Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020)
IEEE Xplore Part Number: CFP20P17-ART; ISBN: 978-1-7281-5821-1

System for Infant Cry Emotion Recognition using DNN

Alishamol K. S., Fousiya T. T., Jasmin Babu K., Sooryadas M., and Leena Mary
(email:ksalisha333@gmail.com;fousiyamolu1998@gmail.com;jasminebabu1998@gmail.com; sooryadevadasan@gmail.com; leena.mary@gecidukki.ac.in)
Dept. of Electronics and Communication
Government Engineering College
Idukki

Abstract— **In general, a child, especially a neonate, requires the utmost care and attention of parents or elders. But nowadays, parents are very busy with their work and hardly find time to take care of infants. This may lead to physical and mental health issues for infants as well as their parents. The proposed system identifies the emotion of an infant through cry analysis. Infant cry signal vary according to emotion, in terms of pitch, formants, and energy. These variations can be represented by using various parametric features such as mel-frequency cepstral coefficients (MFCC). In this work, a deep neural network (DNN) classifier is employed for infant cry emotion recognition. Using different features, infant cries are categorized into three classes, namely, 'hunger', 'sleep' and 'discomfort'. Apart from recording the infant cry, the proposed baby care system also monitors the body temperature, humidity of diaper and outdoor temperature by using appropriate sensors. These values and the recognized cry emotion are sent to the parent's mobile via the GSM module of Arduino UNO. The proposed system then activates immediate soothing measures such as playing music and/or playing parent's voice.**

Keywords— Infant cry analysis, infant cry emotion recognition, MFCC, deep neural network

I. INTRODUCTION

Nowadays, parents are very busy with their work. Their busy work and less enough time to care for babies will make diseases to infants. So, it is essential to have a monitoring system for infants. Plenty of infant monitoring systems are available in the market, but at very high prices. It only measures the external parameters of the babies and judges the cry reason. So, they all have low accuracy and high price. In this paper, we introduce a system for monitoring an infant's cry emotion by classifying the cry audios using deep neural networks (DNN) and measuring the external parameters using sensors.

Infants express their emotions through cries. This emotion can be interpreted (up to some extent) by the mother or others who are in close contact with the infant. Infant cry is very different for 'angry' or 'sleepy' emotions and it may vary for each baby. The main causes of infant cry may be due to hunger, sleepiness, and discomfort.

Infant cry analysis is of research interest for many years. A crying sound is the coordinated output of the work of several brain regions controlling respiration and vibration of the vocal cord [1]. The differences in the infant's cry patterns are studied for many purposes. Apparent behavior of an infant arising from his/her needs is expressed through cry [2]. Understanding the cause of infant cry is essential to identifying their needs very quickly in most cases. The theory that underlies most acoustic analysis of cry sounds is the sound-filter theory [3]. The infant cry falls in the most sensitive range of the human auditory sensation. Their crying style helps the parents to understand the condition of the baby. When a sick baby is crying, it will be different from its normal cry. In this paper, we discuss the analysis of infant cry and present a system for cry emotion detection. In order to address problems related to infant care, a smart baby care system is proposed here. The system also monitors the baby's body temperature, humidity, and outdoor temperature to assess. Therefore, the performance of a system that will decide the reason for an infant cry should be reliable [4]. A Cry is a term associated with the total duration of the acoustic signal, including all inspirations and expirations, from the beginning of the emission until the infant keeps quiet. In the past, technological aspects imposed restrictions to develop efficient analysis techniques [5] of the infant's cry.

Physiological variables such as facial expressions, muscular tonus, sleep, and suction abilities has been researched as parameters to estimate the needs of the infant [6]. However, during the latest years, the variable "cry" has been used in the major part of the studies [7].

The proposed system identifies the emotion of a baby, using appropriate feature extraction followed by classifiers. Deep Neural Network (DNN) is used as classifier here for cry emotion detection. Fundamental frequency (F_0), short time energy (STE), linear prediction (LP), and fast fourier transform (FFT) are used for cry analysis at the initial stages of the work. Along with the cry signal, the system records different parameters such as body temperature and diaper humidity, in order to ensure the cry emotion. By combining these parameters along with the output of the DNN classifier, the reason for the cry is determined and appropriate soothing measures are taken.

