Speech Emotion Recognition Using Deep Learning

Jakir Hasan - 2018331057

30 December 2022

1.1 Abstract

In this study, we have presented a deep learning-based implementation for speech emotion recognition (SER). The system uses a Multilevel Perceptron Classifier. The proposed model has been applied to SUBESCO dataset. The experimental results reveal that the model with MLPClassifier achieves better performance. The proposed model has attained a weighted accuracy (WAs) of 77% for the SUBESCO dataset.

1.2 Index Term

SER, MLPClassifier, SUBESCO

1.3 Introduction

Though a lot of work has been done in the area of Speech Emotion Recognition (SER) , there is still a lot of scope to improve. Researchers have developed many tools and techniques to classify emotions from speech signals. These tools and techniques are mainly developed for languages such as English, German, French etc. But there is not much progress for low resource languages such as Bangla.

Keeping these in mind we have implemented a deep learning based Multi Level Perceptron Classifier (MLPClassifier) to classify human emotions from Bengali audio speech signals. We used the SUBESCO dataset developed by SUST.

We also developed a web application where users can provide audio files or record audio and can show predicted emotion.

1.4 Problem Definition

Speech signals contain a lot of valuable information in it. Extracting the right features from speech signals is tricky and difficult. For predicting emotions we have to extract features which give high quality information. So, the main problem is to extract desired features from audio and develop a model as well as a system which will take the audio as input and provide predicted emotion as output.

1.5 Background/Related Work

Though there are a lot of work done in the field of SER for other languages and not much done for Bangla language. In 2018 Rahman proposed a Dynamic time warping assisted SVM emotion classifier for Bangla words. [3] In 2017, Badshah proposed a CNN architecture consisting of convolutional layers. [1] Satt presented an SER model which calculates log-spectrograms as feature vectors. [4] Chen used deltas and delta deltas of log mel-spectrogram for emotion identification [2]

1.6 Methodology

For input features we extracted mel-frequency spectrograms from audio as they provide highly valuable information. We used Python's Librosa module to extract this feature using built in libraries.

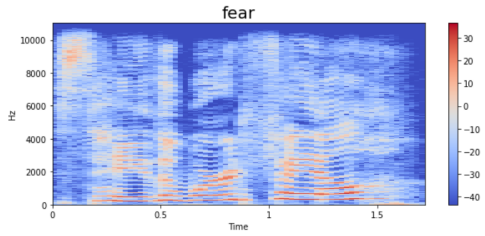

Figure 1.1: Spectrogram

We give the audio file path as input and get a feature as a numpy array.

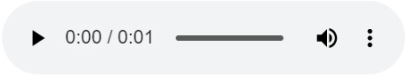

Figure 1.2: Audio

Then these features are feeded into the deep learning based MLPClassifier for training.
The model learns from audio features. After training, when new audio is provided to the model it successfully predicts its emotion.

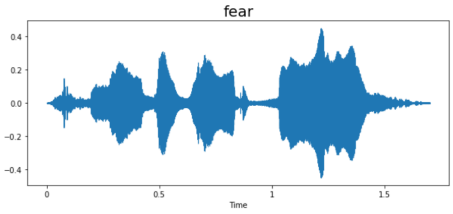

Figure 1.3: Wave Plot

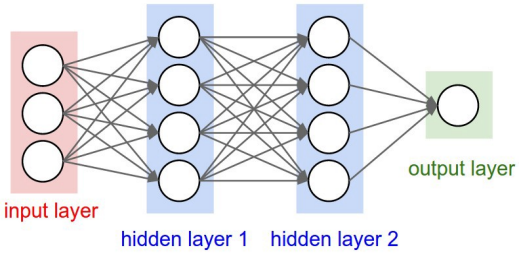

Figure 1.4: Deep Learning Model

1.7 Result/Analysis/Accuracy

Our proposed model attained an overall accuracy of 77%.

We have also developed a web application for our project. Where user can select as well as record audio files and after giving that file to the system it will predict appropriate emotion.

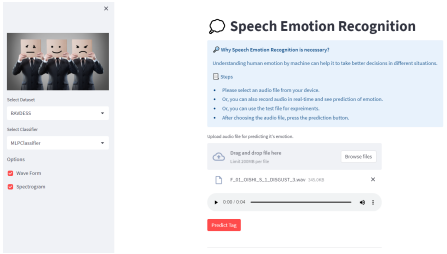

Figure 1.5: Web Application 1

Here is the Web Application for selecting and recording audio for prediction of emotion.

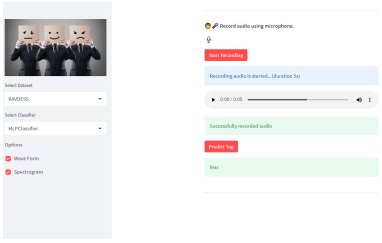

Figure 1.6: Web Application 2

1.8 Conclusion

Our developed system can be used for different sectors such as call centers, tele-communication, tele-medicine etc. Also our proposed model can be used for Bangla Speech Emotion Recognition (SER). We hope, our work will encourage others to improve Bangla SER techniques and applications.

Bibliography

[1] Abdul Malik Badshah, Jamil Ahmad, Nasir Rahim, and Sung Wook Baik. Speech emotion recognition from spectrograms with deep convolutional neural network. In *2017 international conference on platform technology and service (PlatCon)*, pages 1–5. IEEE, 2017.

[2] Mingyi Chen, Xuanji He, Jing Yang, and Han Zhang. 3-d convolutional recurrent neural networks with attention model for speech emotion recognition. *IEEE Signal Processing Letters*, 25(10):1440–1444, 2018.

[3] Md Masudur Rahman, Debopriya Roy Dipta, and Md Mahbub Hasan. Dynamic time warping assisted svm classifier for bangla speech recognition. In *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, pages 1–6. IEEE, 2018.

[4] Aharon Satt, Shai Rozenberg, Ron Hoory, et al. Efficient emotion recognition from speech using deep learning on spectrograms. In *Interspeech*, pages 1089–1093, 2017.