Speech Emotion Recognition

[1] Athira S [2] Dr. T K Bijimol

[1] PG Scholar of Amal Jyothi College of Engineering, Kanjirapally, Kerala

[2] Assistant Professor of Amal Jyothi College of Engineering, Kanjirapally, Kerala

[1]athirasjaimol@gmail.com [2] tkbijimol@amaljyothi.ac.in

***Abstract*–The in-depth learning based on the focus on speech recognition and natural language processing has become very popular. With the focus system, the relevant encoding context vectors make up the major chunk ofthe decoding content construction, simultaneously minimizing the effect of the irrelevant ones. Driven by this idea, a methodology is proposed in this work for the active selection of sub pronounced representations to construct discriminative pronunciation representations. In comparison with the basic standard of a model based on unified focus, a focused model improves weighted accuracy by 1.46% in the emotion classification work. Furthermore, the selection distribution has evolved to a better understanding of the sub-pronunciation piece of an emotional value by users.**

***Keywords*: Attention mechanism, speech emotion recognition, recognizing speech emotion**

I. INTRODUCTION

Emotions play an important role in our daily lives for effective communication. These are often encoded and transmitted through various forms of human behavioral signals, including speech, facial expressions, and body language. Despite the progress on the field, this is a challenging mission with considerable room for improvement here.

Here, we will use libraries librosa, sound file, and sklearn to build a model using an MLP classifier. It can recognize emotions from sound files. We will load the data, extract the features from it, and then divide the dataset into training and testing sets. Next, we will test an MLP classifier, train the model and calculate the accuracy.

II. LITERATURE REVIEW

[1] In this paper, we argue that singing is more emotional than speech. We evaluate classifiers on different feature sets,feature types, song, and speech emotion recognition. Three feature sets: GeMAPS,pyAudioAnalysis, librosa; Two feature types, low-level descriptors and high-level statistical functions.When using the same method, the results show no significant difference between the song and the speech data. Comparisons of the two results reveal that the song is more emotional than speech.

[2] In this paper, we propose a system that will analyze the speech signals and gather the emotion from the same efficient solution based on combinations. This system solely served to identify emotions present in the signal or speech using concepts of deep learning and algorithms of machine learning (ML). Using the above mentioned, the system will determine the eight emotions present in the speech signal; anger, sadness, joy, neutrality, calmness, fear, hatred, surprise.

[3] In this study, we developed a framework that integrates three distinctive classifiers: an in-dept neural network (DNN), a convolution neural network (CNN), and a recurrent neural network (RNN). The framework was used to differentiate between the four distinct emotions (i.e., angry, joy, neutrality and sadness). specific-level outputs of frame-level low-level descriptors (LLDs), segment-level mel-spectrograms (MS), and high-level statistical functions (HSFs) on LLDs are transmitted exclusively RNN, CNN, and DNN.

III. METHODOLOGY

In this prediction system, we use Jupyter notebook that makes it easy to set up, access, and share. It helps visualize maps, charts using third-party Python packages. It helps create models of machine learning algorithms and solve problem.

Figure 1: Jupyter Notebook

DOI: 10.5281/zenodo.6369556
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

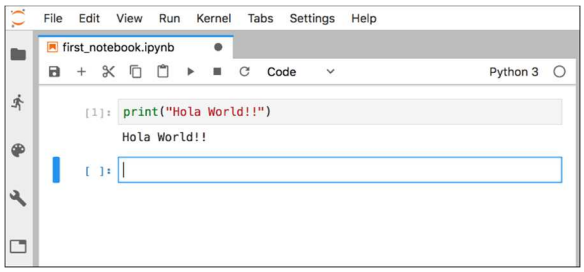


Figure 2: Notebook

(a) DATASET

It is collection of data that will be used for analyzing emotions through the recognition of speech from different users. The dataset is an audio format that shows the dataset of the prediction of the emotions by speech. The is from Kaggle website and Data Flair Training.

