

REVIEW

Open Access

A review of infant cry analysis and classification



Chunyan Ji, Thosini Bamunu Mudiyansele, Yutong Gao and Yi Pan*

Abstract

This paper reviews recent research works in infant cry signal analysis and classification tasks. A broad range of literatures are reviewed mainly from the aspects of data acquisition, cross domain signal processing techniques, and machine learning classification methods. We introduce pre-processing approaches and describe a diversity of features such as MFCC, spectrogram, and fundamental frequency, etc. Both acoustic features and prosodic features extracted from different domains can discriminate frame-based signals from one another and can be used to train machine learning classifiers. Together with traditional machine learning classifiers such as KNN, SVM, and GMM, newly developed neural network architectures such as CNN and RNN are applied in infant cry research. We present some significant experimental results on pathological cry identification, cry reason classification, and cry sound detection with some typical databases. This survey systematically studies the previous research in all relevant areas of infant cry and provides an insight on the current cutting-edge works in infant cry signal analysis and classification. We also propose future research directions in data processing, feature extraction, and neural network classification fields to better understand, interpret, and process infant cry signals.

Keywords: Infant cry, Pathological cry, Infant cry classification, Infant cry detection, Machine learning

1 Introduction

About 130 million babies are born globally each year. Taking good care of newborns is a big challenge, especially for first time parents. Following the suggestions from other parents and books is not enough to solve the problems in practice. The main reason is because it is difficult to understand the meaning of the infant cries. Infants communicate with the world through crying. Experienced parents, caregivers, doctors, and nurses understand the cries based on their experiences. Young parents get frustrated and have trouble calming down their babies because all cry signals sound the same to them. Accurately interpreting infants' cry sound can help parents take better care of their babies. Research on infant cry started as early as 1960s when Wasz-Hockert research group identified the four types of the cries (pain, hunger, birth, and pleasure) auditorily by trained nurses [1]. In

the early years, researches have determined that different types of cries can be differentiated auditorily by trained adult listeners. But training human perception for infant cry is much harder than training machine learning models. In Mukhopadhyay's study, the highest classification accuracy by training a group of people to recognize some cry sounds is 33.09% while machine learning algorithm based on spectral and prosodic features can recognize the same set of data and reach 80.56% accuracy [2]. Building smart machines to understand infant cry leads the way to build intelligent robot caregivers in the future. Besides understanding infants' daily life needs, disease prediction is another critical task in infant cry research. Since infants' vocal tract and breathing system are affected by some diseases, the cry signals of unhealthy infants contain unique characteristics that differ from healthy cry signals. Known examples of such diseases include deaf, autism, and asphyxia, etc. Analyzing pathological cry signals to identify diseases is a non-invasive and fast method that can save infants' lives, especially in the areas that

*Correspondence: yipan@gsu.edu
Georgia State University, 25 Park Place, Atlanta, USA

lack of medical equipment and expertise. In the early years of infant cry research, many works have focused on classifying normal and pathological cry signals. In Saraswathy's review [3], 34 papers on classification of normal and pathological cry signals published from 2003 to 2011 are listed. The works include identifying diseases such as hypo-acoustic, asphyxia, hypothyroidism, hyperbilirubinemia, cleft palate, etc.

Infant cry research involves data collection, cry signal processing, feature extraction and selection, and classification. Due to the sensitivity of cry data, it has been difficult for researchers to acquire data needed. Researchers either record cry clips by themselves or ask permissions for datasets from other authors. Most databases are recorded in hospital, Neonatal Intensive Care Unit (NICU), home, and clinics, etc. by recording in real time or by setting up electronic recording devices close to the infants' crib for long period of time. Signal processing is a must to remove background noises and perform cry segmentation to build cry databases. Once the database is available, feature extraction is the step to extract features from different domains of the cry signals. Features extracted from time domain, cepstral domain, or prosodic domain, etc. represent different aspects of the cry signal. Selecting the most appropriate features and reducing the feature dimensions are another task to build effective classification models. Applying appropriate machine learning models for specific cry features is vital for classification or detection accuracy. As the second Artificial Intelligence (AI) winter ends in 1990s [4], neural networks emerge as a popular method in infant cry research. Neural networks are computing system, containing interconnected neurons, inspired by biological brain system. Input vectors, neurons, weights, activation functions, and output are the main elements in a neural network. Each neuron has a value computed in the forward propagation process based on the weights of each connection and bias of each layer. Activation functions are used to achieve nonlinearity in the network. The back propagation is the key algorithm to train the model and minimize the loss function, which evaluates how well the model fits the dataset. During the 2000s, most methods adopted in infant research are related to neural networks including scaled conjugate gradient neural network, multi-layer perceptron, general regression neural network, evolutionary neural network, probabilistic neural network, neuro-fuzzy network, and Time Delay Neural network, etc. Hidden Markov model and Support Vector Machine (SVM) were also adopted in the 2000s. In the recent decade, many traditional machine learning methods, such as SVM, K-Nearest Neighbor (KNN), Gaussian Mixture Model (GMM), fuzzy classifier, logistic regression, K-means clustering, and Random Forest, are applied to pathological cry classification, cry reason classification, and cry sound detection. In the

same period, novel neural network architectures are used pervasively in industry and research. Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), CNN-RNN, Capsule Net, Reservoir Network, and neuro-fuzzy networks open a new chapter in infant cry research.

This survey reviews infant cry research mainly focusing on the signal processing techniques and machine learning methods developed in the past decade. We first review typical databases used in the research, then introduce pre-processing approaches of infant cry signals, and describe a diversity of features either in time domain or in frequency domain as well as suprasegmental features of infant cry signals. We focus on reviewing the state-of-the-art methods using KNN, SVM, GMM, and CNN-based algorithms for classification and detection. We provide a list of resources for the researchers who are interested to work in this domain, and finally we make a point of the future work in this research area.

2 Data acquisition

As shown in Fig. 1, automatic infant cry research generally involves five stages: data acquisition, pre-processing, feature extraction, feature selection, and classification. Discovering novel methods in any of the stages can help improve the performance of the final classification accuracy.

The data acquisition stage includes recording the infant cry sounds and labeling. Most databases are recorded in hospitals or homes, labeled by doctors, nurses, or parents. Digital recorders are placed close to infants and are either operated on the spot to capture the cry signals one by one or left on to record the sound events around the infants for a long period of time. Infant sound is a short-term stationary signal, and it is assumed to be more stationary because of infants' lack of full control of the vocal tract. Due to the limitation of resources and sensitivity of infant cry data collection process, the total amount of infant cry database is very limited. From the previous review papers [3, 5, 6], we can see that the most commonly used database in infant cry research is Baby Chillanto database [7]. Baby Chillanto database was collected by the National Institute of Astrophysics and Optical Electronics, CONACYT Mexico [8]. It contains five types of cry signals including deaf, asphyxia, normal, hungry, and pain. Each cry is equally segmented into 1-s long and the total number of cries is 2268. Another database used in multiple literatures is named Dunstan Baby Language database [9], which is extracted from the Dunstan baby video tutorial presented by Priscilla Dunstan who invented the Dunstan Baby Language theory. There are several versions of Dunstan Baby Language database since authors extracted the audio clips in their own ways. The version described in [9] consists of 315 wave files, sampled at 16 kHz, with a variable length between 0.3 and 1.6 s. Each utterance is a word of infant

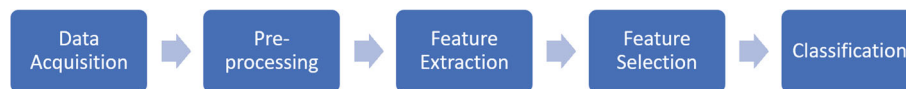


Fig. 1 Five stages of infant cry research

speech corresponding to one of the five “Dunstan words,” which were translated as “Neh” = hungry, “Eh” = need to burp, “Oah” = tired, “Eairh” = low belly pain, and “Heh” = physical discomfort.

Many databases are self-recorded for research. Researchers need to contact other authors to check availability of desired databases. One database named Donate A Cry [10] is available online, but it is not well labeled and only one literature is found using this database. Table 1 shows the commonly used databases in recent research. Some databases are recorded in the Neonatal Intensive Care Unit (NICU), pediatric clinics, or baby-sitting environments [11–24]. Some cry audio signals online are also collected in [25]. Some synthetic databases are created by the authors in order to compare the performances of the proposed methods on real databases and synthetic databases [11, 18, 26, 27]. In Ferretti’s work [18], the CNN detects the cry signal better on the synthetic database than the real database. It shows that the automatic detection and classification of real-time infant cry is still challenging because the real-time environment may exist many types of complications that can affect the quality of the cry signals. Synthetic databases can be generated by adding noises to clean cry recordings or combining different cries together. Training models on synthetic databases can avoid requiring a large amount of data to be acquired in sensible environments such as NICUs [18].

From Table 1, we can see that most datasets are with limited samples. The average size is 2983 and only one database is close to 20,000 samples. Due to the sensitivity of collecting the cry data, especially pathological cry signals, small dataset size is one of the challenges in infant cry research. Data augmentation techniques are used to artificially increase the data size. Zhang et al. created new waveform images from training datasets by transforming these waveform images into slightly faster or slightly slower waveforms for the purpose of increasing training datasets to overcome overfitting problem [12]. In [43], several data augmentation techniques, such as noise variation, signal intensity variation, tonality variation, and spectrogram’s size alteration, were used to artificially increase either the number of audio signals or the number of spectrograms. The experimental results showed that these data augmentation methods cannot lead to accuracy improvement. The reasons lie in the fact that the limited data cannot capture the diversity of variations within infant cry signals.

3 Signal processing and feature generation

3.1 Pre-processing

The main tasks in pre-processing stage are denoising and audio segmentation. The complication of the recording environment leads to unclean infant cry signals. In a neonatal care unit, besides infant cry signals, there could be many kinds of sounds such as footsteps, adult’s speech, air-conditioner sound, alarm sound, etc. To detect or classify cry signals accurately, cleaning up the recorded data at the pre-processing stage is a crucial step. To clean up a signal, the first task is denoising, which removes the background sounds such as speech, fan, footstep, etc. Turan and Erzin applied high-pass FIR filter to remove the speech sound and low frequency noise in the recording [41]. Ferretti et al. reduced coherent noise source by a filter-and-sum beamformer and uses OMLSA post-filter to reduce the residual diffuse noise [18]. In [16], Gu et al. used optimized Blackman window to handle each frame signal, which is the result after the endpoint detection. The signal noise is significantly reduced after filtering.

