

Predicting the Reason for the Baby Cry Using Machine Learning

Chaithra lakshmi C¹, Aravinda B², Deeksha, Deeksha³, Sadhana⁴

Sahyadri College of Engineering & Management, Mangaluru, India

Email id: chaithra.cheral@gmail.com

DOI: <http://doi.org/10.5281/zenodo.2656353>

Abstract

Crying is the primary sound child makes when he/she enters the world. It is a response to address different circumstances including pain, hungry, lonely, discomfort etc. Parents or guardians endeavour to recognize and timely address the reason for the infant cry before crying sets in. Anyway first time they often fail. This prompts to dissatisfaction and feeling of helplessness. This paper proposes the system to predict the reason for the baby cries using machine learning. To predict the causes, this system takes baby crying audio as input, processes it, gives causes and suggestions about the actions to be taken, as output.

Keywords: *Machine learning, infant cry, MFCC, KNN, python_speech_features.*

INTRODUCTION

Communication is very important in life. There are many means of communication for adults. But crying is the only way of communication for infants. Infant crying gives various information's about health of the baby and also what its needs. It is a response to address different circumstances including pain, hungry, lonely, discomfort etc. Parents or guardians endeavour to recognize and timely address the reason for the infant cry

before crying sets in. Anyway first time they often fail. This prompts to dissatisfaction and feeling of helplessness.

Parents have to be quick enough to find the reason and to take suitable action. The interpretation of those cries differs with involvement, to such an extent that interpreting a child's needs can be a speculating amusement for inexperienced parents who, in time, appear to build up a comprehension of what the conveyed

needs are. If there should arise an occurrence of pediatricians, they can recognize different sorts of cries relying upon the example in each kind of cry. Be that as it may, they must be master enough to give precise answer. In case of babysitters interpreting the reasons for the baby cry among many babies will be a tedious task for the care takers and also time consuming.

This paper proposes a machine learning process including steady dataset of age from newborn child cries and choosing proper audio features, with promising trial results for upgrading the monitoring of babies. To predict the causes, this system takes baby crying audio as input, processes it, gives causes and suggestions about the actions to be taken, as output.

This paper uses machine learning technique to classify the causes. First objective of the paper is to train the system. Main objective of the project is to predict the reasons for the baby cry which is given as input to the system. It is done by training the system to recognize the baby cry. When an audio is given as input, system checks whether an audio is of baby cry or not. System is also trained to extract required features from the audio file and to classify them into different categories.

First step of the project is to collect the dataset consisting of baby cry recording and the reason for the baby cry. Second step is to convert audio files into the wav format which can be processed by most of the audio libraries.

Next step is to identify the features to be extracted from the audio files. Once the features are identified, next step is to extract the identified features from the audio files of dataset with labeling. Next step is to train the algorithm with extracted features. Once the training of the algorithm is done, next step is to test the algorithm. Final step is designing interface for the application and testing the application on real-time. System classifies the audio into 5 reasons: hungry, belly pain, need for burping, discomfort and tired. This project is helpful for the Parents, Care takers, guardians and the people who want to know the reason for the baby cry. Constraint of the project is, it can predict reasons for the baby cry for the baby within 2 years.

Objectives of the proposed system

One of the objectives of the system is to train the system to recognize the baby cry. When an audio is given as input, system checks whether it is of baby cry or not. Another objective is to train the system to

extract required features from the audio file and to classify them into different categories. Main objective of the system is to predict the reasons for the baby cry which is given as input to the system.

LITERATURE REVIEW

Yuki Mima and Kaoru Arakawa proposed method for determining why the baby is crying by analyzing the frequency range of the infant's voice. They estimated causes such as hunger, discomfort and sleepiness. These causes were later used for classification of the baby voice. The proposed method was applicable for the babies aged 2-4 months old. Initially they extracted some of the characteristics from the baby cry and then using the characteristics rule based method was used for further classification; they proposed a method to determine how infants are feeling [1].

Rami Cohen, Yizhar Lavner proposed an algorithm for detecting baby cry automatically. Algorithm was used to identify physical harm to babies such as scenario in which infants parent leave their child in vehicle. The algorithm was based on extracting features from the baby cry. Features like pitch parameters, short term energy and frequency cepstral co-efficient were extracted from the baby cry signal.

And for classification they have used k-nearest neighbor algorithm. They evaluated performance of the algorithm in order to check robustness of knn algorithm in the presence of different type of noise such as car horn. Their research mainly focused on detecting the physical harm or danger to the babies when their parent leaves their child in the vehicle [2].

In this paper Ypusra Abdulaziz, and Shariffah Mumtazah syed ahmed proposed an architecture for cry recognition system. Proposed architecture was used to differentiate and to identify between two different cries of the baby such as pain and non- pain cry.

Their system of recognizing the baby cry was based on feed forward neural network, which was trained with gradient algorithm. In their research they extracted features such as and linear prediction cepstral coefficient (LPCC) and Mel frequency cepstral coefficient (MFCC) from the baby cry signals. Accuracy of their research varied from 57% to 75%. Important observation of their research was accuracy of the system gives better results for MFCC features than LPCC features [3].

In this research J. Saraswathy, M. Hariharan, Sazali Yaacob and Wan

Khairunizam analysed various works on baby cry signals. It is difficult to predict cry signals to identify for what the baby cries for. In their research they have used noninvasive technique for the analysis of cry signals which has shown promising result. They have used signal boundary detection and LPCC to extract various features from the cry signals. Their study mainly focused on significant work on the baby cry as well as extraction of acoustic feature from the infant's cry by using extraction techniques like MFCC and LPCC [4].

Kevin Kuo analyzed extraction methods like Signal boundary detection and linear predictive coding coefficients (LPCC) to detect the features from the baby cry. Reasons for the baby cry were recognized. Reasons such as need for attention, hungry and wet diaper were recognized. Qualitative matching technique was proposed to know the similarity between the coefficients of training vectors and LPCC features. They concluded that LPCC feature extraction technique is the best method of feature extraction technique as compared to other methods. To classify the extracted features from the baby cry they have used LVQ i.e. learning vector quantization [5].

Jose Orozco Garcia and Carlos A. Reyes Garcia proposed recognition techniques for the infant's cry in order to classify two different cry types i. e pathological cry and a normal cry of the baby. In their research features from the cry signals were extracted by using Mel-Frequency Cepstrum method and acoustic characteristics which was obtained after feature extraction technique was then fed into the classifier. Classifier used in their research is feed forward neural network [6].

The main objective of the paper by Jose Orozco, Carlos A. Reyes-Garcia was to classify normal and pathological cry of deaf babies. For the experiment the automatic recognition system was developed. Feature extraction was done by the Linear Prediction technique where acoustic characteristic was obtained and feed forward neural network was used as classifier which was trained with different learning methods among that Scaled Conjugate Gradient algorithm showed better result [7].

Pritam Pal, Ananth N. Iyer and Robert E. Yantorno analysed the reason for the baby cry by observing the facial image and sound of cry. Here two types of module were used: image processing and sound

processing module. The image processing module was used to know the state of face expression and its combination to analyze the reason for cry. The sound processing module was used to analyze the data from fundamental frequency and Kmeans method used for clustering. The decision level fusion system was used for fusion of image processing and sound processing module [8].

Ioana-Alina Banica, Horia Cucu, Andi Buzo, Dragoş Burileanu, and Corneliu Burileanu developed a fully automatic system to differentiate between different type of cries. Feature extraction was done by MFCC technique where acoustic characteristic was obtained and Gaussian Mixture Models and i-vectors were used as classifier. The experiment was performed on the database comprised of 6 types of cries of 127 babies [9].

To differentiate between different type of cries, Ioana-Alina Bnic, Horia Cucu, Andi Buzo, Drago Burileanu and Corneliu Burileanu developed a fully automatic system. Feature extraction was done by MFCC technique where acoustic characteristic was obtained and Gaussian Mixture Models and i-vectors were used as classifier. The experiment was performed

on the database comprises of 5 types of cries of 39 babies [10].

Parents will be worried when baby started crying. To know the actual reason for crying, an automatic recognition system was developed by Chuan-Yu Chang, Jia-Jing Li. The experiment was performed to know three types reasons like hungry, pain and sleepy from baby cry. The deep learning algorithm like convolutional neural networks (CNN) was used for extraction and classification [11].

J. Saraswathy, M. Hariharan, Wan Khairunizam, N. Thiyagar developed a system in which 3 types of cry was analysed such as normal cry, cry of deaf babies and baby with asphyxia. Feature extraction was done by Short-time Fourier transform technique were time-frequency was analysed. Two types of radial basis neural network: Probabilistic Neural Network (PNN) and General Regression Neural Network (GRNN) were used for classification [12].

The experiment by Shubham Asthana, Naman Varma and Vinay Kumar Mittal was performed by collecting dataset consisting of 6 types of cry to analyse and to determine the reason for baby cry. Fundamental frequency energy was the

feature which was extracted and analysis was done by using signal processing methods like linear prediction and auto correlation technique. The causes for baby cry was known using spectrograms. The fundamental frequency was high for baby cry when compared to adults [13].

Mrs. Sameena Bano and Dr. K.M. RaviKumar described a novel approach to identify a baby physiological state and its needs. In this work normal infant cry signal of ages 1day to six months old was used. The infant cry signal was segmented by using Pitch frequency and the features like Short-time energy, Harmonicity Factor (HF) and Harmonic-to- Average Power (HAPR) were extracted.

MFC (mel frequency cepstrum) coefficients was computed over MATLAB. KNN classifier using Pitch, Short-time energy, Harmonicity Factor (HF) and Harmonic-to- Average Power (HAPR) were used to classify the normal infant cry signal. Percentages of results obtained Decoding baby talk supports the mother's built-in intuition about knowing and responding to their baby's needs, and physician to treat infant early [14].

The clues about the baby cry causes will be carried in the signals of the baby cry.

By using these clues a mother or a guardian can be able to identify the reasons for the baby cry. In this paper Shivam Sharma, Pruthvi Raj Myakala, Rajasree Nalumachu, Suryakanth V. Gangashetty and V. K. Mittal. proposed causes such as pain and discomfort from the cry signals. Features like frequency analysis and acoustic features, signal energy were derived from the signal. STFT and LPA techniques were used for this purpose [15].

Research by Pruthvi Raj Myakala, Rajasree Nalumachu, Shivam Sharma and V. K. Mittal was based on acoustic analysis of cry signals from the baby in order to detect physical state of the infant. Their system was useful for the parents to verify whether baby is crying by informing the care taker who can alert the parent by sending details through app. Their system involved a smart robot which can fetch live video streaming. And it is also used to send text message to the parent about the state of the baby. Their system contains different sensors, SPU and information transmission unit [16].

Infant's cry is used to represent the physical state such as emotions of the baby. Baby's cry may consist of several reasons such as pain, discomfort or

disturbance. In order to predict these reasons Aomar Osmani, Massinissa Hamidi and Abdelghani Chibani proposed an efficient technique which will be used by the parents or the guardians. It used machine learning technique to do the processing. They extracted low level and high level features from the baby cry signals. Those signals were classified using ensemble learning methods [17].

In this context, Rodica Ileana Tuduce, Horia Cucu, and Corneliu Burileanu focused on creating an automatic system which was able to distinguish between different infant needs based on crying. They extracted various sets of paralinguistic features from the baby-cry audio signals and trained various rule-based or statistical classifiers [18].

METHODOLOGY

Dataset

Dataset is collected from the Speech Technology Course in KTH (Royal Institute of Technology Sweden). Dataset consists of around 100 audio files of baby cry. Each audio file is labeled with reason for the cry, age and gender of the baby.

Architecture of proposed system

Fig 1 shows the software architecture of the system. First step of the proposed system is to collect the dataset consisting of baby cry recording and the reason for the baby cry. Once the dataset is collected, next step is to convert audio files into the wav format which can be processed by most of the audio libraries. Converted audio files undergo noise cancellation process. Next step is to extract the required features from the audio files.

Extracted features are stored in csv files. Extracted features are given as input to the training system. Recorded audio file of baby cry also undergo noise cancellation and feature extraction. Extracted feature of audio file is compared with the extracted feature of dataset. Comparison is done by the classifier. Classifier classifies the audio file into different reasons. Output of the system is reason for the given baby cry and suggestions on action to be taken.

