Applied GIS

ISSN: 1832-5505
Vol-9 Issue-04 Oct 2021

BABY CRY CLASSIFICATION USING MACHINE LEARNING ALGORITHMS

BALAJI SUNIL CHANDRA, ASSISTANT PROFESSOR, hod.cse@svitatp.ac.in

JANGILI RAVI KISHORE, ASSISTANT PROFESSOR, jangiliravi.kishore1@gmail.com

VIJAYA BHASKAR MADGULA, ASSISTANT PROFESSOR, vijaya.bhaskar2010@gmail.com

Department of CSE, Sri Venkateswara Institute of Technology, N.H 44, Hampapuram, Rapthadu, Anantapuramu, Andhra Pradesh 515722

Abstract— Cries are a way for kids to communicate how they're feeling. The natural periodic tone and voice shift of a baby cry are distinguishing features. Parents may remotely watch their newborn in critical situations using cry detection. Applications such as remote infant monitoring rely on the ability to detect a baby cry in voice signals. Scholars who examine the relationship between patterns of baby cry signals and other developmental characteristics also find this capability significant. The goal of this sound recognition research is to extract features and classify them based on the pattern of the sounds. For feature extraction, we use MFCC, and for classification, we employ K-Nearest Neighbour (K-NN). One common categorization approach for audio data is K-Nearest Neighbour (KNN). Out of all the classifiers tested, the KNN classifier performed the best.

Keywords— Signal Patterns, Speech signal Processing, Feature Extraction, MFCC, K- Nearest Neighbor

I. INTRODUCTION

For a number of years, scientists have studied cry signals, often known as cry patterns. Researchers and analysts have discovered that a newborn's scream signals may provide very specific information on the baby's emotional and physiological well-being.

This issue may be approached in several ways. Using facial characteristics is one way to detect lethargy. Feature extraction takes into account the face expressions in this image-based method. When individuals are tired, they usually yawn or shut their eyes. Taking these characteristics into account allows us to forecast the driver's condition. Alert and sleepy pictures make up the dataset. Several years of study and analysis went into the system's design, which is comprised of the fo signals or cry patterns. Researchers and analysts have discovered that a newborn's scream signals may provide very specific information on the baby's emotional and physiological well-being.

According to World Health Organisation data, over 40% of newborn fatalities occur during the first thirty to fifty days of a baby's life. Within the first week after delivery, 72 percent of newborn fatalities occur, and if the reason is known far earlier, up to 66 percent of infant lives may be spared. Methods that assist us recognise the early warning symptoms of poor baby cleanliness and health may significantly lower the newborn mortality rate. Our thesis's primary objective is to design and build a trustworthy system that can diagnose illnesses using just ultrasonic imaging. The first step in creating this kind of system is identifying the trustworthy cry patterns or components in an input waveform. It is likely that the NCDS system becomes confused if the input voice signal includes extraneous sounds in addition to the cry signal. Therefore, developing an automated detection system that can precisely scan the inspiratory and expiratory portions of a cry pattern is the greatest difficulty in the design and implementation of a diagnostic system. Development of automated audio segmentation of expiratory and inspiratory components of child cries was a valuable outcome of extensive study on illnesses and cry signals and their link. A fully automated system that aids in illness comprehension would be very beneficial and easier to implement if we are able to separate audio cry signals and analyse critical portions of a previously recorded sound signal. We can certainly use this method to back up our conclusions while trying to decipher baby screams. Because of this, we can identify the signs sooner and perform the required actions efficiently and affordably. Recent research on baby cries has shown that there are a variety of needs that babies express by crying, including but not limited to hunger, exhaustion, unpleasant emotions, discomfort, and other similar issues. Medical professionals, researchers, and students can learn to recognise patterns in infant cries and use that information to predict how much food and water the baby will need. This is great for babies, but it can be a major hassle for parents who aren't up to the task. This project presents a data collection of eight distinct newborn cries that may be used to train an artificial technique for infant cry categorization. Therefore, the primary goal is to identify the meaning of the baby's cry by extracting

ISSN: 1832-5505
Vol-9 Issue-04 Oct 2021

relevant characteristics from the cry audio signal, which is the baby's cry, and then testing the unknown cry signal with the categorised trainer.

