A NOVEL SENTIMENTS ANALYSIS MODEL USING PERCEPTRON CLASSIFIER

Alex M. Goh and Xiaoyu L. Yann

ABSTRACT:

Speech Emotion Recognition, abbreviated as SER, is used to recognize sentiments of humans and the related affective states from dialog. This is exploiting on the fact that vocal sound often reflects underlying emotion through tone and pitch. Emotion recognition is a rapidly growing research domain in recent years. Unlike humans, machines are genderless and lack the abilities to observe and display sentiments. But human-computer interaction can be improved by implementing automated emotion recognition, thereby reducing the need of human intervention. Here basic emotions like calm, happy, fearful, disgust etc. are analyzed from emotional speech signals. Machine learning techniques like Multilayer Perceptron Classifier (MLP Classifier) which is used to categorize the given data into respective groups which are non-linearly separated. Mel-frequency cestrum coefficients (MFCC), chroma and Mel features are extracted from the speech signals and used to train the MLP classifier.

Index Terms*: Speech Emotion Recognition, Sentiment Analysis, Machine Learning, Artificial Intelligence, MLP, MFCC.*

Reference to this paper should be made as follows:

Alex M. Goh and Xiaoyu L. Yann, (2021), "A Novel Sentiments Analysis Model Using Perceptron Classifier" Int. J. of Electronics Engineering and Applications, Vol. 9, No. 4, pp. 01-10, DOI 10.30696/IJEEA.IX.IV.2021.01-10

Biographical notes:

Alex M. Goh *was born in Malaysia, completed his Bachelor and Master degree from Multimedia University, Malaysia and Ph.D. from Curtin University, Malaysia, currently he is working as Data Scientist in Multimedia University, Multimedia University, Australia.*

Xiaoyu L. Yann *was born in China, He completed the bachelor's and master's degree from Hebei Normal University. He is currently pursuing PhD with Curtin University, Malaysia, His research interest includes Artificial Intelligence and power grid operation analysis.*

1. INTRODUCTION

The beauty of nature is that every human-being is able to express his/her emotions and feelings using languages. In natural language processing, sentiment analysis helps in identification and organization of sentiments of people. With the advent of internet and emergence of social media people are becoming more expressive. Sentiment analysis is one of the powerful techniques to allow business firms and their owners to get information about their customers' views. For any business to succeed, customer satisfaction is the first step. The way consumers express their feelings for a particular product or service may become critical for survival as it influences other customers. For entrepreneurs, it becomes a necessity to be aware of their customers' views to take timely and correct decisions. One of the most vital aspects of various activities in 21st century world is data. Nowadays, the civilization is increasing its dependency on digital information. Most of digital information is stored in computers, hence making the society more dependent on computers in everyday life. The user population has also increased manifolds over the period of time and so has the data generated. Recommendation systems are search and decision tools which filter & provide user relevant information. Because of the presence of plethora of information on web, the presence of such filtering tools has become important. The filtering tools help users find interesting items and saves both their time and energy. At present, speech emotion recognition is an emerging crossing field of artificial intelligence and artificial psychology; besides, it is a popular research topic of signal processing and pattern recognition. The research is widely applied in human-computer interaction, interactive teaching, entertainment, security fields, and so on. Speech emotion processing and recognition system is generally composed of three parts, the first being speech signal acquisition, then comes the feature extraction followed by emotion recognition. The most propitious technique for speech recognition is the neural network-based approach. Artificial Neural Networks, (ANN) are biologically inspired tools for information processing. Speech recognition modelling by artificial neural networks (ANN) doesn't require any prior knowledge of speech process and this technique quickly became an attractive substitute to HMM (Hidden Markov Model). RNN can learn the sublunary relationship of Speech – data & is capable of modelling time dependent phonemes. The conventional neural networks of Multi- Layer Perceptron (MLP) type have been gradually in use for speech recognition and also for various other speech processing applications. Speech recognition is the process of converting an sound signal, captured by microphone or a telephone, to a set of characters. They can also serve as the input to further linguistic processing to achieve speech understanding, a subject covered in section. Speech recognition performs tasks that similar with human brain [5].