IV. IMPLEMENTATION

Firstly, open your Jupyter notebook through your command prompt console and then rename it to the purpose of the work.

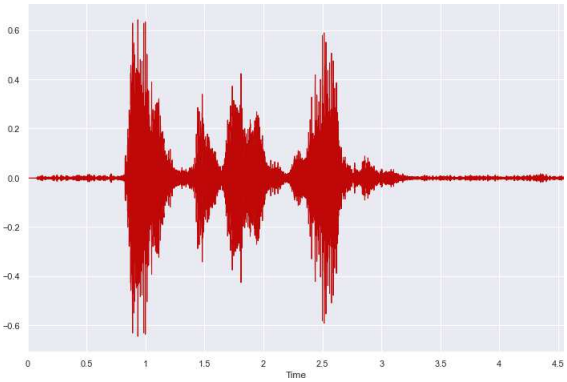


Figure 3: Anger

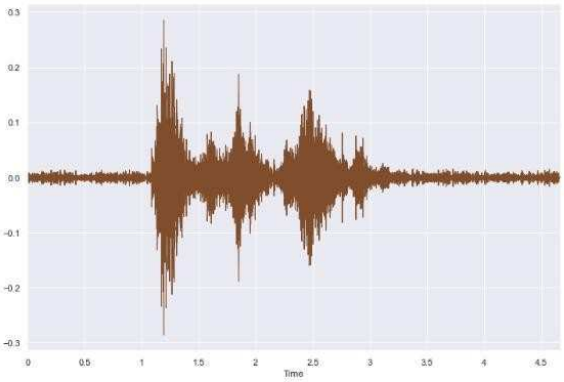


Figure 4: Disgust

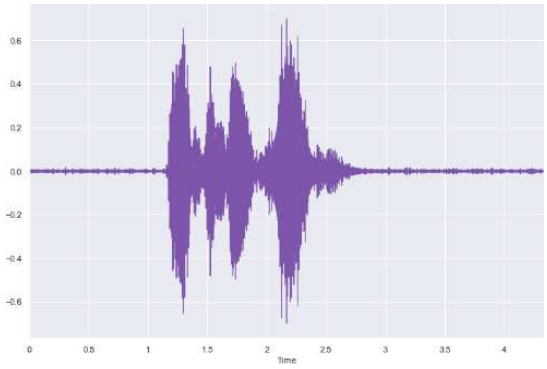


Figure 5: Fear

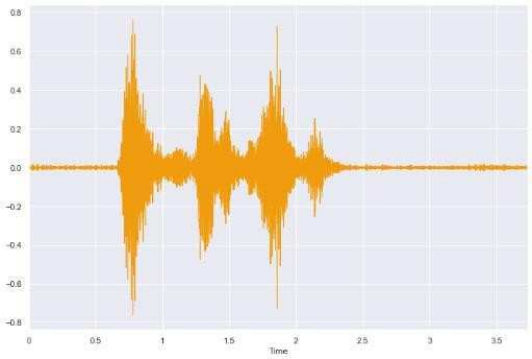


Figure 6: Happiness

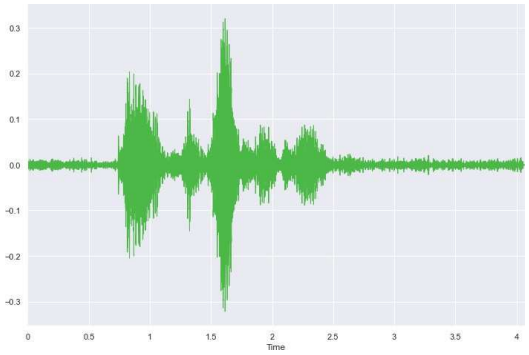


Figure 7: Neutral

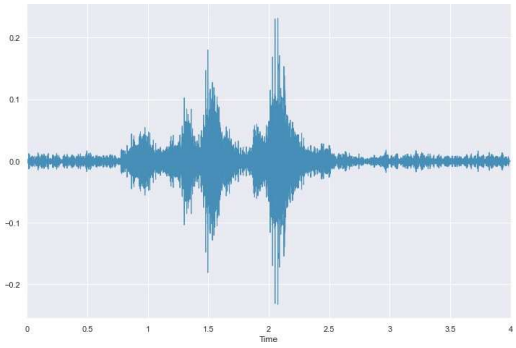
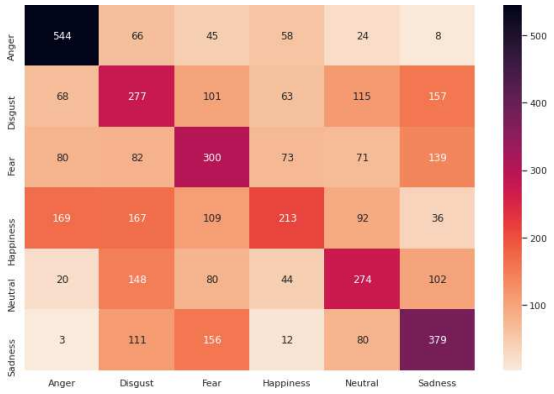


Figure 8: Sadness

DOI: 10.5281/zenodo.6369556
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

V. RESULTS

Through the above illustration, we now know that speech recognized via the procedure is characterized into 6 emotions namely, Anger, Disappointment, Fear, Happiness, Neutral and Sadness. We learned to recognize emotions from speech. We used an MLP Classifier for this and made use of the sound file library to read the sound file, and the librosa library to extract features from it and the model delivered an accuracy of 72.4%.



VI. CONCLUSION

In this work, despite the fact that only a suitable amount of data is used to be used through such a model, we took the help of a way to reduce the over-fitting. The initial experiment results shows that the such a mechanism-based system outstands a system without the same. Moreover, we also found out that this selection distribution is not just associated to the frame energy curve but has a scoring more different than audio property evolution associated to feelings.

VII. FUTURE ENHANCEMENTS