The remaining part of the paper is organized as follows: Section 2 discusses the methodology followed in this work for infant cry emotion recognition. The design of the system is described in Section 3. Section 4 describes the experiments conducted using the collected database, followed by conclusion in Section 5.

II. METHODOLOGY

The proposed system distinguishes the infant's cry from other voices (parents' voice) and identifies the emotional status, using appropriate feature extraction followed by classifiers. Along with the cry, the system records body temperature and diaper humidity, as additional clues for identifying the reason for the cry. The system takes temporary measures to soothe the infant by playing music and a parent's voice. Finally, the system sends a message or makes a call to the guardian informing the status, with the help of a GSM module in order to ensure human attention. If the baby cries, the system takes images and sends it to the parent's mobile via GMAIL.

Figure 1 illustrates the different stages in the development of the proposed system. Data collection is done for training the models for infant cry emotion recognition. A portable audio recorder can be used for this purpose. Different types of infant cries should be identified and recorded. Several examples belonging to multiple infants (both male and female) containing different emotions are required for training the models.

Analysis of infant cries is then carried out for identifying features relevant for infant cry emotion recognition. Analysis of baby cry is based on some features like pitch, short-time energy, formants, etc. The Spectrogram is also investigated for possible detection of emotions. Possibility of using spectral features such as linear prediction coefficients (LPCC), Mel-frequency cepstral coefficients (MFCC), and source features such as LP residual is also investigated.

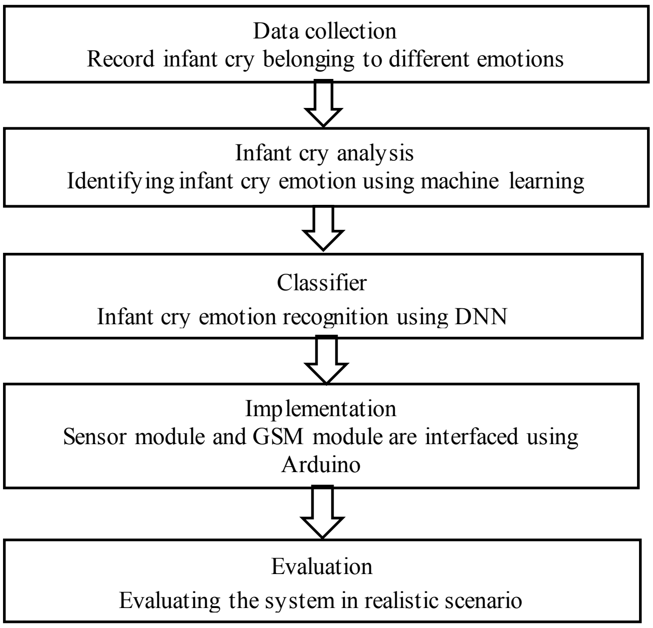


Figure 1. Block schematic illustrating different stages in the development of proposed infant cry recognition system

A. Data Collection

For this work, infant cry samples are collected from different babies belonging to three emotions, namely, 'hunger', 'discomfort' and 'sleepy'. Audios of babies are collected from houses, hospitals and various websites. A portable audio recorder was used for this purpose and was placed at a distance of 10-25 cm from the infant's mouth. All audios are collected using the microphone and recorder records the audios of cries in a standard format, .wav file. The cries of babies are monitored for 12 seconds to detect the cause of the cry with maximum accuracy.

Table 1 shows the details of the infant cry database collected for this work. Cries belonging to multiple infants (15 male babies and 16 female babies) containing three different emotions are included in this set. About 600 files of 6-second average durations were collected from 31 babies of different regions. For feature extraction, 6-second duration recorded cry signals were converted into 1-second duration with a sampling rate of 16KHz. Out of 600 files in the dataset, 450 was reserved for training the models and remaining 150 files for testing the emotion recognition accuracy of the system.