Now a days many companies want to know the emotions of the customers based on their products, either the customers satisfy with the products or not. For that purpose, finding of emotion through MLP classifier plays a important role.

Traditional emotional feature extraction was based on the analysis and comparison of all kinds of emotion characteristic parameters, selecting all the emotional characteristics with high emotional resolution for the purpose of feature extraction. The traditional approach concentrates on the analysis of the features in the speech like time construction, amplitude construction, and frequency construction, etc. Speech time construction refers to the emotion speech pronunciation differences in time. Different emotions have different types of pronunciation time periods which can be recognized and analyzed by closely examining few datasets. Such variations can also be found in the frequency and amplitude of the parameters of respective audio signals. This method, however is the basic concept of categorizing emotions from speech, it also has many drawbacks like time taken is high, judging criteria may vary, and complex programming is required. There are also many models which were proposed earlier to improve the predicting accuracy of the SERS. For example, we have Support Vector Machine (SVM), which is a classifier that mathematically computes the parameters of the audio signal to be able to predict the emotion. This model has been very successful in the domain of SER. But the main disadvantage with SVM's is that it can only classify the data into two classes; either class 1 or 2. And other disadvantages include processing time, noise leading to errors in prediction and low accuracy [5].

The underlying emotion in speech is reflected in voice through tone and pitch. Here elicit types of emotions such as sad, happy, neutral, angry, disgust, surprised, fearful and calm are considered. In this the emotions in the speech are predicted using neural networks. Multi-Layer Perceptron Classifier (MLP Classifier) is used for the classification of emotions. RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song dataset) is the dataset used.

2. LITERATURE SURVEY

The authors in [6], suggested the EPMRS (emotion aware personalized music recommendation system) for determining relationship between the user data and music. For this they combined the results of two different approaches: DCNN (Deep Convolutional Neural Networks) and WFE (Weighted Feature Extraction). For classification, DCNN technique was employed to extract latent features from music data. WFE approach was used for the development of implicit user rating for music, so that to extract the link between the user data and music data. To train the EPMRS, they employed million songs dataset (MSD). They created Android and iOS apps to collect accurate user experience data on the EPMRS. The anonymous user response also revealed that the EPMRS adequately reflects their musical preferences. The EPMRS used the user-to-song interaction to suggest music to users based on their current mood. Artificial intelligence (AI) is a wide spread range of computer science that creates smart machines capable of executing tasks which usually require human intelligence. A number of computer programs use AI[2]. Market Simulators, economic planners and logic systems are a few domains where AI is used. Artificial intelligence stimulates human intelligence in machines in order to make them smarter and capable of make decisions and thinking like a human where it is used in cases of problem solving, learning and reasoning [3]. Artificial intelligence is classified in 4 types namely: Relative machines, limited memory, theory of mind and self – awareness. Being said Relative machines tends to attend the most basic principles of AI which is using its intelligence to recognize and react to the problem. Limited memory refers to capability of an AI to store antecedent data and predictions, with it, Machine learning become a bit complex. Theory of mind AI refers to the ability to predict actions of oneself and others i.e. we can anticipate how one behaves in a specific circumstance, so basically making a model efficient enough to understand and feel like living beings and emotions that guide our decisions. Self-awareness is the one where machines/systems can portray itself i.e. being aware of themselves as well as their internal states [4].

In this paper the authors have taken voice signals as input to the deep neural network, whereas the voice consists of many features, from that they extracted useful features. These features are used to detect the emotion. They implemented various pre-processing methods from that karas based deep neural network using python libraries is the one in which they got the accuracy of 68.5%. As a result 30.1% accuracy more than the support vector machine(SVM) [1].

