International Journal for Modern Trends in Science and Technology, 9(02): 184-187, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0902033
Available online at: http://www.ijmtst.com/vol9issue02.html

Emotion Recognition Through Speech

Mr. B.Dasaradha Ram [1] | A.Lahari [2] | G.Sasi Priyanka [2] | B.Sudarshan [2] | B.Leela Madhava Rao [2]

[1]Associate Professor, Department of CSE, NRI Institute of Technology, India
[2]B.Tech Student, Department of CSE, NRI Institute of Technology, India

To Cite this Article
Mr. B.Dasaradha Ram, A.Lahari, G.Sasi Priyanka, B.Sudarshan, B.Leela Madhava Rao. Research on Emotion Recognition Through Speech, International Journal for Modern Trends in Science and Technology 2023, 9(02), pp. 184-187. https://doi.org/10.46501/IJMTST0902033

Article Info
Received: 02 January 2022; Accepted: 04 February 2023; Published: 14 February 2023

ABSTRACT

One of the most talked-about areas of human-computer interaction (HCI) is emotion identification in spoken language. The field of emotion recognition from spoken language has seen a surge in recent years, with several academics actively working to construct such systems. This is done in an effort to create more human-like computer systems and improve HCI and the user interface. You may categorize characteristics as either elicited, prosodic, or spectral. Emotions in human speech are classified using several approaches, such as the Hidden Markov Model (HMM), the Gaussian Mixtures Model (GMM), the Support Vector Machine (SVM), the Artificial Neural Network (ANN), and the K-nearest neighbor method (KNN) An evaluation of several speech emotion recognition (SER) systems is provided in this section. The many means by which we might communicate our emotions are discussed, along with their theoretical underpinnings and classifications. In order to complete this research, a SER system is built that makes use of a number of distinct classifiers and feature extraction strategies. Different classifiers are trained based on speech signals' derived mel-frequency cepstral coefficients (MFCC) and modulation spectral (MS) properties. To identify the most important subset of features, feature selection (FS) was used. The problem of emotion categorization was applied to a number of different machine learning models. To begin, seven feelings are placed into distinct categories using a recurrent neural network (RNN) classifier. Later, their results are compared to those of multivariate linear regression(MLR) and support vector machines (SVM), two popular methods in the area of emotion identification for spoken audio data. The databases in Berlin and Spain serve as the experimental set.

KEY WORDS: Feature-driven, Speech emotion recognition, Support vector machine (SVM), Hierarchical Classifier, Mean of the Log-Spectrum (MLS)

1. INTRODUCTION

A significant function served by human beings is that of their emotions. The amount of interaction that takes place between humans and computers is growing at an alarming rate. It is necessary for them to communicate with one another in a natural way in order to make this interaction more productive and successful. The goal of the computer should be to reply depending on the user's feelings as they are detected by the computer. In order to satisfy these requirements, the computer must be able to deduce our feelings from either our vocal or visual expressions. Speech is a very important component of human contact. Within the scope of this study, we endeavored to analyze a person's emotional state by analyzing their speech. The engagement may be made more efficient by identifying the speaker's emotions from their words. When we incorporate these kinds of models into it, it will create advances in the

systems that recognize invoices. When we test this model using the participants' native language, we should see an increase in the dataset's overall effectiveness. The ability to read people's emotions via their speech may be beneficial in a variety of ways, all of which are outlined in this article.

Several distinct feature extraction strategies are taken into account within the framework of the emotion identification from speech system, and SVM classification and MLP classifiers are constructed for the purpose of achieving a higher level of precision. This classifier serves as the basis for the subsequent machine learning model. By analyzing its past taught and test sets, it is able to forecast human emotions. The Ryerson Audio-Visual Database of Emotional Speech and Song, often known as RAVDESS, is the source of the data that was utilized to train the model. This database comprises audio recordings representing a variety of emotions.

2. LITERATURE SURVEY

The full review of speech emotion recognition can be found in [9], which discusses the parameters of the dataset as well as the speech emotion recognition research classifier choice. [10] investigates a number of different acoustic characteristics of speech and examines a number of different classifier algorithms, both of which are important in the exploration of more contemporary approaches to emotion identification.

