning F_0–F_3 in Hz): CU length, F_0 median, F_0 mean, F_0 standard deviation (F_0 std), F_0 minimum (F_0 min), F_0 maximum (F_0 max), number of estimated F_0 values, F_1 median, F_1 mean, F_1 standard deviation (F_1 std), F_1 minimum (F_1 min), F_1 maximum (F_1 max), F_2 median, F_2 mean, F_2 standard deviation (F_2 std), F_2 minimum (F_2 min), F_2 maximum (F_2 max), F_3 median, F_3 mean, F_3 standard deviation (F_3 std), F_3 minimum (F_3 min), F_3 maximum (F_3 max).

Classification

The procedure used for the classification of preterm and term infant cries is referred to as data mining. Once the features of the data set are estimated, the following steps are performed: class allocation, attributes selection, classification, and performance evaluation. In data mining, the information is arranged in a table. Each row corresponds to a CU that is described by 22 parameters (features described by 22 attributes) extracted with *BioVoice* and one feature that represents the class: preterm (PN) and term newborns (TN). Class allocation is reported in the first and the last half of the last column, respectively, for TN and PN. Each column represents a property (attribute). In other words, the rows of the table contain the parameters for each CU and the columns the attributes. The rows are the features that must be trained. The size of the table is thus N $\times$ 23, where N is the number of CUs.

Figure 1 shows the data mining process used for classification.

After the first three steps (audio recording, set up of the data set, and extraction of features) are carried out, the table with attributes and classes is set up, the selection of the best attributes is carried out, and finally the classification is performed.

The procedure was carried out using algorithms implemented in Waikato Environment for Knowledge Analysis (WEKA).[42] WEKA is an open-source software issued under the GNU General Public License. It is a collection of machine learning algorithms for data mining tasks. WEKA contains tools for data preprocessing, classification, regression, clustering, association rules, and visualization. In this work, algorithm for attributes selection and data classification were used.

Class allocation.
Each CU is assigned to a class: either preterm newborn (PN) or term newborn (TN).

Select attributes.
To reduce the processing time, it is desirable to reduce the size of the attribute vector (22 acoustic parameters estimated in each CU) without degrading the efficiency of the classification performance.

In WEKA, to select attributes, one has to choose a search method and an attribute evaluator. The search method consists of a search algorithm. In this work, we used the genetic search method on the basis of a genetic algorithm. Genetic algorithms

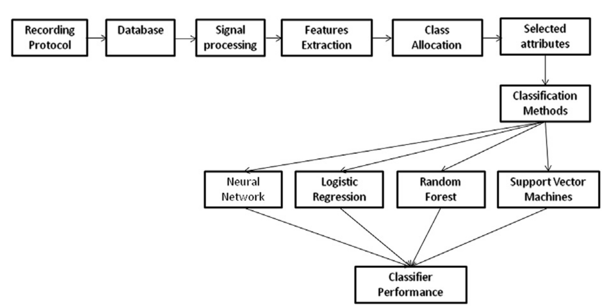

FIGURE 1. Flowchart of the data mining process.

are a family of computational models inspired by biological evolution which encode a potential solution to a specific problem on a simple chromosome-like data structure and apply recombination operators to the structures to preserve relevant information. Each candidate solution has a number of attributes that can change or be altered.[42,43]

The attribute evaluator specifies the weight of each attribute for the correct classification of CUs. We applied the default Attribute Evaluator in WEKA named CfsSubsetEval. This method evaluates the worth of a subset of attributes by considering the individual predictive ability of each one along with the degree of redundancy between the attributes. Selected attributes are thus highly correlated with the class (PN and TN) and have low correlation with each other. The 22 attributes estimated with *BioVoice* on the whole data set underwent the WEKA procedure for the selection of the best ones. The first step was thus the select attributes command with the option CfsSubsetEval for attribute evaluator and genetic search for the search method.

Classification methods. With classification, the learned models from training data are evaluated using a different data set to determine whether the models can be generalized to new cases.[44,45]

To compare the two classes of infants (PN and TN) with WEKA, the classification was performed applying the following methods: Random Forest, MLP, SVM, and Logistic curve that are shortly described in the following sections.