Extraction of speech emotion features are the key to speech emotion recognition. Here author use both CNN and RF, in which CNN used as the feature extractor to extract the speech emotion features from the normalized spectrogram and the Random Forest is used to classify the speech emotion features. This result of experiment states that the CNN-RF model gives best accuracy compared with CNN model and also improved the record sound command box of Nao and applied the CNN-RF model to Nao robot. Finally, Nao robot became an intelligent human-computer interaction by figuring out the human psychology through speech emotion recognition and also know about people's happiness, anger, sadness and joy [2].

This paper works on extended research on emotion recognition for Indonesian language. Previous researches for speech emotion recognition on Indonesian language are using Support Vector Machine (SVM), Feed Forward Neural Network (FFNN) and Long Short-Term Memory (LSTM) are experimented to model emotions. The results shows that LSTM outperform SVM and FFNN.LSTM obtains 65.9% for average F1 measure with using acoustic and lexical feature making it 5% higher than the best SVM in this experiment [3].

Clarity and intelligibility in speech signal demands removal of noise and interference associated with the signal at the source. This poses further challenge when the speech signal is colored with human emotions. In this work, the authors have taken a novel step to improve the emotional speech signal adaptively before classification. Most popular adaptive algorithms such as Least mean square (LMS), Normalized least mean squares (NLMS) and Recursive least square (RLS) have been put to test to obtain improved speech emotions. Neural network based Multilayer perceptron (MLP) classifier is used to recognize fear speech emotion as against neutral voices using effective Linear Prediction coefficients (LPCs). The accuracy has improved to approximately 77% with improved signal. The increased accuracy of this signal has been witnessed with the RLS algorithm as against the noisy signal with corresponding algorithm [4].

3. METHODOLOGY

The below Fig 1 is the block diagram of SER using MLP classifier. At first, we have to load the dataset after that normalization can be done, because of lows and highs in the dataset and then extract the features (MFCC, Mel, Chroma) from the dataset. Give this feature to train the MLP Classifier model after that test the model then we will get the accuracy as output.

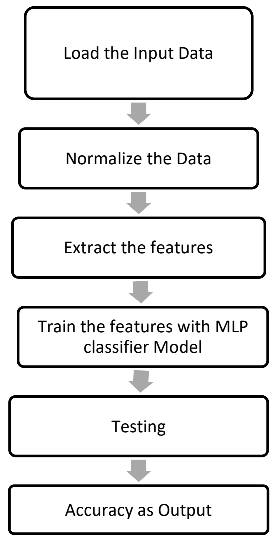

Fig 1: Data flow Diagram of SER Using MLP Classifier

RAVDESS dataset has recordings of 24 actors, 12 male actors and 12 female actors, the actors are numbered from 01 to 24. The male actors are odd numbered and female actors are even numbered. The emotions contained in the dataset are as sad, happy, neutral, angry, disgust, surprised, fearful and calm expressions. The dataset contains all expressions in three formats, which are: Only Audio, Audio-Video and Only Video. Since our focus is on recognize emotions from speech, our model is trained on Audio-only data. Two fixed statements are vocalized by all the 24 actors for all the 8 emotions, with each statement repeated twice. All emotional expressions are uttered at two levels of intensity: normal and strong, except for the 'neutral' emotion, it is produced only in normal intensity. Thus, the portion of the RAVDESS, that we use contains 60 trials for each of the 24 actors, thus making it 1440 files in total. The dataset is labelled in accordance with the decimal encoding. Ever file has a unique filename. The filename is made up of 7-part numerical identifier, the 3rd numerical part of the filename denotes a label to the corresponding emotion. The emotions are labelled as follows: 01-'neutral', 02-'calm', 03-'happy', 04-'sad', 05-'angry', 06- 'fearful', 07-'disgust', 08 -'surprised' [6]. The fig 2 below depicts the emotions in the Dataset.

