Chaos, Solitons and Fractals 167 (2023) 112972

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Optimal tuning of support vector machines and *k*-NN algorithm by using Bayesian optimization for newborn cry signal diagnosis based on audio signal processing features

Salim Lahmiri [a,b,*], Chakib Tadj [b], Christian Gargour [b], Stelios Bekiros [c,d]

[a] Department of Supply Chain and Business Technology Management, John Molson School of Business, Concordia University, Montreal, Canada
[b] Department of Electrical Engineering, École de Technologie Supérieure, Montreal, Canada
[c] FEMA, University of Malta, MSD 2080 Msida, Malta
[d] LSE Health, London School of Economics and Political Science (LSE), London WC2A2AE, UK

ARTICLE INFO

Keywords:
Newborn cry
Mel-frequency cepstral coefficients
Auditory-inspired amplitude modulation
Prosody
Support vector machines
k-Nearest neighbors
Bayesian optimization

ABSTRACT

Recently, the number of machine learning models used to classify cry signals of healthy and unhealthy newborns has been significantly increasing. Various works have already reported encouraging classification results; however, fine-tuning of the hyper-parameters of machine leaning algorithms is still an open problem in the context of newborn cry signal classification. This paper proposes to use Bayesian optimization (BO) method to optimize the hyper-parameters of Support Vector Machine (SVM) with radial basis function (RBF) kernel and *k*-nearest neighbors (*k*NN) trained with different audio features separately or combined; namely, mel-frequency cepstral coefficients (MFCC), auditory-inspired amplitude modulation (AAM), and prosody. Particularly, the chi-square test is applied to each set of features to retain the ten most significant ones used to train optimal classifiers. The accuracy, sensitivity, and specificity of each experimental model are computed following the standard 10-fold cross-validation protocol. One of the contributions is an improvement over previous works on newborn cry signal classification used to distinguish between healthy and unhealthy ones over the same database, in terms of performance. The best model is the SVM trained with AAM ten most significant features achieved 83.62 % ± 0.022 accuracy, 59.18 % ± 0.0469 sensitivity, and 93.87 % ± 0.0190 specificity followed by *k*NN trained with ten most features from MFCC, AAM, and prosody to obtain 82.88 % ± 0.0144 accuracy, 55.34 % ± 0.0350 sensitivity, and 94.42 % ± 0.0075 specificity. These results outperformed existing works validated on the same database. In addition, optimally tuned SVM and *k*NN are fed with a restricted number of selected patterns so as the processing time for training and testing is significantly limited. This means that the RBF-SVM-BO classifier trained with AAM ten most significant features is more able to distinguish between healthy and unhealthy newborns.

1. Introduction

The analysis of voice acoustics was considered in various biomedical engineering applications; including analysis of cry records and classification by using machine learning models to discriminate between full-term and preterm infants [1], assessment of voice quality recovery in patients suffering from cysts and polyps treated with micro-laryngoscopic direct exeresis [2], characterizing neonatal disease; namely, hypoxic-ischemic encephalopathy; based on spectral features [3], and investigating the effect of vowel context on voice quality measured by cepstral peak prominence smoothed and sample entropy [4].

Besides, the medical diagnosis of newborn diseases based on cry signal analysis and classification is a cost-effective and non-invasive solution that may provide promising accurate performance. Indeed, newborn cry signal analysis and classification is an up-to-date approach to build computer-aided diagnosis (CAD) systems used to distinguish between healthy and unhealthy newborns. In this regard, several CAD

* Corresponding author at: Department of Supply Chain and Business Technology Management, John Molson School of Business, Concordia University, Montreal, Canada.
E-mail addresses: salim.lahmiri@concordia.ca (S. Lahmiri), chakib.tadj@etsmtl.ca (C. Tadj), christian.gargour@etsmtl.ca (C. Gargour), stelios.bekiros@um.edu.mt, s.bekiros@lse.ac.uk (S. Bekiros).

https://doi.org/10.1016/j.chaos.2022.112972
Received 2 November 2022; Accepted 30 November 2022

systems have been proposed and evaluated by biomedical engineering researchers as the topic is attracting a growing interest.