In this study [11], the authors studied the possibility of deducing future responses from emotional voice signals by using a variety of classifiers and basing their analysis on the detection of a subject's current emotional state. In [11], classification techniques like as K-NN and Random Forest are used in order to appropriately categorize the user's emotional state. The field of data science has seen an explosion in the development of recurrent neural networks, which aim to answer a wide variety of issues. Deep recurrent neural networks (RNNs) like as LSTM and Bi-directional LSTM that have been trained on auditory characteristics are used in [12]. A wide variety of CNNs are now being used and educated for the purpose of speech emotion identification, as seen in [13]. The use of filter banks and deep convolutional neural networks (Deep CNN) [14] to infer emotion from voice inputs reveals a good accuracy rate, which leads one to believe that deep learning may also be utilized for emotion detection. Image spectrograms combined with deep convolutional networks, as described in [15], may likewise be used to do speech emotion identification.

The many different types of classifiers that were used in order to categorize the numerous speech alternatives. There are several different classifiers that are used, including the Gaussian Matrix Model, the Hidden Markov Model, and the Support Vector Machine. B. Yang and M. Lugar presented an artwork wherever the feeling elicited by the music selections was found.

It was in every way consistent with the principles of music theory. It required two separate pitch intervals to complete. It required two pitch intervals that were completely distinct from one another. the occurrences that square measure the explanation behind a consonant or dissonant perception were identified as a result. They are able to analyze these harmonic choices in a way that is far more realistic. Yashpal performs songs by Chavan, M. L. Dhore, and Pallavi, Yes, I am aware that the speech alternatives have been projected, such as Mel Frequency ceptrum Coefficients and Mel Energy Spectrum Dynamic Coefficients.

They used the utterances of speech that somebody's voice made as an input, and then they retrieved a variety of possibilities from those utterances. The "support vector machine," often known as the SVM, was the kind of classifier that was used for the purpose of categorizing feelings. The LIBSVM was used in order to classify different feelings. J. Sirisha Devi, Y. Srinivas, and Shiva Prasad Nandyala developed text-dependent speaker identification with a sweetening of probing the emotion of the speaker before using the hybrid FFBN and GMM methods. This was accomplished.

Before a speech sample [2, 4, 6, 9] can enter the processing phase, it must first be run through a gender reference database that is kept specifically for the purpose of gender recognition. The statistical method [5] is the one that is used, with pitch being the attribute that is used for gender identification [9]. Using the reference database, it was possible to determine both a lower and an upper limit pitch for the male and female samples. [14] After then, the input human speech sample was segmented into frames, with each frame lasting 16 milliseconds. This was done so that the frame level may be classified in the next phases.

The major characteristic that was used for recognizing emotions was the MFCC (Mel Frequency Cepstral Coefficient), which was determined for each frame. The

MFCCs of different emotions, such as happy, sad, angry, and neutral, are kept in a reference database [14], which is also maintained.

The MFCCs of the frames were compared with the MFCCs that were kept in the reference database, and the distance between the frames that were similar was determined. One is able to categorize the analysis frame as either angry, cheerful, or normal depending on the distance between the analysis frame and the reference database. The results are shown using the frame count for various emotional states.

3.PROPOSED SYSTEM

The creation of computers that can communicate with people via the interpretation of speech is a step in the right direction toward the creation of systems that are meant to have intelligence comparable to that of humans. The area of automated voice recognition within the realm of artificial intelligence has been actively engaged in the development of robots that interact with human beings via speech. Understanding spoken language is one of the most complicated processes that the human brain is capable of, despite the fact that it is the most prevalent and efficient method of human communication. Numerous pieces of information, including gender, words, dialect, emotional state, and age, may be extracted from a voice signal and used for a variety of purposes. Inspeech processing, one of the mostarduous tasks for the researchers is speech emotion recognition. When exploring human-computer recognition, SER is the most interesting to look into. It indicates that the system should be able to comprehend the feelings of the user, since these feelings will determine how the system should behave. A well-developed framework that incorporates all of these modules is required to carry out a variety of activities, including the translation of speech to text, the extraction of features, the selection of features, and the categorization of those characteristics in order to determine emotions. The process of classifying characteristics is still another tough aspect of the work, and it requires the training of a variety of emotional models in order to execute the classification in an effective manner.

