IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13584

Infant Cry Analysis

Viraj Malusare[1], Aneesh Mote[2], Amar Yele[3], Asif Shaikh[4], Asst. Prof. Nitisha Rajgure[5]

Department of Computer Engineering, Zeal College of Engineering and Research, Pune, India[1-5]

Abstract: In this research, we have proposed a machine learning model that works on Random Forest Classifier, which extracts the MFCCs(Mel-Frequency Cepstral Coefficients) from baby cries and utilizes these features for predictions such as hungry, belly-pain, burping, tired and discomfort. This research can help the parents, caregivers to determine the exact reason behind the crying baby and suggesting the necessary actions to be taken further depending upon the baby cry.

Keywords: MFCCs(Mel-Frequency Cepstral Coefficients), FFT(Fast Fourier Transform), ML(Machine Learning), DL(Deep Learning), LSTM(Long Short Term Memory).

I. INTRODUCTION

As humans have their language of communication similarly babies try to express their feelings or needs by their cries. Different types of cries convey different needs or has different meaning. According to our research the five basic cries in Dunstan language are Neh (Hungry), Eh (Needs to burp), Heh (Discomfort), Eairh or Eargghh (Belly-Pain), Owh or Oah (Tired). Usually parents predict\ the reason behind baby crying using their experiences which is not always correct and can result in waste of time and improper care to the infant. Advanced machine learning models like Random Forest Classifier extracts the features of baby cries using MFCC and predicts the actual reason behind baby crying. This is far more better than general traditional approach which relies on trial and error based techniques. Baby cries can be identified in real time on real audio data which can be recorded using devices microphone to predict the main reason behind baby crying. This can become beneficial for the caregivers, parents, babysitters, siblings to assist babies as well as the infant as they will get required care.

II. LITERATURE RIVIEW

A. Yun-Chia Liang, Iven Wijaya, Ming-Tao Yang "Deep Learning for Infant Cry Recognition"[2022].In this research authors used deep learning algorithms like CNN and LSTM to recognize the needs of the infant. ANN was used for the comparison between the developed and pre existing models. Features were extracted using this algorithms using MFCC to detect the reason behind the baby cry .

B. Xuewen Yao, Megan Micheletti, Mckensey Johnson, Edison Thomaz and Kaya de Barbaro,"Infant Cry Detection in Real World Environments"[2022].Several machine learning algorithms were used by authors for classification of data generated in inlab and real-world environments.CNN model was developed with SVM as a classifier which provided the accuracy of 0.613.

C. Chunyan Ji, Thosini Bamunu Mudiyanselage, Yutong Gao and Yi Pan "A review of infant cry analysis and classification"[2021].In this research authors provided their research on infant cries and their crying patterns which could be further used by scientists, developers and researches to continue their work or make further improvements as dataset used in this research was not precise to provide accurate predictions

D. Sweta Bhattacharya, P.M. Durai Raj, Kathiravan "An Efficient classification of Neonates cry using extreme gradient boosting assisted Grouped vector network"[2021]. Authors of the above paper used the the extreme gradient boosting-powered grouped-support-vector networkto classify the baby cries and predict the exact reason behind baby crying. The model also provide better accuracy but was limited for only three classes of crying.

E. Kathleen Wermke, Michael P Robb, Philip J Schluter, "Melody complexity of infants' cry and non-cry vocalisations increases across the first six months" [2021]. This research aim to analyze the cry and non-cry vocalizations using various different approaches such as frequency analysis, pitch analysis and pattern recognition. Also various techniques from musical field of engineering are considered for further feature extraction.

F. K. S. Alishamol; T. T. Fousiya; K. Jasmin Babu; M. Sooryadas "System for Infant Cry Emotion Recognition using DNN"(2020). The proposed research by these authors identify the emotions and feelings of the infant with the help of extracting MFCC from the audio and using deep neural network(DNN).

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 5, May 2024
DOI: 10.17148/IJARCCE.2024.13584

G. Lichuan Liu1, Wei Li1, Xianwen Wu1, Benjamin X. Zhou2, "Infant Cry Language Analysis and Recognition: An Experimental Approach" [2019]. The main aim of the above research made by authors was to utilize the audio patterns in infant cry using techniques such as linear predictive coding (LPC), linear predictive cepstral coefficients (LPCC), Bark frequency cepstral coefficients (BFCC), and Mel frequency cepstral coefficients (MFCC). These features could be used to train several Neural networks as well as ML models to create a system to analyze and classify different types of baby cries.

III. METHODOLOGY

A. Data Collection
Data is collected from various verified open sources and is labelled accordingly depending upon requirements and inputs provided in model. Donate a cry corpus , chatterbaby , data from hospitals and several others sources are used for data collection.

B. Feature Extraction

1) Preprocessing
Preprocessing of audio files mainly contains the use of following libraries and techniques.
Library: This is the library which is basically used for audio processing, it handles the audio data. It is used to load the audio files and also used for resampling.

Sampling Rate: The audio is basically a continuous signal of audio. Sample is basically the number of sample per second in a continuous audio file. Sampling basically resamples the audio signals to create a different audio signals which are necessary for FTTs for further computations.