Indeed, most of the existing works used standard audio processing features typically estimated in frequency, time and cepstral domains to train machine learning (ML) classifiers to distinguish between healthy and unhealthy cry signals of newborns [5–15]. Very early studies include detection of hypoxia-related disorder using Radial Basis Function Neural Networks (FBFNN) with 85 % accuracy [5] and detection of asphyxia with a classification accuracy of 95.86 % by combining principal component analysis and support vector machines [6]. Relevant studies found in the last decade were also interesting. For instance, the authors employed the Multiple Mixtures of Gaussian (MMG) algorithm to segment newborn cry records; then, Mel Frequency Cepstral Coefficients (MFCC) are estimated from the segmented records and used to train Hidden Markov Models (HMM) which achieved 83.79 % accuracy [7]. In [8], the authors examined the association between acoustic measurements (resonance frequencies; for instance) and pathologies to identify the most relevant ones. They found that the distributions of acoustic cry acoustics statistically and significantly vary with the pathology of newborn. In another study, dynamic and static features derived from Mel-Frequency Cepstral Coefficients (MFCC) of both expiratory and inspiratory cry vocalizations were used to train various classifiers; namely, multilayer perceptron (MLP) using the back-propagation algorithm, probabilistic neural networks (PNN) and a support vector machine (SVM) [9]. The MLP, SVM with MLP kernel and PNN achieved respectively 91.68 %, 90.41 %, and 89.93 % maximum accuracy rate depending on the Gaussian Mixture Model adaptation method used in experiments. In [10], the authors considered the problem of cry signal segmentation and classification to distinguish between expiratory and inspiratory phases. In this regard, the original cry signal is decomposed in time and frequency domains from which features are extracted and used to train Gaussian mixture models (GMM) and HMM. They respectively achieved 8.9 % and 11.06 % error classification rates. The effectiveness of acoustic features (fundamental frequency glide and resonance frequencies dysregulation) and conventional features (mel-frequency cestrum coefficients) on the performance of PNN when used to classify healthy and pathological cry signals was examined in [11]. The best result obtained is 88.71 % for the correct classification of healthy preterm newborns and 82 % for correct classification of unhealthy full-term newborns.

In recent studies, short-term and long-term features from different timescales were combined to train SVM to distinguish between healthy newborn and those suffering from respiratory distress syndrome to achieve 68.40 % correct classification rate [12]. In an interesting study [13], deep learning feedforward neural networks (DFFNN), linear SVM, Naïve Bayes (NB), and PNN were trained with cepstrum-based coefficients of the original newborn cry signals and validated on expiration and inspiration sets using ten-fold cross-validation protocol. The DFFNN, SVM, NB, and PNN respectively yielded to a correct classification rate of 99.92 % $\pm$ 0.00, 61.15 % $\pm$ 0.04, 58.11 % $\pm$ 0.01, and 56.71 % $\pm$ 0.01 on expiration set. Besides, when validated on inspiration set, the DFFNN, linear SVM, NB, and PNN respectively obtained 100 %, 59.57 % $\pm$ 0.01, 55.46 % $\pm$ 0.02, and 52.63 % $\pm$ 0.05 correct classification rate.

Very recently, the authors in [13] extended their work to compare the performance of various deep learning neural networks; including the DFFNN, long short-term memory (LSTM) neural networks, and convolutional neural networks (CNN) under various neural networks architectures [14]. The highest accuracy was respectively obtained by CNN, DFFNN and LSTM. In [15], PNN and linear SVM were trained by three different feature sets including MFCC set, auditory-inspired amplitude modulation (AAM) set, and prosody set composed of tilt, intensity, and rhythm features. The linear SVM outperformed the PNN in terms of accuracy under MFCC (76.50 % versus 68.90 %), AAM (75.75 % versus 70.70 %), and prosody (61.50 % versus 52.10 %). The fusion of MFCC and AAM yielded to the best correct classification rate obtained by linear SVM (78.70 %) and PNN (77.90 %). In another very recent study, the authors in [16] compared the performance SVM under various kernels, different models of decision trees, and variants of discriminant analysis by using data in [14] and principal component analysis for dimension reduction of MFCC features set. The SVM with quadratic kernel and trained with all features achieved the highest F-score (86 %) on expiration set whilst the quadratic discriminant analysis trained with tilt features set yield to the highest F-score (83.90 %) in inspiration set.