Random Forest. Random Forests are weighted combinations of tree classifiers that use a random selection of attributes to build the decision taken at each node. Random forests are built by combining through averaging the predictions of several decision trees, each one trained in isolation. The forest consists of K trees. In this work, $k = 10$. The method selects a random number N of attributes for all trees. After the trees have grown, new samples are classified by each tree and their results are combined, giving a membership probability for each class.[46]

Multilayer Perceptron. The architecture of an MLP[47–49] consists of a three-layer feed-forward neural network: one input, one hidden, and one output layer. With WEKA, the user can select

- the number of hidden layer;
- the learning rate factor, that is the amount the weights are updated;
- the momentum applied to the weights during updating;
- the number of epochs through all the records in the training set;
- the validation threshold, to terminate validation testing: its value dictates how many times in a row the validation set error can get worse before training is terminated;
- the number of random seeds, to initialize the random number generator: random numbers are used for setting the initial weights of the connections between nodes and also for shuffling the training data.

In this work, the parameters for the MLP are: hidden layer = 1; learning rate = 0.3; momentum = 0.2; number of epochs = 500; validation threshold = 20; and random seed = 0.

Support Vector Machines. Kernel methods are a class of algorithms for data mining whose best known member is the SVM. The Sequential Minimal Optimization (SMO) is one of the most popular algorithms for classification by SVM[50] and is implemented in WEKA. To obtain proper probability estimates, SMO uses the option that fits logistic regression models to the outputs of the SVM. In our work, a polynomial kernel is used.

Logistic regression. Logistic regression is a well-known technique for classification that describes the relationship between a dichotomous dependent variable (classes) and a set of independent variables (features).[42] In this work, the WEKA default parameters for the maximum number of iterations and the ridge were used.

Classifier performance

The evaluation of the performance of classifiers is an important point in pattern recognition because it helps to understand the quality of an algorithm and to adjust its parameters. There are several metrics for evaluating the predictive performance of classifiers. Here, we applied the following: receiver operating characteristic (ROC), precision, and F-measure.

ROC curve is widely used to calculate the trade-off between TP and false positive (FP) rates. Precision helps to find how many of the classified cases are correct thus giving a measure of the performance. Performance can be measured through the so-called F-measure that is the harmonic mean of precision and sensitivity.

RESULTS

In this section, the results of the analysis and classification of the CUs in the recorded cry signals are presented according to the flowchart in Figure 1.

The first step was the automated detection of the CUs with the Voiced/Unvoiced (V/UV) selection procedure implemented in *BioVoice*. As explained in the section Materials and methods, the acoustical analysis was performed on the same number of 1662 CUs both for PN and TN newborns. On each CU, the attributes F_0, F_1, F_2, and F_3 are estimated with *BioVoice* that also provides their minimum, maximum, mean, median, and standard deviation. Thus, a total of 22 acoustic parameters (attributes) are obtained.

Table 1 summarizes the mean values of the 22 parameters over the whole data set for both TN and PN newborns. The results show that all the parameters have lower mean values in TN with respect to PN. The t test confirms a statistically significant difference between the two groups.

The WEKA command classify was used to build a model of the relationships between the set of attributes and the corresponding class (TN and PN). A procedure consisting of several trials was applied, with a twofold aim: minimize the number of significant attributes and optimize the percentage of success in the classification.

With WEKA, the select attributes command was applied with the option CfsSubsetEval for attribute evaluator and genetic search for the search method. The procedure resulted in the following eight attributes out of the full set: mean of F_0,

TABLE 1.
Mean Values of the 22 Attributes Estimated With *BioVoice* on a Set of 1662 CUs for Term (TN) and Preterm (PN) newborns