DETAILS OF DATASET						
EMOTIONS	**TOTAL NO. OF AUDIOS**	**NO. OF MALE BABIES**	**NO. OF FEMALE BABIES**	**NO. OF DATA FOR TRAINING**	**NO. OF DATA FOR TESTING**	**DURATION OF AUDIOS (s)**
HUNGER	188	8	9	142	53	6
DISCOMFORT	206	6	12	153	51	6
SLEEPY	206	7	8	155	46	6

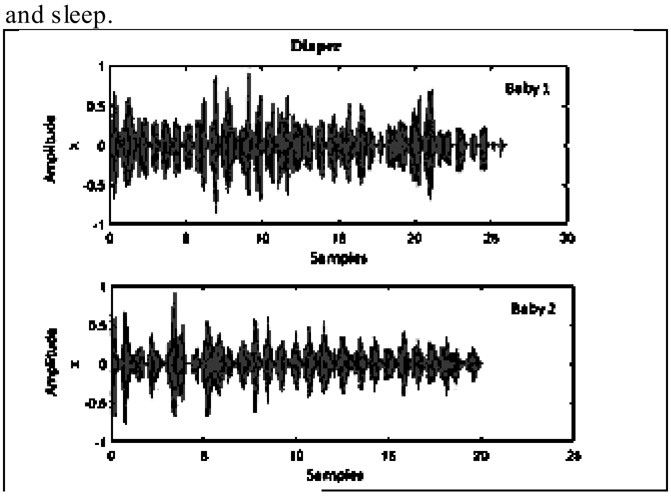
TABLE I. DETAILS OF INFANT CRY DATABASE COLLECTED

B. Infant Cry Analysis

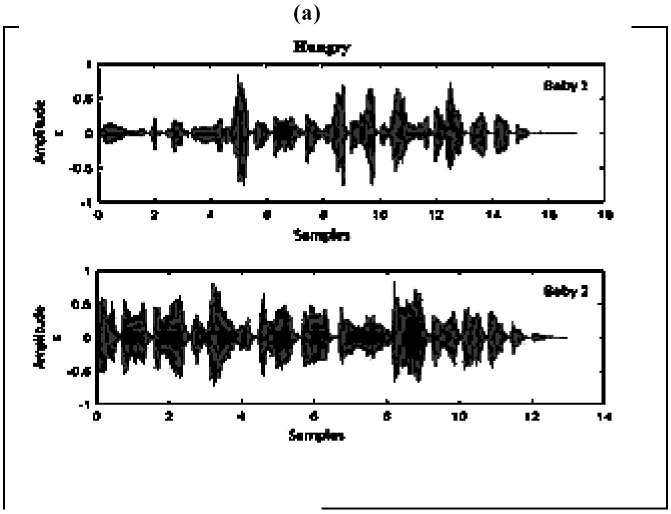
Collected data are analyzed using Wavesurfer and MATLAB. Wavesurfer is an open-source tool often used to visualize, analyze sound in several ways. It also can be used to delete unwanted and noisy regions of the audio collected. It can display waveform, spectrogram, fundamental frequency (F_0), power, or formant plots. This stage involves an analysis of various features like variation of F_0, short-time energy (STE), formants, etc. Then collected data was manually labeled into 'sleepy', 'hungry', and 'discomfort' categories.

The fundamental frequency is an important feature in the study of speech and audio signals. The fundamental frequency of infant cries can be determined using a method based on the autocorrelation of adjacent frames. An earlier study tried to differentiate cries based on the pain, sadness, hunger, fear, and other known causes [8] whereas a new, cheaper hearing screening method was worked out for the early detection of hearing disorders [9]. It has been observed that just by listening to the cry sound, the mother can identify the needs of an infant, though the precision in this decision is context and time-dependent [10]. Various reasons for an infant's cry include hunger, discomfort, colic pain, fever, or any other diseased conditions [11].