Besides the design of CAD systems for analysis and classification of newborn cry signals, other recent studies focused on use of statistical mechanics methods for analysis and characterization of healthy and unhealthy cry signals of newborns [17,18]. Indeed, statistical mechanics-based measures are useful to describe nonlinear dynamics in the underlying cry signal for better understanding of its physiology [17,18]. For instance, approximate entropy (*AppEn*) and correlation dimension (*CD*) were estimated in cepstrum domain of the original newborn cries [17]. Then, Student *t*-test, *F*-test, and two-sample Kolmogorov-Smirnov test were applied to estimated populations of *AppEn* and *CD* to check whether they are different across healthy and unhealthy cries of newborns. It was found that *AppEn* and *CD* are statistically different across healthy control and unhealthy newborns for both expiration and inspiration sets. In [17], bootstrap wavelet leaders method was employed to examine multifractals in newborns and Student *t*-test was applied to check presence of differences between cries of healthy and unhealthy subjects. It was found that newborn cry signals show strong evidence of high complexity under healthy conditions than under unhealthy conditions and that the distributions of multifractal features are statistically dissimilar across healthy and unhealthy newborns. To sum up, the authors in [17,18] concluded that complexity measures derived from statistical mechanics greatly help understanding oscillations in cries of healthy and unhealthy newborns under expiration and inspiration conditions.

Bring in mind that most recent studies [15,16] have already reported encouraging classification results; however, accuracy still needs to be improved. One of the approaches to achieving a more effective CAD system for newborn cry signal analysis and classification is to adopt an efficient algorithm for fine-tuning of the classifier. Indeed, the main goal of implementing an optimization algorithm is to determine the values of the optimal training parameters to faster and improve the learning ability of the classifier. Thus, the optimal values of the hyper-parameters have a positive impact on the performance of the classifier.

In this regard, the main purpose of the current study is to design various CAD systems to distinguish between healthy and unhealthy newborns based on analysis and classification of the acoustics of their cries. Specifically, we propose to use Bayesian optimization (BO) method to optimize the hyper-parameters of the SVM with radial basis function (RBF) kernel and *k*-nearest neighbors (*k*NN), all trained with different audio acoustic features separately or combined; precisely, MFCC, AAM, and prosody. More specifically, only the most significant patterns from each set are selected by a statistical filter (Chi- square; for instance) are used to train each optimal classifier.

We rely of acoustic patterns as they are good descriptors of any sound, easy to measure and to interpret, and was successfully applied in previous works dealing with newborn cry analysis and classification [7–12,15,16]. Besides, we consider the SVM [19,20] thanks to its ability to minimize the upper bound on the generalization error based on the structural risk minimization principle [20] which makes it very successful in various biomedical engineering problems; including, detection of heart murmur [21], hemorrhage in retina [22], Parkinson's disease [23], and Alzheimer's disease [24]. Also, the *k*NN algorithm [25] is considered in the current as it allows for local approximation of any function by learning non-linear decision boundaries and while being flexible. In this regard, the *k*NN algorithm was found to be effective in diagnosis of gastric cancer [26], Parkinson's disease [27], hypertension [28], and seizure [29]. For better tuning of the SVM and *k*NN algorithm, the BO algorithm [30] is adopted to determine their respective optimal

key parameters thanks to its ability to update the prior belief in light of new information to produce an updated posterior belief to find promising minima; hence, to statistically and robustly approximate the objective function. The BO is fast, effective and was successful in optimization of classifiers in various biomedical engineering applications; including, detection of Parkinson's disease in patient's voice [23], malarial cell [32], arrhythmia in electrocardiogram [33], COVID-19 in chest X-ray image [34], and diabetes [35]. Finally, the Chi-square test is employed as statistical filter to identify the most significant patterns from each set of acoustic features separately to faster information processing by each optimized classifier and improve its accuracy. It was found to be effective in identification of significant patterns with application to Parkinson's disease diagnosis [36], gene selection [37], and schizophrenia identification [38].

To sum up, the contributions of the current study are as follows:

i. To design, implement, and compare various optimal CAD systems for diagnosis of newborn based on automatic analysis and classification of cry audio features.
ii. Apply a statistical filter for most significant features selection to allow fast convergence of the classifier along with reduction in system complexity.
iii. To fine tune key parameters of classifiers by using Bayesian optimization. Hence, the accuracy of the optimal classifier is expected to improve.
iv. The performance of the optimal CAD system can easily be interpreted from a physiological perspective as main involved features in improvement of accuracy can be identified.
v. To test the efficacy of designed optimal CAD systems on a large data set considered in very recent studies. Hence, the performance of our best optimal CAD system can be compared to the most recent models tested on the same database.

The rest of the manuscript is organized as follows. Section 2 describes the different techniques for vocal feature extraction, feature selection, the classifiers, Bayesian optimization, and performance measures. The dataset and results are provided in Section 3. Finally, discussion of the obtained results and conclusion are presented in Section 4.