Pre-emphasis Filter: This filter is used to amplify the high frequencies in the audio signal to balance the spectrum and improve the signal-to-noise ratio. The formula used is:

$y[n]=x[n]-\alpha\cdot x[n-1]$y[n]=x[n]−α·x[n−1] where α is the pre-emphasis coefficient (usually set to 0.97).

2) Framing
Frame Size: Each frame from the audio signal taken is of 25 milliseconds which is generally short for pattern recognition and frequency analysis to overcome this problem we make use of strides.

Frame Stride: Frames strides is basically the distance between the different frame sizes. Frame strides overlap with a 10-millisecond stride which reduces the distance and dimensions of the frames making it better for further computations.

Padding: The length of audio frames has impact on our feature extraction hence to make sure that frames are of equal length padding is applied which is set to Zero also known as Zero padding.

3) Windowing and FFT
Windowing: A Hamming window is applied to each frame which basically used to separate the input audio signals into segments of required time basically we can call them temporal segments .

FFT: The Fast Fourier Transform is implemented on these frames and computed further to convert this temporal frames to the frequency domain.

Power Spectrum: The power spectrum is applied and computed by taking the magnitude of the FFT and squaring it.

4) Mel Filter Bank and MFCC
Mel Filter Bank is applied on these generated audio signals to decompose these audio signals into separate Frequency domains or bands in a mel frequency scale, which is basically a non linear conversion of a audio signal through which human ears can process the sound or audio waves and hear and perceive the incoming audio data.

The formula for the following Mel Filter Bank is given as :

$M(f)=2595\cdot\log_{10}(1+f700)$M(f)=2595·log10(1+700f)

Then further the filters are applied to the power spectrum to produce Mel-frequency cpectral coefficients.

Discrete Cosine Transform (DCT): The logarithm of the Mel-filtered spectra is taken to compress the dynamic range to make these values transform into a specified range.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13584

The DCT is applied to the log Mel spectra obtained in the previous procedures. DCT is further used to extract the Mel-Frequency Cepstral Coefficients (MFCCs), which are used as features for further analysis. Typically, the first 12-13 coefficients are retained. These coefficients serve as the features that help us understand and analyze the sound data.

5) Classification Model

The MFCC features are extracted from all the audio files from our collected datasets. These features are further flattened and stored in a list which is used for relational representations. Corresponding labels of array [0,1,2,3,4] for different audio files such as hungry, belly pain, burping, tired and discomfort are generated based on the subfolder depending upon these features and each type of audio file found in that particular folder. The features and labels are combined into a pandas DataFrame and saved as a CSV file which is used for training out data set. The data is loaded back, and features (X) and labels (y) are separated. The dataset is split into training and testing sets using a 85-15(in our case) split ratio to evaluate the model's performance. A Random Forest Classifier is initialized with specific parameters (n_estimators=200, criterion="entropy", max_depth=32) and trained on the training set.

IV. CONCLUSION

The Random Forest Classifier when trained upon the MFCCs extracted from the baby cry dataset successfully results in the Precision of 0.93, Recall of 0.90 and F1 Score of 0.89. The Accuracy score obtained by training the Random Forest Classifier is 90.48%. The Classification report and Confusion which we got upon evaluating our Model are as follows :

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4
1	0.86	1.00	0.92	6
2	1.00	0.33	0.50	3
3	1.00	1.00	1.00	6
4	0.67	1.00	0.80	2
accuracy			0.90	21
macro avg	0.90	0.87	0.84	21
weighted avg	0.93	0.90	0.89	21

Confusion Matrix:

[[4 0 0 0 0]
[0 6 0 0 0]
[0 1 1 0 1]
[0 0 0 6 0]
[0 0 0 0 2]]

The Random Forest Classifier performed well with an accuracy of 90.48% on the test set which is even better than some deep leaning models. The precision, recall, and F1-score we got upon evaluating Random Forest Classifier are 0.93, 0.90 and 0.89 which are also satisfactory. The developed model shows a great accuracy on all the classes of the test data. When tested on real-time data the model successfully predicts the type of baby cry which helps assist the parent, caregivers or babysitters. As compared to neural networks which requires large dataset, Random Forest Classifier provides better accuracy on small dataset. To improve the model, additional data could be collected. Further tuning of the model parameters might also yield better performance.

REFERENCES

[1] A. Chittora, H. A. Patil, in International Conference on Text, Speech, and Dialogue, 9302. Significance of unvoiced segments and fundamental frequency in infant cry analysis, (2015), pp. 273–281. https://doi.org/10.1007/978-3-319-24033-6_31

[2] A. A. Dixit, N. V. Dharwadkar, in Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing, ICCSP 2018. A survey on detection of reasons behind infant cry using speech processing, (2018), pp. 190–194. https://doi.org/10.1109/ICCSP.2018.8524517.

