Human Voice Emotion Recognition Using Multiplayer Perceptron

By **D.Femi and S.Thylashri**

Rishikesh
ANN Presentation & Application

March 6, 2024

1. Introduction

This Study presents the effect of the use of different regularization methods on language features to identify emotions.
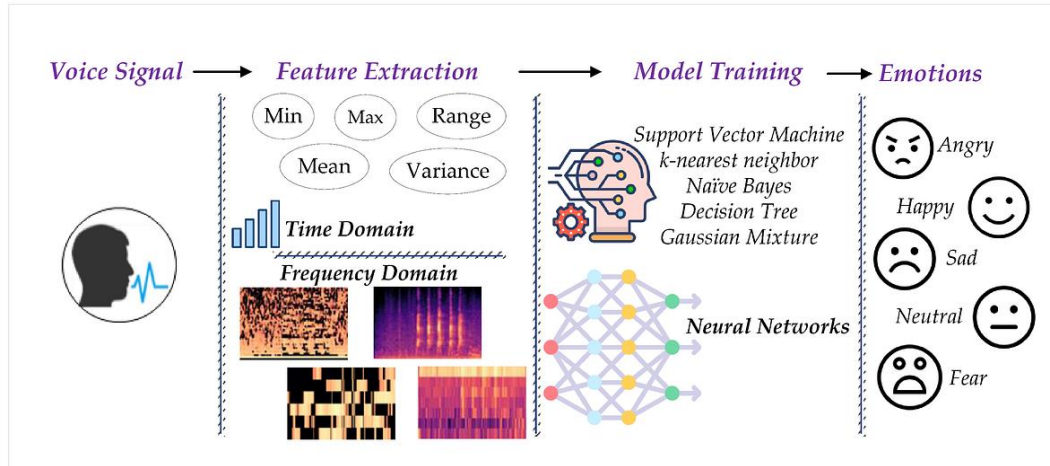
- Ryerson Audio-visual Database
- Speech Signals - Time, frequency, and Cepstral domain features
- Neutral, Calm, Happy, Sad, Angry, Afraid, Disgusted & Startled
- Improve Man-Machine Interface, child's development and so on

- Use of Computer-generated gestures to extract emotional characteristics
- Reliability and integration of all emotions
- Need to identify emotions in source speech and integrate them in a target speech.
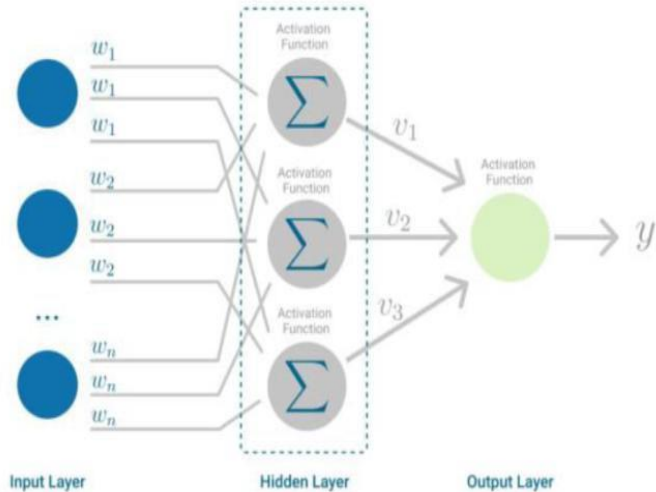
- Multilayer Perceptron (MLP) as a machine learning algorithm
- RAVDESS dataset (Ryerson Audio-Visual Dataset of Emotional Speech and Songs)
  - 7356 short audio files + transcription   (Train:Test) = (75:25)
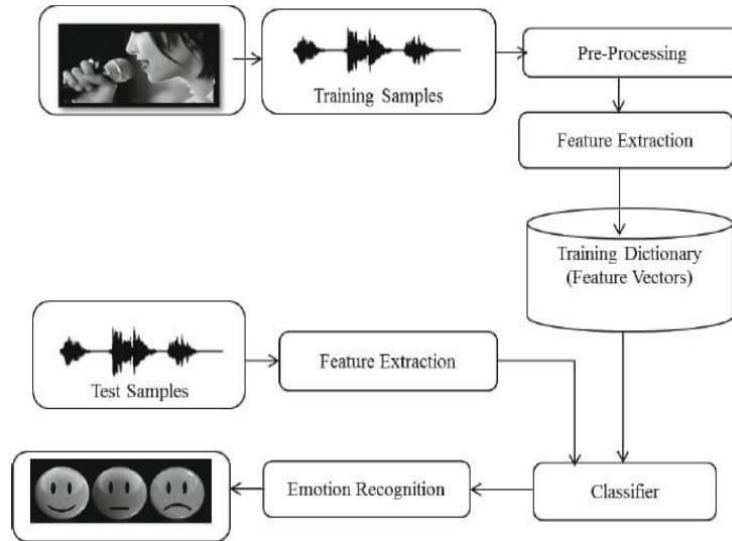
- Pre-processing: Data Visualization and balancing
- Feature extraction: Time, Frequency, and Cepstral-domain features
    1-dimensional
- Training and classification: MLP neural network with back propagation learning algorithm

- Multilayer Perceptron (MLP) as a feed forward neural network
- Input layer, output layer, and hidden layer
- Training using back propagation algorithm
- ML used for classification, recognition, and prediction
- The trained model will classify: Neutral, Clam, Happy, Sad, Angry, Surprise

```
[+] Number of testing samples: 168
[+] Number of features: 180
[*] Training the model...
Accuracy: 81.52%
```

```
              precision    recall  f1-score   support

       angry       0.84      0.84      0.84        61
       happy       0.72      0.56      0.63        41
     neutral       0.52      0.62      0.57        21
         sad       0.70      0.78      0.74        45

    accuracy                           0.73       168
   macro avg       0.69      0.70      0.69       168
weighted avg       0.73      0.73      0.73       168

[[51  6  1  3]
 [ 6 23  5  7]
 [ 1  2 13  5]
 [ 3  1  6 35]]
```

Training the model

Confusion Matrix
- derived the accuracy and confusion matrix by comparing predictions after training.

7. References

D. Femi and S. Thylashri, "Human Voice Emotion Recognition Using Multilayer Perceptron," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-4, doi: 10.1109/ICSES55317.2022.9914336. keywords: {Emotion recognition;Time-frequency analysis;Mood;Neural networks;Speech recognition;Multilayer perceptrons;Feature extraction;Human Computer Interaction;Multilayer Perceptron},

Link: https://ieeexplore.ieee.org/document/9914336

TEXAS TECH UNIVERSITY SYSTEM™