

## Automated Speech Recognition System to Detect Babies' Feelings through Feature Analysis

Sana Yasin<sup>1</sup>, Umar Draz<sup>2,3,\*</sup>, Tariq Ali<sup>4</sup>, Kashaf Shahid<sup>1</sup>, Amna Abid<sup>1</sup>, Rukhsana Bibi<sup>1</sup>, Muhammad Irfan<sup>5</sup>, Mohammed A. HunEIF<sup>6</sup>, Sultan A. Almedhesh<sup>6</sup>, Seham M. Alqahtani<sup>6</sup>, Alqahtani Abdulwahab<sup>6</sup>, Mohammed Jamaan Alzahrani<sup>6</sup>, Dhafer Batti Alshehri<sup>6</sup>, Alshehri Ali Abdullah<sup>7</sup> and Saifur Rahman<sup>5</sup>

<sup>1</sup>Faculty of Computing, University of Okara, Okara, 57000, Pakistan

<sup>2</sup>Department of Computer Science, University of Sahiwal, Sahiwal, 57000, Punjab, Pakistan

<sup>3</sup>Computer Science Department, COMSATS University Islamabad, Lahore Campus, 57000, Pakistan

<sup>4</sup>Department of Computer Science, CUI, Sahiwal Campus, Sahiwal, 57000, Pakistan

<sup>5</sup>Electrical Engineering Department, College of Engineering, Najran University Saudi Arabia, Najran, 61441, Saudi Arabia

<sup>6</sup>Pediatric Department, Medical College, Najran University Saudi Arabia, Najran, 61441, Saudi Arabia

<sup>7</sup>Consultant of ORL&HNS and Facial Plastic Surgery, College of Medicine, Najran University Saudi Arabia, Najran, 61441, Saudi Arabia

\*Corresponding Author: Umar Draz. Email: sheikhumar520@gmail.com

Received: 06 February 2022; Accepted: 13 May 2022

**Abstract:** Diagnosing a baby's feelings poses a challenge for both doctors and parents because babies cannot explain their feelings through expression or speech. Understanding the emotions of babies and their associated expressions during different sensations such as hunger, pain, etc., is a complicated task. In infancy, all communication and feelings are propagated through cry-speech, which is a natural phenomenon. Several clinical methods can be used to diagnose a baby's diseases, but nonclinical methods of diagnosing a baby's feelings are lacking. As such, in this study, we aimed to identify babies' feelings and emotions through their cry using a nonclinical method. Changes in the cry sound can be identified using our method and used to assess the baby's feelings. We considered the frequency of the cries from the energy of the sound. The feelings represented by the infant's cry are judged to represent certain sensations expressed by the child using the optimal frequency of the recognition of a real-world audio sound. We used machine learning and artificial intelligence to distinguish cry tones in real time through feature analysis. The experimental group consisted of 50% each male and female babies, and we determined the relevancy of the results against different parameters. This application produced real-time results after recognizing a child's cry sounds. The novelty of our work is that we, for the first time, successfully derived the feelings of young children through the cry-speech of the child, showing promise for end-user applications.

**Keywords:** Cry-to-speak; machine learning; artificial intelligence; cry speech detection; babies



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1 Introduction

The brains of babies work similarly to those of adults when exposed to identical painful stimuli. To express their feelings, babies “speak” by crying. For example, the feeling of pain is normally expressed by babies by making a sound similar to “ouch”. Similarly, when the baby cannot read the environment but knows that something is not right, a unique means is required for communication. However, exactly understanding a baby’s emotions and needs is challenging, requiring the understanding of cries that show only small differences. The cry is a dynamic and advanced signal with a unique characteristic that alerts parents to the need for action. According to Wasz-Hockert, the four basic types of infant cries express pain, hunger, pleasure, and birth. For caretakers and parents to analyze these signals, the appropriate meaning of different cries must be understood. Many studies have been conducted to identify disease, such as hearing issues, respiratory problems, central nervous system impairment, etc., through cry speech. Cry speech has also been investigated to identify pain, issues, sleeping, and fear. Convolution neural networks (CNNs) have been used for these types of tasks as artificial intelligence (AI) has advanced. Several research findings have provided a limited understanding of the relationship of newborn crying with various issues. The research has focused on the association between the baby’s cry and parents’ (or any health care provider) understanding of the needs of the babies so they can ultimately take appropriate action. This is a baby’s self-support mechanism with its parents’ associations. Cry-speech researchers have not fully explored how a baby quickly tells the parents what they actually need and when it happened. In this study, we constructed an application that would be helpful to parents and doctors for detecting the feelings of a child. This application recognizes the child’s cry voices in a real-time environment. Our findings show that infants experience pain similar to adults, but infants cannot express their feelings of pain, joy, hunger, and fear. Babies experience many common health troubles such as colds, coughs, fevers, and vomiting, which can easily be investigated through cry-speech recognition. Babies also experience skin problems, such as diaper rash or cradle cap (and the baby cries when they want to be changed), which is only expressed through cry-speech level. In Japan, the number of births has shown a downward trend [1], which is considered a critical problem. One of the reasons for the lower birth rate may be the lack of a childcare guide as a social tool. In many cases, the parent works outside of the home, so a healthcare worker or babysitter is left to care for the child. As a result, a young parent frequently takes care of the infant by themselves. However, they may not have the opportunity to learn suitable childcare practices. In this situation, speech recognition technology (SRT) would provide a useful tool as a mobile application. With our developed app, when the child cries, the speech cry recognition app just needs to be opened and the start button pressed. This application recognizes the child’s cry voice and provides a result as a text describing why the baby is crying.

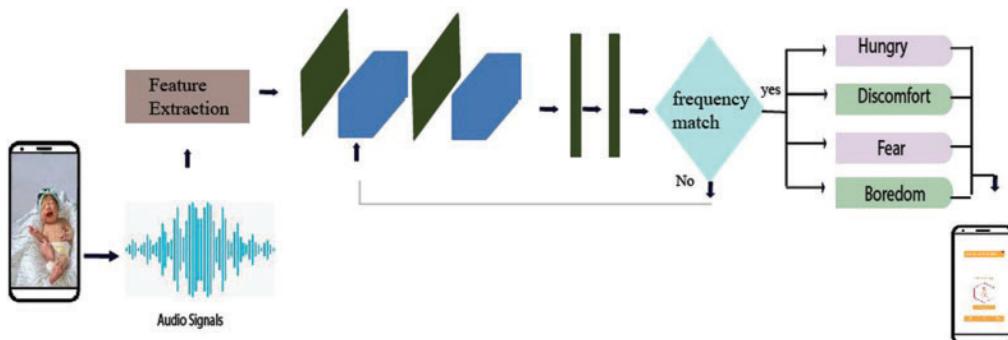
The new era of medical science and IT experts have produced advances in understanding children’s needs, especially when they are crying. In this research, our specific and main concern was finding a solution to the long-standing abovementioned problems. To capture brain signals, we used electroencephalography (EEG) in this study. EEG reads the movement of ions through the brain. We first trained the computer to encode a binary message. A value of one was assigned for movement of hands and one for feet. The movement creates specific electrical signals in the brain that were picked up by the EEG and emailed to three computers 5000 miles away. We analyzed the best spots on the receiver’s brain to stimulate little flashes in the peripheral of human visions. When the receiving computer received the email with the binary code, it was sent to a robot-assisted transcranial magnetic stimulator (TMS).