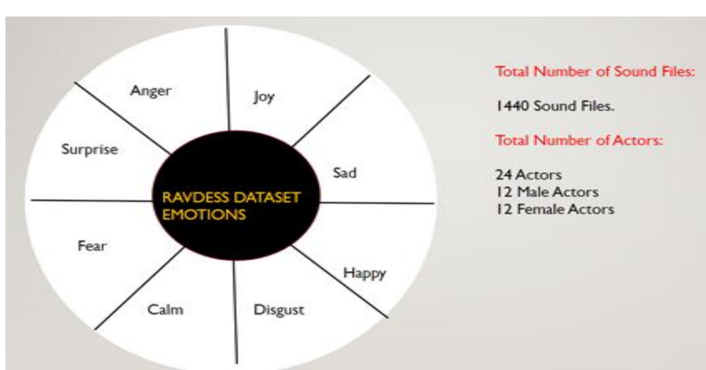

Fig 2: Emotions in the Dataset

A Multi-Layer Perceptron (MLP) is a network made up of perceptron. It has an input layer that receives the input signal, an output layer that makes predictions or decisions for a given input, and the layers present in between the input layer and output layer are called hidden layer. There can be many hidden layers, the number of hidden layers can be changed as per requirement. In the proposed methodology for Speech Emotion Recognition, the Multi-Layer Perceptron Network will have one input layer, of (300,) and (40,80,40) hidden layers and one output layer. The input layer will take as input, the five features, that are extracted from the audio file. The extracted five features being, Mel Frequency Cepstral

Coefficients, Mel Spectrogram Frequency, Chroma, Tonnez and Contrast. The hidden layer uses an activation function to act upon the input data and to process the data. The activation function used is logistic activation function. The output layer brings out the information learned by the network as output. This layer classifies and gives output of the predicted emotion, according to the computation performed by the hidden layer. [6].

4. ALGORITHM

Multi-layer Perceptron Classifier (MLP Classifier) relies on an underlying Neural Network to perform classification. MLP Classifier implements a Multi-Layer Perceptron (MLP) algorithm and trains the Neural Network using Back propagation.

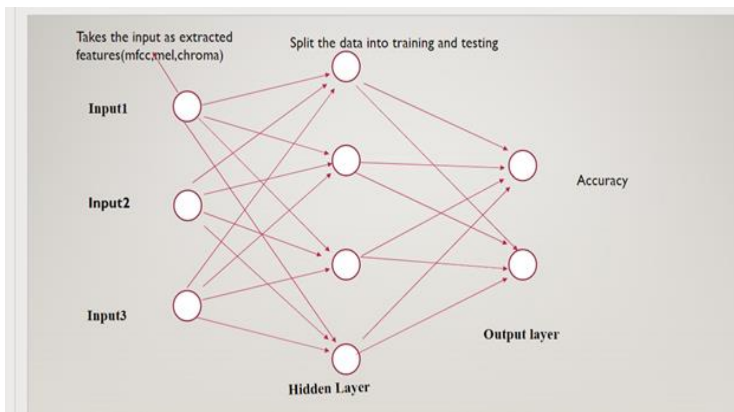

Fig 3: MLP Classifier

Above fig 3 show the MLP classifier. Building the MLP Classifier involves the following steps:

Initialise the MLP Classifier by defining and initiating the required parameters.

• Data is given to the Neural Network to train it.

• The trained network is used to predict the output.

• Calculate the accuracy of the predictions.

5. FEATURE EXTRACTION

Voice frequently reflects hidden feeling through tone and pitch. The objective of feature extraction is to reveal applicable feature from discourse signals as for feelings. Three features are extracted from the discourse signals given as information. The three features are MFCC, Mel, Chroma.

- MFCC: Mel Frequency Cepstral Coefficients (MFCC) is utilized to recover the sound from the given wav audio file by utilizing distinct hop length and HTK-styles Mel frequencies. Pitch of 1 kHz tone and 40 dB over the perceptual discernible edge is characterized as 1000 mels, utilized as reference point. The MFCC gives a Discrete Cosine Change (DCT) of a genuine logarithm of the transient vitality showed on the Mel recurrence scale [6].