II. Literature Survey

There are various methods for detecting drowsiness. Some of the approaches which are used in this domain are discussed here.

A. Components of infant cry audio signal

Both the inspiration (INSV) and expiration (EXP) portions of an infant's cry, which include vocalisation, are crucial and essential components of the audio signal. Developing a strategy that can efficiently search for INSV and EXP inside a certain cry signal is one of the primary obstacles encountered by this sort of technology. Due to the presence of both voiced and unvoiced components in a typical audible baby cry, the cry identification issue differs from unvoiced, voiced segmentation[1][2].

B. Voice Activity Detection(VAD)

The primary challenge with using VAD (Voice Activity Detection) modules to detect cries recorded in very noisy home environments is not easily solved, according to the literature put out by Kuo (2010). VAD modules deal with the problem of searching for speech patterns in other auditory active regions of an audio signal. Silence, sound, or even a doorbell alert are all examples of possible additional auditory active patterns. A crucial quantity, the Signal to Noise Ratio "SNR ", has the potential to produce several unacceptable mistakes. Real-time voice transmission, automated speech recognition, telephones, and other digital resources all rely on VAD[3]. Two fundamental and crucial strategies are included in some of the most popular and extensively used VAD techniques: feature extraction and decision making. Signal properties that permit energy and cepstral coefficient computation, zero-crossing ratio (ZCR) calculations, wavelet-and-entropy transforms (WFTs) introduced in 2008 by Wang and Tasi, decision rule computation based on frame-by-frame data, and extremely basic thresholding rules proposed by Juang et al. For the purpose of cry signal segment identification in 2009, we used the well-known VAD algorithms developed by Rabiner-Sambur and the G.729b approach.

C. Findings

The findings were:

- It is hard to select the threshold settings in a noisy domestic environment.
- While data acquisition, the Traditional VAD module is unable to differentiate between EXP and[6] INSV (cry signal segments) and recorded speech signal segments.

Traditional VAD modules are unable to distinguish expiration (EXP) from inspiration (INSV) parts of a cry audio signal[5]

Statistical approach is a good solution to avoid restricting the problem of adjusting thresholds. That is why due consideration is given to statistical model-based approaches proposed by AbouAbbas in 2015b, 2015c modules.

D. Existing system

There are systems to detect whether a sound file provided is a Baby cry or not. The techniques used LFCC (Linear Frequency Cepstral Coefficients) for feature extraction. There is also a system to classify the reasons for baby cry and in this system various classifiers are used to classify the reasons from the pre-classified data set.

III. Proposed System

A. The Dataset

The first step is collection of the dataset. We considered some of the datasets for this system.

A dataset which contains audio files of baby cries collected by speech research institutes namely donate-a-cry-corpus has been collected. This dataset contains audio samples of many infants captured under different situations over several situations. This dataset contains 8 sets of audio files ie., Awake, Belly pain, Burping, Discomfort, Hug, Hungry, Sleepy and Tired.