2. Methods

The purpose of the current study is to design various CAD systems to distinguish between healthy and unhealthy newborns based on analysis and classification of acoustics of their respective cry signals. The acoustic features are categorized into three sets: MFCC, AAM, and prosody. The Chi-square (χ^2) is applied to each acoustic features set to determine the most 10 significant features used to train the classifiers. The SVM and *k*NN algorithm are chosen as main classifiers two distinguish between health conditions of the newborns. The key parameters of each classifier are optimized by using BO algorithm. Fig. 1 shows the flowchart of the proposed CAD systems. The methods are described next.

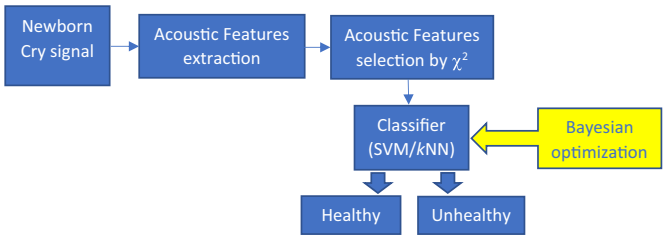

Fig. 1. Diagram of the proposed CAD system for newborn cry signal analysis and classification.

2.1. Acoustic features and selection

To characterize each newborn cry signal, four three of features are extracted; namely, mel frequency cepstral coefficients (MFCC), auditory-inspired amplitude modulation (AAM), and prosody. The MFCC have the merit to identify and track timbre fluctuations in a sound, AAM is able to characterize the rate of change of long-term speech, and prosody provides the melody and the parsing of speech. They are briefly presented afterward.

- Mel frequency cepstral coefficients (MFCC)

The MFCC are well-known popular short-term acoustic features useful to determine the critical bandwidth used by human auditory system to recognize a different tone based on the Mel scale. The latter is defined as follows:

$$M(f) = 1125 \times log\left(1 + \frac{f}{700}\right) \tag{1}$$

Here, f and $M(f)$ denote the frequency value of the signal and its corresponding Mel value respectively. To obtain MFCC, four steps should be performed: (*i*) framing (voice signal is broken down into overlapping frames), (*ii*) windowing (each frame is multiplied by a Hamming window), (*iii*) applying Fast Fourier transform (convert the signal to the frequency domain and calculate its periodogram), (*iv*) applying Mel filter banks (compute the average of each spectral power density contained in each filter and computes its logarithm), and (*v*) converting from cepstral to temporal domain by calculating the inverse discrete Fourier transform. More details can be found in [12].

- AAM features

First, the newborn cry signal is processed by short-time discrete Fourier transform (STDFT). Second, the squared magnitudes of the resulting acoustic frequency components are categorized into 27 subbands. Third, a second transform is performed across time for each of the 27 sub-band magnitude signals. Fourth, a band-pass filter is applied to the grouped squared modulation frequencies. Finally, logarithm transform is applied for compression purpose. More technical details on AAM are found in [39].

- Prosody features

Prosody set is composed of tilt feature subset, intensity feature subset, and rhythm feature subset. Tilt features are estimated based on two parameters: A_t and D_t. They are defined as follows:

$$A_t = \left(\frac{|A_r| - |A_f|}{|A_r| + |A_f|}\right) \tag{2}$$

$$D_t = \left(\frac{|D_r| - |D_f|}{|D_r| + |D_f|}\right) \tag{3}$$

where A_f is the amplitude of the contours of the fundamental frequency F_0 when they are descending and A_r is their amplitude when they are ascending. Likewise, D_f is the length of the contours of F_0 when they are descending and D_r is their length when they are ascending.

Besides, the intensity represents the height of the audio signal. Specifically, the intensity of an audio signal is used to embody its height by measuring the energy of volume in a waveform. Intensity of the audio signal is given by:

$$Intensity = 10 \times log\left(\sum_{n=1}^{N} A^2(n)w(n)\right) \tag{4}$$

where w and A are respectively the window and the amplitude.

Finally, the rhythm feature subset includes two main parameters; namely, the raw pairwise variability index (*rPVI*) and the normalized one (*nrPVI*). Both are useful to quantify the rhythm in a given audio by expressing the level of variability in successive measurements. They expressed as follows:

$$rPVI = \left(\frac{\sum_{k=1}^{M-1} |d_k - d_{k+1}|}{m-1} \right) \tag{5}$$

$$nrPVI = 100 \times \left(\frac{\sum_{k=1}^{M-1} \left| 2 \times \frac{d_k - d_{k+1}}{d_k - d_{k+1}} \right|}{m-1} \right) \tag{6}$$

In addition to *rPVI* and *nrPVI*, six other features are computed; namely, the standard deviation of the expiration signal, the standard deviation of the expiration signal divided by mean length, number of expirations in each cry signal, duration of expiration, range of expiration, average of all expirations in one signal cry signal.