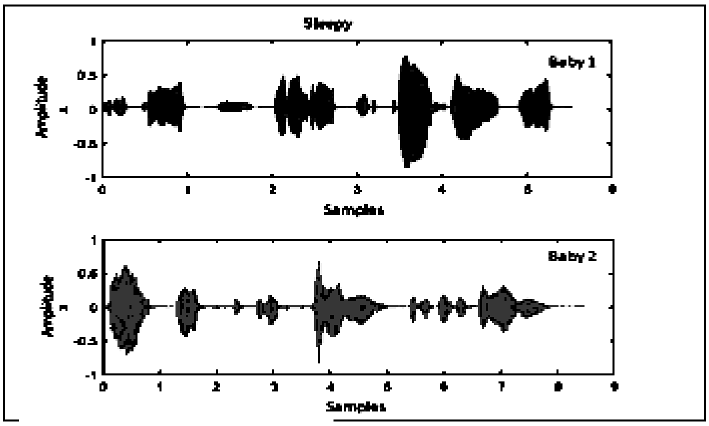
For spectral analysis of cry signal, we used linear prediction (LP) analysis and Fast Fourier transform (FFT) of Wavesurfer. In LP analysis, a sample of the signal is represented as a linear combination of past samples, which decomposes the signal into two highly independent components, the vocal tract parameters (LP coefficients) and glottal excitation. Spectral features represented by mel-frequency cepstral coefficients (MFCC) are obtained from FFT spectrum by placing bandpass filters on the mel scale. Framing, windowing, filtering, and cosine transformation are the important steps involved in the extraction of MFCC. Figure 2(a) – (c) show the cry waveforms of two babies corresponding to discomfort, hunger and sleep.



(a)



(b)



(c)

Figure 2. Waveform representing cry signals of two babies due to (a) diaper humidity/discomfort (b) hunger and (c) sleep

Diaper-related crying starts with a high amplitude of crying. It is a pattern of short duration silence and crying. The high amplitude cry signal followed by a short high-pitched inspirational whistle as shown in Figure.

Hunger related crying has a pattern of silence and crying. Some low amplitude crying patterns have more duration compared to high amplitude crying patterns. The duration of the silence is very small compared to the crying pattern.

Sleep-related crying is quite different from the previous cry signals. The duration of the cry signal is longer and it is starting from very low amplitude. The alternate cry patterns are completely different from each other compared to their amplitude. The cry signal starts with loudness and slowly decreases.

It has been shown that linear prediction Cepstral coefficients (LPCC) are useful for recognizing the causes of cry from instances [12]. The mean fundamental frequency value along with the minimum and maximum fundamental frequency value for each frame is also used to observe the cries of infants with different heart disorders [13].

The following observations are made through the analysis:

- F_0 related features vary with the cause of cry, namely, hunger, pain, normal, and pathological health condition.
- In hunger-related cry signal, repeated respiration can be noticed, which can be analyzed using short-time energy (voiced and unvoiced region of the signal can be easily detected)
- Each baby shows some variations in crying for the same reason. For example, even when if the baby 1 and baby 2 are crying due to hunger, the waveforms will be somewhat different. But certain similarities

can be observed within the same emotion class

- It is observed that spectral parameters provide cues for classifying cry emotions.

B. Classifier

The cry signal is classified using a deep neural network (DNN) based classifier [14]. Capsule neural networks are one such machine learning system that imitates the neural system and develops the structures based on the hierarchical relationships [15]. It has been shown that DNN provides better accuracy compared to other networks [16]. DNN is trained to classify the cry signal into three categories: 'hunger', 'discomfort', and 'sleepy'. Along with features extracted from the cry, data collected using a sensor module (such as body temperature, environment temperature and humidity of diaper) are also used for accurate emotion recognition.

C. Implementation and Evaluation

The system consists of two components. The first component is a laptop, which acts as a control module as well as processing module. The laptop collects voice signals using a microphone, identifies the emotional states of infants, and works collaboratively with other modules as well. The system is also acting as a processing module, which responds to the infant's cry. The second component is the sensor module, which collects parameters connected to causes of an infant's cry. Finally, the system should be evaluated in realistic environments for monitoring infants.

III. SYSTEM DESIGN

As shown in Figure 3, the system consists of a microphone, a sensor module, and a processing module. A microphone is employed to record the infant cry signals. The sensor module is meant for collecting an infant's body temperature, outdoor temperature and diaper humidity. After identifying the emotion using the trained classifier on the laptop, it will be displayed on the screen. The processing module (with the assistance of GSM) is meant to tell parents about the status of the infant. Temporary soothing measures like playing music and playing parent's voices are done using a laptop.

The sensor module is used to measure the body temperature of the baby, diaper humidity, and outdoor temperature. The DHT11 sensor is cost effective and it is used to measure the temperature of the baby and humidity of the diaper. The measures taken from the sensor is not accurate. DHT11 is reliable. LM35 sensor is used to measure the outdoor temperature and it is also cost effective. These sensors are set to some threshold values. If the parameters related to babies exceeded the threshold value, messages are sent to the parent's phone for alerting them. Using the processing module soothing actions can be done by playing music or parent's voices. Music and parent's voices are played using the audio system. GSM module sent messages to the parent's mobile about various parameter of the baby. Alerting messages informs the state of the baby. If the baby starts to cry, the images are captured using a webcam and send to the mail.

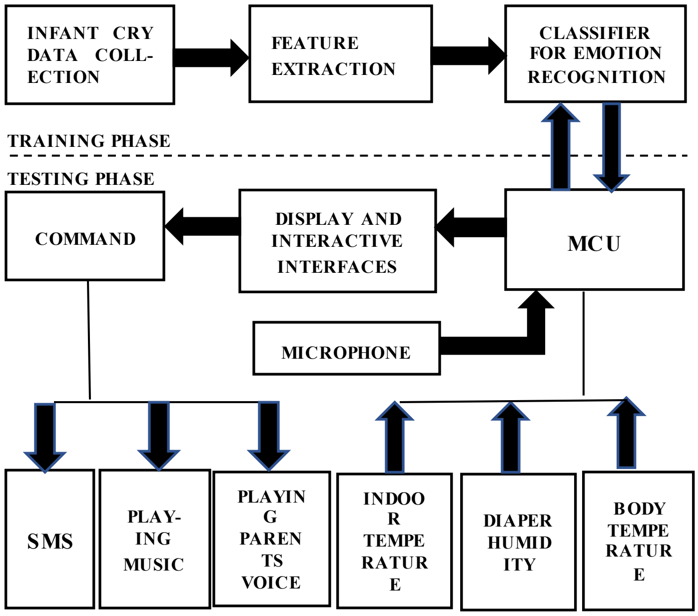


Figure 3. Block schematic diagram of the proposed system

A. Software Design

If the infant is crying, the system will recognize the cause of cry using cry signals and other parameters from the sensor. It contains two types of programs: one for Arduino and the other one DNN training/testing using python (Spyder). The software part of the system consists of the following programs:

- Program to load .wav audios
- Program to play audios
- Program for interfacing Arduino
- Program for capturing the image of the baby
- Program for emotion recognition

B. Hardware Design

The hardware part of the system consists of a microphone, a sensor module, and a processing module. A microphone is employed to record the infant cry signals. The sensor module is meant for collecting an infant's body temperature, outdoor temperature, and diaper humidity. After identifying the emotion, it will be displayed on the screen. The processing module (with the assistance of GSM) is meant to inform parents about the status of the infant. Temporary soothing measures like playing music and playing parent's voices are done. The DHT11 sensor is often used to identify the body temperature and humidity of the diaper. LM35 sensor is often used to detect the outdoor temperature.

IV. EXPERIMENTS

A. Training and Testing of DNN Classifier

The two main phases of the proposed system are training and testing. Data collection, feature extraction, and training of classifiers are the three main sub-phases of training. Data collection is required for the training and testing phase as a prerequisite. Feature vectors are extracted to detect three emotions of the baby.

Deep learning performs training to the computer framework to directly classify the tasks from the documents available either in the form of the text, image, or the sound [17]. For cry emotion recognition, a deep neural network (DNN) model is trained using MFCC features extracted from training samples. The DNN classifier is optimized for better accuracy by changing parameters such as window step size (winstep), epoch, hidden layer, learning rate and window length (winlen). Table II shows the optimization of parameters of DNN (with RELU activation function for the neurons) for improved accuracy of the classifier.

TABLE II. OPTIMIZATION OF PARAMETERS OF DNN WITH RELU ACTIVATION FUNCTION

USING RELU ACTIVATION FUNCTION					
WINSTEP (ms)	EPOCH	HIDDEN LAYER	LEARNING RATE	WINLEN (ms)	ACCUR-ACY
0.09	10	1	0.09	0.01	0.5759
0.03	10	1	0.01	0.01	0.6209
0.009	10	2	0.01	0.09	0.8070
0.009	15	2	0.001	0.08	0.8090
0.001	10	3	0.001	0.08	0.8278
0.001	10	3	0.001	0.09	0.8427

The following are some of the observations:

- High value of winstep will reduce accuracy
- Winstep too low will produce considerable increase in accuracy, but training time increases
- Using ReLU activation function for neurons and with (winstep=0.001, epochs=10, winlen=0.09 and learning rate=0.001), our system yield an accuracy of 84.27%

Table III shows the accurcy optimization of the system using sigmoid activation function for the neurons in DNN.

TABLE III. OPTIMIZATION OF PARAMETERS OF DNN WITH SIGMOID ACTIVATION FUNCTION

USING SIGMOID ACTIVATION FUNCTION					
WINSTEP (ms)	EPOCH	HIDDEN LAYER	LEARNING RATE	WINLEN (ms)	ACCUR-ACY
0.09	10	1	0.09	0.01	0.6853
0.03	10	1	0.01	0.01	0.7583
0.009	10	2	0.01	0.09	0.8236
0.009	15	2	0.001	0.08	0.8479
0.001	10	3	0.001	0.08	0.8278
0.001	10	3	0.001	0.09	0.8927

- Comparing to ReLU, DNN with sigmoid activation function provided better accuracy
- Using the sigmoid activation function and with (winstep=0.001, learning rate=0.001, epochs=10, winlen=0.09), our system yields an accuracy of 89.27%

When the baby starts to cry, the system automatically records the baby cry using a microphone. It identifies the reason for the cry and displays the reason for the cry in the console window. Along with this, the sensors (interfaced with Arduino) placed in the diaper of the baby measure the body temperature, diaper humidity, and outdoor temperature. Arduino is connected to the laptop and the output taken from the sensors is also displayed on the Console window of the Spyder. If the body temperature of the baby exceeds 37.2°C, then it may due to fever or any other disease. The humidity of the diaper must be in the range of 40%-60%. The baby feels discomfort if the outdoor temperature exceeds 100F. Figure 4 shows the result as displayed on the console window.

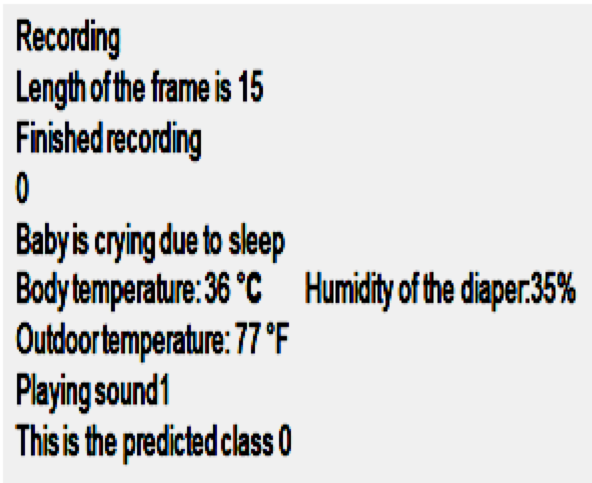


Figure 4. Result displayed on the console window.

Predicted class indicates the reason for the cry, where 0-sleep, 1 -hunger and 2- discomfort.

If the body temperature of the baby, humidity of the diaper and the outdoor temperature exceeds certain values, system combines this information with that of the DNN classifier and then alerts the parents by sending SMS via GSM module as illustrated in Figure 5.

Figure 5. Screenshot of message received

Table IV shows the results of evaluation conducted.

