Speech Emotion Recognition

[1]J.S.V.S. Hari Priyanka, [2]P. Harshitha, [3]A. Vineeth, [4]S. Bhavya, [5]M.

Sai Divya

[1]Assistant Professor,

[2,3,4,5] Student,

[1]Department of Information Technology,

[1]Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, India
[1]priyapatnaik.hari@gmail.com , [2]harshithapadmanabhuni@gmail.com, [3]vineeth.appikatla28@gmail.com ,
[4]bhavya.suvarna227@gmail.com , [5]mulagadadivya@gmail.com

Abstract

Introduction

This is a assignment approximately Speech Emotion Recognition. For numerous years now, the increase with inside the subject of Artificial Intelligence (AI) has been accelerated. AI, which become as soon as a topic understood via way of means of laptop scientists only, has now reached the residence of a not unusual place guy with inside the shape of sensible systems. The improvements of AI have engendered to numerous technology related to Human-Computer Interaction (HCI) [1]. Aiming to expand and enhance HCI strategies is of paramount significance due to the fact HCI is the front-give up of AI which tens of thousands and thousands of customers experience. Some of the prevailing HCI strategies contain communique via touch, movement, hand gestures, voice and facial gestures [1]. Among the distinct strategies, the voice-primarily based totally sensible gadgets are gaining recognition in a extensive variety of packages. In a voice-primarily based totally device, a laptop agent is needed to absolutely recognize the human's speech percept with the intention to correctly choose up the instructions given to it.

Speech Emotion Recognition is difficult to put in force a number of the different additives because of its complexity. Furthermore, the definition of an sensible laptop device calls for the device to imitate human behavior. A placing nature precise to people is the capacity to modify conversations primarily based totally at the emotional country of the speaker and the listener. Speech emotion popularity may be constructed as a category trouble solved the use of numerous device gaining knowledge of algorithms. This assignment discusses in element the numerous strategies and experiments done as a part of enforcing a Speech Emotion Recognition device

1. Literature Survey

[1]. In the paper Speech primarily based totally Emotion Recognition the use of Machine Learning via way of means of Girija Deshmukh provided three feelings with three characteristic vectors. The common device is split widely into dataset formation, pre-processing, characteristic extraction and category. Both male and girl samples are taken. Classification of local languages and their popularity is likewise effectively finished.

[2]. In the paper Speech Emotion Recognition via way of means of S Lalitha, provided 7 feelings with a superb common popularity rates. This become finished in extraction and classifier modules. The researchers located that SVM classifier has yielded higher overall performance due to the minimal structural danger minimization.

[3]. In the paintings Speech Emotion Recognition Based on Deep Belief Network via way of means of Peng Shi provided non-stop version and the discrete version of speech popularity device. In discrete version expresses feelings and for dimensional emotion version, emotion is a factor withinside the multidimensional non-stop emotion space. This paper makes speech emotion popularity experiments with eight feelings. SVM and ANN are used to make pattern analysis.

[4]. In the cutting-edge paintings Speech Emotion Recognition primarily based totally on Interactive Convolutional Neural Network via way of means of Huihui Cheng [4], targeted on enhancing the overall performance of conventional CNN on SER via way of means of providing an ICNN.

2. Methodology

The first segment includes dataset preparation. In our case, the dataset for the speech emotion popularity device is the speech samples and the traits are extracted from those speech samples the use of LIBROSA the use of the capabilities like MFCC, Chroma and Mel-Spectogram. Next we specify the feelings so that it will be diagnosed from the audio documents . The feelings expected are neutral, calm, happy, sad, angry, fearful, disgust, surprised. The feelings also are diagnosed primarily based totally at the gender .The overall variety of audio documents utilized in our assignment are 1561. These audio documents are first gone through characteristic extraction after which we divide the dataset into education and checking out datasets and use those datasets into Machine Learning Model. The algorithms used are Light Gradient Boosting Machine Classification set of rules and MLP classifier for this assignment. For Increasing the accuracy balloting classifiers are used with the aggregate of LightGBM and MLP.