- Acoustics selection by χ^2 test

Since each acoustic features set is large and may negatively affect the processing time and accuracy of the diagnosis, the classifiers will be trained with the ten most significant features from each features set which are selected by a statistical filter; namely, Chi-square test (χ^2). It is chosen thanks to its robustness with respect to the distribution of the data as it does not assume any distribution and is fast to compute. The feature selection process seeks to generate a score for each acoustic feature by counting its frequency in training unhealthy and healthy class samples separately and then finding a function of both. The χ^2 statistic is calculated as follows:

$$\chi^2 = \sum_{i-1} \sum_{j-1} \left(\frac{O^{ij} - E^{ij}}{E^{ij}} \right) \tag{7}$$

where O is the frequency that feature is observed and E is the frequency that it is expected. Larger value of χ^2 statistic suggests significant confirmation that two features are different.

2.2. SVM and kNN classifiers

The SVM [19,20] seeks to find a hyperplane $w.\Phi(x) + b = 0$ to separate the features vector x from classes $+1$ (unhealthy) and -1 (healthy) with a maximal margin. Here, w is a weight vector, Φ is a mapping function, and b a bias. The decision frontier of classes y is written as:

$$y = sign\left(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x) + b \right) \tag{8}$$

In this study, the kernel K is set to be the radial basis function (RBF) as it is a local function, which is flexible and effective in approximation of short variations in a nonlinear function. The RBF is given by:

$$K(x_i, x_j) = exp\left(-\gamma \|x - x_i\|^2 \right) \tag{9}$$

where γ is the width of the RBF.

Besides, the *k*NN algorithm [25] is typically a non-parametric instance-based learning algorithm. It assigns a class to an unclassified point following a majority rule based on the *k*-nearest neighbors in the training set. As a result, the class majority among the *k*NN produces a prediction for a new point. For instance, the *k*NN of point x_i are all k nearest neighbor points x_j in a subset of datasets D defined as follows:

$$NN_k(x_i) = \left\{ x_j \in D, d(x_i, x_j) \leq d(x_i, q) \right\} \tag{10}$$

where q is the kth nearest neighbor of point x_i and $d(x_i,x_j)$ is a distance function.

2.3. Bayesian optimization

The Bayesian optimization (BO) [30] uses an acquisition function to find both regions where the model believes the objective function to be low and regions where uncertainty is high. Let consider $f(x)$ be the objective function and the expected improvement function $EI(x,Q)$ be the acquisition function used to evaluate the feasibility of a point x based on the posterior distribution function Q. The expected-improvement function ($EI(x,Q)$) is expressed as follows:

$$EI(x, Q) = E_Q\left[max\left(0, \mu_Q(x_{best}) - f(x)\right) \right] \tag{11}$$

where x_{best} is the location of the lowest posterior mean and $\mu_Q(xbest)$ is the lowest value of the posterior mean.

The BO is employed to find optimal values of the structural parameters of the SVM and the width of the RBF. Also, it is employed to find the optimal distance metric and the optimal k for the *k*NN algorithm. The BO technique is employed through 10-fold cross validation to find the optimal parameters. More details on BO method can be found in [30].

3. Data and experimental results

We used a private dataset [15,16] that consists of cry signals recorded from newborns by using a two-channel sound recorder. The sampling frequency of the recorded signal is 44.1 kHz and its length varies between two to three minutes. Cry signals are recorded from 763 healthy newborns and 320 unhealthy ones. More details on the dataset can be found in [15,16]. For illustration purpose, Fig. 2 displays examples of recoded healthy and unhealthy cry signals from two different newborns.

For experiments, we consider four different vectors of features. The first one is composed of 190 MFCC features, the second is composed of 200 AAM features, the third one is composed of 38 prosody features, and the fourth is the larger one which includes all MFCC, AAM, and prosody acoustic patterns. Univariate feature ranking for classification using Chi-square test is performed to each vector of features to obtain 10 best features that explain most of variability in each one of them by using 5 % statistical significance level. The ranking of features is shown in Fig. 3 for each category of acoustic patterns. Then, each optimal classifier is trained either with best AAMF vector, best MFCC vector, best prosody vector, or best features selected from the large vector composed of