We need to learn more about childhood physiology and brain to further comprehend mental illnesses. With this knowledge, we can better treat malignancies, epilepsy, and addiction, as well as

discover more about our evolutionary history. This is why scientists are attempting to create a digital brain map. In the future, a small portion of brain tissue may be able to be uploaded as part of a long-term effort. Children make noises, they have facial expressions and body language. A child and a grown up may experience the same pain disorder but the ways in which they express their pain are distinctive [2]. An adult can speak and explain their feelings to a health practitioner or other adults, so the problem can be easily understood. However, a baby is not able to speak, so they adopt different styles of crying as expression.

The main novelty of our study is that we are the first to determine the emotions of a child in real time through the cry/voice. We observed the frequencies of the different cries of children and concluded that if the cry frequency range lies between 700 and 1100, the baby is crying due to severe pain. We successfully identify the emotions of babies through the sound of their crying.

Starting at approximately age three or four years, children can reliably use pain scales, which extensively increases the accuracy of an assessment [3]. A child's self-opinions (what the child says) may be considered along with behavioral and physiological signs. Many of these pain scores can be reliably used at home and in the clinic. In a previous study, children rated the similarities in facial expressions. Control children regarded discrete feelings as such, children that had been neglected distinguished fewer emotions, and physically abused children identified the maximum number of emotions [4]. Fig. 1 represents the workflow of the proposed application for determining how babies feel through cry-speech recognition.



**Figure 1:** Workflow of proposed application, starting from feature extraction and frequency matching through to the detection of the various feelings of babies (e.g., hunger, discomfort, fear, and boredom) through cry-speech recognition

## 2 Related Work

Speech recognition and system translation equipment are two of the main uses of artificial intelligence. Both permit our phrases to be understood by a device, enabling the provision of actionable instructions, used either to manipulate gadgets, such as Google Home Clever Speaker (GHCS), or allowing communication with a person who speaks a different language [5]. Advances in AI generation tools (AIGTs) has enabled the deciphering of babies' cries and discerning their meaning, for example, the development of a baby crying detector (BCD), which is one of many demonstrations available on STBLESensor on iOS and Android mobile devices. The app works in tandem with SensorTile (ST) Field, ST's most powerful sensor platform. As the name implies, the BCD detects whether a baby is crying [6]. ST records field facts in 16 kHz samples using its MP23ABS1 MEMS microphone. It then processes the signal before passing it to a system with a set of rules working on the host STM32L4R9.

If the machine detects crying, the LED on the sensor board turns green, and the smartphone receives a Bluetooth alert. This method paved the way for the BCD through the combination of AI and technology.

Chatter Baby was developed by researchers at the University of California, Los Angeles. The app is primarily based on a set of rules that allow the algorithm to precisely learn each child's crying method and relay these data to the parents [7] with extremely high accuracy, reducing the guesswork required by first-time parents when reacting to their baby's crying. To create the Chatter Baby app, Anderson et al. imported 2000 audio samples of infant cries [8]. They then used AI-based algorithms to identify and explain the distinction between pain-, hunger-, and fussiness triggered cries:

The education was performed via extracting many acoustic functions from our database of pre-classified cries... Pain cries had been taken at some point of vaccinations and ear piercings. We classified other cries using the parent-nomination and a mother panel which includes veteran mothers who had at the least kids. Only cries that had three unanimous ratings were used to train our algorithm, which changes and improves frequently. We used the acoustic capabilities to train a machine getting to know the set of rules to predict the most likely cry reason. Within our sample, the set of rules was about 90 percent correct to flag ache, and over 70 percent accurate standard.

Infant crying can trigger a variety of emotions and behavioral responses in others, with selfless care-giving being a more prevalent response than aggressiveness [9]. In 60 children ranging in late preschool to preadolescence in age, responses to the cries of younger newborns were investigated. From a neighboring room, each participant heard both a short and a full-length tape-recorded cry of an infant, then a mother entered, searching for her (previously) "crying" child's bottle. Each participant was then questioned [10]. As a result, children's emotions and behaviors were assessed in reaction to simulated, real-life, and hypothetical distress. Self-evaluations of empathy, verbalized intents to help, actual helpful responses, and observers' rankings of negative emotion were all common responses to screams in children of all ages. Furthermore, with increasing participant age, prosocial and behavioral interventions substantially increased. The subsequent prosocial conduct that required direct engagement with the child was inversely connected with expressions of negative emotion. The majority of children's emotions and actions were unaffected by whether they heard short or longer screams. The respondents could, however, tell the difference between these screams and expressed "theories" about how they were affected by the cries. Neural networks (NNs) are tools used for improving our understanding of complex brain activities. To achieve this goal, NNs must be neurobiologically feasible [11].

Baby cry researchers have examined the differences between regular and pathological (deaf or hearing-loss babies) cries [12–14] or tried to separate conditional cries [15], such as soreness from needles, worry caused by jack-in-the-box toys, and frustration owing to constraints. A commercially available tool called Cry Translator (CT) claims to be able to detect five cry types: hunger, sleep, soreness, strain, and boredom. The cry type is determined using a standard set of principles that consider the child cry's pattern in terms of loudness, pitch, tone, and inflection [16,17]. In this study, we employed this popular approach to distinguish individual baby cry indicators and found that a set of criteria can be applied to correctly identify a baby's cry meaning 70% of the time [18,19].

The parent of a child may make educated guesses about their infant's needs based on their cry, facial features, body language, and so on, but they not always be correct. At times the parents make the wrong guess regarding their infant's cry and may end up endangering their child. The newborn may cry for a variety of reasons, including exhaustion, stress, not being able to locate their parent, blood loss, abdominal pain, and so on, but the three crucial cries indicate hunger, pain, and extreme

pain. Infants employ typical mouth noises that may be grouped into the following categories: (1) earth (gas), (2) neh (hungry), (3) new (sleepy), (4) heh (uncomfortable), and (5) eh (exhausted) (need to burp). The cry of a small child contains a large amount of information about their physical status, which may be decoded using a one-of-a-kind system and a deep learning algorithm. Deep learning is highly useful for recognizing and classifying a toddler's cry, according to most recent studies. As such, our primary goal in this study was to classify toddler cries. Using deep learning approach to speech is not always simple because methods such CNN are used for pattern recognition; to determine styles speech, the input must first be converted to a suitable format, after which deep learning algorithms are implemented and outcomes are obtained. A feature extraction approach must be used in the conversion. Currently, a variety of strategies can be employed to identify and type digital baby cries: CNN, MFCC, GMM-UBM, SVM, LPCC, decision tree with multilayer perceptron (DTMLP), etc. [20]. Scientists think that child screams are similar to those of adults; hence, the findings of studies on toddler cry recognition should be compared with those of adult voice recognition. The acoustic information of a toddler's scream revealed substantial variances between various types of cries, which may aid in the recognition of a toddler's cry using system learning [21–23]. Tab. 1 lists the studies that considered voice recognition, as well as results and accuracies of the methods.

**Table 1:** Comparison of different expression detection studies that emphasize the technology used and its associated results with accuracy level

No.	Name	Year	Tools and methods	Result	Accuracy
1	Automatic analysis of facial expression	2017	Dynamic time-wrapping, global alignment kernel, classification of trajectories in system++(m)	Deep covariance descriptors and deep covariance trajectories for facial expression recognition	65%
2	Facial expression detection of amused, persuaded, informed, sentimental, and inspired emotions	2016	Database, convolutional neural network (CNN), model validation, confusion matrix for true and predicted emotion counts	When more pooling is added, the feature maps become increasingly abstract	58%
3	Facial expression perception	2016	Social intelligence, black and white photographs	Perceived neural stimuli and only using compensatory strategies, autistic subjects could understand only theoretical meaning of emotional expressions	50%

(Continued)

**Table 1:** Continued

No.	Name	Year	Tools and methods	Result	Accuracy
4	Review of dynamic dataset for facial expression research	2017	International Joint Conference on Artificial Intelligence (IJCAI), International Conference on Machine Learning (ICML), Real-World Affective Faces (RAF)	Broadly describes understanding of facial expressions and their meanings	60%
5	Computer recognizes and responds to user emotions; theoretical and practical implications	2002	Python, HCI	Users to not only focus on information processors; as a normal human being; emotional needs and productivity must be considered	75%
6	Machine analysis of facial expressions	2007	Python, Google Collab, Jupyter Notebook	CNN	80%
7	Recognizing human facial expressions with machine learning	2018	Intertools, matplotlib, Python 3.5 or above.	Test loss: 1.868 Validation loss: 0.973	Threshold accuracy: 86.98% Test accuracy: 0.705
8	Recognition of facial expressions and measurements of level of interest from video	2006	Python, hidden Markov model (HMM), HCI	Briefly describes facial expressions video and other factors that will be helpful now and near future	90.9%
9	Agreement on basic emotions	1992	Biological expressions and experience of emotions, Cannon bard theory of emotions, James Lange theory of emotions, two-factor theory of emotions	A deep study of basic emotions such as crying, happiness, sadness, anger, etc.	70%
10	Infant crying and parent well-being (ICPW) [22]	2021	Screening tool developed to provide efficient and effective indication of infant crying and soothing	Satisfactory diagnostic outputs for extremely high level of crying	Mean comparison of positive and negative screens: 0.57 to 1.16

(Continued)

**Table 1:** Continued

No.	Name	Year	Tools and methods	Result	Accuracy
11	Comparison of different classification techniques [23]	2021	Logistic regression, SVM, random forest, and KNN were used on donate-a-cry corpus	Develop screening instruments on the basis of auditory characteristics of crying	MFCC, STACF, GTCC = 84%

### 3 Materials and Methods

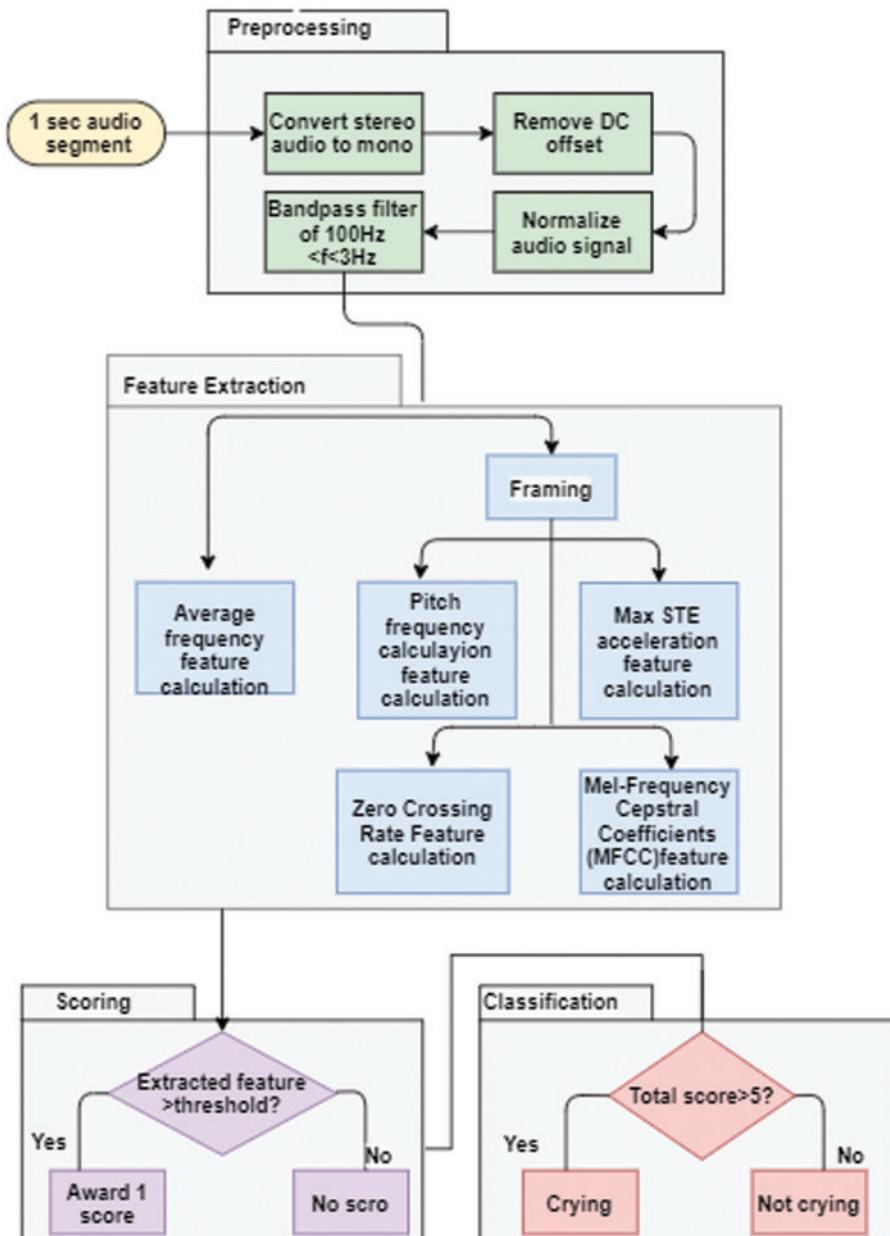
Speech recognition, also known as automatic speech recognition (ASR), is an interdisciplinary subfield of computer science and computational linguistics in which approaches and technology are developed to allow computers to recognize and translate spoken language into text [3]. Voice recognition systems (VRSs) require training, which entails a single person reading text or isolated language into the machine. The technology then analyzes the person's unique voice and uses it to improve speech recognition, resulting in higher accuracy. Speaker-dependent systems rely on training. Recognizing the speaker can help to simplify the system used for translating speech in structures that have been trained on the crying voice of a specific child.

#### 3.1 Hidden Markov Model (HMM) and HTK

HMMs are used in the most advanced ASR systems available today. HMMs are constructed by programmers with expertise in statistical signal processing, acoustic modeling, and signal segmentation [12]. For any ASR system, a sequence of statistical models that explains the diverse acoustic activities of the sounds must be built and adapted.

Speech, song, baby screams, and other sounds may be represented as a sequence of characteristic vectors (temporal, spectral, or both), and HMMs may be useful in creating and implementing time-varying spectral vector sequences [13,24]. An HMM generates a series of observations with the values  $O = O_1, O_2$ . The evaluation, decoding, and training problems must all be solved in HMMs. HMMs, the aforementioned issues, and the proposed solutions have all been well-studied [14,25–30].

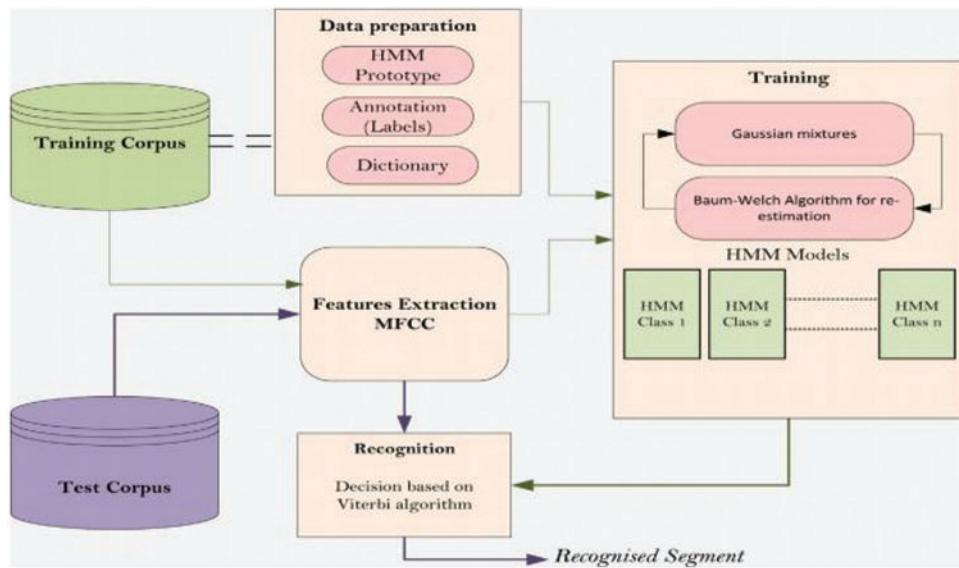
The Viterbi set of rules is a decoding approach for determining a machine's highest probable destiny state based just on its current state. To estimate the HMM parameters, the Baum Welch set of rules, which is an iterative process, is used. The Hidden Markov Model Toolkit (HTK) is used to build and manage HMMs by decoding unknown observations using training observations from a valid corpus. It has preinstalled library modules and tools written in C. Although the HTK has only been used in speech research, it is flexible enough to facilitate the creation of a variety of HMM systems [12,31]. Using the HTK, the audio characteristics of baby cries can be extracted so that they can be distinguished from screams for a specific reason [15,32] and [16,33]. Fig. 2 depicts the flowchart of a baby cry recognition system including preprocessing, feature extraction, classification, and scoring. The preprocessing is performed through stereo audio motion analysis, the results of which are directly converted into mono audio format and the DC offset is removed. The filtered and normalized audio signals with threshold bandpass signals are then categorized. Framing maximum STE is applied to accelerate the feature and average frequency calculations. Scoring and classification are applied to extract the features from cry speech as per the set threshold values.



**Figure 2:** Flowchart of the proposed method for detecting baby cries with its classification, scoring, and feature extraction criteria

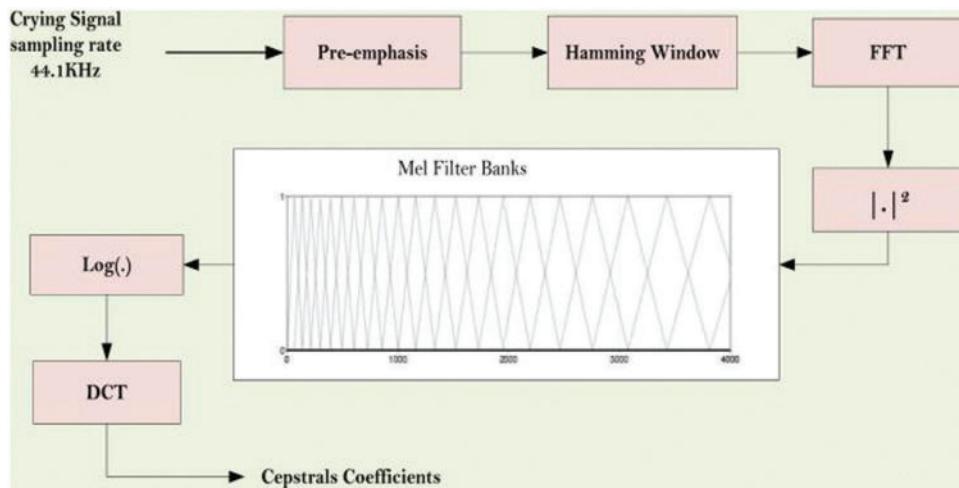
### 3.2 Database

The database we used in this study contains a several hours of audio recordings made by parents of babies taken from Google. The babies were in their first 6 months of life. The recordings consist of various types of sounds such as crying, parents talking, etc. The sampling frequency of the recording ( $F_s$ ) is 15,000 Hz. Fig. 3 provides the details of the training and testing data.

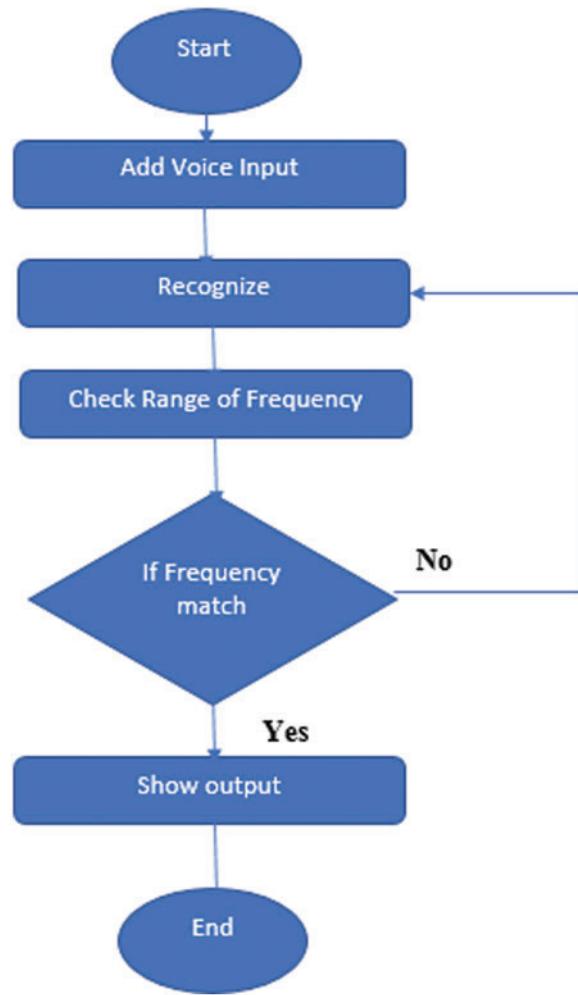


**Figure 3:** Automatic infant cry segmentation system architecture for recognizing segments using an HMM on a training and testing corpus

The data preparation, feature extraction, training, and recognition modules comprise the cry segmentation system architecture. Fig. 4 depicts a block diagram of the proposed system and Fig. 5 outlines the process.



**Figure 4:** Mel frequency cepstral coefficients (MFCCs) extraction from recorded audio signals



**Figure 5:** Process flow of baby cry-speech recognition system

### 3.3 Fundamental or Pitch: Frequency

The Baby Speech Recognition app identifies the four types of cries using specific frequency ranges:

**Pain Cry:** If the frequency ranges is between 700 and 1100 Hz, the baby is crying due to pain. If the sound does not lie in this frequency, the process continues.

**Hunger Cry:** If the frequency ranges between 400 and 600 Hz, then the baby is crying due to hunger.

**Fear Cry:** If the frequency ranges between 50 and 400, then the baby is crying due to fear.

**Happy Cry:** If the frequency ranges between 600 and 700, then the baby is happy.

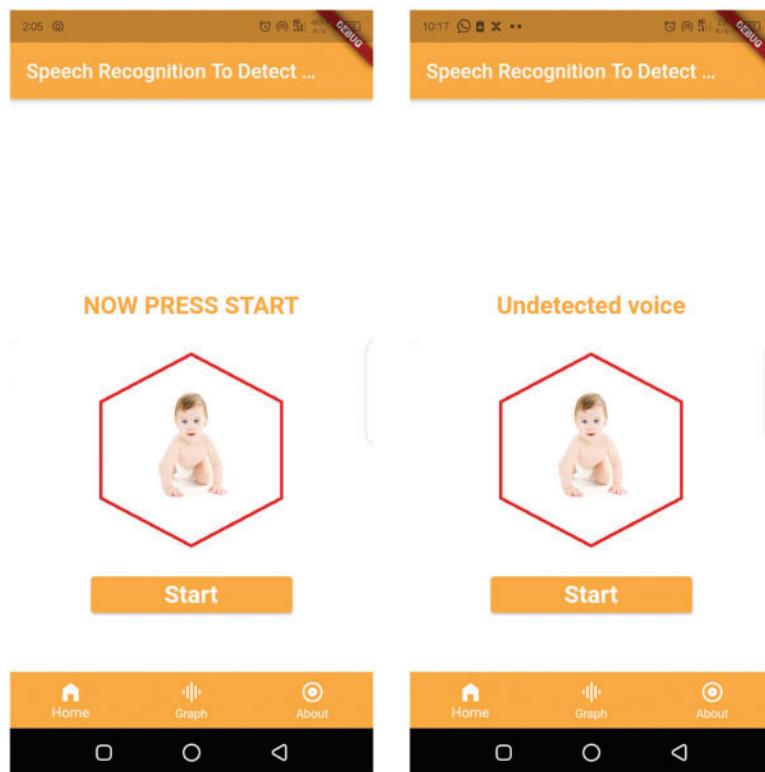
**Undetected Voice:** If the frequency of the voice does not lie in any of the above frequency ranges, then it shows that the voice is undetected.

#### 4 Results

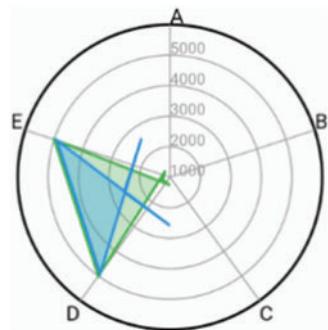
The proposed automatic child cry identification method intelligently classifies a baby's cries. We built and implemented an automatic infant crying detection method in this study, based on the frequencies of babies' crying. The method is adaptable, accurate, efficient and works in real-time according to the test results [34–38]. The automatic monitoring system intelligently recognizes the baby cries [39,40]. The proposed method can be used for improving the efficiency of medical staff and reducing their working intensity. Tab. 2 lists the accuracy of cry detection for upset (due to pain), hunger, fearing, and happy emotions. In addition, Fig. 6 shows the baby cry-speech recognition window and undetected voice notification. Figs. 7–10 show the states of upset (due to pain), hunger, fear, and happy, respectively. Figs. 11–14 show the panel window used for the fear, happy, upset, and hungry states along with their graphical analyses, respectively. In addition, Tab. 3 presents the baby cry-speech detection with pitch and deviation for the different frequency ranges.

**Table 2:** Accuracy of the recognition rates of baby's crying using a continuous cry

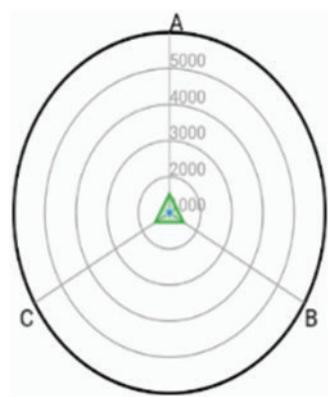
Feeling	Accuracy
Pain	70%
Hungry	98%
Fear	75%
Happy	85%



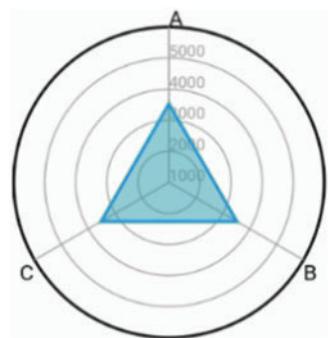
**Figure 6:** Baby cry-speech recognition system window and undetected voice panel



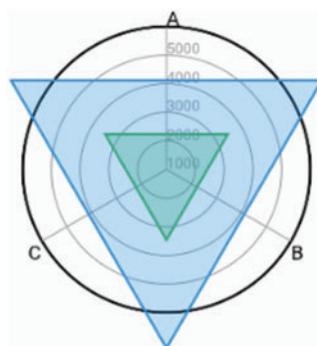
**Figure 7:** Baby feel upsets due to pain