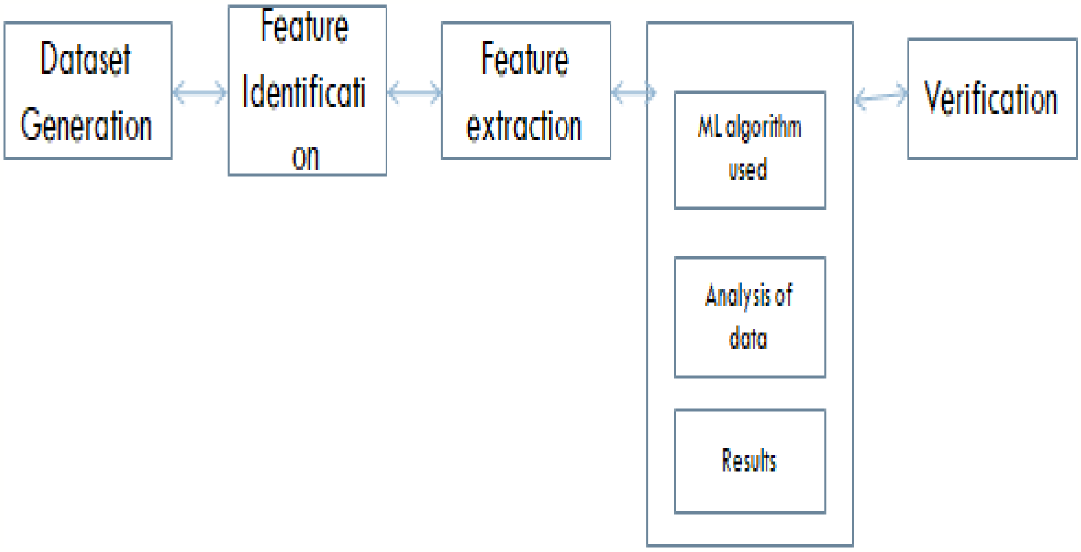


Figure 1. Block diagram of Methodology

3. Algorithms:

3.1 LightGradient Boosting Machine Classification (LightGBM)

Light Gradient Boosted Machine, or LightGBM , is an open-supply library that gives an green and powerful implementation of the gradient boosting set of rules. LightGBM extends the gradient boosting set of rules via way of means of including a form of automated characteristic choice in addition to that specialize in boosting examples with large gradients. This can bring about a dramatic speedup of education and stepped forward predictive overall performance. As such, LightGBM has come to be a de facto set of rules for device gaining knowledge of competitions while running with tabular facts for regression and category predictive modeling tasks. As such, it owns a percentage of the blame for the multiplied recognition and wider adoption of gradient boosting strategies in general, together with Extreme Gradient Boosting (XGBoost). The LightGBM boosting set of rules is turning into greater famous via way of means of the day because of its pace and efficiency. LightGBM is capable of deal with massive quantities of facts with ease. But preserve in thoughts that this set of rules does now no longer carry out nicely with a small variety of facts points.

LightGBM Algorithm:

1. Load the facts and break up the facts into education and checking out facts sets.

2. Import LightGBM and additionally import classification_report from sklearn.metrics and cross_val_score from sklearn.model_selection.

3. Create a variable named lgb_params and specify the parameters which might be utilized in LightGBM

4. Specify the version internal a fashions dictionary withinside the shape of 'lgb':lgb.LGBMClassifier(**lgb_params)

5. Now for locating accuracy we used np.imply of cross_val_score with parameters internal it are version, x_train, y_train .By the use of this facilitates in locating out the accuracy.

6. The confusion matrix is constructed each for education and checking out facts the use of the approach confusion_matrix() with parameters test/educate values and expected values .

The accuracy acquired varies among 65-70% the use of this LightGBM version to our dataset.

3.2 MLP (Multilayer Perceptron)

A multilayer perceptron (MLP) is a category of feed ahead synthetic neural network (ANN). An MLP includes at the least 3 layers of nodes: an enter layer, a hidden layer and an output layer. Except for the enter nodes, every node is a neuron that makes use of a nonlinear activation function. MLP makes use of a supervised gaining knowledge of approach referred to as lower back propagation for education.[2][3] Its more than one layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish facts that isn't always linearly separable. The MLP includes 3 or greater layers (an enter and an output layer with one or greater hidden layers) of nonlinearly-activating nodes. Since MLPs are absolutely connected, every node in a single layer connects with a positive weight to each node withinside the following layer.

MLP Algorithm:

1. Load the dataset and break up it into education and checking out datasets.

2. Import MLP Classifier from sklearn.neural_network additionally import classification_report from sklearn.metrics and cross_val_score from sklearn.model_selection.

three. Create a variable named mlp_params and specify the parameters which might be utilized in MLP .