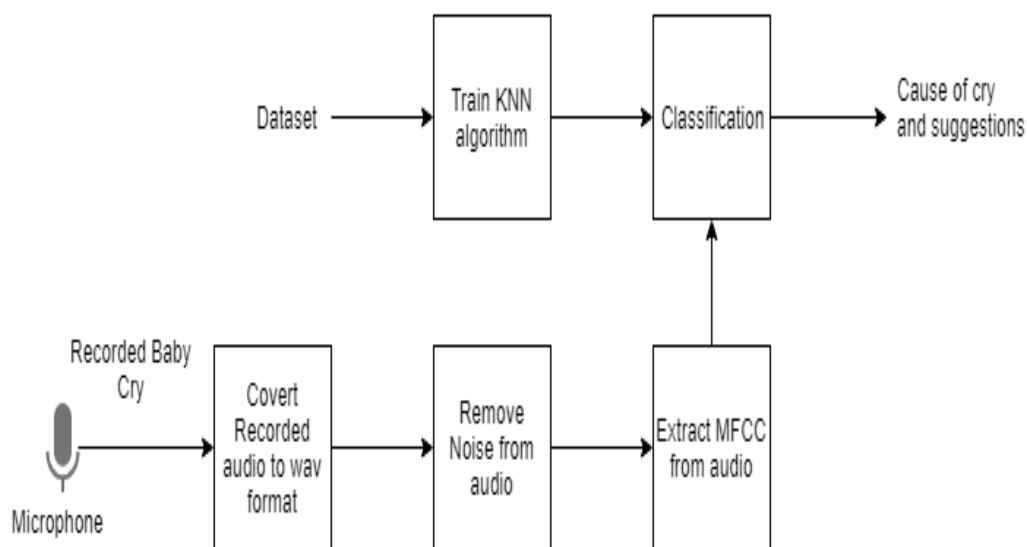


Fig 1 Software architecture

Mel Frequency Cepstral Coefficient

Mel Frequency Cepstral Coefficient (MFCC) is an audio feature commonly used in speech recognition systems. While calculating MFCC we will make audio signal into short frames. Then we take power spectrum which helps in identifying the different frequencies present in audio. Then we take logarithm of that value which decrease the difference in loudness and makes easy to compare. Then we apply DCT coefficients to get the final result. But however we are using a python library called `python_speech_features` which has a function “`mfcc`” that returns MFCC of the audio file.

K-Nearest Neighbor Algorithm (KNN)

KNN algorithm takes MFCC extracted from audio files as input data. KNN is a

lazy learning algorithm. So it stores the input as training data. When a MFCC of recorded baby cry is given to the algorithm, it calculates the distance of a given MFCC from all the entries in training data. KNN selects k number of entries from data set having least distance from MFCC of recorded baby cry. Algorithm finds the maximum occurring reason in the selected k entries. The maximum occurring reason is given as the output.

RESULTS

Expected outcome of the system is, once the baby cry audio is given as input, system will give reasons for the baby cry as output. It also gives suggestions on necessary actions to be taken. System is expected to be more efficient with less

delay. Initially dataset was in 3gp format. This dataset is converted into wav format by using Audio Segment class of pydub library. Pydub requires libav to be installed as dependency. Wav can be processed by most of the audio processing libraries. Then MFCC is extracted from these audio files using the python_speech_features library. The extracted features with labelling of reason are stored as a csv file so that it is not needed to extract MFCC every time and the process can be made faster. Fig 2 shows the data stored in csv file.

Despite the fact that higher order coefficients indicates speak levels of spectral details, contingent upon the sampling rate and estimation strategy, 12 to 20 cepstral coefficients are ordinarily ideal for speech analysis. Choosing more coefficients results in greater multifaceted nature in the models. However, choosing number of coefficients differs according to the application. It depends on the algorithm used. Fig 3 shows a graph of MFCC vs time for a single audio file.