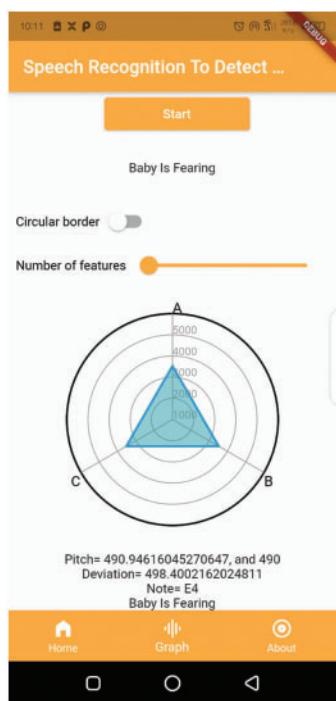
**Figure 8:** Baby is hungry



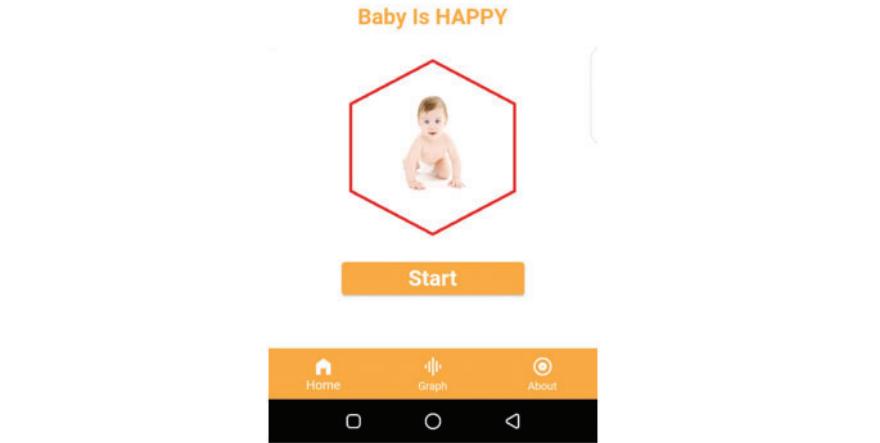
**Figure 9:** Baby is afraid



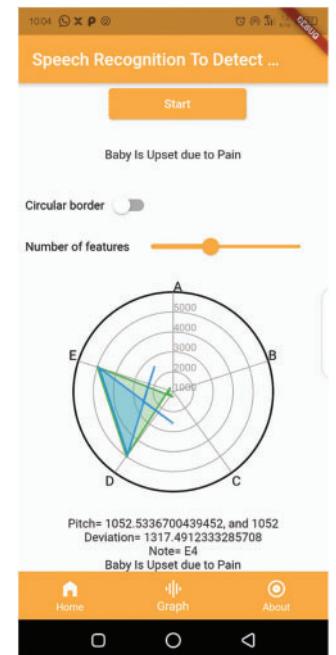
**Figure 10:** Baby is happy



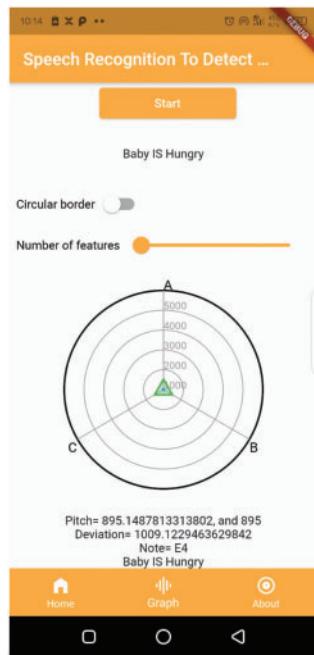
**Figure 11:** Baby cry speech recognition system window: fear state



**Figure 12:** Baby cry speech recognition detection window: happy state



**Figure 13:** Baby cry speech recognition system window: pain



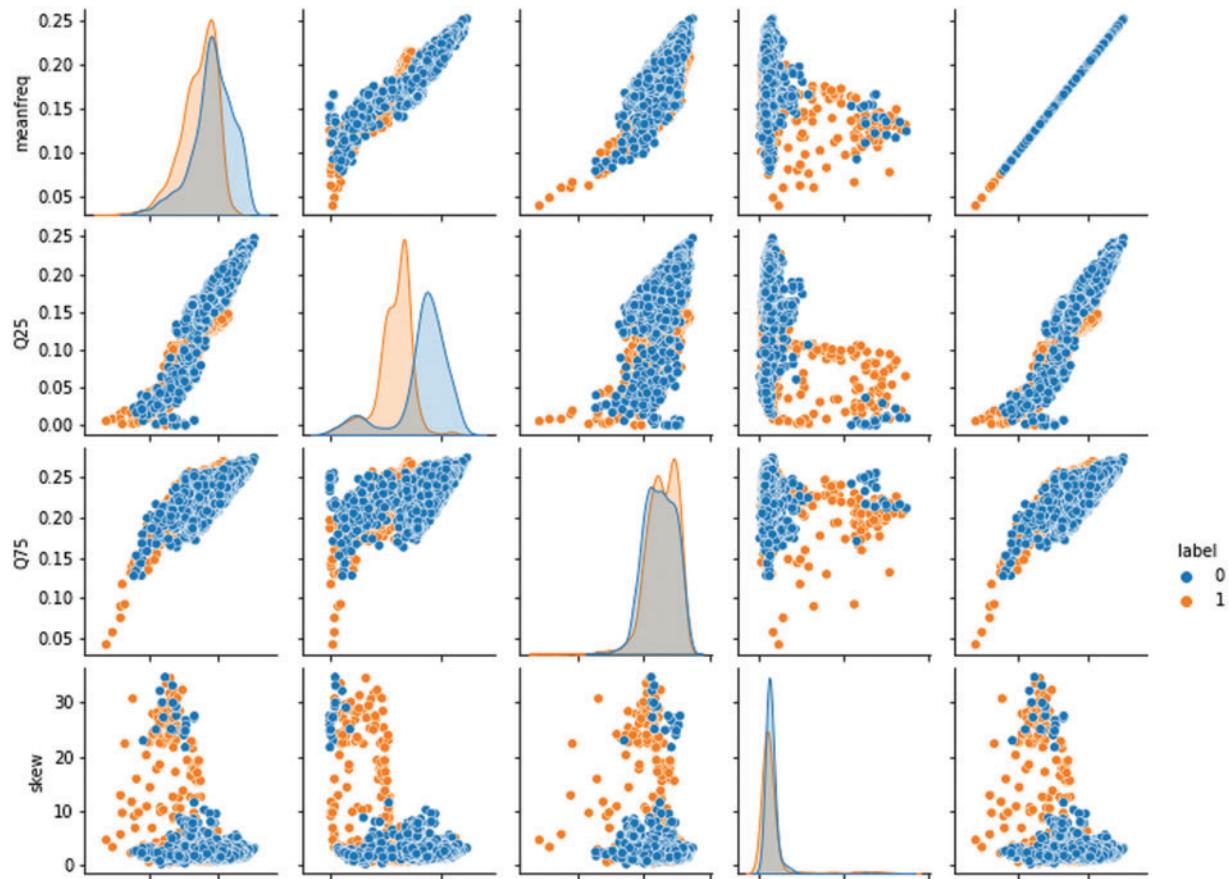
**Figure 14:** Baby cry speech recognition system window: hunger

**Table 3:** Detection of baby feelings using pitch and deviation in different frequency ranges

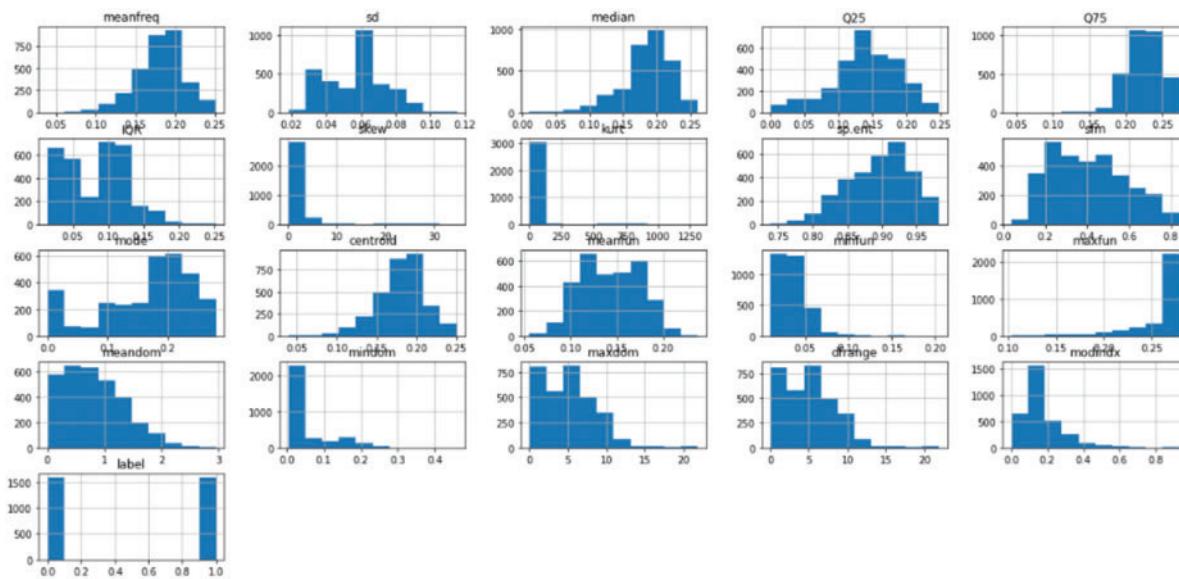
CRY	Frequency	Pitch	Deviation	Note
Pain	700 to 1100	1052.5336700439452 and 1052	1317.4912	E4
Hunger	400 to <600	895.1487813313802 and 895	1009.1229	E4
Fear	50 to <400	490.94616045270647 and 490	498.4002	E4
Happy	600 to <700	1491.7787148611885 and 1491	2249.9583	E4
Undetected	Frequency does not lay in the above ranges	1109.5246887207031 and 1109	969.964	E4

## Feature Analysis

We plotted the different feature analyses of the dataset to visualize the effect of our method. We used the meanfreq analysis, Q25, Q75, skew, and centroid to evaluate if the method successfully labeled types of baby cry with the targeted objective value. We labeled the targeted value that successfully measured and produced the results are label as one; otherwise, we used a value of zero for untargeted values. We provide a detailed description in Fig. 15, which shows that the plot skewness with the quartiles are weakly related to the parameters of cry-speech, unlike the mean frequency with the centroid, which are strongly related. For the same analysis, we drew a bar chart to further explain the function of the dataset, and how cry frequency gradually increases and decreases according to the pitch of the sound. In the analysis, 50% of the participants were male babies and 50% were female babies. The bar chart in Fig. 16 shows the outliers: the data did not contain many outliers. The features that contained the most outliers were skewness and kurtosis. We also can found that the data were not biased as they were split evenly between the classes. Tab. 3 shows that the success rate of baby feeling detection and the pitch deviation.



**Figure 15:** Plotted relationships between the dataset features and skewness



**Figure 16:** Outliers and the features, which are split evenly between the classes (50% boys; 50% girls)

## 5 Conclusions and Future Work

In this study, we identified the emotions of babies using the sound of their crying. Our developed application can further help doctors and parents to identify their child's needs through cry-speech recognition. For this, the relevant persons (parents, doctors, etc.) can easily determine the child's condition through these emotions to provide the appropriate solutions. Our application represents the baby's emotions through their different types of cry. We successfully used the proposed technique of identifying a baby's emotions using their cries. This method, based on AI, has real-world applications. When assuming that the emotions were sadness, hunger, happiness, and pain, the average baby emotions recognition accuracy of the method was 80%. We will employ our proposed approach to determine children's emotions on many more children in the future because the emotions represented by a child's cry may vary. We will also include additionally frequencies to help distinguish more emotions to provide further help for parents and doctors.

**Acknowledgement:** The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the research collaboration funding program grant code (NU/RC/SERC/11/5).

**Funding Statement:** This research was funded by the Deanship of Scientific Research, Najran University, Kingdom of Saudi Arabia, grant number NU/RC/SERC/11/5.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 2, pp. 254–272, 1981.