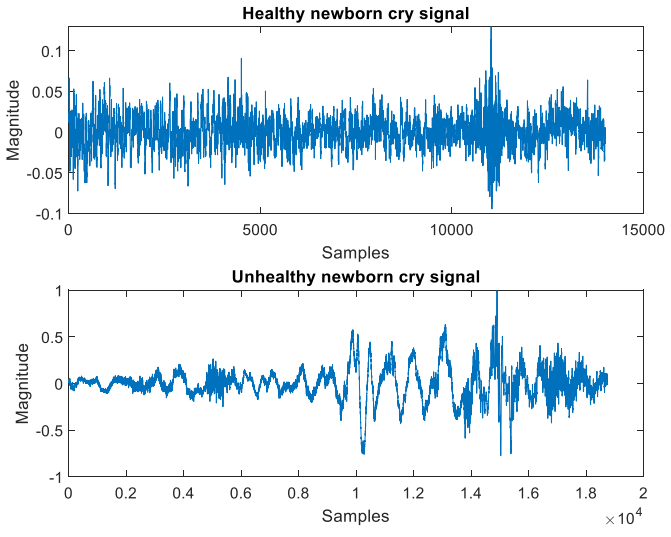

Fig. 2. Examples of cry signals recorded from healthy and unhealthy newborns.

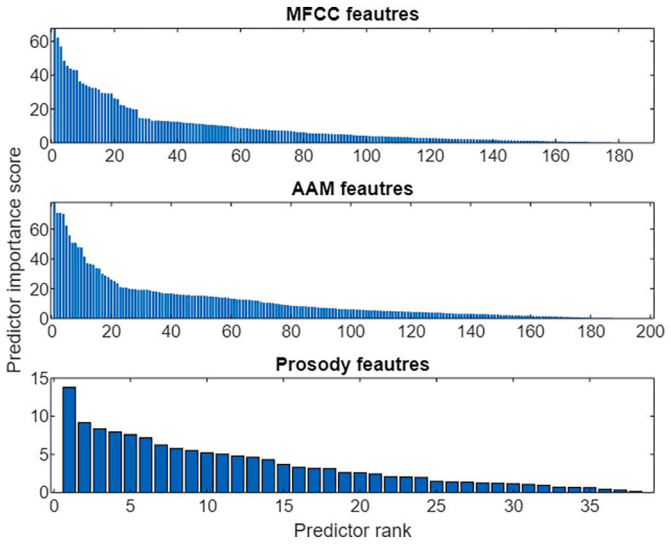

Fig. 3. Ranking of acoustic features by type for selection of the most significant ones. The higher is the predictor importance, the higher is rank. Initially, there are 190 MFCC features, 200 AAM features, and 38 prosody features. The Chi-square statistical test is applied as filter to select from each acoustic features set the ten most significant features used to train each optimal classifier.

AAMF, MFCC, and prosody feature set. Ten-fold cross-validation is employed and average and standard deviation of accuracy, sensitivity (correct classification rate of unhealthy infant records) and specificity (correct classification rate of unhealthy infant cry records) are reported in Table 1.

Accordingly, the obtained performance by each optimal classifier trained with a specific set of features. As shown, the optimal SVM trained with AAM features yielded to the highest accuracy (correct classification rate) 83.62 % ± 0.0229 followed by the optimal *k*NN trained with AAM, MFC and prosody (82.88 % ± 0.0144). The least accuracy is obtained by *k*NN trained with prosody features (70.43 % ± 0.0007) and SVM trained with prosody features (70.65 % ± 0.0025).

Besides, the highest sensitivity (correct classification of healthy newborns) is achieved by the optimal SVM trained with AAM features (59.18 % ± 0.0469) followed by the optimal *k*NN trained with AAM, MFC and prosody (55.34 % ± 0.0350). Finally, the latter system achieved the highest specificity (correct classification of unhealthy newborns) 94.42 % ± 0.0075 followed by the optimal SVM trained with AAM features (93.87 % ± 0.0190).

It is worth to mention three interesting observations. First, integrations of all three different features sets considerably improves the accuracy of the optimal *k*NN. Also, it improves the accuracy of the optimal SVM compared to the one trained with MFCC or prosody features. Another interesting observation is the fact that AAM features allow the optimal *k*NN to outperform the one trained with MFCC or prosody features. Finally, training and testing the best systems requires 0.5161 s by the optimal SVM trained with AAM and 0.1713 s by the optimal *k*NN trained with AAM, MFC and prosody. Hence, these two best CAD systems are fast and can be implemented for real applications.

With comparison to previous works where various CAD systems have been proposed to distinguish between healthy and unhealthy cry signals of newborns, our best CAD system achieved 83.62 % ± 0.0229 accuracy. Hence, it outperformed a very recent study validated on the same database [15] where the linear SVM and PNN achieved 76.50 % and 68.90 % respectively when trained with MFCC, 75.75 % and 70.70 % respectively when trained with AAM features, 61.50 % and 52.10 % respectively when trained with prosody features and 78.70 % and 77.90 % respectively when trained with combination of MFCC and AAM.