TABLE IV. EVALUATION RESULTS

EXPERIMENTAL RESULT					
BABY	BODY TEMPER -ATURE (°C)	HUMIDITY OF THE DIAPER (%)	OUTDOOR TEMPER -ATURE (F)	PREDI -CTED CLASS	REASONS OF THE CRY
BABY 1	37	33	99.3	1	HUNGER
BABY 1	36.8	38	95	2	SLEEPY
BABY 2	38.5	42	90	0	DISCOMFORT
BABY 2	36.7	36	91	2	SLEEPY
BABY 3	37	48	86	1	HUNGER

V. CONCLUSION

The proposed system is primarily focused on the design of an infant monitoring system based on cries. To alleviate the worries of busy younger parents, this system will be helpful for monitoring the infant cry emotion. When the signal is recognized as infant cry, MFCC features are extracted and it is classified using the DNN model. The predicted class is obtained from DNN, along with other parameters from sensors that helps the system to judge the reason for the cry and then to take appropriate measures (playing music or parents' voices). Finally, it sends a message to the guardian via GSM module.

REFERENCES

[1] L. L. LaGasse, A. R. Neal, and B. M. Lester, "Assessment of infant cry: acoustic cry analysis and parental perception," Mental retardation and developmental disabilities research reviews, vol. 11, no. 1, pp. 83-93, 2005.

[2] Y. Skogsdal, M. Eriksson, and J. Schollin, "Analgesia in newborns given oral glucose," Acta Paediatrica, vol. 86, no. 2, pp. 217-220, 1997.

[3] S. Möller, & R. Schönweiler, "Analysis of infant cries for the early detection of hearing impairment," SpeechCommunication, 28, pp. 175-193, 1999.

[4] M. Petroni, A.S. Malowany, C.C. Johnston, and B.J. Stevens, "A New, Robust Vocal Fundamental Frequency (F0) Determination Method for the Analysis of Infant Cries," Computer-Based Medical Systems, 1994, Proceedings of IEEE Seventh Symposium on Computer-based Medical Systems, pp. 223-228, June 1994.

[5] Q. B. Xie, R. K. Ward and C. A. Laszlo, "Automatic assesment of infants' levels-of-distress from the cry signals" IEEE Trans. On Speech and Audio Proc., vol 4Jo.4, pp.

[6] Y. Skogsdal, M. Eriksson and J. Schollin, "Analgesia in Newborns given oral glucose" Acta Paediatrics, vol. 86, pp.

[7] J. Lawrence, D. Alcock, P. Mcgrath,, J. Kay, B. MacMurray and C. Dulberg, "The development of a tool to assess neonatal pain," Neonatal Network, vol. 6, pp. 59-65, 1993.

[8] Yousra Abdulaziz and Sharrifah Mumtazah Syed Ahmad, "Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients," In International Conference on Information Retrieval & Knowledge Management (CAMP), pp. 260–263, IEEE, 2010.

[9] Gyorgy V arallyay Jr, "The melody of crying," International Journal of Pediatric Otorhinolaryngology, pp. 1699–1708, 2007.

[10] Hemant A Patil, Anshu Chittora, "Data collection and corpus design for analysis of neonatal and pathological infant cry," IEEE International Conference on Asian spoken language research and evaluation, November 2013.

[11] Linda L. LaGasse, A.Rebecca Neal, and Barry M. Lester, ''Assessment of Infant Cry: Acoustic Cry Analysis and Parental Perception," Mental Retardation and Developmental Disabilities Research Reviews 11, PP. 83-93, 2005.

[12] Lin Zhong, Yanxiong Li, Xin Wei, Guanglong Li, Zili Wang, Yewen Jiang, "System design for monitoring infant speech emotion," IEEE International Conference on Control & Automation, Hangzhou, China, June 12-14, 2013.

[13] S. Chandralingam, T. Anjaneyulu, and K. Satyanarayana, "Estimation of fundamental and formant frequencies of infants cries; a study of infants with congenital heart disorder," Indian Journal of Computer Science and Engineering, pp. 574–582, 2012.

[14] G. Dahl, T. Sainath, G. Hinton. "Improving deep neural networks for lvcsr using rectified linear units and dropout," ICASSP, British columbia: IEEE, 2013, pp. 8609-8613.

[15] Bashar, A. (2019),"Survey on evolving deep learning neural network architectures" ,Journal of Artificial Intelligence, 1(02), 73-82.

[16] Y. LeCun, Y. Bengio, and G. Hinton, ''Deep learning,'' Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[17] Vijayakumar, T. (2019),"Comparative study of capsule neural network in various applications",Journal of Artificial Intelligence, 1(01), 19-27.