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⫻ Peer-reviewed & Refereed journal ⫻ Vol. 13, Issue 5, May 2024

DOI: 10.17148/IJARCCE.2024.13584

[3] Chunyan Ji, Thosini Bamunu Mudiyanselage, Yutong Gao & Yi Pan "A review of infant cry analysis and classification" (2021).Springer

[4] C. Ji, S. Basodi, X. Xiao, Y. Pan."Infant sound classificatio.n on multi-stage CNNs with hybrid features and prior knowledge" (2020). Springer

[5] D. Ferretti, M. Severini, E. Principi, A. Cenci, S. Squartini, in 2018 26th European Signal Processing Conference (EUSIPCO). Infant cry detection in adverse acoustic environments by using deep neural networks, (2018), pp. 992–996. https://doi.org/10.23919/EUSIPCO.2018.8553135.

[6] Hemant A. Patil, "Infant Identification from Their Cry" (2009).

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

[8] J. Saraswathy, M. Hariharan, S. Yaacob, W. Khairunizam, in 2012 International Conference on Biomedical Engineering (ICoBE). Automatic classification of infant cry: a review, (2012), pp. 543–548. https://doi.org/10.1109/ICoBE.2012.6179077.

[9] 9.K. S. Alishamol; T. T. Fousiya; K. Jasmin Babu; M. Sooryadas"System for Infant Cry Emotion Recognition using DNN"(2020).IEEE.

[10] L. Liu, Y. Li, K. Kuo, in 2018 International Conference on Information and Computer Technologies, ICICT 2018. Infant cry signal detection, pattern extraction and recognition, (2018), pp. 159–163. https://doi.org/10.1109/INFOCT.2018.8356861.

[11] L. Liu, W. Li, X. Wu, B. X. Zhou, Infant cry language analysis and recognition: an experimental approach. IEEE/CAA J. Autom. Sin.6(3), 778– 788 (2019). https://doi.org/10.1109/JAS.2019.1911435.

[12] M. V. Varsharani Bhagatpatil, An automatic infant's cry detection using linear frequency cepstrum coefficients (LFCC). Int. J. Sci. Eng. Res.5(12), 1379– 1383 (2014).

[13] M. Hariharan, L. S. Chee, S. Yaacob, Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network. J. Med. Syst.36(3), 1309– 1315 (2012). https://doi.org/10.1007/s10916-010-9591-z.

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

[15] R. I. Tuduce, M. S. Rusu, H. Cucu, C. Burileanu, in 2019 42nd International Conference on Telecommunications and Signal Processing, TSP 2019. Automated baby cry classification on a hospital-acquired baby cry database, (2019) IEEE, pp. 343–346. https://doi.org/10.1109/TSP.2019.8769075.

[16] S. Bano, K. M. Ravikumar, in Proceedings of the IEEE International Conference on Soft Computing and Network Security, ICSNS 2015. Decoding baby talk: a novel approach for normal infant cry signal classification, (2015), pp. 24–26. https://doi.org/10.1109/ICSNS.2015.7292392.

[17] S. P. Dewi, A. L. Prasasti and B. Irawan, "Analysis of LFCC Feature Extraction in Baby Crying Classification using KNN," 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS), Bali, Indonesi, 2019

[18] Sweta Bhattacharya,P.M. Durai Raj,Kathiravan." An Efficient classification of Neonates cry using extreme gradient boosting assisted Grouped vector network"(2021).Hindavi Journal of Healthcare Engineering.

[19] Shivam Sharma, Pruthvi Raj Myakala, Rajasree Nalumachu , "Acoustic analysis of infant cry signal towards automatic detection of the cause of crying" (2017). DBLP

[20] S. Yamamoto, Y. Yoshitomi, M. Tabuse, K. Kushida, T. Asada, Recognition of a baby's emotional cry towards robotics baby caregiver. Int. J. Adv. Robot. Syst.10: (2013). https://doi.org/10.5772/55406.

[21] S. Orlandi, C. A. Reyes Garcia, A. Bandini, G. Donzelli, C. Manfredi, Application of pattern recognition techniques to the classification of full-term and preterm infant cry. J. Voice. 30(6), 656–663 (2016). https://doi.org/10.1016/j.jvoice.2015.08.007.

[22] S. P. Dewi, A. L. Prasasti, B. Irawan, in Proceedings - 2019 IEEE International Conference on Signals and Systems, ICSigSys 2019. The study of baby crying analysis using MFCC and LFCC in different classification methods, (2019), pp. 18–23. https://doi.org/10.1109/ICSIGSYS.2019.8811070.

[23] Wermke K, Robb MP, Schluter PJ." Melody complexity of infants' cry and non-cry vocalizatios increases across the first six months"(2021). National Institute of Health

[24] Xuewen Yao, Megan Micheletti, Mckensey Johnson, Edison Thomaz and Kaya de Barbaro,"Infant Cry Detection In Real World Environments"[2022].Proc IEEE Int Conf Acoust Speech Signal Process. 2022.National Institute of Health.

[25] Yun-Chia Liang , Iven Wijaya , Ming-Tao Yang "Deep Learning for Infant Cry Recognition"[2022].Int. J. Environ. Res. Public Health 2022.MDPI.