Table 1
Experimental results.

CAD systems	Accuracy	Sensitivity	Specificity	Processing time (s)
AAM + SVM	83.62 % ± 0.0229	59.18 % ± 0.0469	93.87 % ± 0.0190	0.5161
MFCC + SVM	72.37 % ± 0.0091	24.49 % ± 0.0177	92.46 % ± 0.0164	5.1411
Prosody + SVM	70.65 % ± 0.0025	0.0000	1.0000	0.8269
AAM + MFCC + Prosody + SVM	81.74 % ± 0.0183	55.33 % ± 0.0413	92.80 % ± 0.0096	0.7969
AAM + *k*NN	80.07 % ± 0.0162	48.80 % ± 0.4880	93.14 % ± 0.0151	0.4490
MFCC + *k*NN	74.07 % ± 0.0090	33.29 % ± 0.0156	91.17 % ± 0.0077	0.2349
Prosody + *k*NN	70.43 % ± 0.0007	0.0000	1.0000	0.1763
AAM + MFCC + Prosody + *k*NN	82.88 % ± 0.0144	55.34 % ± 0.0350	94.42 % ± 0.0075	0.1713

4. Conclusion

To explore the effectiveness of various automatic systems in analysis and classification of newborn cry signals to distinguish between healthy and healthy ones, our work designed and compared different CAD models involving nonlinear SVM and *k*NN classifiers optimized by using Bayesian optimization and trained by Chi-square based selected features from MFCC, AAM, prosody or combination of those selected features. This is the first study to design and compare these CAD systems for detection of unhealthy newborn cry signals. The best model is the SVM trained with AAM followed by *k*NN trained with combination of MFCC, AAM, and prosody. Our best model outperformed most existing works validated on the same database while being considerably fast to perform. As being effective and explainable, the proposed CAD system can be promising for diagnosis of newborns based on their cry signals in clinical milieu.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgment

This research is partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2016-05067].

References

[1] Orlandi S, Reyes Garcia CA, Bandini A, Donzelli G, Manfredi C. Application of pattern recognition techniques to the classification of full-term and preterm infant cry. J Voice 2016;30:656–63.
[2] Manfredi C, Bocchi L, Cantarella G. A multipurpose user-friendly tool for voice analysis: application to pathological adult voices. Biomed Signal Process Control 2009;4:212–20.
[3] Satar M, Cengizler C, Hamitoglu S, Ozdemir M. Investigation of relation between hypoxic-ischemic encephalopathy and spectral features of infant cry audio. J Voice 2022. https://doi.org/10.1016/j.jvoice.2022.05.015.
[4] Selamtzis A, Castellana A, Salvi G, Carullo A, Astolfi A. Effect of vowel context in cepstral and entropy analysis of pathological voices. Biomed Signal Process Control 2019;47:350–7.
[5] Poel M, Ekkel T. F 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.

