International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 03 | March - 2021 ISSN: 2582-3930

Speech Emotion Recognition using Machine Learning

Harshvardhan Tiwari, Pankaj Yadav and Kshitija Acham, Prof. Sarang Dube y

Information Technology Department, ABM SP's Anantrao Pawar College of Engineering and Research, Pune -411009

***Abstract*- Emotions plays a very important role in human mental life. It is nothing but the medium of expression of one's perspective to others . Speech Emotion Recognition is defined as extracting the emotional state of the speaker from their respective speech signal. There are few emotions like Anger, Happiness, Sadness, Neutral, Fearful, Disgust, Surprise, Calm etc. can be recognized by using this system. Emotion Detection from speech signal requires feature extraction and Classifier training. This feature extraction contains feature vectors consist of speech signals which provides speakers features such as tone, pitch, energy etc. these are the features which are used to train the classifier model to recognize the speakers emotions accurately. Mel-Frequency Cepstrum Coefficient(MFCC) and Modulation Spectral(MS) features are extracted from the speech signals and they are used to train different classifiers . This system have wide area of applications and one of them is Health Care Units, in these units many raw data is collected under a specific techniques . This technique includes speech signals conversion into wave form, utterance level feature extraction, emotion classification, extracting database recognition, alert signal creation through cloud is the sequence of steps to be followed.**

***Keywords*-** Speech Signal, Mel-Frequency Cepstrum Coefficient(MFCC) and Modulation Spectral(MS), Feature Extraction, Classifier Training, Emotion Classification, Database Recognition.

I. INT RODUCT ION

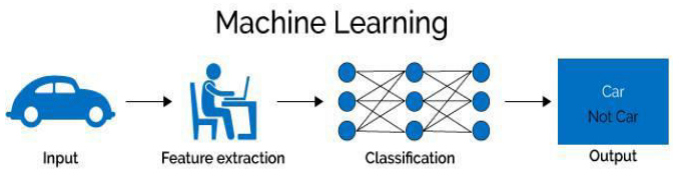
Over the most recent couple of years, Human Emotion Speech recognition systemhas been considered as the most imperative applications compared to other biometric-based systems . It plays a vital role in human life. The emotional information hidden in speech is an important fact of interaction between humans because, it provides feedback in communication. Speech emotion Recognition can be called as the fetching of the appropriate emotional state of the speech uttered by the person from their speech signal.

There are emotions - like Neutral, Anger, Surprise, Fear, Happiness, and Sadness which can detect by any intelligent system with finite computational resources which can be trained to identify or synthesize as per required. In case of direct face to face interaction the emotion could be recognized via facial expressions and body languages whereas, if communication is made through a medium between persons residing far apart from one another, the prediction of emotion is hard and may be un - efficient.

The Relationship between man and machines has become a new trend of revolutionary technology such that machines now have to respond by considering The human emotional levels . In the recent years human-computer interaction has become more interesting. Machine learning provides algorithms to build lots of analytical models, helping computers to learn from data. It helps us to match and understand the feelings of others by conv eying our feelings and giving feedback to others . Emotional displays convey considerable information about the mental state of an individual. This has opened up a new research field called automatic emotion recognition, having basic goals to understand and retrieve desired emotions .

Machine learning - : Machine learning is a concept that a computer program can learn and adapt to new data without any human interference. Machine learning is a field of artificial intelligence (AI) that keeps a computer's built-in algorithms current regardless of changes in the worldwide economy.

The signal level processing, artificial intelligence and machine learning technologies have boosted the machine intelligence, so that the machines gained the capability to understand human emotions . Considering the aspects of speech processing and pattern recognition algorithms an intelligent and emotions specific man-machine interaction can be achieved which can be harnessed to design a smart and secure automated home as well as commercial application.



II. LIT ERAT URE WORK

Following are the research papers we studied for the Speech Emotion recognition system.

