Applied GIS

ISSN: 1832-5505

Vol-10 Issue-01 Feb 2022

The Use of Machine Learning Algorithms for the Classification of Babies' Cries

D. SANKAR, Assistant Professor, shankar.dasari126@gmail.com

VIJAYA BHASKAR MADGULA, Professor, vijaya.bhaskar2010@gmail.com

KUMMARA RANGA SWAMY, Assistant Professor, rangaswamy.kumara@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. There is a natural periodic tone and variation in voice that distinguish a baby cry from others. Parents may remotely watch their newborn in critical situations using cry detection. Scholars who examine the association between patterns of baby cry signals and other developmental factors find the detection of baby cries in speech signals to be an essential step in applications like remote infant monitoring. 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, K-Nearest Neighbour, Feature Extraction, MFCC, and Speech Signal Processing

I. INTRODUCTION

Researchers have been studying cry signals or patterns for a while now. The cry signals may provide precise representations of the mental and physical conditions of babies, according to scholars and experts. This issue may be approached in several ways. Facial characteristics may be used as a tool to detect lethargy. Facial expressions are taken into account using Feature extraction in this image-based approach. When individuals are tired, they usually yawn or shut their eyes. We can forecast the driver's condition by taking these characteristics into account. 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. The cry signals may provide precise representations of the mental and physical conditions of babies, according to scholars and experts. 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. Creating or implementing a trustworthy technique that allows us to comprehend ailments just via cry sound examination is, to be more specific, the superior aim of our thesis. The first step in creating this kind of system is identifying the trustworthy cry patterns or components in an input waveform. Input speech signals that include sounds other than the cry signal alone likely cause confusion in the NCDS system. 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. It was possible to automatically partition the expiratory and inspiratory portions of newborn cries after extensive study on the relationship between cry signals and disorders yielded some beneficial findings. It could be much easier and more useful to create an entirely automated system that aids in the study of illnesses if we can separate audio cry signals and analyse crucial portions of a previously recorded sound signal. We can certainly use this method to back up our conclusions while trying to decipher baby screams. We can identify the signs sooner and take action efficiently and affordably because of this. Recent research on baby cries has highlighted a variety of needs that babies express via 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 offers a data set of eight distinct newborn cries to develop an artificial system for cries categorization.

Therefore, the primary goal is to identify the meaning of the baby's cry by extracting 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

The important and key components of an infant cry audio signal are inspiration and expiration parts with vocalization and hear-able inspiration(INSV) and expiration (EXP). The vital challenges faced in this type of system is implementing a method that can effectively search INSV, EXP exactly within a respective cry signal. The problem of cry detection is different from unvoiced, voiced segmentation because a typical hear-able infant cry audio signal contains each of the unvoiced and voiced parts[1][2].

B. Voice Activity Detection(VAD)

The literature proposed by Kuo, 2010, has stated thatthe main problem is detection of cry sound using a system recorded in a lot of noisy domestic local environment is not too easy to be solved by VAD (Voice Activity Detection) modules, VAD deals with the problem of searching or finding speech patterns from other auditory active regions of a considered audio signal. The other auditory active patterns may be any type like silence, noise or a doorbell warning. The Signal to Noise Ratio “SNR " is a key parameter and it might result in a lot of unwanted errors. VAD is vital in several audio communication systems like automatic speech recognition, telephones, other digital resources, and transmission of speech in real time[3]. Some of the Common and very widely used VAD methods contain two basic and important methods: Feature Extraction and Decision making. Features of a signal that allows computation of energy, cepstral coefficients, ZCR, Marzinzik and Kollmeier, proposed spectrum analysis in 2002, Wang and Tasi, proposed wavelet ad entropy transforms in 2008 ; Juang et al., proposed decision rule computation based on frame-by-frame and very simple rules for thresholding. In 2009, After applying familiar VAD algorithms from Rabiner-Sambur and G.729b method in detection of cry signal parts or segments.

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 distinguishexpiration (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

ISSN: 1832-5505
Vol-10 Issue-01 Feb 2022

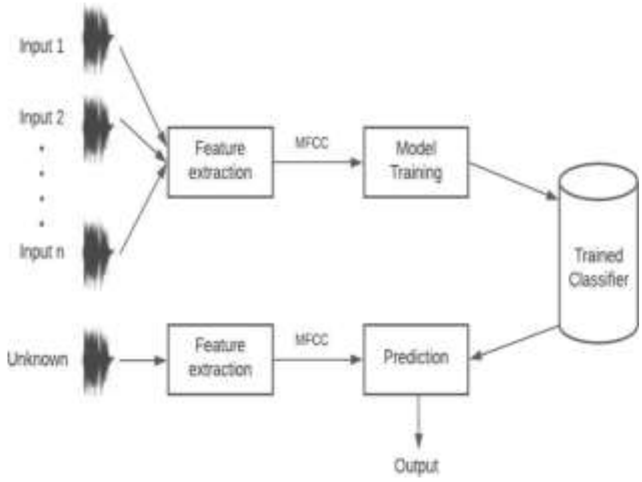

Fig. 1. System Architecture

C. Feature Extraction

At first, the input is taken. The input is usually an audio file which is to be processed further.

The speech signal contains an oversized number of data which reflects the emotional characteristics, gender classification and therefore the speaker's identity. Every speech and speaker has special individual characteristics which are embedded in their speech utterances. Feature extraction from the speech signal is the signal processing forepart process which converts the human speech into some useful parametric representation. Feature extraction

plays an important role within the overall performance of a speech recognition moreover as a speaker recognition system. an honest feature extraction technique must capture the important characteristics of the signal and may also discard some irrelevant attributes. Feature extraction is the process of keeping useful information from the signal while discarding redundant and unwanted information . Feature parameters play a big role to differentiate speeches likewise as speakers from one another. These parameters are useful for analysis in various speech allied applications such as speech recognition, speaker recognition, speech synthesis and speech coding. Therefore, we tried to extract some prominent characteristics from the speech to satisfy the performance of our proposed system.

D. Mel frequency cepstral coefficients(MFCC)

MFCC is one in every one of the foremost popular featureextraction techniques employed in automatic speech or speaker recognition systems using the Mel scale which is based on the human ear scale[11]. it's supported the nonlinear human perception of the frequency of sounds. These coefficients represent audio supported perception. they're derived from the Mel frequency cepstrum.[14] The spectral information can after that be converted to MFCC by passing the signals through band pass filters where higher frequencies are artificially boosted, so applying an inverse Fast Fourier Transform (FFT) thereon . It combines the benefits of the cepstrum analysis with a perceptual frequency scale supported critical bands.[17] As a result, theupper frequencies are becoming more prominent. Since the Mel frequency cepstrum can represent a listener's response system clearly, therefore MFCC is usually considered to be the simplest available approximation of the human ear.

.

E. K-Nearest Neighbor(KNN)

K-Nearest Neighbor is one in every of the best Machine Learning algorithms supported Supervised Learning technique.The K-NN algorithm assumes the similarity between the new case/data and available cases and puts the new case into the category that's most kind of like the available categories.K-NN algorithm stores all the available data and classifies a brand new information supported the similarity. This implies when new data appears then it is easily classified into a well suited category by using K- NN algorithm. The K-NN algorithm will be used for Regression also as for Classification but mostly it's used for the Classification problems. K-NN could be a non-parametric algorithm, which implies it doesn't make any assumption on underlying data. It is additionally called a lazy learner algorithm because it doesn't learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

After the cepstral coefficients are extracted from the audio sample, they are stored in an array and the mean of it is calculated. When the model is applied on the data, according to the value of k, the classification category is decided and is given as output.*Naive Bayes*

The Naive Bayes algorithm could be a supervised learning algorithm, which is predicated on Bayes theorem and used for solving classification problems. It's mainly employed in text classification that has a high-dimensional training dataset. Naive Bayes Classifier is one amongst the easy and best Classification algorithms which helps in building the fast machine learning models which will make quick predictions.It is a probabilistic classifier, which implies it predicts on the idea of the probability of an object. Some popular samples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

ISSN: 1832-5505
Vol-10 Issue-01 Feb 2022

F. Support Vector Machine

Support Vector Machine or SVM is one amongst the foremost popular Supervised Learning algorithms, which is employed for Classification additionally as Regression problems. However, primarily, it's used for Classification problems in Machine Learning.The goal of the SVM algorithm is to form the most effective line or decision boundary which will segregate n-dimensional space into classes so we are able to easily put the new information within the correct category within the future. This best decision boundary is termed a hyperplane. SVM chooses theintense points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed as Support Vector Machine.

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 ofthe 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.

B. Comparison among KNN, Naive Bayes and

```
[[-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

C.

D. 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.

E. K value in KNN process

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

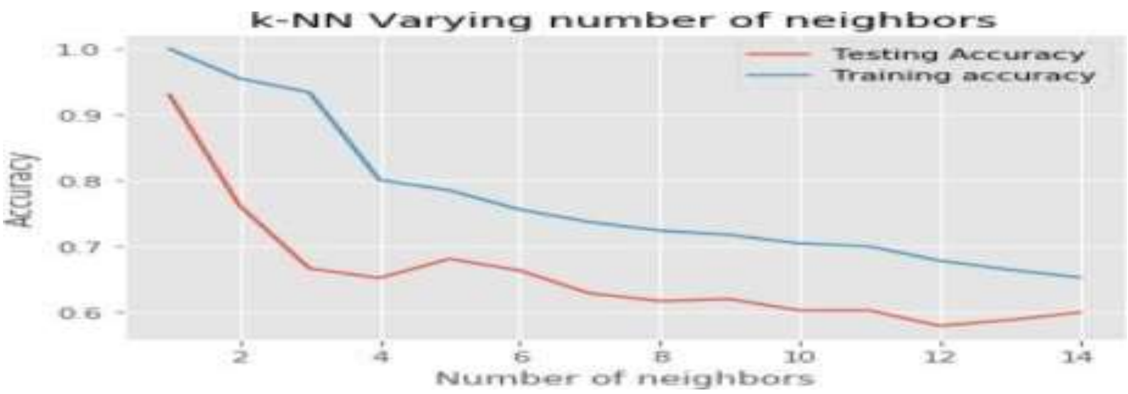

Fig.3. Accuracies for different of K

CONCLUSION

Using eight previously defined categorization categories, the primary objective is to identify the source of a scream given an audio sample as input.

ISSN: 1832-5505
Vol-10 Issue-01 Feb 2022

The audio files come from a variety of sources and were captured in loud locations. We gained a better understanding of feature extraction preprocessing and how to interpret baby cries as a result of this study, which should help with newborn care. At k=2, we achieved 76% accuracy with the KNN classifier, which correctly classifies most of the time and provides the right explanation. To implement, Python versions higher than 3.6 were used.The suggested approach can identify the source of a cry given an audio recording.The same principle may be used in areas such as remote baby control and medical applications to determine the cause of a baby's cries and provide appropriate treatment.

REFERENCES

[1] "Automatic methods for infant cry classification" presented at the 2016 International Conference on Communications (COMM) in Bucharest by Bănică, H. Cucu, A. Buzo, D. Burileanu, and C. Burileanu, pages 51–54. According to [2], Lina Abou-Abbas, Chakib Tadj, and Hesam Alaie Fersaie. An article titled "A fully automated approach for baby cry signal segmentation and boundary detection of expiratory and inspiratory episodes" was published in the Journal of the Acoustical Society of America in 2017 and can be found on pages 1318–1331.

In the 2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE) in Halifax, Canada, Bou-Abbas, Alaie, and Tadj presented a paper titled "Segmentation of voiced newborns' cry sounds using wavelet packet based features." The paper can be found on pages 796 to 800.

[4] Jonas Beskow, Kåre Sjölander, and themselves. An open-source voice tool called Wave-surfer was presented at the Sixth International Conference on Spoken Language Processing in the year 2000.

"Infant Cry Analysis and Detection" by R. Cohen, 2012, pp.2-6.

[6] "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, S. Sharma, P. R. Myakala, R. Nalumachu, S. V. Gangashetty, and V. K. Mittal.

In their 2016 paper "Application Development for Recognising Type of Infant's Cry Sound," Limantoro, Fatihah, and Yuhana (pp.157-161) discuss this topic.

ACOUSTIC ANALYSIS OF BABY CRY, no. May 2013, by R. P. Balandong.

The article "An Automatic Infant's Cry Detection Using Linear Frequency Cepstrum Coefficients (LFCC)" was published in 2014 and can be found in volume 5, issue 12, pages 1379–1383.

[10] "Testing the Universal Baby Language Hypothesis - Automatic Infant Speech Recognition with CNNs," 41st International Conference on Telecommunications and Signal Processing (TSP 2018), pp. 1-4, 2018, E. Franti, I. Ispas, and M. Dascalu.

Dewi, S. P., Prasasti, A. L., and Irawan, B. (2019) in their publication. Analysing Infant Cry Utilising MFCC and LFCC in Various Classification Approaches: A Study. Pages 19–24 of the 2019 IEEE ICSigSys International Conference on Signals and Systems. Publisher: IEEE Bandung.

[12] From the Journal of the Acoustical Society of America, "Audio signal processing device," 2006, vol. 120, no. 6, p. 3445.

The article "Pitch Analysis of Infant Crying" was published in 2013 in the International Journal of Digital Content Technology and Its Applications and was written by G. L. -, Y. H. -, L. Y. -, and M. N. -.

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

Referenced in [15] "Spoken word recognition using mfcc and learning vector quantization" by E. C. Djamal, N. Nurhamidah, and R. Ilyas in 2017's International Conference on Electrical Engineering and Computer Science and Informatics, volume 4, issue 9, pages 250–255.

An article titled "Perbandingan Sistem Perhitungan Suara Tepuk Tangan dengan Metode Berbasis Frekuensi dan Metode Berbasis Amplitudo" was published in 2013 in the Journal of Computer Science in Agri-Informatika and can be found on pages 29–37.

The article "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition" was published in 2013 by the International Journal of Advanced Research in Engineering Technology.

The article "Infant Cry Detection System with

ISSN: 1832-5505
Vol-10 Issue-01 Feb 2022

Automatic Soothing and Video Monitoring Functions" was published in 2017 in the Journal of Engineering Technology Open University of Sri Lanka and was written by G. V. I. S. Silva and D. S. Wickramasinghe.
For example, in their 2016 paper "System propose for Be acquainted with newborn cry emotion using linear frequency cepstral coefficient," Jagtap, Kadbe, and Arotale discuss the use of a linear frequency cepstral coefficient in a system for identifying the emotional state of a baby's cries.