- MEL: The Mel scale relates evident repeat, or pitch, of an unadulterated tone to its real assessed recurrence. Individuals are incredibly improved at perceiving little changes in pitch at low frequencies than they are at high frequencies. Solidifying this scale makes our features arrange even more eagerly what individuals listen [6].

- CHROMA: The Chroma Feature is a descriptor, which speaks to the tonal substance of a melodic sound signals in a consolidated form. Therefore, Chroma highlight can be taken as significant essential for elevated level semantic examination, similar to harmony acknowledgment or consonant closeness estimation. A better nature of the extricated chroma include empowers much better outcomes in these elevated level assignments. The chroma is figured by including the log-repeat size range across octaves. The coming about plan of chroma vectors is known as chroma-gram [6].

6. IMPLEMENTATION

Using the MLP classifier to predict emotions from the fed input. We get results using the extracted five features. We send the ranging five features to the model. Using the features independently and passing it altogether we get a great deviation of the prediction emotion, as a single featured parameter is not enough to come up with an efficient prediction. The Ravdess dataset is passed to the MLP Classifier to train the model, we split the dataset into a 75:25 ratio, i.e.; the training and testing dataset. The dataset consists of the audio samples of 24 professional actors with North American accent. Eight types of emotions are covered. The Classifier is being used as it is efficient for time series-based data, in our case the audio that we will be predicting the emotion. The Fig 4 below shows the training process.

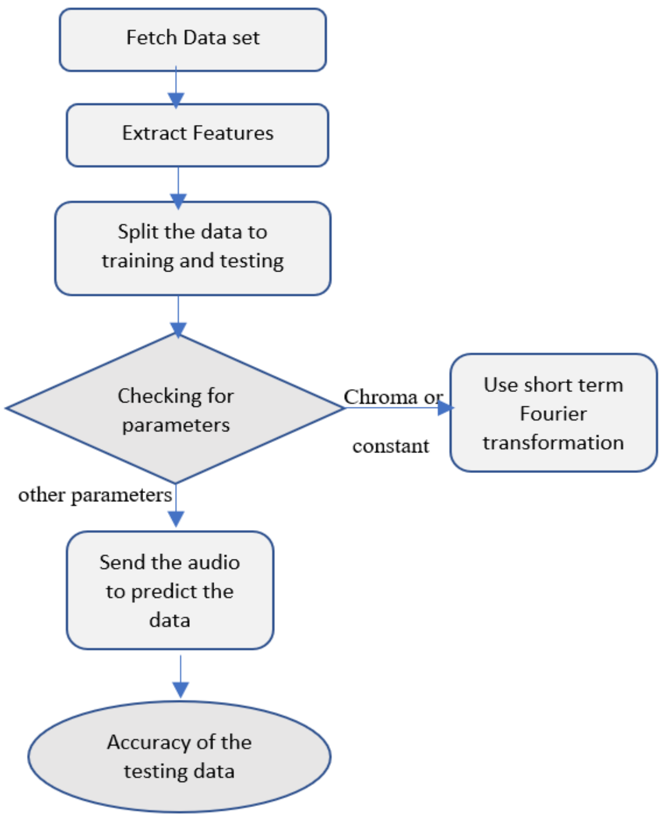

Fig 4: Training process workflow

We then use the rest of the 25% of the remaining data set for test in-order to predict the emotion. Fig 5 shows the workflow of the testing process.