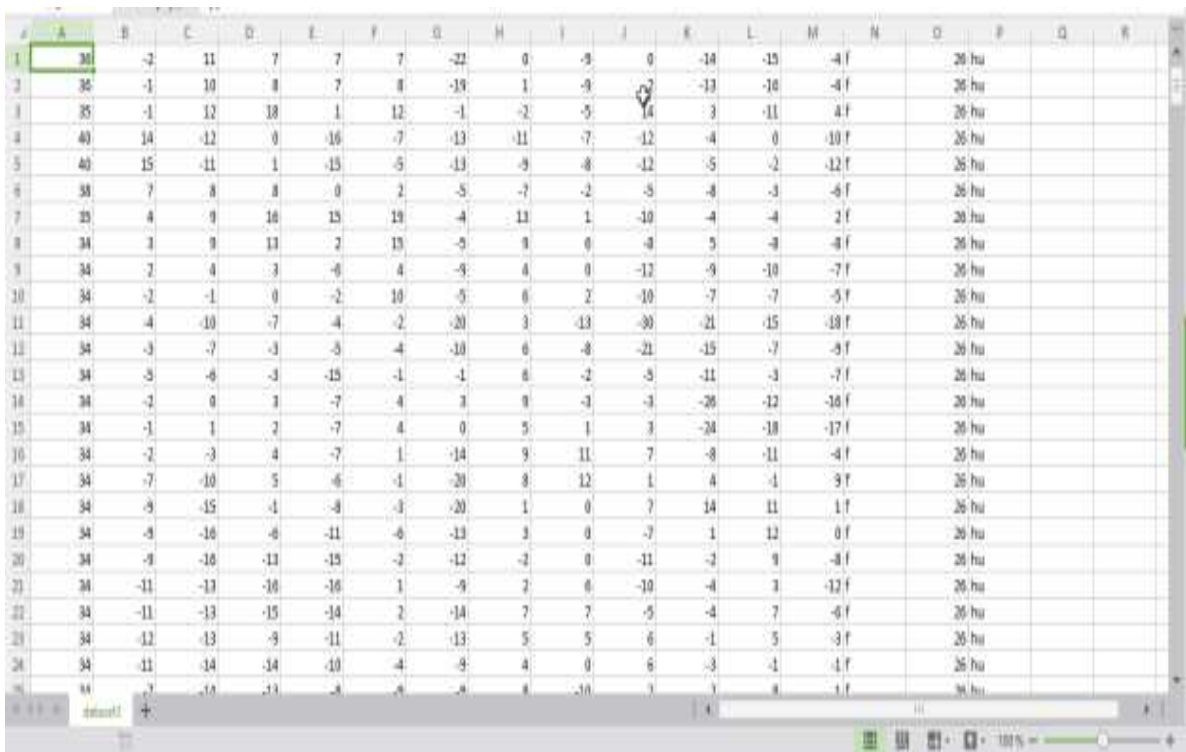


Figure:-2 Data stored in CSV

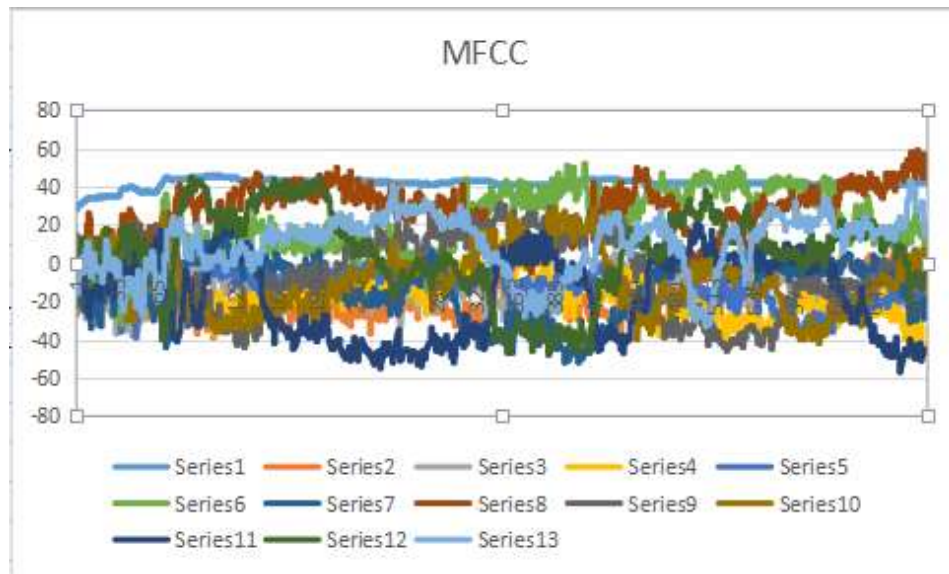


Figure:-3 MFCC of a single audio file

This paper considers 13 co-efficients. This paper uses KNN algorithm to classify the audio of baby cry. Every line from the csv file is read and parsed in to an array. This array is used to train knn algorithm and to get trained model. This model is used to classify the input audio file. Difficult task in KNN algorithm is to choose the value for k. K value depends on application and nature of the dataset. We have experimented with different k values and found that value 5 gives highest accuracy. Fig 4 shows the graph plot of k value vs accuracy.