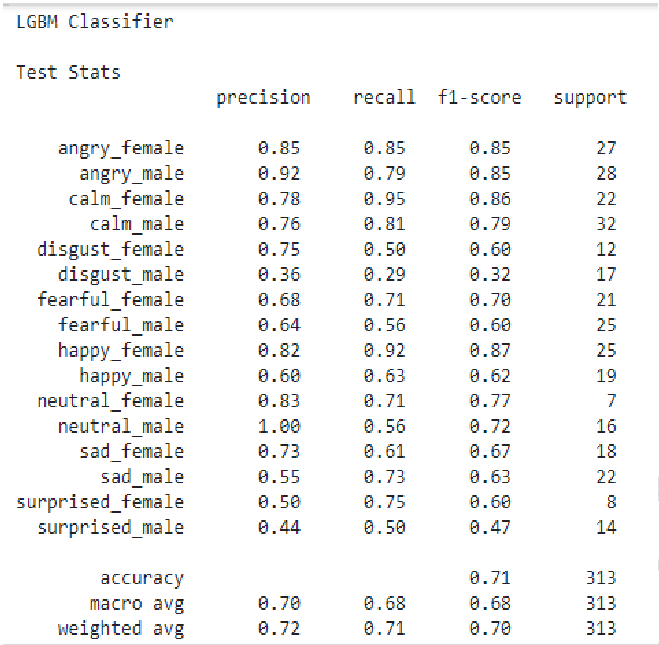
4. Specify the version internal a fashions dictionary withinside the shape of ‘mlp’:MLPClassifier (**mlp_params)

five. Now for locating accuracy we used np.imply of cross_val_score with parameters internal it are version, x_train, y_train .By the use of this facilitates in locating out the accuracy.

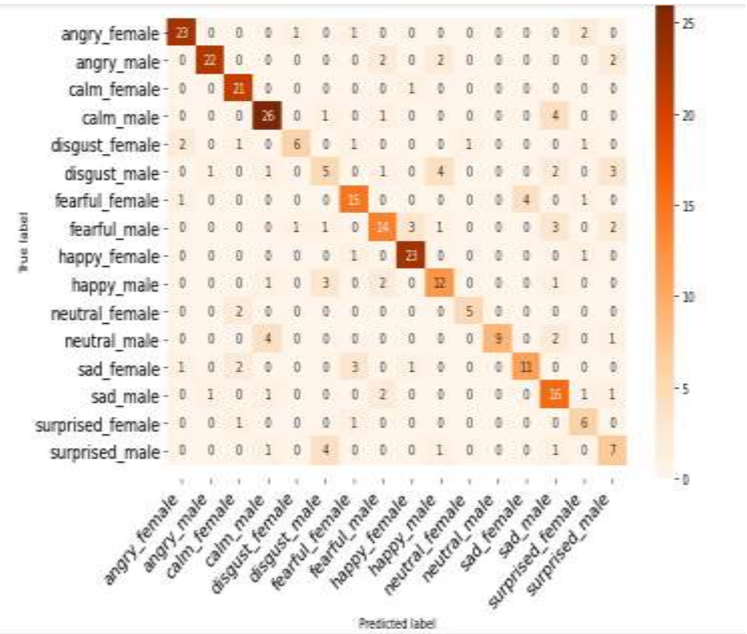
6. The confusion matrix is constructed each for education and checking out facts the use of the approach confusion_matrix() with parameters test/educate values and expected values .

The accuracy acquired varies among 78-88% the use of this MLP version to out dataset

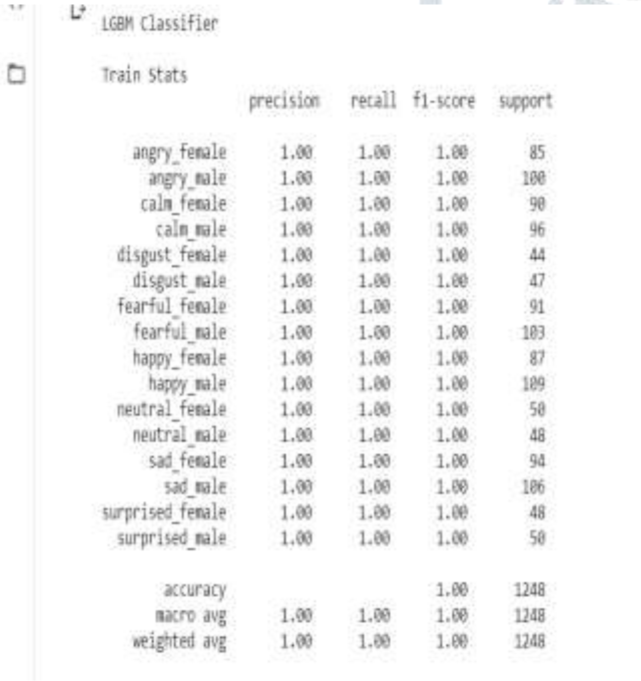
4. Experimental Results:



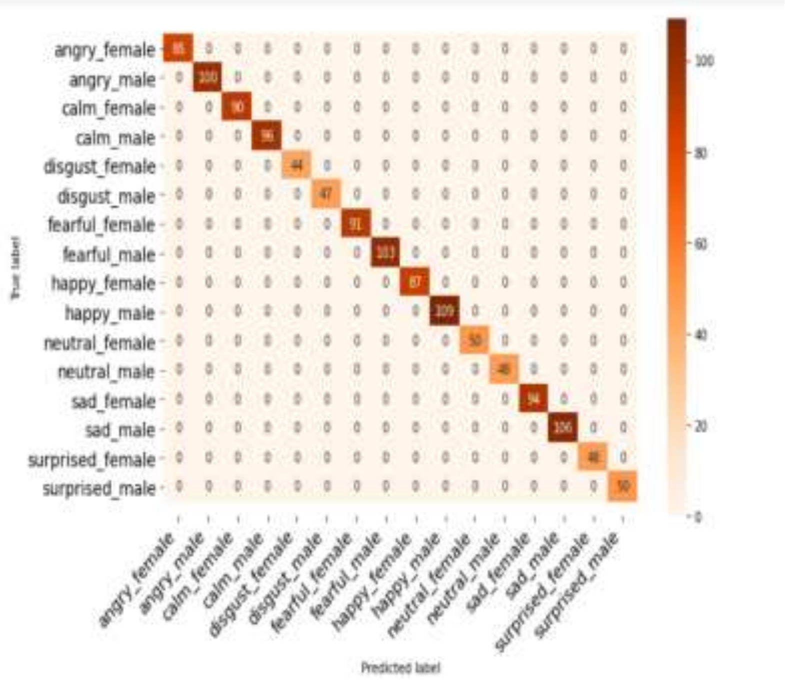
```
LGBM Classifier

Test Stats
                  precision    recall  f1-score   support

    angry_female       0.85      0.85      0.85        27
      angry_male       0.92      0.79      0.85        28
     calm_female       0.78      0.95      0.86        22
       calm_male       0.76      0.81      0.79        32
  disgust_female       0.75      0.50      0.60        12
    disgust_male       0.36      0.29      0.32        17
  fearful_female       0.68      0.71      0.70        21
    fearful_male       0.64      0.56      0.60        25
    happy_female       0.82      0.92      0.87        25
      happy_male       0.60      0.63      0.62        19
  neutral_female       0.83      0.71      0.77         7
    neutral_male       1.00      0.56      0.72        16
      sad_female       0.73      0.61      0.67        18
        sad_male       0.55      0.73      0.63        22
surprised_female       0.50      0.75      0.60         8
  surprised_male       0.44      0.50      0.47        14

        accuracy                           0.71       313
       macro avg       0.70      0.68      0.68       313
    weighted avg       0.72      0.71      0.70       313
```