B. System Architecture

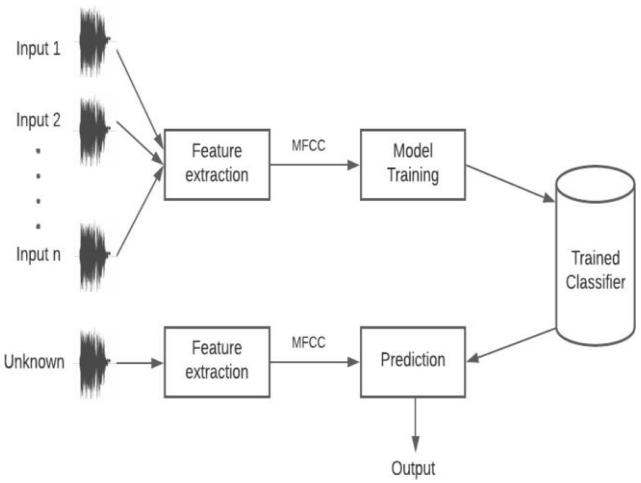

Fig. 1. System Architecture

C. Feature Extraction

The first step is to collect the data. Typically, an audio file destined for additional processing serves as the input. The speaker's identity, gender, and emotional traits are all reflected in the massive amounts of data found in the voice signal. There are unique qualities inherent in every speaker's voice and every speech they deliver. An important first step in signal processing is feature extraction, which takes raw human speech and transforms it into a more usable

ISSN: 1832-5505
Vol-9 Issue-04 Oct 2021

parametric representation. Extracting features functions as a speaker recognition system and plays a significant part in speech recognition overall performance. An ethical feature extraction method will take note of the signal's key features while leaving out any superfluous ones. The goal of feature extraction is to extract meaningful information from a signal while removing noise and irrelevant details. When it comes to distinguishing between speakers and their speeches, feature parameters are crucial. Many applications related to speech make use of these factors for analysis, including speaker identification, voice synthesis, speech coding, and speech recognition. Since we wanted our suggested system to work well, we focused on extracting key features from the speech.

D. Mel frequency cepstral coefficients(MFCC)

When it comes to automated speech or speaker identification systems that use the Mel scale—a scale based on the human ear—MFCC is among the most prominent feature extraction methods.[11]. The nonlinearity of human hearing in relation to sound frequency has been validated. The audio-supported perception is represented by these coefficients. Mel's frequency cepstrum is their source.[14] Band pass filters artificially increase higher frequencies, allowing the spectrum information to be transformed to MFCC. Then, an inverse Fast Fourier Transform (FFT) is applied to the filtered signal. It combines the advantages of cepstrum analysis with important bands backed by perceptual frequency scales.[17] The outcome is an increased emphasis on the higher frequencies. The Mel frequency cepstrum is often thought of as the most basic model of the human ear because of how well it represents the listener's response system.

.

E. K-Nearest Neighbor (KNN)

Among the top Machine Learning algorithms that rely on the Supervised Learning method is K-Nearest Neighbour.By comparing the new case/data to the existing cases, the K-NN algorithm classifies it into the category that is most similar to the existing ones.The K-NN algorithm compiles all the existing data and uses similarity to categorise new data. This means that the K-NN algorithm is able to effortlessly sort newly-arrived data into an appropriate category. Although the K-NN technique is most often used to classification issues, it will also be utilised for regression. One possible interpretation of K-NN is that it is a non-parametric method, meaning it does not assume anything about the input data. This method is also known as a lazy learner since it retains the information and only uses it when classifying, rather than instantly learning from the training set.
Arithmetic is performed by taking the mean of an array that contains the cepstral coefficients taken from the audio sample. The model decides the categorization category to deliver as output when applied to the data, taking the value of k into account.

F. Naive Bayes

One kind of supervised learning method that uses the Naive Bayes algorithm to solve classification issues is one that is based on the Bayes theorem. Its primary use case is text categorization using a dataset that has several dimensions. When it comes to developing rapid machine learning models that can produce quick predictions, one of the easiest and best algorithms to use is the Naive Bayes Classifier.Being a probabilistic classifier, it bases its predictions on the concept of an object's likelihood. Popular applications of the Naïve Bayes Algorithm include article classification, sentiment analysis, and spam filter.

G. Support Vector Machine

One of the most well-known supervised learning techniques, Support Vector Machine (SVM) is used for both classification and regression tasks. Nevertheless, its main use is in Machine Learning classification issues.The SVM algorithm's main purpose is to find the best decision boundary or line that will divide n-dimensional space into classes. This will allow us to easily classify fresh information in the future. Hyperplanes are the finest choice boundaries. To aid in the creation of the hyperplane, SVM selects the intense points/vectors. The approach is dubbed Support Vector Machine because these extreme examples are called support vectors.

IV. Results

A. Mel Frequency cepstral Coefficients (MFCC)

These are the cepstral coefficients obtained when a random audio file is employed for testing. There are 40 cepstral coefficients in total and every value is different. There are positive and negative values.[14] Positive value of the cepstral coefficient implies that the bulk of spectral energy is concentrated in low frequency regions. Negative value of the cepstral coefficient means the spectral energy is concentrated in high frequency regions.[15][17] Here, the amount of cepstral coefficients is 40 as shown in Fig.2. because it yields better results.

```
[[-83.377686    33.08731    -42.55658     22.141876   -12.549919
   23.294508   -22.59206     14.99546      1.4717332    4.209373
   -0.74977094 -14.92513      0.9372652    8.224226    -0.12192553
   13.989015     6.5319715    1.9878587   -8.569176    -6.5270405
    0.10054475   4.7355175   -4.119027    -3.9245353   -5.563842
    3.058156    -2.8110352   -2.5545697   -0.23786436   7.003273
    3.155918    -1.0060205   -2.7972405    4.107716    -0.52387434
   -1.1573135   -1.2756749    1.5718873   -3.6838982   -0.8891317 ]]
```

Fig.2. Mel frequency Cepstral Coefficients

B. Comparison among KNN, Naive Bayes and SVM

The accuracy obtained when Naive Bayes is applied on the data is 45% and in the case of SVM, the accuracy is 42% and the highest accuracy that is obtained is 76.16% when KNN is used for classification.

ISSN: 1832-5505
Vol-9 Issue-04 Oct 2021

C. K value in KNN process

76.16% when using K value is 2. Moreover, the smallest accuracy obtained is in between 50% and 60% when the K value is in between 4 and 10. This can be influenced by the training data and testing data. For this system, the highest observed accuracy is 76.16% and is observed when the K value is 2.

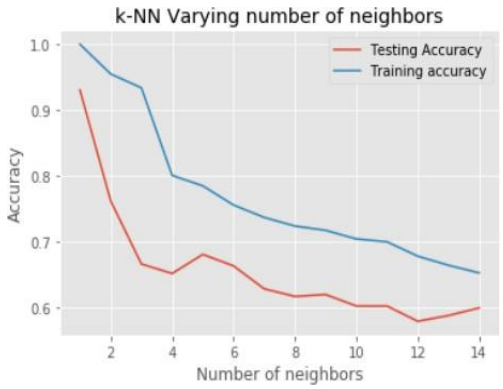

Fig.3. Accuracies for different of K

V. Conclusion

Using eight previously defined categorization categories, the primary objective is to identify the source of a scream given an audio sample as input. The audio files come from a variety of sources and were captured in loud locations. As a result of working on this project, we now understand the significance of preprocessing in feature extraction and how to interpret baby screams, which may help with newborn care. We achieved 76% accuracy at k=2 using the KNN classifier, which correctly identifies and provides the right explanation most of the time. The implementation was carried out using Python versions higher than 3.6.Using an audio file as input, this suggested system may identify the cause of a cry.Remote baby controlling and medical applications that aim to identify the cause of a baby's scream may both benefit from this same principle.

References

[1] Bănică, H. Cucu, A. Buzo, D. Burileanu and C. Burileanu, "Automatic methods for infant cry classification", International Conference on Communications (COMM),Bucharest, 2016, pp.51-54.

[2] Abou-Abbas, Lina, ChakibTadj, and HesamAlaieFersaie. "A fully automated approach for baby cry signal segmentation and boundary detection of expiratory and inspiratory episodes" ,The Journal of the Acoustical Society of America 142.3 (2017):1318-1331.

The comparison result of the K value for the classification process is shown in fig.3. The highest accuracy obtained is

[3] Bou-Abbas, L., Alaie, H., and Tadj, C. "Segmentation of voiced newborns' cry sounds using wavelet packet based features" ,in 2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE), Halifax, Canada, 2015(a), pp.796–800.

[4] Sjölander, Kåre and Jonas Beskow. "Wave-surfer - an open source speech tool",Sixth International Conference on Spoken Language Processing.2000.

[5] R. Cohen, "Infant Cry Analysis and Detection," 2012, pp.2–6.

[6] S. Sharma, P. R. Myakala, R. Nalumachu, S. V. Gangashetty, and V. K. Mittal, "Acoustic analysis of infant cry signal towards automatic detection of the cause of crying," 2017 7th Int. Conf. Affect. Comput. Intell. Interact. Work. Demos, ACII 2017, vol. 2018– January, pp. 117–122,2018.

[7] W. S. Limantoro, C. Fatihah, and U. L. Yuhana, "Application Development for Recognizing Type of Infant ' s Cry Sound," 2016, pp.157–161.

[8] R. P. Balandong," ACOUSTIC ANALYSIS OF BABY CRY", no. May.2013.

[9] V. V Bhagat Patil and P. V. M. Sardar, "An Automatic Infant's Cry Detection Using Linear Frequency Cepstrum Coefficients (LFCC)," vol. 5, no. 12, pp.1379–1383, 2014.

[10] E. Franti, I. Ispas, and M. Dascalu, "Testing the Universal Baby Language Hypothesis - Automatic Infant Speech Recognition with CNNs," 2018 41st Int. Conf. Telecommun. Signal Process. TSP 2018, pp. 1–4, 2018.

[11] Dewi, S. P., Prasasti, A. L., & Irawan, B. (2019). The Study of Baby Crying Analysis Using MFCC and LFCC in Different Classification Methods. The 2019 IEEE International Conference on Signals and Systems (ICSigSys) (pp. 19-24). Bandung: IEEE.

[12] R. M. Aarts, "Audio signal processing device," J. Acoust. Soc. Am., vol. 120, no. 6, p. 3445, 2006.

[13] G. L. -, Y. H. -, L. Y. -, and M. N. -, "Pitch Analysis of Infant Crying," Int. J. Digit. Content Technol. its Appl., vol. 7, no. 6, pp. 1072–1079, 2013.

[14] X. Zhou, D. Garcia-romero, R. Duraiswami, C. Espy- wilson, S. Shamma, and A. Motivation, "Linear versus Mel Frequency Cepstral Coefficients for Speaker Recognition," pp. 559–564, 2011.

[15] E. C. Djamal, N. Nurhamidah, and R. Ilyas, "Spoken word recognition using mfcc and learning vector quantization," Int. Conf. Electr. Eng. Comput. Sci. Informatics, vol. 4, no. September, pp. 250–255, 2017.

[16] P. K. Sari, K. Priandana, and A. Buono, "Perbandingan Sistem Perhitungan Suara Tepuk Tangan dengan Metode Berbasis Frekuensi dan Metode Berbasis Amplitudo Comparison of Applause Calculation Systems using Frequency- Based Method and Amplitude-Based Method," J. Ilmu Komput. Agri-Informatika, vol. 2 Nomor 1, pp. 29–37, 2013.

[17] N. Dave, "Feature Extraction Methods LPC , PLP and MFCC In Speech Recognition," Int. J. Adv. Res. Eng. Technol., vol. 1, no. Vi, pp. 1–5, 2013.

[18] G. V. I. S. Silva and D. S. Wickramasinghe, "Infant Cry Detection System with Automatic Soothing and Video Monitoring Functions," J. Eng. Technol. Open Univ. Sri Lanka, vol. 5, no. 1, pp. 36–53, 2017.

[19] S. Jagtap, P. K. Kadbe, and P. N. Arotale, "System propose for Be acquainted with newborn cry emotion using linear frequency cepstral coefficient," Int. Conf. Electr. Electron. Optim. Tech. ICEEOT 2016, pp. 238–242, 2016.