Audio segmentation task is commonly performed using Voice Activity Detection (VAD). VAD technique is widely used in speech recognition to detect the human speech in audio signals. Researchers also use it to detect the infant cry and remove the silent duration in a sample recording. VAD also faces the challenge of separating the cry and noise. Pan et al. uses it to detect the presence or absence of baby cry in a noisy environment to improve the overall baby cry recognition rate [56] and it is used to detect the sections of the audio with sufficient audio activity [57]. In [41], authors implemented a basic VAD algorithm, which uses short-time features of audio frames and a decision strategy for determining sound and silence frames. Sometimes researchers also manually cut the samples to remove the silent part and the voice interference part, and only the continuous crying part of the sound was retained [51].

3.2 Feature extraction

Infant cry signal differs from adult speech. Figure 2 gives a comparison of spectrograms between infant sound and adult speech. We can see that the variations within waveform and spectrogram are quite different, especially in the areas of energy, intensity, and formants. In general, infant cry is a combination of vocalization, silence, coughing, choking, and interruptions, which includes a diversity of acoustic and prosodic information at different levels. It is the only way for babies to communicate with the world.

Table 1 Main databases used in literatures

Database	Creator	Data	Papers
Baby Chillanto (2004)	National Institute of Astrophysics and Optical Electronics, CONACYT Mexico	Total 2268 (deaf 879, asphyxia 340, normal 507, hunger 350, pain 192)	[7, 15, 28–35]
Dustan Baby Language(2018)	Extracted from Dunstan Baby Language DVD by authors	Total 315 (hungry 56, sleepy 106, burping 55, belly pain 37, discomfort 61)	[9], different versions [36–38]
Donate A Cry (2015)	https://github.com/gveres/donateacry-corpus	Total 457 (hungry 382, tired 24, burping 8, belly pain 16, discomfort 27)	[39]
CRIED (2018)	INTERSPEECH 2018 computational paralinguistic challenge	Total 2838 (neural 2292, fussing 368, crying 178)	[40–42]
iCOPE (2019)	https://www.infantcope.org/	Total 113 (pain 42, no pain 71)	[43]
ChatterBaby (2020)	https://chatterbaby.org/	Total 1071 (fuzzy 171, hungry 167, pain 353, colic 380)	[44]
SPLANN (2015)	“Sf. Pantelimon” Emergency Clinical Hospital within the SPLANN research project (project coordinator SOFTWIN SRL) Bucharest, Romania	Total 13373 (colic 225, eructation 505, discomfort 2210, hungry 5536, pain 4404, pathology 459, tired 34)	[45–47]
Self-recorded database (2019)	Recorded by first-time parents	Total 19691 (hungry 6263, sleepy 4927, wet diaper 3056, pain 5445)	[48]
Self-recorded database (2019)	Recorded in NICU of a hospital	Total 971 (hungry 422, sleepy 79, diaper 137, attention 151, discomfort 182)	[49]
Self-recorded database (2016)	Collected from National Taiwan University Hospital	Total 2169 (hungry 586, pain 723, sleepy 860)	[50]
DA-IICT Cry (2018)	Self-recorded	Total 1190 (normal 793, asphyxia 215, asthma 182)	[32]
Autism database (2019)	Self-recorded	Total 84 (normal 64, autism 20)	[51]
Hypothyroid database (2009)	Self-recorded	Total 88 (normal 45, hypothyroid 20)	[52, 53]
Ankyloglossia with Deviation of the Epiglottis and Larynx (ADEL) database (2010)	Self-recorded	Total 39 (normal 22, ADEL 17)	[54, 55]

Feature extraction is the stage to extract the discriminative features from the audio signals and later feed into the machine learning algorithms. It is one of the most vital parts of a machine learning process [58]. Performing feature extraction task either in time or frequency domain addresses the fundamental work of baby cry analysis and processing. Time domain features, such as zero-crossing

rate, amplitude, and energy-based features, etc., is simple and straightforward to compute. While time domain features are not robust enough to cover the variations within infant cry signals and the features are sensitive to background noises, the frequency domain features have strong ability to model the characteristics within infant cry signals. The commonly used MFCCs, LPCCs, and LFCCs

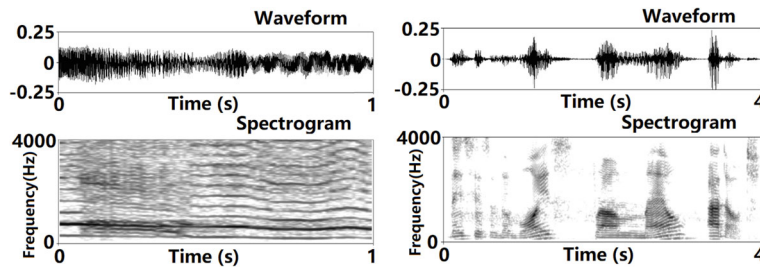


Fig. 2 Adult speech vs. infant cry signal in time and frequency domain

have proven better performance than using time domain features. On the other hand, it is shown that infant cry signal is rhythmic and has cyclic changes due to the natural interruption and breath. The high-level information, such as prosodic features, are important to improve the discriminative ability within signals. Therefore, attaching prosodic domain features together with time or frequency domain is capable for capturing both physical and physiological information. In addition, spectrogram is an image that is a time-frequency representation of an audio clip. It is known that spectrogram has a strong ability to present the signal and include both acoustic and prosodic information.

Figure 3 depicts the main categories of the audio features that are applied to research related to speech, music, and environmental sounds. Acoustic and prosodic features are commonly used for infant cry detection and classification. Cepstral domain features, prosodic features, and image-based features are widely used in speech

processing and infant cry processing with a proportion over 70% research articles. In this section, we review feature extraction approaches in the latest research work. The detailed explanation and algorithms of audio features can be found in [58] and [59].

3.2.1 Cepstral domain features

Mel-frequency cepstral coefficient (MFCC) is widely used in speech recognition. It is a cepstral representation of the audio signals. Researchers use it to test proposed approaches [17, 29, 49, 52, 57, 60–62] and often use it for baseline experiments [13, 15, 22, 31, 37, 63]. Liu et al. used MFCC along with two other cepstral features Linear Prediction Cepstral Coefficients (LPCC) and Bark Frequency Cepstral Coefficients (BFCC) for infant cry reason classification. The result showed that the BFCC with a neural network model produces the best recognition rate of 76.47% [13]. The main idea of LPCC is to remove the redundancy from a signal and tries to predict next values

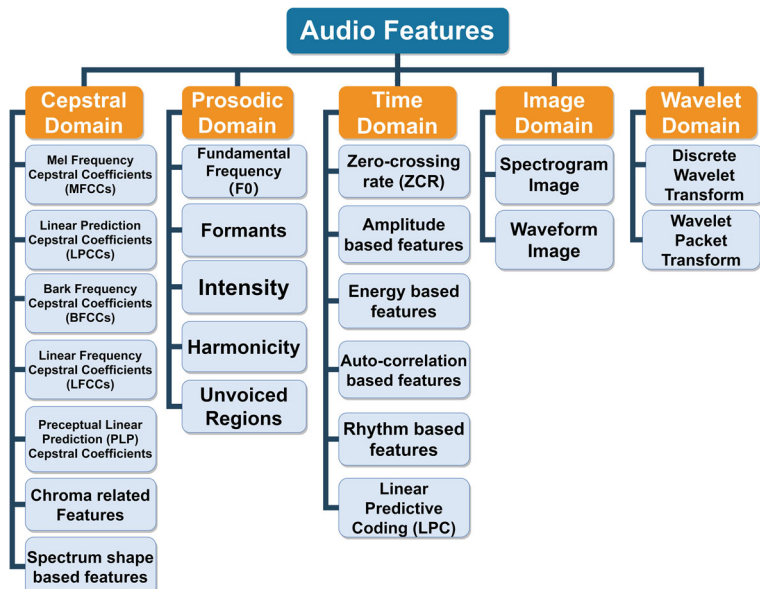


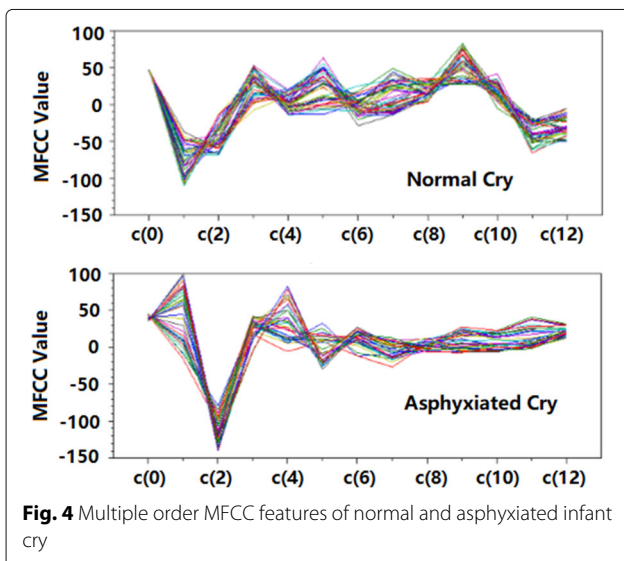
Fig. 3 Main audio feature categories

by linearly combining the previous known coefficients. It is used in [16] for cry detection. Linear Frequency Cepstral Coefficients (LFCC) extraction process is similar to MFCC extraction. The difference is that it uses a linear filter-bank instead of the Mel filter-bank [37, 64]. In [22] and [65], the authors showed that LFCC performs better than MFCC in discriminating high frequency audio signals such as female voice and baby cry signals. In [24], Singh et al. explored the residual MFCC and implicit LP residual features that represent excitation source information. Researchers have also tried other cepstral features such as Fast Fourier Transform (FFT) [23, 66], Log-Mel feature [11, 18], Mel Scale [43], Constant-Q Chromagram [43], Log-mel spectrum [12], and delta spectrum [12].

According to auditory perception models, MFCC coefficients are more robust than other coefficients such as LPC coefficients. In our previous work [15], MFCC features of normal and abnormal infant cry signals within a certain frame combined with 12 orders were plotted in a space. It is observed that the acoustic features of normal infant cry signals are quite different from the asphyxiated ones as shown in Fig. 4. It indicates that the value range and tendency of acoustic features of normal and asphyxiated infant cry are different.

3.2.2 Prosodic domain features

It is shown that infant cry is made of four types of sound: one coming from the expiration phase, a brief pause, and a sound coming from the inspiration phase followed by another pause. Variations in intensity, fundamental frequency (F0), formants, and duration are typical acoustic cues that carry prosodic information about infant cry and speech [13, 67]. It is shown that the above prosodic features are efficient to identify the types of infant cry.



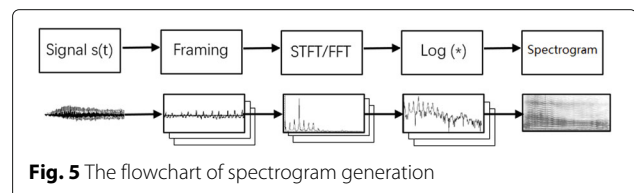
Adult F0 ranges between 85 and 200 Hz while infant crying F0 is characterized by its high F0 250–700 Hz. F0 is commonly computed using an autocorrelation-based method provided by Praat [68].

Our previous work [15] has shown that combining weighted prosodic features with MFCC features help improve the classification accuracy in a deep learning model. Other researchers have also found that F0 is critical in identifying infant cry signals [40]. Chittora and Patil used F0 to calculate unvoiced segments ratio and found out unvoiced percentage in a cry is an important parameter for analysis of infant cry [19]. Orlandi et al. used mean, median, standard deviation, and minimum and maximum of F0 and F123 to exploit differences between full-term and preterm infant cry [21]. In 2017, Torres et al. used three handcraft features (voiced/unvoiced counter, consecutive F0, and harmonic ratio accumulation) to show comparable detection performance but resulting in 20 times lower computational cost than standard MFCCs with no additional memory cost [27].

3.2.3 Image domain features

Spectrogram is an image that is a time-frequency representation of an audio clip. It is known that spectrogram has a strong ability to present the signal and include both acoustic and prosodic information. Spectrogram can be extracted through framing, FFT, and calculating the log of the filtered spectrum steps illustrated in Fig. 5. Feeding spectrograms into classifiers can solve the problem of different cry signals having different durations. Instead of using zero padding to achieve same length of feature vectors, normalization is applied in the process of spectrogram generation, which produces the same size images without changing the original signal. Besides feeding the spectrogram into CNN [9, 35, 48, 50] and capsule neural network [41], researchers take extra step to use the spectrogram image to retrieve extra features such as Local Binary Pattern (LBP), Local Phase Quantization (LPQ), and Robust Local Binary Pattern (RLBP) [43] to help improve the classification performance.

Waveform image represents the pattern of sound pressure amplitude in the time domain. It is also used in deep learning models such as AlexNet to achieve above 90% accuracy on identifying the asphyxia cry [28, 30]. In our previous work, we use Praat to generate images containing



the prosodic feature lines including F0, intensity, and formants. The prosodic feature images CNN model is good at identifying certain types of cry signals. Combining it with spectrogram CNN and Waveform CNN produces 5% better accuracy on Baby Chillanto database and 4% on Dunstan Baby Language database [69].

3.2.4 Other relevant domain features

Other domain features used in infant cry research include time domain features such as zero-crossing rate, short-time energy, and voiced-unvoiced regions, etc. Zero-crossing rate is the rate at which the signal passes zeros and changes signs. It can be used in conjunction with short-time energy to detect endpoints of speech utterances, hence to detect the existence of the cry sound from other sounds happening in the environment [17, 67]. Since the amplitude of an audio signal varies with time, the short-time energy can serve to differentiate voiced and unvoiced segments. It is used in [20, 57, 70] for infant cry detection and classification. Torres et al. used voiced-unvoiced counter, which counts all frames having a significant periodic content, as one of the features for cry detection [27]. Linear Predictive Coding (LPC) serves as a time domain measure of how close two different waveforms are and it is used for infant cry classification in [13, 49, 71].

Wavelet Transform is a method to convert the audio signal into time-frequency domain. The waveform packet transform was used in asphyxia classification research and reached high accuracy of 99% with neural network models [72]. It also performs well in infant cry reason classification. The Discrete Wavelet Transform MFCC (DWT-MFCC) features work well with SVM and neural network architectures [31, 33, 51, 73].

Researchers also calculate the statistical natural parameters of the data such as mean frequency, standard deviation, and third quartile range, etc. to help infant cry detection and classification [39]. Feature extraction is a critical step in audio processing. Besides aforementioned Praat software, feature extraction tools such as LibROSA library [74] and OpenSMILE toolkit [75] have made audio feature extraction easier.

3.3 Feature selection

Feature selection is the process of selecting a subset of features from the original features extracted from the audio signals using the feature extraction techniques. The objective is to reduce the dimensionality of the features without reducing classification accuracy. Less features require less computational resources, and hence make building smart infant cry detection and classification devices possible and affordable in the future. The original features may also contain some redundant information that prevents effectively differentiating the different types of cry signals.

Selecting the right features to fit the specific need of the task may also improve the classification accuracy. This section reviews some feature selection methods applied to the infant cry research. F-ratio method was used to select the top 20 MFCC features. The coefficients that have significant importance have higher F-ratio scores [63]. In 2013, Yamamoto et al. used Principal Component Analysis (PCA) to reduce the dimensionality of FFT features [23]. Forward variable Selection Method (FSM) was applied to infant cry classification by Wang in 2010 [55] and Okada et al. proposed Iterative FSM (IFSM) based on cross-validation concept in 2011 [54]. Later, Binary Particle Swarm Optimization (BPSO) was used to remove the redundant features and keep the significant features from MFCC coefficients in [61, 62]. Orlandi et al. used a software called Biovoice to extract 22 features from the cry signal and then used a genetic algorithm-based search method to select the best features to feed to the classifiers [21]. In 2016, Wahid et al. compared five feature selection methods: OneR, ReliefF, Fast Correlation-Based Filter (FCBF), Consistency-Based Subset Evaluation (CNS), and Correlation-Based Feature Selection (CFS). It is proven that the feature selection techniques were able to greatly reduce the feature space, hence to reduce computational time. Most selection technique can also improve the performance of the neural network classifier [76].

In 2019, Tuduze et al. utilized three Best Feature Selection (BFS) approaches to exclude irrelevant features and redundant features and tested them with 35 classifiers. The feature set is reduced from over 6000 features to 500 and the result shows that BFS can improve the classification accuracy for some classifiers [45]. Feature selection techniques remove the features irrelevant to the specific task, so it can reduce the feature space, save computational time, and improve classification accuracy.

4 Infant cry classification

With data cleaned and segmented and features extracted, selected, and normalized, finding the appropriate classifier is the most important stage in the machine learning process. In this section, we review some popular machine learning methods and applications used in infant cry classification in the past decade.

4.1 Infant cry classification models

4.1.1 Traditional machine learning classifiers

A Support Vector Machine

The most popular probabilistic classifier used in infant cry classification is Support Vector Machine (SVM) [26, 40, 43]. The types of SVM include multi-class SVMs [25], linear, and RBF kernels binary SVM [31]. The features fed into the SVM include temporal features, prosodic features, and cepstral features. In 2017, Onu et al. compared SVM to other non-

linear classifiers like neural networks on asphyxia classification and concluded that SVMs are designed to work effectively with limited examples and high-dimensional data [29]. In 2015, Chang et al. used the incremental SVM learning model, which keeps adding new data into the dataset in each training step, producing more than 18% better accuracy than the original SVM model on infant cry classification based on FFT features [77].

B **K-Nearest Neighbor**

KNN is a well-known pattern recognition method used in classification. There are k nearest neighbors in the feature space. The goal is to assign test sample to the class that its nearest neighbor belongs to. If k is greater than 1, the nearest neighbor is selected based on the number of nearest neighbors. In the case of infant cry classification, researchers used Euclidean distance, Minkowski distance, and other methods to measure the distance between two sample feature vectors. Feature vectors selected are usually MFCC and LFCC [20, 22, 37, 64]. Cohen and Lavner used KNN algorithm, in which each frame is classified either as a cry or not a cry, and then the sample is classified to be cry signal if more frames in the sample is identified as cry [57].

C **Gaussian mixture model**

GMM is a probabilistic model that assumes the datapoints are in Gaussian distribution of some mean and variance. The idea is to learn the parameters to model the provided training data as mixture of several Gaussian distributions. Then the test data can be classified by the trained model. Expectation Maximization (EM) algorithm is used for finding the maximum likelihood estimates of the parameters under GMM-based structures [24, 39]. In 2016, Banica et al. used GMM-UBM method to classify Dunstan baby cries. The universal background model (UBM) is a GMM model that is trained on large amount of general cry signals with no specific labels. The classification accuracy of the GMM-UBM with MFCC achieved 70% on Dunstan baby cries [38] and 50.6% on SPLANN database [47]. GMM-UBM is also used by Alaie et al. to classify healthy cries and pathological cries. The Boosting Mixture Learning adaptation method proposed outperforms the MAP algorithm [78]. In 2019, Sharma et al. compared the GMM clustering to hierarchical clustering and K-means clustering on cry features and showed the GMM model produces the best result with least amount of overlapping datapoints with a certain database [39]. It is shown that GMM-based classifiers are sensitive to environment and cannot lead to satisfied results especially with limited training data.

D **Fuzzy classifier**

Fuzzy logic systems have been used in many applications such as transmission systems, power systems, and wireless network routing [79]. It is also used in infant cry classification. Selected features are converted into fuzzy values in the fuzzification step, certain fuzzy membership functions are used, and fuzzy rules are defined. In [66], Kia et al. used fuzzy classification to detect infant cry signals from laughter signals. In [71], Rosales-Pérez et al. used fuzzy decision tree, fuzzy decision forest, fuzzy KNN, and fuzzy relational neural network classifier for pathological cry classification. Type-2 fuzzy pattern matching algorithm is used in [80] to classify asphyxia, normal, and hyperbilirubinemia. It also outperforms SVM and logistic regression classifier on classifying hunger and pain [81].

E **Logistic regression classifier**

Logistic regression classifier is a low-complexity supervised algorithm, and it is usually used as a referencing experiment for infant cry research. Lavner et al. used it to show that CNN performs better on cry detection [17] and Orlandi et al. used it to compare with many other classifiers, in which random forest performed the best on classifying full-term and preterm infant cries [21].