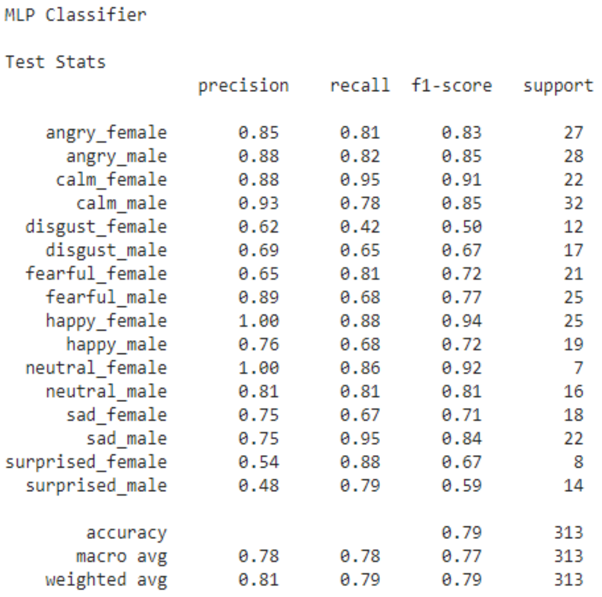
	angry_female	angry_male	calm_female	calm_male	disgust_female	disgust_male	fearful_female	fearful_male	happy_female	happy_male	neutral_female	neutral_male	sad_female	sad_male	surprised_female	surprised_male
angry_female	23	0	0	0	1	0	1	0	0	0	0	0	0	0	2	0
angry_male	0	22	0	0	0	0	0	2	0	2	0	0	0	0	0	2
calm_female	0	0	21	0	0	0	0	0	1	0	0	0	0	0	0	0
calm_male	0	0	0	26	0	1	0	1	0	0	0	0	0	4	0	0
disgust_female	2	0	1	0	6	0	1	0	0	0	1	0	0	0	1	0
disgust_male	0	1	0	1	0	5	0	1	0	4	0	0	0	2	0	3
fearful_female	1	0	0	0	0	0	15	0	0	0	0	0	4	0	1	0
fearful_male	0	0	0	0	1	1	0	14	3	1	0	0	0	3	0	2
happy_female	0	0	0	0	0	0	1	0	23	0	0	0	0	0	1	0
happy_male	0	0	0	1	0	3	0	2	0	12	0	0	0	1	0	0
neutral_female	0	0	2	0	0	0	0	0	0	0	5	0	0	0	0	0
neutral_male	0	0	0	4	0	0	0	0	0	0	0	9	0	2	0	1
sad_female	1	0	2	0	0	0	3	0	1	0	0	0	11	0	0	0
sad_male	0	1	0	1	0	0	0	2	0	0	0	0	0	16	1	1
surprised_female	0	0	1	0	0	0	1	0	0	0	0	0	0	0	6	0
surprised_male	0	0	0	1	0	4	0	0	0	1	0	0	0	1	0	7



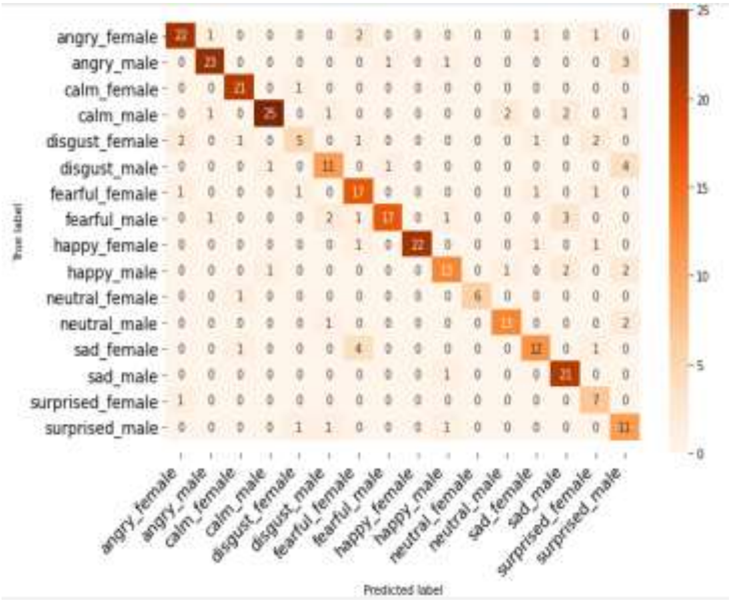
```
LGBM Classifier

Train Stats
                  precision    recall  f1-score   support

    angry_female       1.00      1.00      1.00        85
      angry_male       1.00      1.00      1.00       100
     calm_female       1.00      1.00      1.00        90
       calm_male       1.00      1.00      1.00        96
  disgust_female       1.00      1.00      1.00        44
    disgust_male       1.00      1.00      1.00        47
  fearful_female       1.00      1.00      1.00        91
    fearful_male       1.00      1.00      1.00       103
    happy_female       1.00      1.00      1.00        87
      happy_male       1.00      1.00      1.00       109
  neutral_female       1.00      1.00      1.00        50
    neutral_male       1.00      1.00      1.00        48
      sad_female       1.00      1.00      1.00        94
        sad_male       1.00      1.00      1.00       106
surprised_female       1.00      1.00      1.00        48
  surprised_male       1.00      1.00      1.00        50

        accuracy                           1.00      1248
       macro avg       1.00      1.00      1.00      1248
    weighted avg       1.00      1.00      1.00      1248
```

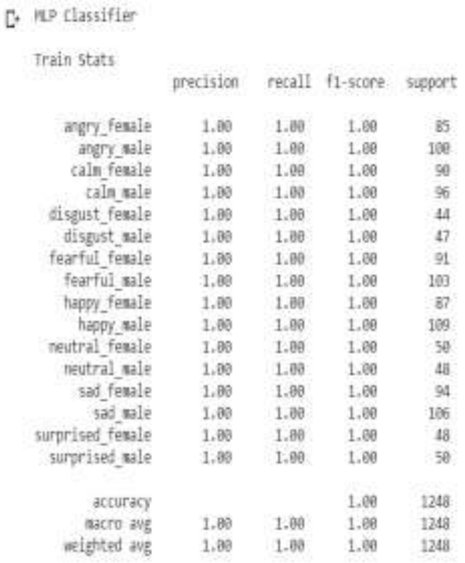


	angry_female	angry_male	calm_female	calm_male	disgust_female	disgust_male	fearful_female	fearful_male	happy_female	happy_male	neutral_female	neutral_male	sad_female	sad_male	surprised_female	surprised_male
angry_female	85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
angry_male	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
calm_female	0	0	90	0	0	0	0	0	0	0	0	0	0	0	0	0
calm_male	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0
disgust_female	0	0	0	0	44	0	0	0	0	0	0	0	0	0	0	0
disgust_male	0	0	0	0	0	47	0	0	0	0	0	0	0	0	0	0
fearful_female	0	0	0	0	0	0	91	0	0	0	0	0	0	0	0	0
fearful_male	0	0	0	0	0	0	0	103	0	0	0	0	0	0	0	0
happy_female	0	0	0	0	0	0	0	0	87	0	0	0	0	0	0	0
happy_male	0	0	0	0	0	0	0	0	0	109	0	0	0	0	0	0
neutral_female	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
neutral_male	0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0
sad_female	0	0	0	0	0	0	0	0	0	0	0	0	94	0	0	0
sad_male	0	0	0	0	0	0	0	0	0	0	0	0	0	106	0	0
surprised_female	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0
surprised_male	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50

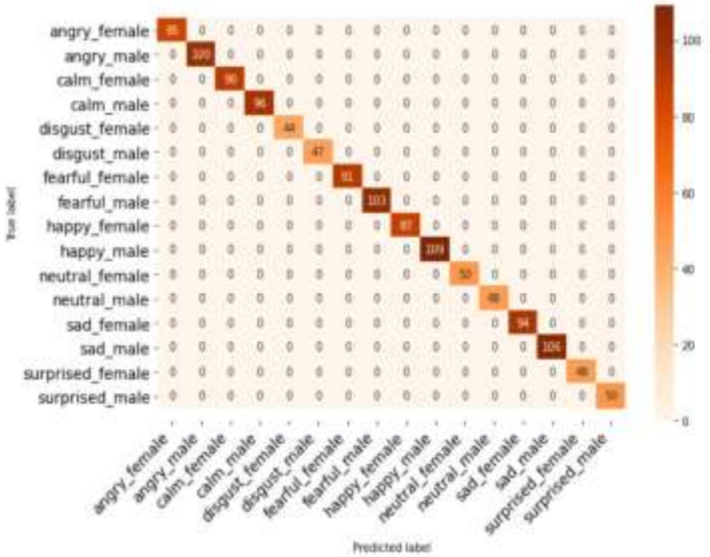


```
MLP Classifier

Test Stats
                  precision    recall  f1-score   support

    angry_female       0.85      0.81      0.83        27
      angry_male       0.88      0.82      0.85        28
     calm_female       0.88      0.95      0.91        22
       calm_male       0.93      0.78      0.85        32
  disgust_female       0.62      0.42      0.50        12
    disgust_male       0.69      0.65      0.67        17
  fearful_female       0.65      0.81      0.72        21
    fearful_male       0.89      0.68      0.77        25
    happy_female       1.00      0.88      0.94        25
      happy_male       0.76      0.68      0.72        19
  neutral_female       1.00      0.86      0.92         7
    neutral_male       0.81      0.81      0.81        16
      sad_female       0.75      0.67      0.71        18
        sad_male       0.75      0.95      0.84        22
surprised_female       0.54      0.88      0.67         8
  surprised_male       0.48      0.79      0.59        14

        accuracy                           0.79       313
       macro avg       0.78      0.78      0.77       313
    weighted avg       0.81      0.79      0.79       313
```





```
MLP Classifier

Train Stats
                  precision    recall  f1-score   support

    angry_female       1.00      1.00      1.00        85
      angry_male       1.00      1.00      1.00       100
     calm_female       1.00      1.00      1.00        90
       calm_male       1.00      1.00      1.00        96
  disgust_female       1.00      1.00      1.00        44
    disgust_male       1.00      1.00      1.00        47
  fearful_female       1.00      1.00      1.00        91
    fearful_male       1.00      1.00      1.00       103
    happy_female       1.00      1.00      1.00        87
      happy_male       1.00      1.00      1.00       109
  neutral_female       1.00      1.00      1.00        50
    neutral_male       1.00      1.00      1.00        48
      sad_female       1.00      1.00      1.00        94
        sad_male       1.00      1.00      1.00       106
surprised_female       1.00      1.00      1.00        48
  surprised_male       1.00      1.00      1.00        50

        accuracy                           1.00      1248
       macro avg       1.00      1.00      1.00      1248
    weighted avg       1.00      1.00      1.00      1248
```



5. Future Scope and Conclusion:

After building numerous fashions, we were given the higher CNN version for the emotion difference task. We reached 71% accuracy from the formerly to be had version. Our version would've achieved higher with greater facts. Also our version achieved thoroughly while distinguishing amongst a masculine and female voice.

Our assignment may be prolonged to combine with the robotic to assist it to have a higher expertise of the temper the corresponding human is in, that allows you to assist it to have a higher communication in addition to it is able to be included with numerous song packages to endorse songs to its customers in keeping with his/her feelings, it is able to additionally be utilized in numerous on line buying packages which includes Amazon to enhance the product advice for its customers. Moreover, in the approaching years we are able to assemble a series to collection version to create voice having distinct feelings. E.g. asad voice, an excited one etc.

6. References:

1. Gil Levi,Tal Hassner; Emotion Recognition withinside the Wild through Convolutional Neural Networks and Mapped Binary Patterns, SC / Information Sciences Institute, the Open University of Israel, 2014.

2. KunHan, Dong Yu, Ivan Tashev; Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine; Department of Computer Science and Engineering, The Ohio State University, Columbus,43210, OH, USA; Microsoft Research, One Microsoft Way, Redmond,98052, WA, USA,2014.

3. Shiqing Zhang, Xiaohu Wang , Gang Zhang, Xiaoming Zhao; Multimodal Emotion Recognition Integrating Affective Speech with Facial Expression ; Institute of Image Processing and Pattern Recognition Taizhou University Taizhou 318000 CHINA, Hunan Institute of Technology Hengyang 421002 CHINA,Bay Area Compliance Labs. Corp. Shenzhen 518000 CHINA, 2014.

4. N. Morgan, ―Deep and extensive: Multiple layers in automated speech popularity,‖ Audio, Speech, and Language Processing, IEEE Transactions on, vol. 20, no. 1, pp. 7–13, 2012.

5. A. Mohamed, G.E. Dahl, and G. Hinton, ―Acoustic modeling the use of deep perception networks,‖ Audio, Speech, and Language Processing, IEEE Transactions on, vol. 20, no. 1, pp. 14–22, 2012.

6. G. Sivaram and H. Hermansky, ―Sparse multilayer perceptron for phoneme popularity,‖ Audio, Speech, and Language Processing, IEEE Transac- tions on, vol. 20, no. 1, pp. 23–29, 2012.

7. Martin W¨ollmer, Angeliki Metallinou, Florian Eyben, Bj¨orn Schuller, Shrikanth Narayanan; Context-Sensitive Multimodal Emotion Recognition from Speech and Facial Expression the use of Bidirectional LSTM Modeling; Institute for Human-Machine Communication, Technische Universit¨at M¨unchen, Germany Signal Analysis and Interpretation Lab (SAIL), University of Southern California, Los Angeles, CA, 201