Through this project, we showed how to improve Machine learning to achieve the underlying emotion from the audio data and some clarity on the human emotion expression through sound. It can be used in a variety of settings such as Call Centre for complaints or marketing, in voice-based virtual assistants or chatbots, in language research. The proper implementation of the speed can be investigated to see if some of the shortcomings of the model can be fixed. Adds more data volume through other augmentation techniques such as time-shifting or speeding up/slowing down the audio or simply finding more such audio clips.

VIII. REFERENCES

[1] S. S. Narayanan and P. Georgiou, “Behavioral signal processing: Deriving human behavioral informatics from speech and language,” Proceedings of IEEE, vol. 101, no. 5, pp. 1203 – 1233, May 2013

[2] C. M. Lee and S. S. Narayanan, “Toward detecting emotions in spoken dialogs,” IEEE Transactions on Speech and Audio Processing, vol. 13, no. 2, pp. 293–
303 [IEEE Signal Processing Society Best Paper Award 2009], Mar. 2005.

[3] M. Grimm, K. Kroschel, E. Mower, and S. S. Narayanan, “Primitives-based evaluation and estimations of emotions in speech,” Speech Communication, vol. 49, no. 10-11, pp. 787–800, Nov. 2007.

[4] T.L. Nwe, S.W. Foo and L.C. De Silva, "Speech emotion recognition using hidden Markov models", Speech Communication, vol. 41, no. 4, pp. 603-623.

[5] M. El Ayadi, M.S. Kamel and F Karray, "Survey on speech emotion recognition: Features classification schemes and databases", Pattern Recognition, vol. 44, no. 3, pp. 572-587, 2011.

DOI: 10.5281/zenodo.6369556
ISBN: 978-93-5607-317-3@2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam