iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.41433

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Automatic Detection of Baby Cry using Machine Learning with Self Learning Music Player System for Soothing

Varsha Dange[1], Tejaswini Bhosale[2]
[1, 2]*Dept .of Multidisciplinary, VIT, Pune*

Abstract: ***Automatic voice detection of baby cry plays an outstanding role in different applications for smart monitoring of smart baby condition. In this proposed model, a baby's cry is being detected and a music player will be played after detection in order to create a soothing environment for the baby. This system employs a machine learning approach to recognize newborn cry sounds in a variety of residential settings under difficult situations. The automatic detection of a baby cry can be used in a variety of situations involving various types of sounds in the environment. The proposed system also provided with alert-based notifications to parents. Also, in commercial used products such as remote monitoring of baby, baby facial recognition as well as in medical applications. Also, evaluate K-Nearest Neighbours (KNN) algorithm for baby cry detection and perform priority queue functions for playing music. In this system, the input consists of the cry sound with k nearest training samples in the database. The output is depended on analysis of the cry-detection performance with KNN which is used for classification.***
Keywords: Baby cry detection, Machine learning, Audio detection, K-Nearest Neighbours, smart monitoring

I. INTRODUCTION

Today's lifestyle of a human being is so busy that is truly affecting the basic livelihood of the human being. In today's hectic world, parents are so preoccupied with their daily professional lives that they don't have enough time to stay at home and care for their children. This makes it difficult for parents to attend to and care for their newborn kid while working. Therefore by considering this, have tried to design a simple system that can help the parents in taking care of the baby.

This system proposed a simple voice detection system that has been implemented practically a device that has the capability to detect an infant's cry and turning on automatically a soothing music for baby sleep with music player. Whenever the baby cries it gets detected by the system with the help of Mic or a microphone. And then in order to do that it turns on the music which creates a sleep mechanism by producing a soothing sound that makes the baby sleep gently. The main program in this system is the implementation of detection of an infant's crying and for playing a piece of particular music on which baby is going to sleep. It recognizes a baby's scream while disregarding other noises like as sneezing, laughing, clapping, unexpected noises, environmental noises, and so on. Nowadays there are some types of baby monitoring systems that consist of wearable devices, some android applications, some wirelessly controlled camera systems, etc. Most of such systems are covered only at home using Wi-Fi or Bluetooth. Due to this condition, employed parents cannot ensure the safety of their babies because they are unable to connect with the child when they're at the workplace. There are few products with remote monitoring facilities. But they are not affordable because they are priced as high-end products. Also, these products are not easy to set up and if it is fixed near to the baby that may cause health hazards due to electromagnetic radiation. But our product is designed for an affordable cost and it can be used from birth to 15 months of babies with the ability to detect the crying immediately and send a notification to warn parents/caregivers while creating a soft sound to soothe the baby covered only home without using Wi-Fi or Bluetooth. Data acquisition, data cleaning, signal processing, feature extraction, selection, and classification are all phases in cry sound detection. It has been difficult to obtain the necessary data because sensitivity of baby cry data. Therefore, in this system own dataset has been created by recording different kinds of audio clips. Voice Signal processing is needed in order to remove background sounds and segment cry data in order to generate databases of baby cry, voice signal processing is required. After the dataset has been obtained, feature extraction will be carried out, which will involve extracting features from various domains of the cry signals Different characteristics of the cry signal are represented by features derived from various domains such as the temporal domain, cepstral domain, and so on. Also, to build effective classification models system need to select the most matched features and reduces the feature dimensions that is a challenging task. Features are vital for categorization or detection accuracy when using appropriate machine learning models for a given baby cry database.

©IJRASET: All Rights are Reserved

For the purpose of performance evaluation an appropriate dataset containing several audio clips of babies under different domestic environments. The proposed system aims to demonstrate a significant performance improvement with the KNN model when it comes to classifying new data and making predictions when compared to standard machine-learning models, notably with the low false-positive rate

II. RELATED WORK

A. Theoretical Background

Voice recognition is the process of using an algorithm to turn our speech stream into a sequence of words. [1]. Because recognition performance is strongly dependent on this step, feature extraction is a key part of recognition of speech. Feature extraction is used to create feature vectors that reflect the input signal in a simple manner. The extraction of features is separated into steps. In the first stage, the signal is subjected to a spectro temporal analysis, which yields raw characteristics defining the power of the spectrum of short speech intervals. In the second step, the enhanced static and dynamic features are obtained. Finally, by converting these expanded feature vectors into more compact and resilient vectors, these features are supplied to the recognizer. [2].

There are number of techniques used for feature extraction. Cepstral Analysis, Mel Cestrum Analysis, Mel-Frequency Cestrum Coefficients (MFCC), Linear Discriminant Analysis (LDA), Fusion MFCC, Linear Predictive Analysis (PLP) Analysis and Perceptually Based Linear Predictive Coding are some of the techniques being comes in use. After feature extraction, the very important step is speech recognition. Basically, there are mainly three approaches to speech recognition.

B. Literature Survey and Similar System

In 2019, Karinki Manikanta, K.P. Soman, M. Sabarimalai Manikandan proposes a “Deep Learning-Based Effective Baby Crying Recognition Method under Indoor Background Sound Environments” They ran into issues with data availability. There aren't many cry sound databases available for free in the literature. As a result, they are attempting to assess ML-based newborn crying sound detection algorithms. [4].

In 2016 Yizhar Lavner, Rami Cohen, Dima Ruinskiy, and Hans Ijzerman proposes a “Baby Cry Detection in Domestic Environment using Deep Learning”. The challenges faced are Preprocessing and feature extraction from the dataset to build the model. As a result, the audio recordings are split into overlapping segments, which are then further divided into frames. [5].

In 2017 Pruthvi Raj Myakala1, Rajasree Nalumachu2, Shivam Sharma3, and V. K. Mittal4 proposes an “An Intelligent System for Infant Cry Detection and Information in Real-Time” in which information transmitting functionality (through MMS, text messages) is hampered by the availability of strong mobile networks in the system's immediate proximity. The latency of various operations is measured. [6].

In 2017 G.V.I.S. Silva, D.S. Wickremasinghe “Baby Cry Recognition in Real-World Conditions” With the help of the proposed cry detection algorithm, it easily identifies the infant's cry and verified it by using KNN with accurate results. To reduce the size of features involved, they used MFCC to obtain the Feature vector. Framing and windowing are followed by FFT and DCT to obtain the features [7].

III. CRY SOUND RECOGNITION METHOD

Infant cry sound detection algorithms which are based on MFCC features and machine learning classifiers like KNN are provided in this work. The cry sound recognition methods are tested against a variety of background sounds found in typical interior household scenarios. The above picture depicts the block diagram of the Machine Learning-based cry sound detection approach, which consists of four steps: data preprocessing, extraction of feature, models for cry, and identification. The following subsections go through each step of the cry sound detection process:

A. Pre-processing

Low-frequency noise components for example micro-phone artifacts, recording instrument biasing, and power-line interference, which get generated by the movements of the sensors and electromagnetic (EM) while the audio sensor is exposed to the environments condition, are commonly corrupted cry sound signals that are recorded. As the result of this, the audio signal is transmitted via a High Pass Filter (HPF) with a threshold frequency of 60 Hz. Following that, the signal is split into frames with three frame lengths: 100 milliseconds, 250 milliseconds, and 500 milliseconds, which are used to evaluate performance. Because of the stochastic nature of cry sound production and the time-varying area of the sound observing zone from acoustic sensors, the intensity of the cry sound varies. In practice, the origin of the sound can be difficult to pinpoint. Regardless of the fact that the function is insensitive to the amplitude rate, amplitude normalization is performed on the zero-mean audio-formed signal to limit changes in the microphone's sensitivity.

B. MFCC Feature Extraction

The MFCC characteristics were used in this investigation to recognize cry sounds signals [3]. MFCC features were retrieved for the cry, music, fan and AC, and speech signals for each of the signals.

1) *Fourier Spectrum:* FFT is used to create window-structured audio frames z[n], which result in energy distributions over frequencies and, as a result, the spectrum's amplitude.
2) *Mel-Frequency Spectrum:* The spectrum's magnitude is formed using 26 band-pass filters with positions on the Mel- scale that are spaced at regular intervals. This concern to linear frequency f expressed as:

(f)=11.25 1n (1+ f) (1)

700

Mel-frequency refers to the log-off (linear frequency) that shows similar auditory effects in people's perceptions.

C. Algorithm

To classify the labeled data, supervised machine learning is used and KNN is one of the approaches to classify the data. The main goal of KNN is to classification of the data based on similarities between varieties of classes.

1) *Steps Involved in KNN*

a) Stores all the available samples and classifies the new data on basis of similarity.
b) Features extracted are used in KNN to make the training data.
c) When new user input (speech) is given, the KNN algorithm identifies the speaker based on the training data.

2) *Music Player:* When the infant cry is detected the system sends an alert message and start playing Music. After playing the music for few seconds, it again checks for baby cry and plays the song with the highest priority.

3) *Steps Involved In Playing A Song With The Highest Priority*

a) To play the song with the highest priority from the list of available songs system used "Priority Queue".
b) When the baby starts crying the song with the highest priority has been popped from the queue and if songs have similar priority, then the song is played on the basis of first come first serve (FCFS).
c) If the baby stops crying after playing a specific song, then the priority of that song gets increased by 1 and the song is pushed back in the queue.

IV. METHODOLOGY OF INFANT CAR DETECTION

The objective of the detection method is to sort incoming audio signals into "cry" and "not cry" categories. The algorithm explores about the audio signal over a wide range of time periods. The figure shows the audio processing algorithm to detect cry signals.Fig.1 shows the block diagram for infant cry detection algorithm.

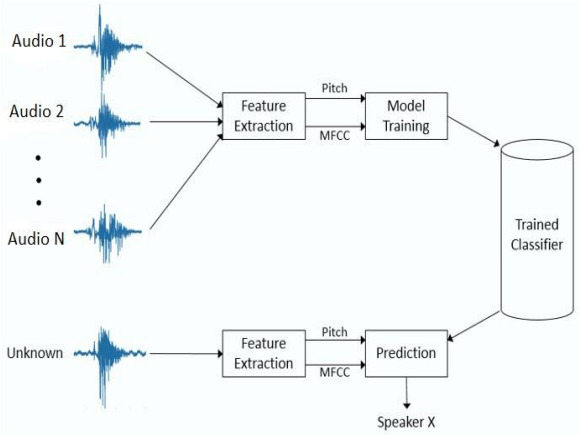


Fig .1 Infant cry detection algorithm –block scheme

1) A KNN is applied and the MFCC features are calculated for each signal.
2) In the training phase of the model, all the audio signals that represent infant cry are labeled as 1 and others as 0.
3) Using a KNN classifier model, each signal is categorized as cry or not cry based on its extracted attributes..
4) After taking the input from the mic, the background noise is removed from the audio signal and the MFCC features are calculated.
5) After taking the input from the mic, the background noise is removed from the audio signal and the MFCC features are calculated.
6) Determined by the attributes The KNN classifier classifies the signal as "Cry" or "Not Cry."

V. PROPOSED SYSTEM

A simple audio detection method for creating a device that detects an infant's scream and plays music automatically. A system that continuously monitors children, detects newborn cry, and alerts parents by text message or alert message that their child is crying. It uses speech signal techniques on hardware components to build a real-time system. To handle real-time audio inputs, extract features, and detect the infant cry signal, the proposed system employs signal processing algorithms.

VI. IMPLEMENTATION

All the audio processing algorithms were first implemented and tested in the python 2020 environment. In order to design and test the system performance, various crying samples are required.

A. Data Collection

A large number of baby cries has been recorded from day- care centers, neighbors, and online databases, and then fed to the computer. The signal is stored on the computer as a lossless WAV file. Cry signals of babies ranges in age between 0-18 months. Collected data mainly has 2 voice domains; the voice of the baby, the voice of the Background Noise. It is assumed all the babies were in healthy, Noise-free environments (not engaged with engines, passers-by, car horns, high hammering sounds, etc.). For negative samples of data uses baby laugh, sneezing, music and, giggles, happy vocals ranging between 0- 18 months baby. There are 200 training data, each of which represents the all-sound forms, including cry and non-cry. 180 testing data are there respectively, including cry and non-cry.

B. Testing the algorithm

1) *Sample Data:* Sample data is the data that use instantly to check the results, according to the accuracy of the system simulations, change the sample size which system has taken as input.
2) ***Training Set:*** It is the set installed in the programmed memory and is used for distance calculation with sample data, 50 data set as the training set. With 40 babies' cry samples and 10 non- cry samples and after that, gradually increased the training set 50 to 80 and 80 to 150 samples and 150 to 200.
3) *Group Matrix:* Domains are defined in this matrix (i.e., cry and non-cry) of each and every data in the training set.
4) *Output:* Features are extracted for each and every sample data in training set. All data in the training set are defined in the Group matrix as it belongs to cry or not. As an example, consider one sample data, this sample is calculated with each and every data in the training set for obtaining the distance values and find out the shortest distance given by which training data, then according to the definition of group matrix algorithm identify the nearest neighbors of the sample data.

C. Soothing system

In this Paper, the Soothing feature is achieved by Music Player which does the task of playing a light/entertaining sound which in result may divert the baby from crying. The Song list contains more than 50 songs. The Song playing criteria is totally dependent on a requirement-based modified Priority Queue Scheduling Algorithm. The care has been strictly taken that no Song will face problem of starvation. If the songs with similar priority are present in the queue, then the song is played on the basis of first come first serve (FCFS). At each iteration after infant cry detection as positive (in favor of baby is crying), a random song is played for specific 'N' Seconds. Then, if played song stops the baby's cry in result Priority Increment is done. In the next cry detection iteration, the highest priority song in the queue sequence is played. And if still, the priority song fails to stop the baby from crying, The priority song is pushed back in the queue and a next song is played. This whole song playing process gets triggered each time the detection is performed.Fig.2 shows flow for music player and notification alert.

©IJRASET: All Rights are Reserved

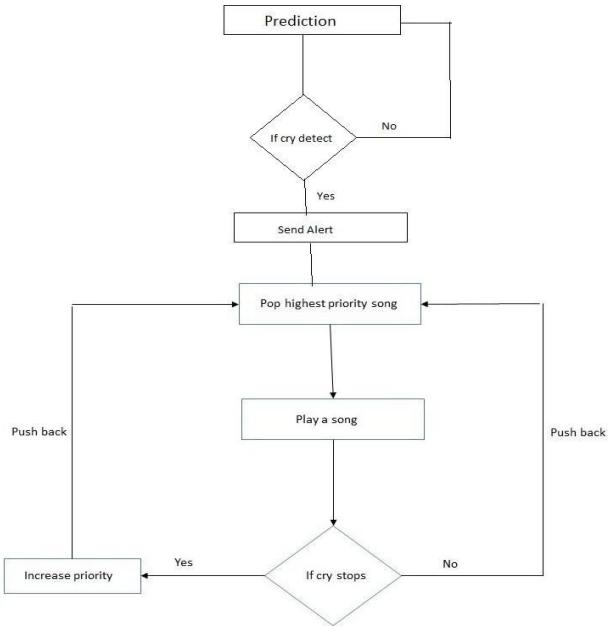


Fig 2. Music Player and Notification alert Flow

The modified and tricky use of Priority Queue has been resulting in efficient performance and accuracy impacting the overall implementation.

VII. PERFORMANCE EVALUATION

A. Dataset

Audio recordings of babies (recorded at 44, 100 Hz) make up the dataset for this study. The sound of the baby was captured in a genuine environment or from clips in the dataset. With 20 different events like Crying, laughing, parent's conversation, wind/fan sounds, door opening/closing, and other events were marked on the recordings.

B. Training and test process

The test corpus contained the remaining 80% of the labeled data, while the training corpus contained only 20%. With an initial learning step of 0.00001, the system has trained KNN structures, which is an adaptive learning rate optimization technique.

C. Result analysis

Our System Records audio for 6 secs at a sample rate of 4100Hz then based on the extracted features system detects either the baby is crying or not and displays the output, if yes then text message is sent to the registered mobile number and plays a song with higher priority, if baby stops crying at any particular song, then the system increases the priority of that song and plays it first when next time the baby is crying.Fig.3 shows output for proposed system and Fig.4 shows SMS received by parent.



Fig.3 Output for proposed system

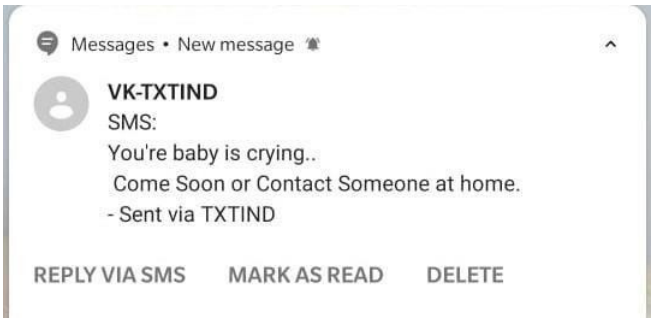


Fig.4 Received SMS

VIII. CONCLUSION

With the help of the proposed cry detection algorithm, it easily identifies the cry and verified it by using KNN with accurate results. By using the KNN model and MFCC feature extraction with the combination of Pitch gives a more promising approach to cry detection. This improved detection as well as recognition accuracy. Cry detection becomes challenging because of the highly variable nature of input voice signals. Voice signals in training and testing parts could be different due to many facts such as:

1) Baby's voice changes with time.

2) Variations in recording environments

Therefore, increasing more training samples of different noises and speeches would give more accurate results. The proposed system has employed machine learning and deep learning methodologies to recognize infant cry sounds in a variety of household contexts, with the goal of designing, evaluating, and analyzing KNN architectures for baby cry identification using traditional machine-learning approaches. In the future, a detailed examination of the faults found in this system, as well as the evaluation of various classification methods and auditory aspects will be carried out. Additionally, these technologies can be implemented on real-time hardware for remote video monitoring applications.

IX. ACKNOWLEDGEMENT

The authors of proposed system would like to thank our Head of the Multidisciplinary Department, Prof Mukund Kulkarni, for guiding us throughout the project. Also, like to express our special thanks of gratitude to our guide teacher, supervisor, Prof Varsha Dange, for encouraging and helping in research work.

REFERENCES

[1] Therese S. S. and Lingam, C. (2013). “Review of Feature Extraction Techniques in Automatic Speech Recognition” International Journal of Scientific Engineering and Technology [Online] Volume No.2, Issue No.6, June pp: 479-484 Available from com/ijset/ publication/v2s6 /paper 606.pdf [Accessed:24/02/2015].

[2] R.Torres, D. Battaglino, and L. Lepauloux, ”Baby cry sound detection: A comparison of hand crafted features and deep learning approach,” in Proc. Int. Conf. on Engineering Applications of Neural Networks, pp. 168-179. Springer, Aug. 2017.

[3] J. O. Garcia, and C. R. Garcia,”Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neuralnetworks,” in Proc. Int. Joint Conf. IEEE on Neural Networks, Vol. 4, pp. 3140-3145, July .2003.