F **K-means clustering**

K-mean clustering represents an unsupervised algorithm mainly used for clustering. Unlabeled data points can be gradually separated into groups based on the mean value and centroid moving. Sharma et al. used K-means clustering to show that the GMM model has better performance differentiating different types of cry [39]. In [22], K-means clustering was used to build a speaker database for speaker recognition.

G **Bagging, boosted trees, and random forest**

Bagging, boosted trees, and random Forest are techniques that perform ensemble decision trees. They all combine multiple decision trees to produce better performance. Experiments have shown that they are powerful on infant cry classification. Osmani et al. showed bagging and boosted trees outperform SVM [67]. Milano et al. compared it to MLP, SVM, Reservoir Network, GMM, and HMM models and showed random forest classifier is next to Reservoir Network [82]. In [21, 45, 83, 84], an open source data mining software named Waikato Environment for Knowledge Analysis (WEKA) is used. Among over 100 classification algorithms implemented in WEKA, random forest outperforms SVM, MLP, logistic regression, and BayesNet, etc. Tuduce et al. tested 40 classifiers in WEKA and the tree classifiers showed the best overall performance comparing to Bayes classifiers, lazy

classifiers, function classifiers, and rule classifiers, etc. [83].

4.1.2 Neural network-based models

Artificial Neural Network (ANN) is a machine learning method. In 1995, Petroni et al. made the first attempt of ANN in infant cry classification [85].

A **Feed Forward Neural Network (FFNN)** is the simplest neural network and Multi-Layer Perceptron (MLP) is a type of FFNN that contains at least three layers. The experiments in [37] and [13] both showed that FFNN's performance was not as good as nearest neighbor classifier based on MFCC features. MLP was used in [52, 61–63] with MFCC for identifying pathological cries. To classify asphyxia, Hariharan et al. used Probabilistic Neural Network (PNN), General Regression Neural Network (GRNN), and Time-Delay Neural Network (TDNN) and achieved above 97% accuracy [33, 86].

B **Convolutional Neural Network** is a deep learning algorithm that has been successfully used in computer vision, language processing, and other domains achieving unprecedented high accuracy. Multi-channel CNN, which accepts multiple channel input, were applied in [11]. Manikanta et al. used 1D CNN on MFCC features for cry detection and the result outperformed feed forward neural network and SVM classifier [25]. In 2019, Le et al. applied transfer learning with CNN on spectrograms on Baby Chillanto database and achieved promising result [35].

C **Long Short-Term Memory (LSTM)** is a type of Recurrent Neural Network (RNN), which has internal states to make accepting sequence of data possible and is known as best neural network for time series data such as language translation and speech recognition. Mark Huckvale fed the low-level signal temporal features into the bidirectional LSTM model and later combine with another two dense-layer neural network in [40]. The LSTM itself and combined network both outperformed the baseline SVM model.

D **CNN-RNN** is a deep learning architecture combining CNN with RNN. It has shown its power in sound detection and classification. In Detection and Classification of Acoustic Scenes and Events (DCASE) 2017 competition, Lim et al. used it to win the first place on detecting the target sounds (baby crying, glass breaking, gunshot) with mixed noisy background [87]. In 2019, Maghfira et al. used CNN-RNN to classify the five types of cries and reached the highest accuracy of 94.97% for Dunstan Baby Languages database [36].

E **Neuro-fuzzy Network** combines fuzzy logic with neural networks and it has been used successfully by

researchers in infant classification. In 2009, Santiago-Sánchez et al. used type-2 fuzzy set to classify asphyxia and hyperbilirubinemia [80]. In 2012, Molaeezadeh et al. proposed a type-2 fuzzy pattern matching classifier and it outperformed SVM and logistic regression classifiers in classifying hunger and pain [81]. In recent years, it is noticed that combining fuzzy systems with neural networks can unite their advantages and evade the disadvantages of both methods. Fuzzy systems require rules while neural networks directly learn from data. Neuro-fuzzy approach was used to classify Dunstan baby language type of cries. Neural networks were trained and the Mandani fuzzy logic was adopted after data normalization to create new "transformed dataset," which is used for final classification step of KNN [88]. The classification accuracy reached 86.25%, which is better than normal neural network model, SVM, and GMM methods.

F **Capsule Network** [89] is a deep learning topology that adds a structure called capsules into the CNN model. As maxpooling in CNN only picks the maximum value within a region and throws away information in certain positions, higher-level capsules cover larger regions of the image and performs routing by agreement instead. CapsNet was applied to classify infants' emotional cry in domestic environments and the accuracy is improved more than 10% over the CNN model with spectrograms [41].

G **Reservoir Network (RN)** is a neural network model derived from RNN. Its input nodes connect to a non-trainable reservoir, which contains connected non-linear units with randomly generated fixed weights. Ntalampiras used RN in infant cry multi-class classification [82] with fused feature sets and showed that RN model outperformed MLP, SVM, random forest, GMM clustering, etc.

Many machine learning methods have been experimented in infant research. Each of them has advantages and disadvantages and no algorithm is the perfect for every dataset and task. Selecting a suitable model to achieve high performance is challenging. To determine the classification ability of the different models, Fuhr et al. experimented differentiating healthy infant cries and cries of infants suffering from several diseases using 12 classifiers including SVM, decision tree, KNN, MLP, etc. The result showed only C5 decision tree and KNN achieved greater than 90% accuracy [90]. Applying many algorithms on the task before selecting the algorithm to use is impractical. Comparing the machine learning algorithms used in infant research, we analyze them from the following aspects. Readers can choose the appropriate algorithm accordingly for their datasets and tasks.

- **Time complexity.** It includes training time and classification time relying on the data size, searching space, and the complexity of coefficients. In general, traditional methods such as SVM, K-means clustering, and GMM-based approaches are relatively simple and straightforward. Smaller sample size is acceptable, which differs from neural network methods. Hence, training time, searching time, and classification time are much less than those of neural network methods. Also, fine tuning in neural network models also requires more developing time.
- **Sample complexity.** It indicates whether the model requires large size of data or not to learn. It depends on the complexity of the data and the complexity of the algorithms. To reach better performance, neural network methods generally require larger sample size for complex searching space than other traditional algorithms. Larger size infant cry databases are needed for deep neural networks.
- **Parametricity.** It indicates if the number of the parameters used in the model is fixed or it varies along when new data is brought in. Linear regression, GMM, and neural networks are parametric methods while KNN and SVM are nonparametric models.
- **Feature complexity.** Features extracted from either time domain or frequency domain have the same abilities to represent the different characteristics of the cry signals in different models. There is no feature complexity difference involved for traditional models or neural network-based models. But using too many features to represent one sample may cause overfitting issue; therefore, selecting the most appropriate features for specific models is critical.
- **Parallelizability.** Parallelizability is a pivotal feature for saving the training time of machine learning methods. Large amount of data in neural networks is associated with high computation cost in both time and space. Parallelizability with Graphics Processing Unit (GPU) computing greatly reduces the training time and made deep learning possible. Other method such as KNN is easy to parallel, but parallelism is tricky if the next step is based on the previous step result such as decision trees.

With current powerful computation environments, methods used in infant cry research can achieve real-time prediction. Due to limited samples in current infant cry databases, training time and testing time have not been highlighted as an issue. There are no very deep models with big data involved in the research yet. At present, the largest dataset has less than 20,000 samples. The small and imbalanced datasets lead to high classification accuracy but low confidence for some of the tasks. To achieve high

performance with high confidence, real big data with real deep learning models are to be explored.

4.2 Infant cry applications

Researchers use different classifiers to perform infant cry processing tasks. In the past decade, most research work continue to pay effort to improve the classification accuracy of infant cry signals including differentiating the pathological cries from the normal cries and understand the meaning behind the cry signals. In this section, we review the significant works on infant cry classification and detection.

4.2.1 Infant cry reason classification

In the early years of infant cry research, more works were performed on automatically differentiating the cries of healthy infants from pathological cries. In recent years, exploring the meaning of the cries attract more research interests. As Table 2 shown, some significant works are done on this topic. It is noticeable that researchers are using different datasets, most of which are self-recorded. With different datasets in similar research, even the classification types are the same, it is unfair to make direct comparison on the performances of the proposed methods. The infant classification remains in challenging stage due to the lack of standard public datasets and the classification accuracy is still relatively low.

4.2.2 Infant pathological cry classification

Infant cry signals have been used to identify many diseases such as asphyxia, hypo-acoustic (hearing disorder), hypothyroidism, hyperbilirubinemia, cleft palate, respiratory distress syndrome, ankyloglossia with deviation of the epiglottis and larynx, etc. Readers can find the related works on pathological cry classification before 2011 in [3]. In the past decade, researchers continue to apply novel methods to classify normal cry and pathological cry. Asphyxia cry is the most popular disease in research. Table 3 shows the latest works on classifying normal cry from asphyxia cry. Researchers have been using the Baby Chillanto database to perform the binary classification. In 2012, Probabilistic Neural Network (PNN) and General Regression Neural Network (GRNN) reached 99% accuracy [34, 86], the latest SVM model reached 97.7% accuracy [31], and the deep learning FFNN model reaches 96.74% accuracy [15].

Besides identifying asphyxia, other types of diseases have also been studied. According to Esposito's review [94], it is shown that the infants' cry signals are useful for early diagnosis of autism spectrum disorder (ASD). In 2012, Orlandi et al. analyzed the cry signals of the high-risk infants whose siblings have already been diagnosed to be ASD. It is noticed that less cry episodes

Table 2 Significant works on infant cry reason classification

First author	Dataset	Features	Classifiers	Best performance
Felipe [43] (2019)	iCOPE (pain vs. no pain)	Mel Scale (MS), MFCC, Constant-Q Chromagram (CQC) Local Binary Pattern (LBP), Local Phase Quantization (LPQ), Robust Local Binary Pattern (RLBP) extracted from spectrogram	SVM	71.68%
Sharma [39] (2019)	Donate A Cry (hungry, burp needed, belly pain, discomfort, tired, lonely, feeling cold/hot, is scared, unidentified)	Mean frequency; standard deviation; median frequency; third quartile range; spectral entropy; kurtosis, skewness, spectral flatness, etc.	K-means clustering, hierarchical clustering, Gaussian mixture models clustering	81.27%
Maghfira [36] (2019)	Dunstan Baby Database (pain, hunger, discomfort, need to burp, belly pain)	Spectrogram	CNN-RNN	94.97%
Franti [9] (2018)	Dunstan Baby Database (pain, hunger, discomfort, need to burp, belly pain)	Spectrogram	CNN	89%
Liu [13] (2018)	NICU recorded (draw attention cry, diaper change needed cry, and hungry)	LPC, LPCC, MFCC, BFCC	Nearest Neighbor Artificial Neural Network	76.4%
Turan [41] (2018)	CRIED	Spectrogram	Capsule Network	86.1%
Osmani [67] (2017)	Dunstan Baby Database (hunger, pain, tiredness, belly pain, need burp)	spectrum, pitch, zero-crossing rate, root mean square, intensity, energy along with their calculated statistics (mean, variance, skewness, etc)	SVM, Bagging Decision Tree, and Boosted Trees,	N/A
Chang [50] (2016)	Collected from National Taiwan University Hospital (hungry, pain, and sleep)	Spectrogram	CNN	78.5%
Bano [20] (2015)	Self-recorded (hungry, need to burp, sleepy, pain, discomfort)	Pitch short-time energy MFCC Statistical properties of MFCC	KNN	86%
Orlandi [21] (2015)	Self-recorded (full term vs. preterm)	CU length, F0 median, F0 mean, F0 standard deviation, F0 minimum, F0 maximum, number of estimated F0 values, F123 median, F123 mean, F123 standard deviation, F123 minimum, F123 maximum	Logistic regression, Multilayer perceptron NN, Support Vector Machine, and Random Forest	87%

Table 2 Significant works on infant cry reason classification (Continued)

First author	Dataset	Features	Classifiers	Best performance
Bhagatpatil [22] (2015)	Self-recorded (pain, hunger, discomfort, need to burp, belly pain)	LFCC, MFCC	K-mean clustering, KNN	91.58%
Rosales-Pérez [71] (2014)	Baby Chillanto (hungry vs. pain)	MFCC, LPC	Fuzzy model	97.96%
Yamamoto [23] (2013)	Self-recorded (discomfort, hungry, sleepy)	FFT	Nearest neighbor	62.1%

occur, F0 is lower, and Formants reach high values for high-risk infants than healthy infants [95]. Although some babies are born with ASD, it is usually diagnosed when they are 2 to 3 years old since the diagnosis involves observing the behavior of children. This leads to the difficulty of the cry signal acquisition for autism babies. In 2019, Wu et al. recorded twenty audio samples of autistic children whose ages are between 2 and 3 years old. They reached 96% accuracy by using SVM classifier with MFCC features [51]. Identifying hypo-acoustic cry signal has been successful in the early years. In 2011, Hariharan’s General Regression Neural Network reached 99% on Baby Chillanto database [96], and in 2009, O.E. Reyes-Galaviz et al. used evolutionary neural network system to reach almost 100% on Mexican-Cuba database [8]. Then in 2014, Rosales-Pérez et al. used fuzzy model

and genetic algorithm to reach 99.42% on Baby Chillanto database [97]. Other types of diseases such as hypothyroidism, respiratory distress syndrome, cleft palate, and ankyloglossia with deviation of the epiglottis and larynx (ADEL) were studied in the early years and were reviewed in [3]. In 2014, Feier et al. studied newborns’ cries within minutes after birth. Random tree and random forest methods were able to classify cries of healthy newborns from premature newborns, newborns with umbilical cord strangulation during birth, and newborns with other pathologies with accuracy above 95% [98].

4.2.3 Infant cry detection

Infant cry detection is considered as a binary classification with cry and not-cry categories. It is another attractive

Table 3 Classification of asphyxia cry from other cries

Literature	Dataset	Features	Classifiers	Performance
Ji [15] (2019)	Baby Chillanto	MFCC with Weighted Prosodic	FFNN	96.74%
Badreldine [31] (2018)	Baby Chillanto	DWT-MFCC	RBR Kernel SVM	97.7%
Zabidi [60] (2017)	Baby Chillanto database with database from University of Milano-Bicocca	MFCC Image	CNN	92.8%
Onu [29] (2017)	Baby Chillanto	MFCC	SVM	85%
Sachin [30] (2017)	Baby Chillanto	Waveforms	AlexNet	92%
Moharir [28] (2017)	Baby Chillanto	Waveforms	GoogleNet,AlexNet	94%
Rosales-Pérez [71] (2014)	Baby Chillanto	MFCC, LPC	Fuzzy Model	90.68%
Saraswathy [33] (2012)	Baby Chillanto	Wavelet Packet Transform	PNN, GRNN	99.04%
Hariharan [86] (2012)	Baby Chillanto	STFT	PNN, GRNN, TDNN, MLP	99%
Hariharan [34] (2012)	Baby Chillanto	Weighted LPCC	PNN	99%
Sahak [91] (2010)	Baby Chillanto asphyxia and self-recorded normal cries	MFCC	SVM	95.86%
Sahak [92] (2010)	Database from University of Milano Bicocca	MFCC	Orthogonal least square (OLS)-based SVM	93.16%
Zabidi [93] (2010)	Baby Chillanto	MFCC	MLP	93.38%

research topic in the latest decade. The goal is to detect the infant cry signal efficiently and accurately in various environments, such as car, home, and hospital, etc., while other sounds happening at the same time. Data is recorded during a long period of time in a certain environment such as home or hospital. The detection algorithm needs to be able to detect the cry sound despite the background sounds happening in the environment. Researchers also propose different methods to build smart cradles, which can detect infant cries and alert the parents while they are away [99–102]. The proposed methods not only target to higher detection accuracy but also consider the price of the baby monitoring system to make it affordable for low income families.

Table 4 shows some recent significant works on infant cry detection. It is seen that neural network-based approaches reach good performance under clean and constrained conditions. On the other hand, with noisy environment and limited training data, classifiers are sensitive at the boundary and easy to be confused and overlapped with noise signals.

5 Challenges and future directions

With the improvement of computational ability and the use of deep learning approaches, the following challenges remain in infant cry research.

- Lack of existing data and scalability of research. Researches are based on different datasets recorded by authors. Therefore, it is difficult to compare the performances of methods experimented on different datasets. The only database shared by some researchers is Baby Chillanto database, which has been around for two decades. The total amount of Baby Chillanto database is 2287 and the largest private database has less than 20,000 samples, which is insufficient for deep learning NN models. Data is the key elements of machine learning, especially deep learning. We notice that although some deep learning methods such as CNN and CNN-RNN are used in infant cry research, the architectures of models are not deep. The main reason is that the deep models underfit the small training dataset and lead to poor performance. To take advantage of deep learning, large-scale databases with sufficient samples covering diverse changes within acoustic and prosodic features of different babies are in need.
- Collecting data and labeling is a time-consuming process and requires skilled labors. Most databases used so far are self-recorded by authors and private to certain people or organizations. Although some online resources are available such as videos on Youtube, which is what Google audioset links to, most cry clips have no relevant labels and many recordings are full

of background noises. To accelerate the progress of building automatic infant cry classifiers, smart cradle systems, and further to build robotic babysitter caregivers, effort to make public comprehensive well-structured and labeled databases are urgently in need. In addition, databases that contain samples from specific babies that can track their cries at different ages are needed. This type of database is essential to study the characteristic of infant cry along with their body development. Setting up recording devices on infants' cradles and recording real-time cry signals using cell phones by caregivers are the main methods used by data collectors. Baby cry translator mobile applications such as ChatterBaby [44] help predict infant cry reasons and made data collection easier. It will be more beneficial to the development of infant cry research if some newly collected datasets can be made public.

- Poor connection between medical professionals and researchers diminishes the ability of interdisciplinary mutual promotion. Researches have proven that classifying infant cry signals is a non-invasive method and can be very helpful in some early disease diagnosis such as asphyxia, autism, cleft palate, and hypothyroidism, etc. But most of the pathological disease researches with infant cry were performed before 2010, and the sizes of the datasets were very small. The difficulties of data acquisition may be the biggest obstacles in this research area. The ethical and legal issues involved in data collection process hinder the development of infant cry research. Cooperation between medical professionals and computer scientists may trigger some opportunities in this life saving research topic.

We are currently building a large infant cry database consisting of cries of infants from 0 to 9 months old. The cry clips are recorded and labeled by parents at home and by doctors and nurses in hospitals using cell phones. It is currently in the data acquisition stage and it is expected to be a database containing over 30,000 samples reaching 50 h of recording, which fits the need of deep neural networks. We are also applying Graph Neural Network (GNN) to infant cry classification. GNN has been used across various domains and the graph can represent the non-Euclidean data with complex relationships between objects. Combined with deep learning, which has proved to perform successfully on Euclidean data, the GNN deep learning model should be able to take advantage of more features and have more discriminating abilities for infant classification tasks. In addition, new deep learning architectures embedded with prior knowledge can also be explored. With more databases available in the future, we believe that more machine

Table 4 Significant works on infant cry detection

Literature	Dataset	Features	Classifiers	Performance
Chang [48] (2019)	Self-recorded (crying with TV, speech, etc.)	Spectrogram	CNN	99.83%
Manikanta [25] (2019)	Recorded in homes (crying with ac, fan, etc.)	MFCC	1D-CNN, FFNN, SVM	86%
Dewi [64] (2019)	Self-recorded (cry and not cry)	LFCC	KNN	90%
Gu [16] (2018)	Self-recorded (crying with laughter, barking, etc.)	LPC	Dynamic time warping algorithm	97.1%
Ferretti [18] (2018)	Real Dataset: recorded in the NICU of a hospital; Synthetic DB: crying with speech, "beep" sounds, etc.)	Log-Mel Coefficients	CNN	86.58% on real dataset, 92.92% on synthetic dataset
Feier [87] (2017)	TUT Rare Sound Events 2017 (crying with "glass breaking", "gunshot", etc.)	Log-amplitude mel-spectrogram	CRNN	85% for baby crying detection, 87% for all three targets
Torres [27] (2017)	Online resources (crying with adult cry, vacuum cleaning, etc.)	Voiced unvoiced counter, Consecutive F0 and harmonic ratio accumulation, MFCC	Support Vector Data Description (SVDD), CNN	AUC 92%
Lavner [17] (2016)	Recorded in domestic environment (crying with speech, door opening, etc.)	MFCC, Pitch, Formants, etc.	CNN	95%

learning methods can be explored in this area. Combining new audio signal processing methods and novel machine learning methods will lead this research to a remarkable future, which will change people's lives by providing affordable infant automatic care-giving.

6 Conclusion

In this paper, we describe the significant research work in infant cry analysis and classification, providing details and resources that are helpful for both researchers and medical professionals who work in this area. It is shown that the limited database resources hinder the development of the infant cry research. Large databases with diverse samples fitting the need of deep neural networks is imperatively desired. The current tendency for feature extraction is to generate a mixed feature set and takes advantages of different domains to achieve better discriminating ability. The relevant research results show promising improvement with combined features. In addition, new neural network-based architectures become the mainstream methods. It proves better robustness and performance than traditional machine learning approaches. In the future, we are interested in creating

a large database, extracting more robust features, combining features with good ratio, establishing novel neural network architectures with the use of prior knowledge as well as other space information from interdisciplinary areas.

Acknowledgements

The authors would like to thank the editors and reviewers for their valuable suggestions.

Authors' contributions

CJ organizes the project and writes the manuscript. TM and YG are responsible for summarizing the papers during 2014–2016 and provide suggestions on the writing. YP provides guidance for the project and gives suggestions on the writing. All authors have read and approved the final manuscript.

Availability of data and materials

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 19 August 2020 Accepted: 8 January 2021

Published online: 05 February 2021

References

1. O. Wasz-Höckert, T. J. Partanen, V. Vuorenkoski, K. Michelsson, E. Valanne, The identification of some specific meanings in infant vocalization. *Experientia*. **20**(3), 154 (1964)

2. J. Mukhopadhyay, B. Saha, B. Majumdar, A. K. Majumdar, S. Gorain, B. K. Arya, S. D. Bhattacharya, A. Singh, in *2013 Indian Conference on Medical Informatics and Telemedicine, ICMIT 2013*. An evaluation of human perception for neonatal cry using a database of cry and underlying cause, (2013). <https://doi.org/10.1109/IndianCMIT.2013.6529410>
3. J. Saraswathy, M. Hariharan, S. Yaacob, W. Khairunizam, in *2012 International Conference on Biomedical Engineering (ICoBE)*. Automatic classification of infant cry: a review, (2012), pp. 543–548. <https://doi.org/10.1109/ICoBE.2012.6179077>
4. L. Floridi, AI and its new winter: from myths to realities. *Philos. Technol.*, 1–3 (2020). <https://doi.org/10.1007/s13347-020-00396-6>
5. A. A. Dixit, N. V. Dharwadkar, in *Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing, ICCSP 2018*. A survey on detection of reasons behind infant cry using speech processing, (2018), pp. 190–194. <https://doi.org/10.1109/ICCSP.2018.8524517>
6. G. Zamzmi, R. Kasturi, D. Goldgof, R. Zhi, T. Ashmeade, Y. Sun, A review of automated pain assessment in infants: features, classification tasks, and databases (2018). <https://doi.org/10.1109/RBME.2017.2777907>
7. O. F. Reyes-Galaviz, E. A. Tirado, C. A. Reyes-Garcia, in *International Conference on Computers for Handicapped Persons*. Classification of infant crying to identify pathologies in recently born babies with ANFIS, vol. 3118, (2004), pp. 408–415. https://doi.org/10.1007/978-3-540-27817-7_60
8. O. F. Reyes-Galaviz, S. D. Cano-Ortiz, C. A. Reyes-García, in *7th Mexican International Conference on Artificial Intelligence - Proceedings of the Special Session, MICAI 2008*. Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies, (2008), pp. 330–335. <https://doi.org/10.1109/MICAI.2008.73>
9. E. Franti, I. Ispas, M. Dascalu, in *2018 41st International Conference on Telecommunications and Signal Processing, TSP 2018*. Testing the Universal Baby Language hypothesis - automatic infant speech recognition with CNNs, (2018), pp. 1–4. <https://doi.org/10.1109/TSP.2018.8441412>
10. GitHub - gveres/donateacry-corporus: an infant cry audio corpus that's being built through the Donate-a-cry campaign - see <http://donateacry.com>. <https://github.com/gveres/donateacry-corporus>. Accessed 07 Aug 2020
11. M. Severini, D. Ferretti, E. Principi, S. Squartini, Automatic detection of cry sounds in neonatal intensive care units by using deep learning and acoustic scene simulation. *IEEE Access*, **7**, 51982–51993 (2019). <https://doi.org/10.1109/ACCESS.2019.2911427>
12. X. Zhang, Y. Zou, Y. Liu, in *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. AICDS: an infant crying detection system based on lightweight convolutional neural network, (2018). https://doi.org/10.1007/978-3-319-94361-9_14
13. L. Liu, Y. Li, K. Kuo, in *2018 International Conference on Information and Computer Technologies, ICICT 2018*. Infant cry signal detection, pattern extraction and recognition, (2018), pp. 159–163. <https://doi.org/10.1109/INFOCT.2018.8356861>
14. S. Sharma, P. R. Myakala, R. Nalumachu, S. V. Gangashetty, V. K. Mittal, in *2017 7th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos, ACIIW 2017*. Acoustic analysis of infant cry signal towards automatic detection of the cause of crying, (2018), pp. 117–122. <https://doi.org/10.1109/ACIIW.2017.8272600>
15. C. Ji, X. Xiao, S. Basodi, Y. Pan, in *Proceedings - 2019 IEEE International Congress on Cybermatics: 12th IEEE International Conference on Internet of Things, 15th IEEE International Conference on Green Computing and Communications, 12th IEEE International Conference on Cyber, Physical and So*. Deep learning for asphyxiated infant cry classification based on acoustic features and weighted prosodic features, (2019). <https://doi.org/10.1109/IThings/GreenCom/CPSCoM/SmartData.2019.00206>
16. G. Gu, X. Shen, P. Xu, in *Proceedings of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2018*. A set of DSP system to detect baby crying, (2018), pp. 411–415. <https://doi.org/10.1109/IMCEC.2018.8469246>
17. Y. Lavner, R. Cohen, D. Ruinskiy, H. Ijzerman, in *2016 IEEE International Conference on the Science of Electrical Engineering, ICSEE 2016*. Baby cry detection in domestic environment using deep learning, (2017). <https://doi.org/10.1109/ICSEE.2016.7806117>
18. D. Ferretti, M. Severini, E. Principi, A. Cenci, S. Squartini, in *2018 26th European Signal Processing Conference (EUSIPCO)*. Infant cry detection in adverse acoustic environments by using deep neural networks, (2018), pp. 992–996. <https://doi.org/10.23919/EUSIPCO.2018.8553135>
19. A. Chittora, H. A. Patil, in *International Conference on Text, Speech, and Dialogue*. Significance of unvoiced segments and fundamental frequency in infant cry analysis, vol. 9302, (2015), pp. 273–281. https://doi.org/10.1007/978-3-319-24033-6_31
20. S. Bano, K. M. Ravikumar, in *Proceedings of the IEEE International Conference on Soft-Computing and Network Security, ICSNS 2015*. Decoding baby talk: a novel approach for normal infant cry signal classification, (2015), pp. 24–26. <https://doi.org/10.1109/ICSNS.2015.7292392>
21. S. Orlandi, C. A. Reyes Garcia, A. Bandini, G. Donzelli, C. Manfredi, Application of pattern recognition techniques to the classification of full-term and preterm infant cry. *J. Voice*, **30**(6), 656–663 (2016). <https://doi.org/10.1016/j.jvoice.2015.08.007>
22. M. V. Varsharani Bhagatpatil, An automatic infant's cry detection using linear frequency cepstrum coefficients (LFCC). *Int. J. Sci. Eng. Res.* **5**(12), 1379–1383 (2014)
23. S. Yamamoto, Y. Yoshitomi, M. Tabuse, K. Kushida, T. Asada, Recognition of a baby's emotional cry towards robotics baby caregiver. *Int. J. Adv. Robot. Syst.* **10** (2013). <https://doi.org/10.5772/55406>
24. A. K. Singh, J. Mukhopadhyay, K. S. Rao, in *2013 Indian Conference on Medical Informatics and Telemedicine, ICMIT 2013*. Classification of infant cries using source, system and supra-segmental features, (2013), pp. 58–63. <https://doi.org/10.1109/IndianCMIT.2013.6529409>
25. K. Manikanta, K. P. Soman, M. Sabarimalai Manikandan, in *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*. Deep learning based effective baby crying recognition method under indoor background sound environments, vol. 4, (2019), pp. 1–6. <https://doi.org/10.1109/CSITSS47250.2019.9031058>
26. G. Joshi, C. Dandvate, H. Tiwari, A. Mundhare, in *Proceedings - 2017 International Conference on Vision, Image and Signal Processing, ICVISP 2017*. Prediction of probability of crying of a child and system formation for cry detection and financial viability of the system, (2017), pp. 134–141. <https://doi.org/10.1109/ICVISP.2017.33>
27. R. Torres, D. Battagliolo, L. Lepauloux, in *International Conference on Engineering Applications of Neural Networks*. Baby cry sound detection: a comparison of hand crafted features and deep learning approach, (2017). https://doi.org/10.1007/978-3-319-65172-9_15
28. M. Moharir, M. U. Sachin, R. Nagaraj, M. Samiksha, S. Rao, *Identification of asphyxia in newborns using GPU for deep learning*, (2017), pp. 236–239. <https://doi.org/10.1109/I2CT.2017.8226127>
29. C. C. Onu, I. Udeogu, E. Ndiomu, U. Kengni, D. Precup, G. M. Sant'anna, E. Alikor, P. Opara, Ubenwa: cry-based diagnosis of birth asphyxia. *Nips*, 2–5 (2017). <https://doi.org/10.1109/1711.06405>
30. M. U. Sachin, R. Nagaraj, M. Samiksha, S. Rao, M. Moharir, GPU based deep learning to detect asphyxia in neonates. *Indian J. Sci. Technol.* **10**, 3 (2017). <https://doi.org/10.17485/ijst/2017/v10i3/110617>
31. O. M. Badredine, N. A. Elbeheiry, A. N. M. Haroon, S. Elshehaby, E. M. Marzook, in *ICENCO 2018 - 14th International Computer Engineering Conference: Secure Smart Societies*. Automatic diagnosis of asphyxia infant cry signals using wavelet based mel frequency cepstrum features, (2019), pp. 96–100. <https://doi.org/10.1109/ICENCO.2018.8636151>
32. H. B. Sailor, H. A. Patil, in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. Auditory filterbank learning using ConvRBM for infant cry classification, (2018), pp. 706–710. <https://doi.org/10.21437/Interspeech2018-1536>
33. J. Saraswathy, M. Hariharan, V. Vijean, S. Yaacob, W. Khairunizam, in *Proceedings - 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications, CSPA 2012*. Performance comparison of Daubechies wavelet family in infant cry classification, (2012), pp. 451–455. <https://doi.org/10.1109/CSPA.2012.6194767>
34. M. Hariharan, L. S. Chee, S. Yaacob, Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network. *J. Med. Syst.* **36**(3), 1309–1315 (2012). <https://doi.org/10.1007/s10916-010-9591-z>
35. L. Le, A. N. M. H. Kabir, C. Ji, S. Basodi, Y. Pan, in *Proceedings - 2019 IEEE 16th International Conference on Mobile Ad Hoc and Smart Systems Workshops, MASSW 2019*. Using transfer learning, SVM, and ensemble classification to classify baby cries based on their spectrogram images, (2019). <https://doi.org/10.1109/MASSW.2019.00028>
36. T. Nadia Maghfira, T. Basaruddin, A. Krishnadi, Infant cry classification using CNN - RNN. *J. Phys. Conf. Ser.* **1528**(1), 012019 (2020). <https://doi.org/10.1088/1742-6596/1528/1/012019>

37. S. P. Dewi, A. L. Prasasti, B. Irawan, in *Proceedings - 2019 IEEE International Conference on Signals and Systems, ICSigSys 2019*. The study of baby crying analysis using MFCC and LFCC in different classification methods, (2019), pp. 18–23. <https://doi.org/10.1109/ICSIGSYS.2019.8811070>
38. I. A. Banica, H. Cucu, A. Buzo, D. Burileanu, C. Burileanu, in *2016 International Conference on Communications (COMM)*. Automatic methods for infant cry classification, (2016), pp. 51–54. <https://doi.org/10.1109/ICComm.2016.7528261>
39. K. Sharma, C. Gupta, S. Gupta, in *2019 10th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2019*. Infant weeping calls decoder using statistical feature extraction and Gaussian mixture models, (2019), pp. 1–6. <https://doi.org/10.1109/ICCCNT45670.2019.8944527>
40. M. Huckvale, in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. Neural network architecture that combines temporal and summative features for infant cry classification in the Interspeech 2018 Computational Paralinguistics Challenge, (2018), pp. 137–141. <https://doi.org/10.21437/Interspeech.2018-1959>
41. M. A. Tugtekin Turan, E. Erzin, in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. Monitoring infant's emotional cry in domestic environments using the capsule network architecture, (2018). <https://doi.org/10.21437/Interspeech.2018-2187>
42. B. W. Schuller, S. Steidl, A. Batliner, P. B. Marschik, H. Baumeister, F. Dong, S. Hantke, F. B. Pokorný, E. M. Rathner, K. D. Bartl-Pokorný, C. Einspieler, D. Zhang, A. Baird, S. Amiriparian, K. Qian, Z. Ren, M. Schmitt, P. Tzirakis, S. Zafeiriou, in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. The INTERSPEECH 2018 computational paralinguistics challenge: atypical & self-assessed affect, crying & heart beats, (2018), pp. 122–126. <https://doi.org/10.21437/Interspeech.2018-51>
43. G. Z. Felipe, R. L. Aguiar, Y. M. G. Costa, C. N. Silla, S. Brahmam, L. Nanni, S. McMurtrey, in *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*. Identification of infants' cry motivation using spectrograms, (2019), pp. 181–186. <https://doi.org/10.1109/IWSSIP.2019.8787318>
44. J. J. Parga, S. Lewin, J. Lewis, D. Montoya-Williams, A. Alwan, B. Shaul, C. Han, S. Y. Bookheimer, S. Eyer, M. Dapretto, L. Zeltzer, L. Dunlap, U. Nookala, D. Sun, B. H. Dang, A. E. Anderson, Defining and distinguishing infant behavioral states using acoustic cry analysis: is colic painful? *Pediatr. Res.* **87**(3), 576–580 (2020). <https://doi.org/10.1038/s41390-019-0592-4>
45. R. I. Tuduce, M. S. Rusu, H. Cucu, C. Burileanu, in *2019 42nd International Conference on Telecommunications and Signal Processing, TSP 2019*. Automated baby cry classification on a hospital-acquired baby cry database, (2019), pp. 343–346. <https://doi.org/10.1109/TSP.2019.8769075>
46. M. S. Rusu, t. S. Diaconescu, G. Sardescu, E. Bartil, in *2015 International Conference on Speech Technology and Human-Computer Dialogue, SpED 2015*. Database and system design for data collection of crying related to infant's needs and diseases, (2015). <https://doi.org/10.1109/SPED.2015.7343077>
47. I. A. Banica, H. Cucu, A. Buzo, D. Burileanu, C. Burileanu, in *2016 39th International Conference on Telecommunications and Signal Processing, TSP 2016*. Baby cry recognition in real-world conditions, (2016), pp. 315–318. <https://doi.org/10.1109/TSP.2016.7760887>
48. C. Y. Chang, L. Y. Tsai, in *Workshops of the International Conference on Advanced Information Networking and Applications*. A CNN-based method for infant cry detection and recognition, (2019). https://doi.org/10.1007/978-3-030-15035-8_76
49. L. Liu, W. Li, X. Wu, B. X. Zhou, Infant cry language analysis and recognition: an experimental approach. *IEEE/CAA J. Autom. Sin.* **6**(3), 778–788 (2019). <https://doi.org/10.1109/JAS.2019.1911435>
50. C. Y. Chang, J. J. Li, in *2016 IEEE International Conference on Consumer Electronics-Taiwan, ICCE-TW 2016*. Application of deep learning for recognizing infant cries, (2016), pp. 1–2. <https://doi.org/10.1109/ICCE-TW.2016.7520947>
51. K. Wu, C. Zhang, X. Wu, D. Wu, X. Niu, in *Proceedings - 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation, YAC 2019*. Research on acoustic feature extraction of crying for early screening of children with autism, (2019), pp. 290–295. <https://doi.org/10.1109/YAC.2019.8787725>
52. A. Zabidi, L. Y. Khuan, W. Mansor, I. M. Yassin, R. Sahak, in *Proceedings - CSPA 2010: 2010 6th International Colloquium on Signal Processing and Its Applications*. Detection of infant hypothyroidism with mel frequency cepstrum analysis and multi-layer perceptron classification, (2010), pp. 140–144. <https://doi.org/10.1109/CSPA.2010.5545331>
53. A. Zabidi, W. Mansor, L. Y. Khuan, I. M. Yassin, R. Sahak, in *2009 IEEE International Conference on Signal and Image Processing Applications*. Classification of infant cries with hypothyroidism using multilayer perceptron neural network, (2009), pp. 246–251. <https://doi.org/10.1109/ICSIPA.2009.5478608>
54. Y. Okada, K. Fukuta, T. Nagashima, in *IMECS 2011 - International MultiConference of Engineers and Computer Scientists 2011*. Iterative forward selection method based on cross-validation approach and its application to infant cry classification, vol. 1, (2011), pp. 49–52
55. X. Wang, T. Nagashima, K. Fukuta, Y. Okada, M. Sawai, H. Tanaka, T. Uozumi, Statistical method for classifying cries of baby based on pattern recognition of power spectrum. *Int. J. Biom.* **2**(2), 113–123 (2010). <https://doi.org/10.1504/IJBM.2010.031791>
56. C. Pan, W. Zhao, S. Deng, W. Wei, Y. Zhang, Y. Xu, in *Proceedings of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. The methods of realizing baby crying recognition and intelligent monitoring based on DNN-GMM-HMM, (2018), pp. 352–356. <https://doi.org/10.1109/IMCEC.2018.8469328>
57. R. Cohen, Y. Lavner, in *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*. Infant cry analysis and detection, (2012), pp. 1–5. <https://doi.org/10.1109/EEEI.2012.6376996>
58. G. Sharma, K. Umamathy, S. Krishnan, Trends in audio signal feature extraction methods. *Appl. Acoust.* **158**, 107020 (2020). <https://doi.org/10.1016/j.apacoust.2019.107020>
59. F. Alías, J. C. Socoró, X. Seviliano, A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. *Appl. Sci.* **6**(5) (2016). <https://doi.org/10.3390/app6050143>
60. A. Zabidi, I. M. Yassin, H. A. Hassan, N. Ismail, M. M. A. M. Hamzah, Z. I. Rizman, H. Z. Abidin, Detection of asphyxia in infants using deep learning ction of asphyxia in infants using deep learning convolutional neural network (CNN) trained on Mel frequency cepstrum coefficient (MFCC) features. *Aust. Ranger Bull.* **4**(1), 768–778 (2017)
61. A. Zabidi, W. Mansor, Y. K. Lee, I. M. Yassin, R. Sahak, in *Proceedings - 2011 IEEE 7th International Colloquium on Signal Processing and Its Applications*. Binary particle swarm optimization for selection of features in the recognition of infants cries with asphyxia, (2011), pp. 272–276. <https://doi.org/10.1109/CSPA.2011.5759886>
62. M. Z. M. Ali, W. Mansor, Y. K. Lee, A. Zabidi, in *Proceedings - 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications*. Asphyxiated infant cry classification using Simulink model, vol. 10, (2012), pp. 491–494. <https://doi.org/10.1109/CSPA.2012.6194778>
63. A. Zabidi, L. Y. Khuan, W. Mansor, I. M. Yassin, R. Sahak, in *2010 2nd International Conference on Computer Engineering and Applications*. Classification of infant cries with asphyxia using multilayer perceptron neural network, vol. 1, (2010), pp. 204–208. <https://doi.org/10.1109/ICCEA.2010.47>
64. S. P. Dewi, A. L. Prasasti, B. Irawan, in *Proceedings - 2019 IEEE International Conference on Internet of Things and Intelligence System, IoTals 2019*. Analysis of LFCC feature extraction in baby crying classification using KNN, (2019), pp. 86–91. <https://doi.org/10.1109/IoTals47347.2019.8980389>
65. S. S. Jagtap, P. K. Kadbe, P. N. Arotale. System propose for Be acquainted with newborn cry emotion using linear frequency cepstral coefficient, (2016), pp. 238–242. <https://doi.org/10.1109/ICEEOT.2016.7755094>
66. M. Kia, S. Kia, N. Davoudi, R. Biniagian, in *2nd International Conference on Innovative Computing Technology, INTECH 2012*. A detection system of infant cry using fuzzy classification including dialing alarm calls function, (2012), pp. 224–229. <https://doi.org/10.1109/INTECH.2012.6457776>
67. A. Osmani, M. Hamidi, A. Chibani, in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*. Machine learning approach for infant cry interpretation, (2018). <https://doi.org/10.1109/ICTAI.2017.00038>
68. Praat: doing phonetics by computer. <https://www.fon.hum.uva.nl/praat/>. Accessed 07 Aug 2020
69. C. Ji, S. Basodi, X. Xiao, Y. Pan, in *International Conference on AI and Mobile Services*. Infant sound classification on multi-stage CNNs with hybrid features and prior knowledge, (2020). https://doi.org/10.1007/978-3-030-59605-7_1