[6] 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. In: Proceeding of the 32nd IEEE EMBS international conference; 2010. p. 6292–5.
[7] Abou-Abbas L, Alaie HF, Tadj C. Automatic detection of the expiratory and inspiratory phases in newborn cry signals. Biomed Signal Processing Control 2015; 19:35–43.
[8] Kheddache Y, Tadj C. Resonance frequencies behavior in pathologic cries of newborns. J Voice 2015;29:1–12.
[9] Alaie HF, Abou-Abbas L, Tadj C. Cry-based infant pathology classification using GMMs. Speech Commun 2016;77:28–52.
[10] Abou-Abbas L, Tadj C, Gargour C, Montazeri L. Expiratory and inspiratory cries detection using different signals' decomposition techniques. Journal of Voice 2017; 31. 259.e13–259.e28.
[11] Kheddache Y, Tadj C. Identification of diseases in newborns using advanced acoustic features of cry signals. BiomedSignal Process Control 2019;50:35–44.
[12] Salehian Matikolaie F, Tadj Chakib. On the use of long-term features in a newborn cry diagnostic system. Biomed Signal Process Control 2020;59:101889.
[13] Lahmiri S, Tadj C, Gargour C. Biomedical diagnosis of infant cry signal based on analysis of cepstrum by deep feedforward artificial neural networks. IEEE Inst Meas Mag 2021;24:24–9.
[14] Lahmiri S, Tadj C, Gargour C, Bekiros S. Deep learning systems for automatic diagnosis of infant cry signals. Chaos Solitons Fractals 2022;154:111700.
[15] Salehian Matikolaie F, Kheddache Y, Tadj C. Automated newborn cry diagnostic system using machine learning approach. Biomed Signal Process Control 2022;73: 103434.
[16] Salehian Matikolaie F, Tadj C. Machine learning-based cry diagnostic system for identifying septic newborns. J Voice 2022. https://doi.org/10.1016/j.jvoice.2021.12.021.
[17] Lahmiri S, Tadj C, Gargour C, Bekiros S. Characterization of infant healthy and pathological cry signals in cepstrum domain based on approximate entropy and correlation dimension. Chaos Solitons Fractals 2021;143:110639.
[18] Lahmiri S, Tadj C, Gargour C. Nonlinear statistical analysis of normal and pathological infant cry signals in cepstrum domain by multifractal wavelet leaders. Entropy 2022;24:1166.
[19] Cortes C, Vapnik V. Support-vector networks. MachLearn 1995;20:273–97.
[20] Vapnik V, Golowich S, Smola A. Support vector machine for function approximation, regression estimation, and signal processing. Adv Neural Information Process Syst 1996;9(1996):281–7.
[21] Lahmiri S, Bekiros S. Complexity measures of high oscillations in phonocardiogram as biomarkers to distinguish between normal heart sound and pathological murmur. Chaos Solitons Fractals 2022;154:111610.
[22] Lahmiri S. Hybrid deep learning convolutional neural networks and optimal nonlinear support vector machine to detect presence of hemorrhage in retina. Biomed Signal Process Control 2020;60:101978.
[23] Lahmiri S, Shmuel A. Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. Biomed Signal Process Control 2019;49:427–33.
[24] Lahmiri S, Shmuel A. Performance of machine learning methods applied to structural MRI and ADAS cognitive scores in diagnosing Alzheimer's disease. Biomed Signal Process Control 2019;52:414–9.
[25] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Transact Information Theory 1967;13:21–7.
[26] Li C, Shi C, Zhang H, Chen Y, Zhang S. Multiple instance learning for computer aided detection and diagnosis of gastric cancer with dual-energy CT imaging. J Biomed Inform 2015;57:358–68.
[27] Hosny M, Zhu M, Gao W, Fu Y. A novel deep learning model for STN localization from LFPs in Parkinson's disease. Biomed Signal Process Control 2022;77:103830.
[28] Parmar KS, Kumar A, Kalita U. ECG signal based automated hypertension detection using fourier decomposition method and cosine modulated filter banks. Biomed Signal Process Control 2022;76:103629.
[29] Lahmiri S, Shmuel A. Accurate classification of seizure and seizure-free intervals of intracranial EEG signals from epileptic patients. IEEE Trans Inst Meas 2019;68: 791–6.
[30] Gelbart M, Snoek J, Adams RP. Bayesian optimization with unknown constraints. https://arxiv.org/abs/1403.5607; 2014.
[32] Dĩker A. An efficient model of residual based convolutional neural network with Bayesian optimization for the classification of malarial cell images. Comput Biol Med 2022;148:105635.
[33] Li H, Lin Z, An Z, Zuo S, Zhu W, Zhang Z, Mu Y, Cao L, Prades García JD. Automatic electrocardiogram detection and classification using bidirectional long short-term memory network improved by Bayesian optimization. Biomed Signal Process Control 2022;73:103424.
[34] Loey M, El-Sappagh S, Mirjalili S. Bayesian-based optimized deep learning model to detect COVID-19 patients using chest X-ray image data. Comput Biol Med 2022; 142:105213.
[35] Joseph LP, Joseph EA, Prasad R. Explainable diabetes classification using hybrid bayesian-optimized TabNet architecture. Comput Biol Med 2022:106178. https://doi.org/10.1016/j.compbiomed.2022.106178.
[36] Ali L, Zhu C, Zhou M, Liu Y. Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection. Expert Syst Appl 2019;137:22–8.
[37] Lee J, Choi IY, Jun C-H. An efficient multivariate feature ranking method for gene selection in high-dimensional microarray data. Expert Syst Appl 2021;166:113971.
[38] Cao H, Duan J, Lin D, Shugart YY, Calhoun V, Wang Y-P. Sparse representation based biomarker selection for schizophrenia with integrated analysis of MRI and SNPs. Neuroimage 2014;102:220–8.
[39] Sarria-Paja M, Falk TH. Fusion of auditory inspired amplitude modulation spectrum and cepstral features for whispered and normal speech speaker verification. Comput Speech Lang 2017;45:437–56.