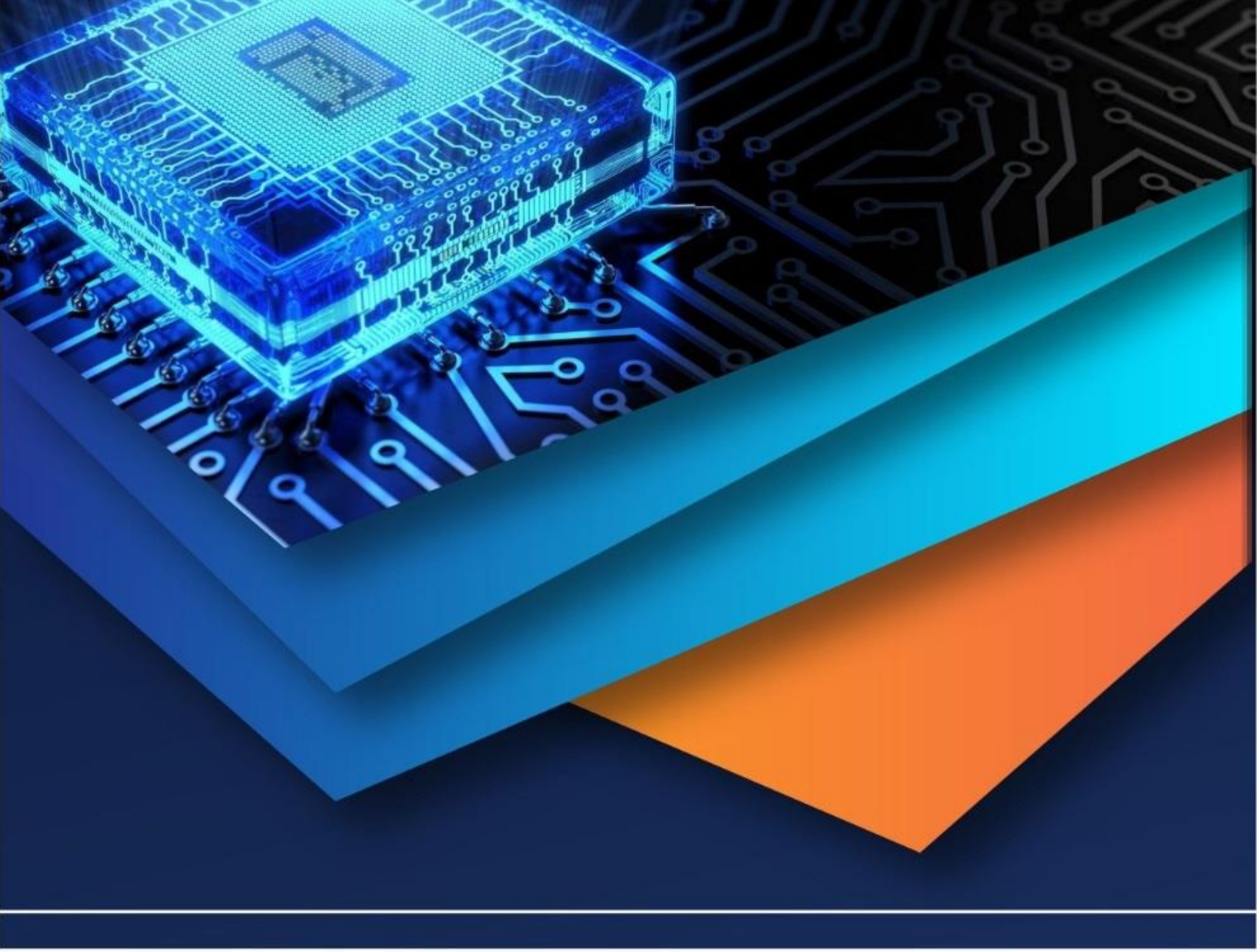
[4] K. Manikanta, K. P. Soman and M. S. Manikandan, "Deep Learning Based Effective Baby Crying Recognition Method under Indoor Background Sound Environments," 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), 2019, pp. 1-6, doi: 10.1109/CSITSS47250.2019.9031058.

[5] Y. Lavner, R. Cohen, D. Ruinskiy and H. Ijzerman, "Baby cry detection in domestic environment using deep learning," 2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE), 2016, pp. 1-5, doi: 10.1109/ICSEE.2016.7806117.

[6] P. R. Myakala, R. Nalumachu, S. Sharma and V. K. Mittal, "An intelligent system for infant cry detection and information in real time," 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), 2017, pp. 141-146, doi: 10.1109/ACIIW.2017.8272604.

[7] Bğnicğ, H. Cucu, A. Buzo, D. Burileanu and C. Burileanu, "Baby cry recognition in real-world conditions," 2016 39th International Conference on Telecommunications and Signal Processing (TSP), 2016, pp. 315-318, doi: 10.1109/TSP.2016.7760887.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)