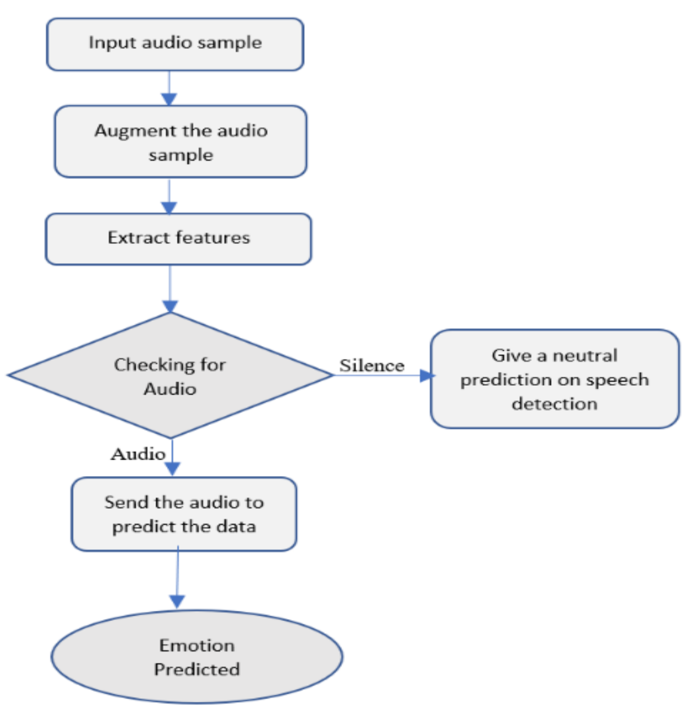

Fig 5: Testing process workflow

The audio is recorded for 15 seconds, by adding a 0.5 seconds gap in the start and end of the audio file in-order to get a good grasp of the audio. The audio will be having a variation in the volume, this will hinder with extraction of the features, to avoid it we normalize to average the volume of the whole sample. The length of the audio will be in a 32-Bit representation, as the number is represented in float format with a ten to the power expression, this has a greater significance as it can represent larger and smaller numbers, thus increasing the range of the audio in terms of DB is from -758 to 770 DB. The MLP-Classifier takes a list of hyper-parameters. The activation function used is the logistic, it is a differential function which helps us to find the slope of a curve at any two points.

7. RESULTS

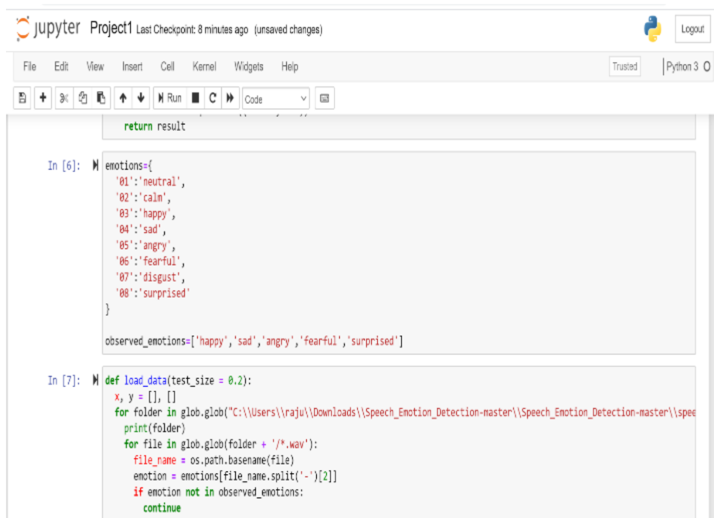

Fig 6: Five Emotions taken for Results

Here as shown in the fig 6 five emotions were taken for prediction. Based on this list we are able to get the accuracy as output.

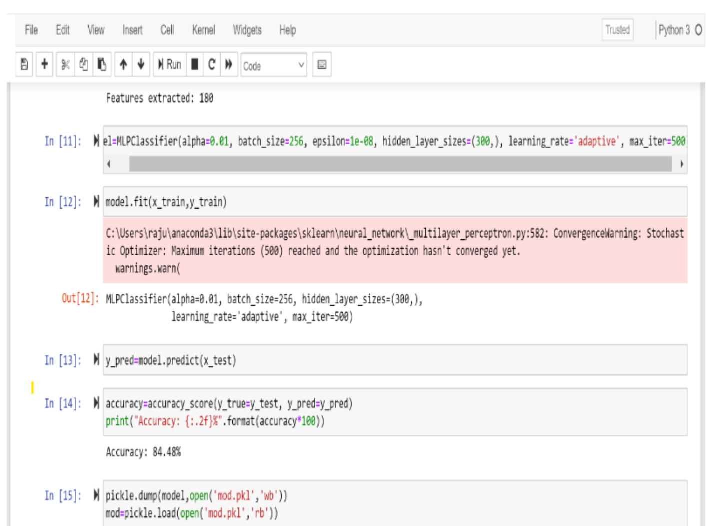

Fig 7: Accuracy Results for five Emotions

As shown in the fig 7, the accuracy obtained is nearly 84% for five emotions (happy, sad, angry, fearful, surprised).