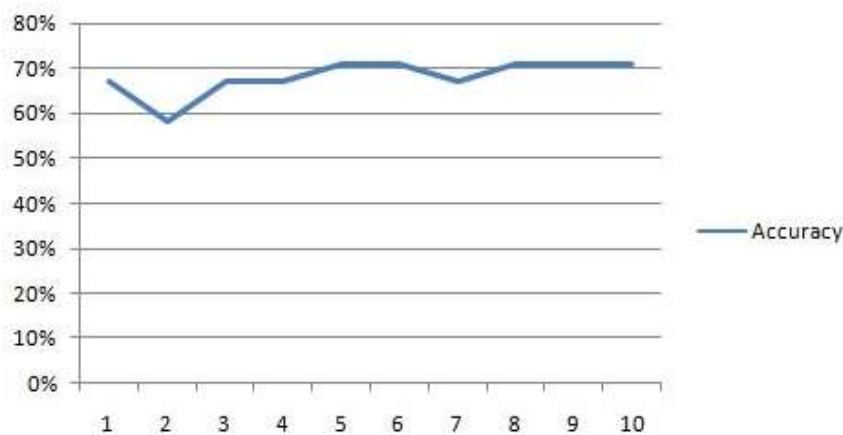


Figure:-4 K value vs accuracy

Since our dataset consists of around 350 audio files for the reason hungry, the trained system was biased. To make equal proportion of audio files for all the reasons, we considered only

31 audio files of the reason hungry. Number of audio files considered for each reason is given in table 1.

Table 1: Reasons, Number of audio files taken for each reason and length of audio

Reason	Number of audio files	Length of audio(sec)
Hungry	31	217
Belly pain	16	111
Need for burping	8	55
Discomfort	27	186
Tired	23	158

Dataset is divided into training and testing sets. Table 2 shows the number of files considered for each set. It also shows the accuracy. Accuracy of the system is 71.42%.

Table:-2 Accuracy of the system

	Total	Training set	Testing set	Accuracy(%)
Number of audio files	106	85	21	71.42
Length of audio(sec)	742	595	147	71.42

CONCLUSION

The purpose of this paper is to help parents to predict the reason for the baby cry. This system takes audio of baby cry as input. MFCC features from input audio will be extracted. Extracted features will be given to the machine learning algorithm. Machine learning technique is used to classify the baby cry into different reasons. This system can be used by parents, guardians, babysitting and baby care takers. This system also gives the

suggestions so that user can take necessary actions.

FUTURE WORK

Accuracy of the proposed system can be improvised. More data can be used for training the system. Better noise cancellation can be used to get better result.

ACKNOWLEDGEMENT

We wish to express our deep and sincere gratitude to our project guide Mrs.

Sadhana, Asst Professor, Dept of CSE for her valuable guidance and support throughout the tenure of the project. We are grateful to Dr. J V Gorbali, Head of the Department of CSE for providing us with right atmosphere in the department, encouragement and support, which made our entire task appreciable. We are indebted to our principal Dr. Srinivasa Rao Kunte and the management of Sahyadri College of Engineering & Management, Mangaluru, for having provided all facilities for the completion of the project. We would also wish to convey our profound thanks to all teaching and non-teaching staff, lab assistants and friends of Department of Computer Science who directly or indirectly helped us in making the project successful.

REFERENCES

1. Yuki Mima, Kaoru Arakawa, "Cause estimation of Younger babies Cries from the frequency analysis of the voice", Department of Computer Sciences Meiji University, Kawasaki 214-8571 Japan, 2006.
2. Rami Cohen, Yizhar Lavner, "Infant cry analysis and detection", IEEE 27th Convention of Electrical and Electronics Engineers in Israel, 2012.
3. Ypusra Abdulaziz, Shariffah Mumtazah syed ahmed, "Infant Cry Recognition System: A comparison of system performance based on Mel frequency and linear prediction", 2010.
4. J. Saraswathy, M. Hariharan, Sazali Yaacob and Wan Khairunizam, "Automatic Classification of Infant Cry: A Review", School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), 02600 Perlis, Malaysia, 2011.
5. Kevin Kuo, "Feature extraction and recognition of infant cries". 2010.
6. Jose Orozco Garcia, Carlos A. Reyes Garcia, "Mel-Frequency Cepstrum Coefficients Extraction from Infant Cry for Classification of Normal and Pathological Cry with Feed-forward Neural Networks", Instituto Nacional de Astrofisica Optica y Electronica (INAOE) Luis Enrique Err

- I, Tonantzintla, Puebla, Mexico, 2003.
7. Jose Orozco, Carlos A. Reyes-Garcia, "Implementation and Analysis of Training Algorithms for the Classification of Infant Cry with Feed-forward Neural Networks", Instituto Nacional de Astrofisica Optica y Electronica (INAOE) Luis Enrique Erro # 1, Tonantzintla, Puebla, Mexico, 2003.
 8. Pritam Pal, Ananth N. Iyer and Robert E. Yantorno, "Emotion Detection From Infant Facial Expressions And Cries" Electrical & Computer Engineering Department, Temple University 12th & Norris Streets, Philadelphia, 2006.
 9. Ioana-Alina Bănică, Horia Cucu, Andi Buzo, Dragoş Burileanu, and Corneliu Burileanu, "Baby Cry Recognition in Real-World Conditions", Speech and Dialogue Research Laboratory, University Politehnica of Bucharest, Romania, 2016.
 10. Andi Buzo, Dragos Burileanu, Corneliu Burileanu, "Automatic Methods for infants cry classification", Speech and Dialogue Research Laboratory, University Politehnica of Bucharest, Bucharest, Romania 2016.
 11. Chuan-Yu Chang, Jia-Jing Li, "Application of Deep Learning for Recognizing Infant cries", Chuan-Yu Chang, Jia-Jing Li National Yunlin University of Science & Technology, Taiwan, 2016.
 12. J. Saraswathi, N. Thiyagar, M. Hariharan, "Infant Cry Classification: Time Frequency Analysis", International conference on control system, computing and engineering, Malaysia, 2013.
 13. Shubham Asthana, Naman Varma, Vinay Kumar Mittal, "Preliminary Analysis of causes of Infant Cry", Electronics and Communication Engineering 1, 2 The LNM Institute of Information Technology 1, 2 Rupa Ki Nangal, Post-Sumel, Via Jamdoli, Jaipur 302 031, Rajasthan, India 3 Indian Institute

- of Information Technology 3Sri City, Satyavedu Mandal, Chittoor District 517 588, Andhra Pradesh, India,2014.
14. Mrs. Sameena Bano, Dr. K.M. RaviKumar, “Decoding Baby Talk: A Novel Approach for Normal Infant Cry Signal Classification”, 2015 International Conference on Soft-Computing and Network Security (ICSNS -2015), Feb. 25 – 27, Coimbatore, INDIA,2015.
15. Suryakanth V.Gangashetty, Pruthvi Raj Myakala, Rajashree Nalumachu, Shivam Sharma, VK Mithal,” Acoustic Analysis of Infant Cry Signal Toward Automatic Detection of the cause of Crying”, Indian Institute of Information Technology Chittoor, Sri City, A.P., India1,2,3 International Institute of Information Technology, Hyderabad, Telangana, India4 CEO, Ritwik Software Technologies Pvt. Ltd., Hyderabad, India5,2017.
16. Pruthvi Raj Myakala,Rajashree Nalumachu,Shivam Sharma, VK Mithal ,” An Intelligent System for Infant Cry detection and information in real time”, Speech and Dialogue Research Laboratory, University Politehnica of Bucharest, Bucharest, Romania, 2016.
17. Aomar Osmani, Massinissa Hamidi, Abdelghani Chibani, “Machine Learning Approach for Infant Cry Interpretation”, International Conference on Tools with Artificial Intelligence, 2017.
18. Rodica Ileana Tuduce, Horia Cucu, and Corneliu Burileanu,”Why is my Baby Crying? An in-depth Analysis of Paralinguistic Features and Classical Machine Learning Algorithms for Baby Cry Classification”, Speech and Dialogue Research Laboratory, University Politehnica of Bucharest, Romania, 2018.

Cite this Article

Chaithra lakshmi C, Aravinda B, Deeksha, Deeksha, Sadhana, (2019).

Predicting the Reason for the Baby Cry Using Machine Learning

“Journal of Artificial Intelligence, Machine Learning and Soft Computing”, 4(1), 11- 25

<http://doi.org/10.5281/zenodo.265635>

AUTHORS' PROFILE

[1] Chaithra lakshmi C, Student

Department: CS&E

College: Sahyadri College of Engineering & Management, Mangaluru, India

[2] Aravinda B, Student

Department: CS&E

College: Sahyadri College of Engineering & Management, Mangaluru, India

[3] Deeksha, Student

Department: CS&E

College: Sahyadri College of Engineering & Management, Mangaluru, India

[4] Sadhana, Assistant Professor

Department: CS&E

College: Sahyadri College of Engineering & Management, Mangaluru, India