Attributes	TN	PN
1. Length (s)	0.89	0.79
2. $F0$ median (Hz)	410.69	449.62
3. $F0$ mean (Hz)	414.09	451.12
4. $F0$ std (Hz)	72.79	87.67
5. $F0$ min (Hz)	242.94	266.94
6. $F0$ max (Hz)	579.13	647.87
7. Number of $F0$ values (a.u.)	86.11	82.21
8. $F1$ median (Hz)	1025.69	1530.54
9. $F1$ mean (Hz)	1062.23	1524.04
10. $F1$ std (Hz)	358.47	364.14
11. $F1$ min (Hz)	498.24	817.45
12. $F1$ max (Hz)	2470.97	2439.56
13. $F2$ median (Hz)	3176.78	3689.65
14. $F2$ mean (Hz)	3392.18	3791.14
15. $F2$ std (Hz)	906.38	914.00
16. $F2$ min (Hz)	2217.97	2354.52
17. $F2$ max (Hz)	6614.65	6642.10
18. $F3$ median (Hz)	5921.64	6822.52
19. $F3$ mean (Hz)	6204.94	6879.19
20. $F3$ std (Hz)	1435.88	1561.55
21. $F3$ min (Hz)	4177.32	4337.12
22. $F3$ max (Hz)	10978.52	11268.45

median, mean, minimum and maximum of F_1, median and mean of F_2, and median of F_3.

Classification was performed through the Random Forest and Logistic Regression methods. For testing, we used the following options: full training set, percentage split at 66%, and 10-fold cross-validation. This step is indeed crucial to select the best classification with different attributes. Classification was applied first both to the whole set of 22 attributes and to a subset of 16 attributes: mean, median, maximum and minimum of F_0, F_1, F_2, and $F3$. This subset was considered as it makes use of the same parameters (mean, median, maximum, and minimum) for both F_0 and F_1–F_3 that is a consistent set of attributes.

In a second step, we used only the eight best attributes listed previously. However, in this case, a worsening in the results for both the two classifiers (Random Forest and Logistic Regression) was found. Thus, to recover the classification accuracy, we added two out of the previously eliminated attributes (median of F_0 and mean of F_3) to the eight parameters. With these 10 parameters, we obtained an improvement in the overall results.

Table 2 summarizes the results obtained with 22, 16, 8, and 10 attributes applying the Logistic Regression and Random Forest methods.

The best results are obtained with 10 attributes and Random Forest with 87.34% of proper classification (highlighted in bold). Notice the small difference with the other sets of attributes and in particular with the case of 22 attributes. Thus, the choice of 10 attributes would be preferable as it allows reducing the computation time.

Tables 3 and 4 summarize the rate of TP and FP results, the precision, the F-measure, and the ROC area obtained with the Random Forest method, respectively, with 22 and 10 attributes. Again, using only 10 attributes instead of 22 does not worsen the results that are comparable ranging between 0.85 and 0.99 for TP and 0.01 and 0.14 for FP.

This result allowed selecting the following 10 best attributes: mean and median of F_0, median, mean, minimum, and maximum of F_1, and median and mean of F_2 and of F_3.

To assess the results, the following classification algorithms were compared: Random Forest, SVMs, MLP, and Logistic regression (Logistic). These methods were tested with the percentage split (66%) option and 10 attributes. Again, the best results were obtained applying the Random Forest classifier though with slight differences with the other approaches. The results are reported in Table 5. The best result 0.941 is highlighted in bold.

On the basis of these results, we carried out a more detailed analysis with the 10 best attributes and the Random Forest classifier. The 1662 CUs for each class (thus 3324 instances) were randomly divided into two groups: a training set (2194 CUs) and a test set (1130 CUs) corresponding to 66% and 34%, respectively, of the whole set, and the WEKA “split“ option was applied.