Fig 8 : Six Emotions taken for Results

```
In [13]: y_pred=model.predict(x_test)

In [14]: accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
         print("Accuracy: {:.2f}%".format(accuracy*100))

         Accuracy: 76.62%

In [15]: pickle.dump(model,open('mod.pkl','wb'))
         mod=pickle.load(open('mod.pkl','rb'))
```

Fig 9: Accuracy Results for Six Emotions

As shown in the fig 8 and fig 9,we got the accuracy as 76% for six emotions (‘happy’, ’sad’, ’angry’, ’fearful’, ’disgust’, ’surprised’).
Accuracy decreased from 84% to 76% when we change the number emotions from 5 to 6.

```
In [6]: emotions={
          '01':'neutral',
          '02':'calm',
          '03':'happy',
          '04':'sad',
          '05':'angry',
          '06':'fearful',
          '07':'disgust',
          '08':'surprised'
        }

        observed_emotions=['happy','sad','angry','fearful','disgust','surprised','neutral']

In [7]: def load_data(test_size = 0.2):
          x, y = [], []
          for folder in glob.glob("C:\\Users\\raju\\Downloads\\Speech Emotion Detection-master\\Speech Emotion Detection-master\\spee
```

Fig 10: Seven Emotions taken for Results

```
In [11]: model=MLPClassifier(alpha=0.01, batch_size=256, epsilon=1e-08, hidden_layer_sizes=(300,), learning_rate='adaptive', max_iter=

In [12]: model.fit(x_train,y_train)

Out[12]: MLPClassifier(alpha=0.01, batch_size=256, hidden_layer_sizes=(300,),
                       learning_rate='adaptive', max_iter=500)

In [13]: y_pred=model.predict(x_test)

In [14]: accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
         print("Accuracy: {:.2f}%".format(accuracy*100))

         Accuracy: 70.83%

In [15]: pickle.dump(model,open('mod.pkl','wb'))
         mod=pickle.load(open('mod.pkl','rb'))
```

Fig 11: Accuracy Results for Seven Emotions

As shown in the fig 10 and fig 11, the accuracy obtained is 70% for seven emotions (‘happy’, ’sad’, ’angry’, ’fearful’, ’disgust’, ’surprised’, ’neutral’).
Accuracy decreased from 76% to 70% when we change the number emotions from six to seven.

```
In [6]: emotions={
          '01':'neutral',
          '02':'calm',
          '03':'happy',
          '04':'sad',
          '05':'angry',
          '06':'fearful',
          '07':'disgust',
          '08':'surprised'
        }

        observed_emotions=['happy','sad','angry','fearful','disgust','surprised','neutral','calm']
```

Fig 12 : Eight Emotions taken for Results

```
In [12]: model.fit(x_train,y_train)

Out[12]: MLPClassifier(alpha=0.01, batch_size=256, hidden_layer_sizes=(300,),
                       learning_rate='adaptive', max_iter=500)

In [13]: y_pred=model.predict(x_test)

In [14]: accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
         print("Accuracy: {:.2f}%".format(accuracy*100))

         Accuracy: 65.62%

In [15]: pickle.dump(model,open('mod.pkl','wb'))
         mod=pickle.load(open('mod.pkl','rb'))
```

Fig 13 : Accuracy Results for 8 Emotions

As shown in the fig 12 and fig 13, we got the accuracy as 65% nearly for 8 emotions (‘happy’, ’sad’, ’angry’, ’fearful’, ’disgust’, ’surprised’, ’neutral’, ’calm’).
Accuracy decreased from 70% to 65% when we change the number emotions from seven to eight.

Table 1: Summary of emotions and Accuracy

Emotions Taken	Accuracy
happy, sad, angry, fearful, surprise - 5 emotions	84%
happy, sad, angry, fearful, disgust, surprised – 6 emotions	76%
happy, sad, angry, fearful, disgust, surprised, neutral – 7 emotions	70%
happy, sad, angry, fearful, disgust, surprised, neutral, calm -8 emotions	65%

As shown in the table 1, the accuracy went down when number of emotions were increasing.

8. FUTURE SCOPE

Further the study can improve the efficiency of this work. The misclassification can be overcome by making additional variations in training and testing ratio of speech trials. Here, only cepstral features have been considered for sentiment recognition. The work can be extended to combine both time domain and frequency domain features along with the proposed features. The algorithms may be tested with different databases.

9. CONCLUSION

This study states that performance of a module is highly dependent on the quality of pre-processing. Mel Frequency Cepstrum Coefficients are very dependable. Every human emotion has been thoroughly studied, analysed and the accuracy has been checked. The results obtained in this study determine that speech recognition is feasible, and that MLPs can be used for any task concerning recognizing of speech and demonstrating the accuracy of each emotion present in the speech.

REFERENCES

[1] K. H. Lee, H. Kyun Choi, B. T. Jang and D. H. Kim, "A Study on Speech Emotion Recognition Using a Deep Neural Network," 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea (South), 2019.

[2] L. Zheng, Q. Li, H. Ban and S. Liu, "Speech emotion recognition based on convolution neural network combined with random forest," 2018 Chinese Control and Decision Conference (CCDC), Shenyang, 2018.

[3] J. J. Lasiman and D. Puji Lestari, "Speech Emotion Recognition for Indonesian Language Using Long Short-Term Memory," 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Tangerang, Indonesia, 2018.

[4] R. Ram, H. K. Palo and M. N. Mohanty, "Recognition of fear from speech using adaptive algorithm with MLP classifier," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, India, 2016.

[5] Sanjita. B. R ,Nipunika. A, Rohita Desai, "speech Emotion Recognition using MLP Classifier",2020 by IJESC

[6] Jerry Joy , Aparna Kannan , Shreya Ram , S. Rama," Speech Emotion Recognition using Neural Network and MLP Classifier",2020 by IJESC.

[7] Anoop Joyti Sahoo, and Rajesh Kumar Tiwari "A Novel Approach for Hiding Secret data in Program Files" International Journal of Information and Computer Security. Volume 8 Issue 1, March 2016,

[8] Abu Salim, Sachin Tripathi and Rajesh Kumar Tiwari "A secure and timestamp-based communication scheme for cloud environment" Published in International Journal of Electronic Security and Digital Forensics, Volume 6, Issue 4, 319-332.

[9] Rajesh Kumar Tiwari and G. Sahoo, "A Novel Watermark Scheme for Secure Relational Databases" Information Security Journal: A Global Perspective, Volume 22, Issue 3, July 2013.

[10] N. Nandal, R. Tanwar and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products", Spat. Inf. Res. 28, 601–607 (2020).

[11] D. M. Thomas and S. Mathur, "Data Analysis by Web Scraping using Python," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 450-454, doi: 10.1109/ICECA.2019.8822022..

[12] R. A. Laksono, K. R. Sungkono, R. Sarno and C. S. Wahyuni, "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes," 2019 12th International Conference on Information & Communication Technology and System (ICTS), 2019, pp. 49-54, doi: 10.1109/ICTS.2019.8850982..

[13] B. Samal, A. K. Behera and M. Panda, "Performance analysis of supervised machine learning techniques for sentiment analysis," 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS), 2017, pp. 128-133, doi: 10.1109/SSPS.2017.8071579