Now we will discuss the second component of emotional speech recognition, which is the database that is used for the training of models. It entails picking out just the characteristics that are most significant in order to express the feelings in an appropriate manner. When all of these modules are combined in the right manner, we are provided with an application that is able to identify a user's emotions and then pass that information on to the system so that it can react accordingly.

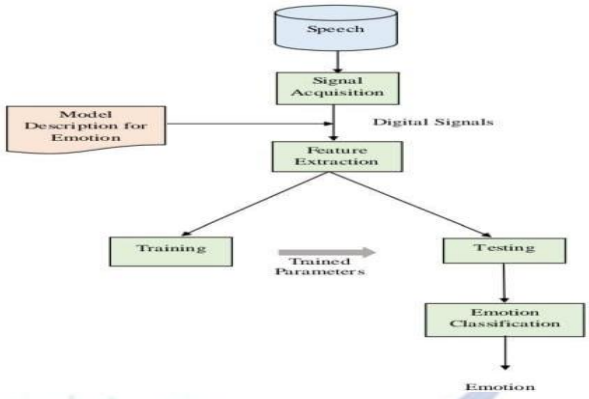

Figure1: Block Diagram for Proposed System

4. RESULTS

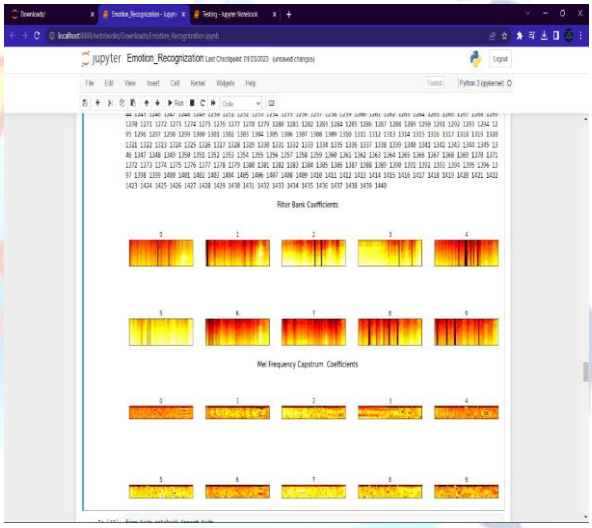

Figure 2: Frequency Coefficients

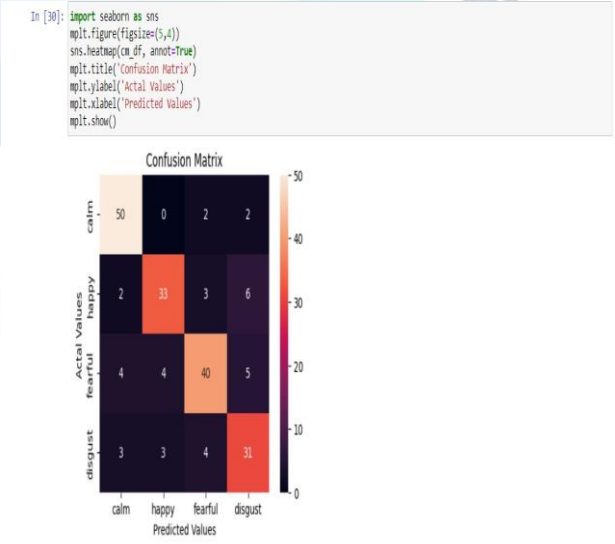

Figure 3: Confusion Matrix for Actual and Predicted Values

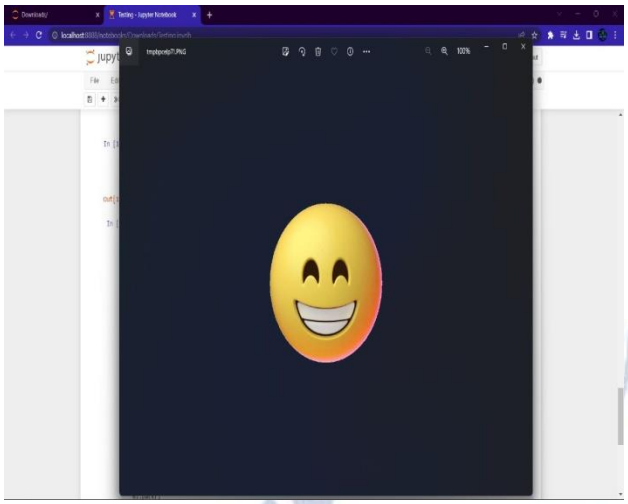

Figure 4: Sample of Human Emotions

5.CONCLUSIONS:

In this paper, we proposed a Speech Emotion Recognition System (SER) that utilizes the Machine Learning Algorithm "MLP" classification approach to categorize a variety of emotions. Therefore, the RAVDESS dataset is parsed in order to get three distinct characteristics referred to respectively as "MFCC," "Chroma," and "Mel." We demonstrated how classifiers and features are used in order to decipher emotional states based on voice inputs. The machine learning models were taught and tested on their ability to distinguish the emotions conveyed by the voice signals based on the extracted characteristics. Through the use of the MLP classifier, the Speech Emotion Recognition System was able to accomplish an accuracy of 80%. A subset of characteristics that are very good at differentiating between people is chosen. In machine learning applications, having an excessive amount of information is not always advantageous, as shown by feature selection algorithms. The machine learning models were taught to distinguish emotional states based on these traits, and then they were tested on their ability to do so. There is still room for improvement in the robustness of emotion identification systems via the combination of datasets and the fusion of classifiers. By combining the results of several emotion detectors into a single detection system, it is possible to evaluate the impact of training multiple emotion detectors. Because the quality of the feature selection has an effect on the emotion recognition rate (a good emotion feature selection technique can identify characteristics representing emotion state fast), one of our goals is to additionally make use of other feature selection methods.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] R. A. S. Nilofer, R. P. Gadhe, R. Deshmukh and P. V. B. Wasghmare, "Automatic emotion recognition from speech signals," International Journal of Scientific and Engineering Research, vol. 6, no. 4, p. 4, April 2015.

[2] A. Rawat and P. K. Mishra, "Emotion Recognition through Speech Using Neural Network," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, no. 5, May 2015.

[3] S. B. R, N. A and R. Desai, "Speech Emotion Recognition using MLP Classifier, "International Journal of Engineering Science and Computing, vol. 10, no. 5, May 2020.

[4] Chavhan, Y., Dhore, M. L., &Yesaware, P. (2010). Speech Emotion Recognition Using Support Vector Machine. International Journal of Computer Applications

[5] C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a hierarchical binary decision tree approach,” Speech Commun., vol. 53, no. 9–10, pp. 1162–1171, Nov. 2011.

[6] S. S. Narayanan, “Toward detecting emotions in spoken dialogs,” IEEE Trans. Speech Audio Process., vol. 13, no. 2, pp. 293–303, Mar. 2005.

[7] J.-H. Yeh, T.-L. Pao, C.-Y. Lin, Y.-W. Tsai, and Y.-T. Chen, "segment-based emotion recognition from continuous Mandarin Chinese speech," Comput. Human Behav., vol. 27, no. 5, pp. 1545–1552, Sep. 2011.

[8] Devi, J. S., Srinivas, Y., &Nandyala, S. P. (2014). Automatic Speech Emotion and Speaker Recognition based on Hybrid GMM and FFBNN. International Journal on Computational Sciences & Applications (IJCSA).

[9] Tin Lay, S Foo, and De Silva, “Speech emotion recognition using hidden Markov models,” 2003, November.

[10] Pazhanirajan, S., & P. Dhanalakshmi, P. (2013). EEG “Signal Classification using Linear Predictive Cepstral Coefficient Features. International Journal of Computer Applications”., November

[11] M. Choubassi, H. Khoury, C. Jabra Alagha, J. Skaf and M. AlAlaoui "Arabic Speech Recognition Using Recurrent Neural Networks" 2004.

[12] Rawat, A., & Mishra, P. K. (2015). Emotion Recognition through Speech Using Neural Network. International Journal of Advanced Research in Computer Science and Software Engineering.