In Table 6, the confusion matrix shows the best results obtained with the Random Forest method. The results with Random Forest classification algorithm are very good both as far as the values of the performance measures, and the statistical

TABLE 2.
Comparison of the Results of Accuracy of the Logistic Regression and Random Forest Classifiers Under Different Testing Options (Full Training Set, 10-Fold Cross-Validation, and 66% Split) and 22, 16, 8, and 10 Attributes

Number of Attributes	Training Logistic, Full Training Set (%)	Classification Logistic, Cross-Validation (%)	Classification Logistic, Percentage Split (%)	Training Random Forest, Full Training Set (%)	Classification Random Forest, Cross-Validation (%)	Classification Random Forest, Percentage Split (%)
22	81.016	80.505	80.796	99.548	86.702	87.079
16	79.663	79.512	81.504	99.398	86.221	86.460
8	78.971	78.880	80.266	99.489	84.988	86.195
10	79.362	79.121	80	99.187	85.649	**87.345**

TABLE 3.
Performance Metrics for the Random Forest Algorithm With 22 Attributes

Classifier	TP Rate	FP Rate	Precision	F-Measure	ROC Area
Random Forest, full training set	0.99	0.01	0.99	0.99	1
Random Forest, 10-fold validation	0.86	0.13	0.86	0.86	0.94
Random Forest, percentage split (66%)	0.87	0.12	0.87	0.87	0.93

TABLE 5.
Comparison of the Performance Measures for the Classification Algorithms Considered (Random Forest, Support Vector Machines [SVMs], Multilayer Perceptron, and Logistic Regression [Logistic]) and the 66% Percentage Split

Classifier	TP Rate	FP Rate	Precision	F-Measure	ROC Area
Random Forest	0.873	0.126	0.874	0.873	**0.941**
Logistic	0.843	0.153	0.847	0.843	0.918
SMO	0.843	0.154	0.846	0.843	0.845
Multilayer Perceptron	0.893	0.106	0.894	0.894	0.934

errors are concerned with 86.6% of correct classification for TN. A slightly lower percentage of correct classification (85.2%) occurs with PN. This might be due to the greater irregularity and lower intensity of crying in PN with respect to TN as well as possible disturbances for the more difficult conditions of recording performed in the intensive care unit.

DISCUSSION

The aim of this work was the assessment of a reliable set of acoustical parameters of newborn cry for the classification of preterm (PN) and term (TN) infants through the application of the most adequate classifier.

The first step, of fundamental importance for the subsequent ones, was the acoustical analysis of the neonatal cry. We applied here the software tool *BioVoice* entirely developed at the Biomedical Engineering Lab, University of Firenze, Italy, that enables the robust analysis of this kind of signals. In addition to avoiding the inevitable human errors due to manual selection, *BioVoice* allows processing a considerable amount of data in a very short time with obvious advantages for clinical applications. It has allowed obtaining in an automatic way a proper distinction between the voiced and nonvoiced parts of a significant number of audio recordings carried out in Florentine pediatric clinics. The acoustical analysis was made on about 7000 CUs. *BioVoice* provided 22 acoustic parameters both in the time domain and in the frequency domain.

With the proposed classification approach, the following 10 best parameters were detected: mean and median of F_0, median, mean, minimum and maximum of F_1, median and mean of F_2 and of F_3. Notice that F_2 and F_3 are among the 10 best attributes. These parameters are seldom considered in literature because of the difficulty of obtaining their reliable estimate, especially in the case of neonatal cry. Thanks to the high resolution and robustness features of *BioVoice* they could be successfully included among the attributes used in the classification procedure proposed here.

The parameter values for PN infants were found generally higher than those of the TN newborns. The t test applied to the whole 22 parameters (including the "best" 10) showed significant differences ($P < 0.05$) between PN and TN cry. In particular, for the 10 selected parameters, all the differences were highly significant ($P < 0.01$) except for F_0 ($P = 0.04$).

We recall that preterm infants were recorded between 35 weeks and 43 weeks of gestational age, close to that of the term newborns. The results thus show that there is a difference between term and preterm newborns even when the preterm reaches a gestational age similar or equal to that of the term infant. This might indicate a delay lasting beyond the normal gestational age in the development of neuromotor control for the preterm baby that thus would require a longer time to fully recover.

We point out that the procedure used for classification, although based on open-source software tool, was arranged and carried out rigorously, comparing several options and solutions. To the authors' knowledge, this was never done before.

TABLE 4.
Performance Metrics for the Random Forest Algorithm With 10 Attributes

Classifier	TP Rate	FP Rate	Precision	F-Measure	ROC Area
Random Forest, full training set	0.992	0.008	0.992	0.992	1
Random Forest, 10-fold validation	0.856	0.144	0.857	0.856	0.927
Random Forest, split (66%)	0.873	0.126	0.874	0.873	0.941

TABLE 6.
Confusion Matrix of the Test Set Classified With Random Forest 66% Percentage Split

Random Forest—Confusion Matrix		
TN	PN	Class
485	75	TN
84	486	PN

This allowed us to make the best choice not only of the parameters but also of the selection criterion for the classification which therefore could be used and standardized for clinical applications of this type of signals. Specifically, the different experiments performed allowed assessing the robust behavior of the Random Forest algorithm in classifying infant cry thanks to its appropriateness for managing high-dimensional data and because it can handle continuous, categorical, and binary data. Given that the best overall results were obtained with the Random Forest classifier, we suggest this pattern recognition model for classification of newborn cry features.

As regards the specific application, the main strengths of the present work are as follows:

- Once the signal is recorded, the acoustical parameters are estimated without any manual intervention;
- Recordings were made according to a specific protocol: they are consistent as regards the newborn age and concern hunger cry;
- The classification procedure is rigorous but easily manageable also by a nonexpert user;
- With reference to the existing literature, results are extremely good with 87% of correct classification.

CONCLUSIONS

The classification of the neonatal cry is a completely noninvasive and inexpensive method that can provide useful clinical information on the neurologic status of the newborn. The research study proposed here was focused on the definition of possible differences in neonatal cry among the category of healthy term newborns and that of premature babies who typically undergo to the risk of neurodevelopmental disorders. To this aim, the most significant acoustical parameters were detected with the *BioVoice* tool, ensuring good discrimination of the characteristics of the neonatal cry between healthy term and preterm infants. The classification was carried out using algorithms implemented in WEKA.

The results showed that this differentiation is achieved with high accuracy on the basis of a limited but specific set of 10 significant acoustical parameters that concern not only the vibration of the vocal folds but also the anatomical and physiological characteristics of the vocal apparatus of the newborn. These 10 cry features might convey important additional information to assist the clinical specialist in the diagnosis and follow-up of possible delays or disorders in the neurologic development due to premature birth in this extremely vulnerable population of patients.

The results show that the Random Forest classifier has the best performance along the different experiments. The high and consistent results obtained with the selected model, supported by the different performance measures applied show that the proposed approach is reliable. When applied to acoustic parameters obtained with a reliable software tool, this classification procedure could provide a valuable support to the perceptive analysis made by the clinician reducing the required amount of time often prohibitive in daily clinical practice. Therefore, the proposed approach might be a first step toward an automated infant cry recognition system for fast and proper identification of risk in preterm babies.

Moreover, it is a first step toward the assessment of normative ranges of the newborn cry acoustical parameters.

A future development could concern additional classification experiments with the proposed methodology and the comparison of infants with neurologic disorders to healthy babies.

Acknowledgments

The authors thank the students A. Palazzo and A. Bigazzi for their contribution in the preliminary steps of this work.

REFERENCES

1. Stefaneli PSZ, Formiga CKMR, Linhares MBM. Early neurobehavioral development of preterm infants. *Psicol Reflex Crit*. 2013;26:202–211.
2. Michelsson K, Michelsson O. Phonation in the newborn, infant cry. *Int J Pediatr Otorhinolaryngol*. 1999;49(Suppl 1):S297–S301.
3. Fuller BF. Acoustic discrimination of three types of infant cries. *Nurs Res*. 1991;40:156–160.
4. Laufer MZ, Horii Y. Fundamental frequency characteristics of infant non-distress vocalization during the first twenty-four weeks. *J Child Lang*. 1977;4:171–184.
5. Reggiannini B, Sheinkopf SJ, Silverman HF, Li X, Lester BM. A flexible analysis tool for the quantitative acoustic assessment of infant cry. *J Speech Lang Hear Res*. 2013;56:1416–1428.
6. Sirviö P, Michelsson K. Sound-spectrographic cry analysis of normal and abnormal newborn infants. *Folia Phoniatr Logo*. 1976;28:161–173.
7. Fort A, Ismaelli A, Manfredi C, Bruscaglioni P. Parametric and non-parametric estimation of speech formants: application to infant cry. *Med Eng Phys*. 1996;18:677–691.
8. Fort A, Manfredi C. Acoustic analysis of newborn infant cry signals. *Med Eng Phys*. 1998;20:432–442.
9. Kheddache Y, Tadj C. Resonance frequencies behavior in pathologic cries of newborns. *J Voice*. 2015;29:1–12.
10. Michelsson K, Järvenpää AL, Rinne A. Sound spectrographic analysis of pain cry in preterm infants. *Early Hum Dev*. 1983;8:141–149.
11. Lester BM. Developmental outcome prediction from acoustic cry analysis in term and preterm infants. *Paediatrics*. 1987;80:529–534.
12. Robb MP, Goberman AM, Cacace AT. An acoustic template of newborn infant crying. *Folia Phoniatr Logo*. 1997;49:35–41.
13. Goberman AM, Robb MP. Acoustic examination of preterm and full-term infant cries-the long-time average spectrum. *J Speech Lang Hear Res*. 1999;42:850–861.
14. Bellieni CV, Sisto R, Cordelli DM, Buonocore G. Cry features reflect pain intensity in term newborns: an alarm threshold. *Pediatr Res*. 2004;55:142–146.
15. Michelsson K, Eklund K, Leppanen P, Lyytinen H. Cry characteristics of 172 healthy 1-to 7-day-old infants. *Folia Phoniatr Logo*. 2002;5:190–200.
16. Manfredi C, Bocchi L, Orlandi S, Spaccaterra L, Donzelli GP. High-resolution cry analysis in preterm newborn infants. *Med Eng Phys*. 2009;31:528–532.
17. Reyes-Galaviz OF, Tirado EA, Reyes-Garcia CA. *"Classification of Infant Crying to Identify Pathologies in Recently Born Babies With ANFIS,"* Computers Helping People with Special Needs. Berlin Heidelberg: Springer; 2004:408–415.
18. Cano OSD, Escobedo DI, Suaste I, Ekkel T, Reyes Garcia CA. *"A Combined Classifier of Cry Units with New Acoustic Attributes,"* Progress in Pattern Recognition, Image Analysis and Applications. Berlin Heidelberg: Springer; 2006:416–425. 4225.
19. Amaro-Camargo E, Reyes-García CA. *"Applying Statistical Vectors of Acoustic Characteristics for the Automatic Classification of Infant Cry,"* Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues. Berlin Heidelberg: Springer; 2007:1078–1085.

20. Poel M, Ekkel T. Analyzing infant cries using a committee of neural networks in order to detect hypoxia related disorder. *Int J Artif Intell Tools*. 2006;15:397–410.
21. Lederman D, Zmora E, Hauschildt S, Stellzig-Eisenhauer A, Wermke K. Classification of cries of infants with cleft-palate using parallel hidden Markov models. *Med Biol Eng Comput*. 2008;46:965–975.
22. Mijovic B, Silva M, Van den Bergh BRH, Allegaert K, Aerts JM, Berckmans D, Van S. Huffel assessment of pain expression in infant cry signals using empirical mode decomposition. *Methods Inf Med*. 2010;49:448–452.
23. Sahak R, Mansor W, Lee YK, Yassin AIM, Zabidi A. "Performance of combined support vector machine and principal component analysis in recognizing infant cry with asphyxia," *Proc. 32nd IEEE EMBS Int. Conf.*, 2010;6292–6295.
24. Zabidi A, Khuan LY, Mansor W, Yassin IM, Sahak R. "Optimization of MFCC parameters using particle swarm optimization for diagnosis of infant hypothyroidism using multi-layer perceptron," *Proc. 32nd IEEE EMBS Int. Conf.*, 2010;1417–1420.
25. Aucouturier JJ, Nonaka Y, Katahira K, Okanoya K. Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models. *J Acoust Soc Am*. 2011;130:2969–2977.
26. Hariharan M, Chee LS, Yaacob S. Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network. *J Med Syst*. 2012;36:1309–1315.
27. Hariharan M, Sindhu R, Yaacob S. Normal and hypoacoustic infant cry signal classification using time-frequency analysis and general regression neural network. *Comput Methods Programs Biomed*. 2012;108:559–569.
28. Etz T, Reetz H, Wegener C. A classification model for infant cries with hearing impairment and unilateral cleft lip and palate. *Folia Phoniatr Logop*. 2012;64:254–261.
29. Alaie HF, Tadj C. Cry-based classification of healthy and sick infants using adapted boosting mixture learning method for gaussian mixture models. *Model Simulat Eng*. 2012;2012:55–65.
30. Singh AK, Mukhopadhyay J, Rao KS, Viswanath K. Classification of infant cries using dynamics of epoch features. *Int J Intell Syst*. 2012;22:351–364.
31. LaGasse LL, Neal AR, Lester BM. Assessment of infant cry: acoustic cry analysis and parental perception. *Ment Retard Dev Disabil Res Rev*. 2005;11:83–93.
32. Orlandi S, Dejonckere PH, Schoentgen J, Lebacq J, Rruqja N, Manfredi C. Effective pre-processing of long term noisy audio recordings. An aid to clinical monitoring. *Biomed Signal Proces Control*. 2013;8:799–810.
33. Díaz MAR, García CAR, Robles LCA, Altamirano JEX, Mendoza AV. Automatic infant cry analysis for the identification of qualitative features to help opportune diagnosis. *Biomed Signal Proces Controll*. 2012;7:43–49.
34. Robb MP, Crowell DH, Dunn-Rankin P. Sudden infant death syndrome: cry characteristics. *Int J Pediatr Otorhinolaryngol*. 2013;77:1263–1267.
35. Robb M, Sinton-White H, Kaipa R. Acoustic estimates of respiration in the pain cries of newborns. *Int J Pediatr Otorhinolaryngol*. 2011;75:1265–1270.
36. Wermke K, Birr M, Voelter C, Shehata-Dieler W, Jurkutat A, Wermke P, Stellzig-Eisenhauer A. Cry melody in 2-month-old infants with and without clefts. Cleft palate craniofac. *Cleft Palate Craniofac J*. 2011;48:321–330.
37. Available at: http://www.kayelemetrics.com. Accessed April 20, 2015.
38. Available at:http://www.fon.hum.uva.nl/praat. Accessed April 20, 2015.
39. Boersma P. In: Podesva R, Sharma D, eds. *Acoustic Analysis in Research Methods in Linguistics*. Cambridge, UK: Cambridge University Press; 2014.
40. Rruqja N, Dejonckere PH, Cantarella G, Schoentgen J, Orlandi S, Barbagallo SD, Manfredi C. Testing software tools with synthesized deviant voices for medicolegal assessment of occupational dysphonia. *Biomed Signal Process Contr*. 2014;13:71–78.
41. Sheinkopf SJ, Iverson JM, Lester BM. Atypical cry acoustics in 6-month-old infants at risk for autism spectrum disorder. *Autism Res*. 2012;5:331–339.
42. Available at: http://www.cs.waikato.ac.nz/ml/weka. Accessed April 20, 2015.
43. Holland J. *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor, MI, USA: University of Michigan Press; 1975.
44. Michie D, Spiegelhalter D, Taylor C. *Machine Learning, Neural and Statistical Classification*. Hemel Hempstead, Hertfordshire, UK: Ellis Horwood; 1994.
45. Jensen R, Shen Q. In: *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*, vol. 8. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2008.
46. Skurichina M, Duin RPW. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Anal Appl*. 2002;5:121–135.
47. Ripley BD. *Pattern Recognition and Neural Networks*. Cambridge, UK: Cambridge University Press; 1996.
48. Bishop CM. *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press; 1995.
49. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev*. 1958;65.
50. Schölkopf B, Smola AJ. *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press; 2002.