- [2] T. Fuhr, H. Reetz and C. Wegener, "Comparison of supervised-learning models for infant cry classification/Vergleich von Klassifikationsmodellen zur Säuglingsschreianalyse," *International Journal of Health Profession*, vol. 2, no. 1, pp. 4–15, 2015.
- [3] G. V. I. S. Silva and D. S. Wickremasinghe, "Infant cry detection system with automatic soothing and video monitoring functions," *Journal of Engineering and Technology of the Open University of Sri Lanka*, vol. 5, no. 1, pp. 36–53, 2017.
- [4] D. Ravichandran, P. Praveenkumar, S. Rajagopalan, J. B. B. Rayappan and R. Amirtharajan, "ROI-based medical image watermarking for accurate tamper detection, localisation and recovery," *Medical & Biological Engineering & Computing*, vol. 59, no. 6, pp. 1355–1372, 2021.
- [5] Y. Skogsdal, M. Eriksson and J. Schollin, "Analgesia in newborns given oral glucose," *Acta Paediatrica*, vol. 86, no. 2, pp. 217–220, 1997.
- [6] S. M. Luddington-Hoe, X. Cong and F. Hashemi, "Hashemi infant crying: Nature, physiologic consequences, and select interventions," *Neonatal Network*, vol. 21, no. 2, pp. 29–36, 2002.
- [7] E. Rayachoti and S. R. Edara, "Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest," *IET Image Processing*, vol. 9, no. 8, pp. 615–625, 2015.
- [8] M. A. T. Turan and E. Erzin, "Monitoring infant's emotional cry in domestic environments using the capsule network architecture," in *Interspeech*. Hyderabad, India, 132–136, 2018.
- [9] C. Zahn-Waxler, S. L. Friedman and E. M. Cummings, "Children's emotions and behaviors in response to infants' cries," *Child Development*, vol. 54, no. 6, pp. 1522–1528, 1983.
- [10] P. R. Myakala, R. Nalamachu, S. Sharma and V. K. Mittal, "An intelligent system for infant cry detection and information in real time," in *Seventh Int. Conf. on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) IEEE*, San Antonio, USA, pp. 141–146, 2017.
- [11] F. Pulvermüller, R. Tomasello, M. R. Henningsen-Schomers and T. Wennekers, "Biological constraints on neural network models of cognitive function," *Nature Reviews Neuroscience*, vol. 22, no. 8, pp. 488–502, 2021.
- [12] B. Bathellier, L. Ushakova and S. Rumpel, "Discrete neocortical dynamics predict behavioral categorization of sounds," *Neuron*, vol. 76, no. 2, pp. 435–449, 2012.
- [13] M. Gales and S. Young, "The application of hidden Markov models in speech recognition," *Foundation and Trends in Signal Process*, vol. 1, no. 3, pp. 195–304, 2007.
- [14] M. A. Bourjaily and P. Miller, "Dynamic afferent synapses to decision-making networks improve performance in tasks requiring stimulus associations and discriminations," *Journal of Neurophysiology*, vol. 108, no. 2, pp. 513–527, 2012.
- [15] N. Brunel, "Dynamics and plasticity of stimulus-selective persistent activity in cortical network models," *Cerebral Cortex*, vol. 13, no. 11, pp. 1151–1161, 2003.
- [16] K. Kuo, "Feature extraction and recognition of infant cries," in *Proc. of the IEEE Int. Conf. on Electro/Information Technology (EIT)*, Normal, IL, USA, pp. 1–5, 2010.
- [17] L. Liu, K. Beemanpally and S. M. Kuo, "Real-time experiments of ANC systems for infant incubators," *Noise Control Engineering Journal*, vol. 60, no. 1, pp. 36–41, 2012.
- [18] L. Liu, K. Kuo and S. M. Kuo, "Infant cry classification integrated anc system for infant incubators," in *IEEE Int. Conf. on Networking, Sensing and Control*, Paris, France, pp. 383–387, 2018.
- [19] L. Liu and K. Kuo, "Active noise control systems integrated with infant cry detection and classification for infant incubators," *Acoustic, Hong Kong*, vol. 131, no. 4, pp. 1–6, 2012.
- [20] A. Sharma and D. Malhotra, "Speech recognition based IICC-Intelligent Infant cry classifier," in *IEEE Int. Conf. on Smart Systems and Inventive Technology*, Tirunelveli, India, pp. 992–998, 2020.
- [21] J. O. Garcia and C. A. Reyes Garcia, "Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks," in *Proc. of the Int. Joint Conf. on Neural Networks*, Portland, USA, vol. 4, pp. 3140–3145, 2003.

- [22] S. Lamidi, C. Tadj and C. Gargour, "Biomedical diagnosis of infant cry signal based on analysis of cepstrum by deep feedforward artificial neural networks," *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 2, pp. 24–29, 2021.
- [23] C. Ji, T. B. Mudiyanselage, Y. Gao and Y. Pan, "A review of infant cry analysis and classification," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 1, no. 8, pp. 1–17, 2021.
- [24] K. Ashwini, P. D. R. Vincent, K. Srinivasan and C. Y. Chang, "Deep learning assisted neonatal cry classification via support vector machine models," *Frontiers in Public Health*, vol. 9, no. 1, pp. 1–10, 2021.
- [25] D. Bhati, "Cry classification," *Journal of Advanced Research in Signal Processing and Applications*, vol. 2, no. 2, pp. 8–10, 2021.
- [26] A. Shankhdhar, V. Kumar and Y. Mathur, "Human scream detection through three-stage supervised learning and deep learning," *Inventive Systems and Control*, vol. 204, no. 3, pp. 379–390, 2021.
- [27] S. Fusi, E. K. Miller and M. Rigotti, "Why neurons mix: High dimensionality for higher cognition," *Current Opinion in Neurobiology*, vol. 37, no. 1, pp. 66–74, 2016.
- [28] S. A. Romanov, N. A. Kharkovchuk, M. R. Sinelnikov, M. R. Abrash and V. Filinkov, "Development of an non-speech audio event detection system," in *IEEE Conf. of Russian Young Researchers in Electrical and Electronic Engineering*, St. Petersburg, Russia, pp. 1421–1423, 2020.
- [29] F. S. Matikolaie and C. Tadj, "On the use of long-term features in a newborn cry diagnostic system," *Biomedical Signal Processing and Control*, vol. 59, no. 2, pp. 1–10, 2020.
- [30] R. Cohen, D. Ruinskiy, J. Zickfeld, H. IJzerman and Y. Lavner, "Baby cry detection: Deep learning and classical approaches, development and analysis of deep learning architectures," *Springer*, vol. 867, no. 1, pp. 171–196, 2019.
- [31] L. S. Foo, W. S. Yap, Y. C. Hum, Z. Kadim, H. W. Hon *et al.*, "Real-time baby crying detection in the noisy everyday environment," in *11th IEEE Control and System Graduate Research Colloquium*, Shah Alam, Malaysia, pp. 26–31, 2020.
- [32] J. Chunyan, M. T. Bamunu, G. Yutong and Y. Pan, "A review of infant cry analysis and classification," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 8, no. 1, pp. 1–12, 2021.
- [33] K. S. Alishamol, T. T. Fousiya, K. J. Babu, M. Sooryadas and L. Mary, "System for infant cry emotion recognition using DNN," in *Third Int. Conf. on Smart Systems and Inventive Technology*, Tirunelveli, India, pp. 867–872, 2020.
- [34] L. Abou-Abbas, C. Tadj and H. A. Fersaie, "A fully automated approach for baby cry signal segmentation and boundary detection of expiratory and inspiratory episodes," *The Journal of the Acoustical Society of America*, vol. 142, no. 3, pp. 1318–1331, 2017.
- [35] F. S. Matikolaie and C. Tadj, "Machine learning-based cry diagnostic system for identifying septic newborns," *Journal of Voice*, vol. 10, no. 1, pp. 1–10, 2022.
- [36] J. Cha and G. Bae, "Deep learning based infant cry analysis utilizing computer vision," *International Journal of Applied Engineering Research*, vol. 17, no. 1, pp. 30–35, 2022.
- [37] F. S. Matikolaie, Y. Kheddache and C. Tadj, "Automated newborn cry diagnostic system using machine learning approach," *Biomedical Signal Processing and Control*, vol. 73, no. 1, pp. 100–110, 2022.
- [38] S. Lahmiri, C. Tadj, C. Gargour and S. Bekiros, "Deep learning systems for automatic diagnosis of infant cry signals," *Chaos Solitons & Fractals*, vol. 154, no. 2, pp. 35–46, 2022.
- [39] X. R. Zhang, W. F. Zhang, W. Sun, X. M. Sun and S. K. Jha, "A robust 3-D medical watermarking based on wavelet transform for data protection," *Computer Systems Science & Engineering*, vol. 41, no. 3, pp. 1043–1056, 2022.
- [40] X. R. Zhang, X. Sun, X. M. Sun, W. Sun and S. K. Jha, "Robust reversible audio watermarking scheme for telemedicine and privacy protection," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.