70. Y. D. Rosita, H. Junaedi, in *Proceedings - 2016 2nd International Conference on Science and Technology-Computer, ICST 2016*. Infant's cry sound classification using Mel-Frequency Cepstrum Coefficients feature extraction and Backpropagation Neural Network, (2017). <https://doi.org/10.1109/ICSTC.2016.7877367>
71. A. Rosales-Pérez, C. A. Reyes-García, J. A. Gonzalez, O. F. Reyes-Galaviz, H. J. Escalante, S. Orlandi, Classifying infant cry patterns by the Genetic Selection of a Fuzzy Model. *Biomed. Signal Process. Control.* **17**, 38–46 (2015). <https://doi.org/10.1016/j.bspc.2014.10.002>
72. M. Hariharan, S. Yaacob, S. A. Awang, Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. *Expert Syst. Appl.* **38**(12), 15377–15382 (2011). <https://doi.org/10.1016/j.eswa.2011.06.025>
73. S. Tejaswini, N. Sriraam, G. C. M. Pradeep, in *2016 International Conference on Circuits, Controls, Communications and Computing*. Recognition of infant cries using wavelet derived mel frequency feature with SVM classification, (2017). <https://doi.org/10.1109/CIMCA.2016.8053313>
74. B. McFee, C. Raffel, D. Liang, D. Ellis, M. McVicar, E. Battenberg, O. Nieto, in *Proceedings of the 14th Python in Science Conference*. librosa: audio and music signal analysis in Python, (2015). <https://doi.org/10.25080/majora-7b98e3ed-003>
75. F. Eyben, M. Wöllmer, B. Schuller, in *Proceedings of the 18th ACM international conference on Multimedia*. OpenSMILE - the Munich versatile and fast open-source audio feature extractor, (2010). <https://doi.org/10.1145/1873951.1874246>
76. N. S. A. Wahid, P. Saad, M. Hariharan, Automatic infant cry pattern classification for a multiclass problem. *J. Telecommun. Electron. Comput. Eng.* **8**(9), 45–52 (2016)
77. C. Y. Chang, Y. C. Hsiao, S. T. Chen, in *Proceedings - 2015 18th International Conference on Network-Based Information Systems, NBIS 2015*. Application of incremental SVM learning for infant cries recognition, (2015), pp. 607–610. <https://doi.org/10.1109/NBIS.2015.90>
78. H. Farsaie Alaie, L. Abou-Abbas, C. Tadj, Cry-based infant pathology classification using GMMs. *Speech Commun.* **77**, 28–52 (2016). <https://doi.org/10.1016/j.specom.2015.12.001>
79. H. Liu, J. Li, Y. Q. Zhang, Y. Pan, in *Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Network*. An adaptive genetic fuzzy multi-path routing protocol for wireless ad-hoc networks, vol. 2005, (2005), pp. 468–475. <https://doi.org/10.1109/SNP-D-SAWN.2005.12>
80. K. Santiago-Sánchez, C. A. Reyes-García, P. Gómez-Gil, in *International Conference on Intelligent Computing*. Type-2 fuzzy sets applied to pattern matching for the classification of cries of infants under neurological risk, (2009), pp. 201–210. https://doi.org/10.1007/978-3-642-04070-2_23
81. S. F. Molaezadeh, M. Salarian, M. H. Moradi, in *The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012)*. Type-2 fuzzy pattern matching for classifying hunger and pain cries of healthy full-term infants, (2012), pp. 233–237. <https://doi.org/10.1109/AISP.2012.6313750>
82. S. Ntalampiras, Audio pattern recognition of baby crying sound events. *J. Audio Eng. Soc.* **63**(5), 358–369 (2015). <https://doi.org/10.17743/jaes.2015.0025>
83. R. I. Tuduce, H. Cucu, C. Burileanu, in *2018 41st International Conference on Telecommunications and Signal Processing, TSP 2018*. Why is my baby crying? An in-depth analysis of paralinguistic features and classical machine learning algorithms for baby cry classification, (2018), pp. 1–4. <https://doi.org/10.1109/TSP.2018.8441363>
84. R. Robu, F. Feier, V. Stoicu-Tivadar, C. Ilie, I. Enătescu, in *2011 15th IEEE International Conference on Intelligent Engineering Systems*. The analysis of the new-borns' cry using NEONAT and data mining techniques, (2011), pp. 235–238. <https://doi.org/10.1109/INES.2011.5954750>
85. M. Petroni, A. S. Malowany, C. C. Johnston, B. J. Stevens, in *IEEE International Conference on Acoustics, Speech and Signal Processing*. Classification of infant cry vocalizations using artificial neural networks (ANNs), vol. 5, (1995), pp. 3475–3478. <https://doi.org/10.1109/icassp.1995.479734>
86. M. Hariharan, J. Saraswathy, R. Sindhu, W. Khairunizam, S. Yaacob, Infant cry classification to identify asphyxia using time-frequency analysis and radial basis neural networks. *Expert Syst. Appl.* **39**(10), 9515–9523 (2012). <https://doi.org/10.1016/j.eswa.2012.02.102>
87. H. Lim, J. Park, K. Lee, Y. Han, in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop*. Rare sound event detection using 1D convolutional recurrent neural networks, (2017), pp. 80–84
88. K. Srijiranon, N. Eiamkanitchat, in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*. Application of neuro-fuzzy approaches to recognition and classification of infant cry, (2015), pp. 1–6. <https://doi.org/10.1109/TENCON.2014.7022296>
89. S. Sabour, N. Frosst, G. E. Hinton, in *Advances in Neural Information Processing Systems*. Dynamic routing between capsules, (2017). <http://arxiv.org/abs/1710.09829>
90. T. Fuhr, H. Reetz, C. Wegener, Comparison of supervised-learning models for infant cry classification / Vergleich von Klassifikationsmodellen zur Säuglingsschreianalyse. *Int. J. Health Prof.* **2**(1), 4–15 (2015). <https://doi.org/10.1515/ijhp-2015-0005>
91. R. Sahak, W. Mansor, Y. K. Lee, A. I. M. Yassin, A. Zabidi, in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*. Performance of combined support vector machine and principal component analysis in recognizing infant cry with asphyxia, (2010), pp. 6292–6295. <https://doi.org/10.1109/IEMBS.2010.5628084>
92. R. Sahak, W. Mansor, Y. K. Lee, A. I. Mohd Yassin, A. Zabidi, in *Proceedings - 2010 3rd International Conference on Biomedical Engineering and Informatics, BMEI 2010*. Orthogonal least square based support vector machine for the classification of infant cry with asphyxia, vol. 3, (2010), pp. 986–990. <https://doi.org/10.1109/BMEI.2010.5639300>
93. A. Zabidi, W. Mansor, L. Y. Khuan, I. M. Yassin, R. Sahak, in *Proceedings of 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2010*. The effect of F-ratio in the classification of asphyxiated infant cries using multilayer perceptron neural network, (2010), pp. 126–129. <https://doi.org/10.1109/IECBES.2010.5742213>
94. G. Esposito, N. Hiroi, M. L. Scattoni, Cry, baby, cry: expression of distress as a biomarker and modulator in autism spectrum disorder. *Int. J. Neuropsychopharmacol.* **20**(6), 498–503 (2017). <https://doi.org/10.1093/ijnp/pyx014>
95. S. Orlandi, C. Manfredi, L. Bocchi, M. L. Scattoni, in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Automatic newborn cry analysis: a non-invasive tool to help autism early diagnosis, (2012), pp. 2953–2956. <https://doi.org/10.1109/EMBC.2012.6346583>
96. M. Hariharan, R. Sindhu, S. Yaacob, Normal and hypoacoustic infant cry signal classification using time-frequency analysis and general regression neural network. *Comput. Methods Programs Biomed.* **108**(2), 559–569 (2012). <https://doi.org/10.1016/j.cmpb.2011.07.010>
97. A. Rosales-Pérez, C. A. Reyes-García, J. A. Gonzalez, O. F. Reyes-Galaviz, H. J. Escalante, S. Orlandi, Classifying infant cry patterns by the Genetic Selection of a Fuzzy Model. *Biomed. Signal Process. Control.* **17**, 38–46 (2015). <https://doi.org/10.1016/j.bspc.2014.10.002>
98. F. Feier, I. Enătescu, C. Ilie, I. Silea, in *2014 International Conference on Optimization of Electrical and Electronic Equipment, OPTIM 2014*. Newborns' cry analysis classification using signal processing and data mining, (2014), pp. 880–885. <https://doi.org/10.1109/OPTIM.2014.6850990>
99. A. F. Symon, N. Hassan, H. Rashid, I. U. Ahmed, S. M. T. Reza, in *4th International Conference on Advances in Electrical Engineering, ICAEE 2017*. Design and development of a smart baby monitoring system based on Raspberry Pi and Pi camera, (2017), pp. 117–122. <https://doi.org/10.1109/ICAEE.2017.8255338>
100. V. Hiremath, P. Venkataratnam, in *International Conference On Smart Technologies For Smart Nation (SmartTechCon)*. Automatic cradle system with measurement of baby's vital biological parameters, (Bangalore, 2017), pp. 480–485. <https://doi.org/10.1109/SmartTechCon.2017.8358420>
101. M. P. Joshi, D. C. Mehete, in *2017 International Conference on Computing, Communication, Control and Automation, ICCUBEA 2017*. IoT based smart cradle system with an Android app for baby monitoring, (2018), pp. 1–4. <https://doi.org/10.1109/ICCUBEA.2017.8463676>
102. W. A. Jabbar, H. K. Shang, S. N. I. S. Hamid, A. A. Almomhammedi, R. M. Ramli, M. A. H. Ali, IoT-BBMS: Internet of Things-based baby monitoring system for smart cradle. *IEEE Access.* **7**, 93791–93805 (2019). <https://doi.org/10.1109/ACCESS.2019.2928481>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.