1. Applying Machine Learning Techniques for Speech Emotion Recognition.
   - Author: K.Tarunika , R.B Pradeeba , P.Aruna
   - Year: 2018
   - Key Point: The research over this idea fetches that this techniques is yet play the vital role in medical and technical field. Voice and face detection will play a wise role in upcoming equipment and systems . The authentication processes also highly lay their concern over this recognition formula. This idea may gain its assert over the vast field of computer science and other related branches by collecting lots of data on the predictable form and lay its root firm to become the indispensable one of the future world.

2. Speech based Emotion Recognition using Machine Learning
   - Author: Girija Deshmukh, Apurva Gaonkar, Gauri Golwalkar, Sukanya Kulkarni
   - Year: 2019
   - Key Point: In this paper, three emotions - anger, happiness, and sadness, were classified using three feature vectors . Pitch, Mel frequency cepstral coefficients, Short Term Energy were the three feature vectors extracted from audio signals . Open source North American English acted speech corpus and recorded natural speech corpus were used as input. The dataset used for training and testing consisted of audio samples in male and female voice and divided in ratio 4:1. The mean method provided greater accuracy over the mode method.

3. Speech Emotion Recognition using Deep Learning Techniques: A Review
   - Authors: RUHUL AMIN KHALIL , EDWARD JONES , MOHAMMAD INAYATULLAH BABAR , TARIQULLAH JAN , MOHAMMAD HASEEB ZAFAR , AND THAMER ALHUSSAIN
   - Year: 2016
   - Key Point: This paper has provided a detailed review of the deep learning techniques for SER. Deep learning techniques such as DBM, RNN, CNN, and AE have been the subject of much research in recent years . These deep learning methods and their layer-wise architectures are briefly elaborated based on the classification of various natural emot ion such as happiness, joy, sadness, neutral, surprise, boredom, disgust, fear, and anger. These methods offer easy model training as well as the efficiency of shared weights .

III. PROPOSED WORK

We are developing the Speech Emotion recognition (AI) app using Machine learning provide the efficient Speech recognition.

The System should be strong enough to be a ble to replace traditional Speech processing system. It should also be able to recognize the emotions which is present in the speakers voice.

We are developing a system which can detect and recognize the Human Emotions . Initially we divide the system into two main parts:

A. Feature Extraction
B. Feature Selection

The main motto of our system is to develop Emotion recognition System which uses machine learning as a key ingredient to Recognize and map a person 's Speech features from a call and then tries to recognize their mental state and their emotions .

A. Feature Extraction

The speech signal contains a large number of parameters that reflect the emotional characteristics . One of the sticking points in emotion recognition is what features should be used. In recent research, many common features are extracted, such as energy, pitch, formant, and some spectrum features such as linear prediction coefficients (LPC), Mel-frequency Cepstrum coefficients (MFCC), and modulation spectral features . In this work, we have selected modulation spectral features and MFCC, to extract the emotional features .

Mel-frequency Cepstrum coefficient (MFCC) is the most used representation of the spectral property of voice signals . These are the best for speech recognition as it takes human perception sens itivity with respect to frequencies into consideration. For each frame, the Fourier transform and the energy spectrum were estimated and mapped into t he Mel-frequency scale. The dis crete cosine transform (DCT) of the Mel log energies was estimated, and the first 12 DCT coefficients provided the MFCC values used in the classification process .

B. Feature Selection

the objective of feature selection in ML is to "reduce the number of features used to characterize a dataset so as to improve a learning algorithm's performance on a given task." The objective will be the maximization of the classification accuracy in a specific task for a certain learning algorithm; as a collateral effect, the number of features to induce the final classification model will be reduced. Feature selection (FS) aims to choose a subset of the relevant features from the original ones according to certain relevance evaluation criterion, which usually leads to higher recognition accuracy. It can drastically reduce the running time of the learning algorithms . In this section, we present an effective feature selection method used in our work, named recursive feature elimination with linear regression (LR-RFE).

IV. ARCHITECTURE DIAGRAM

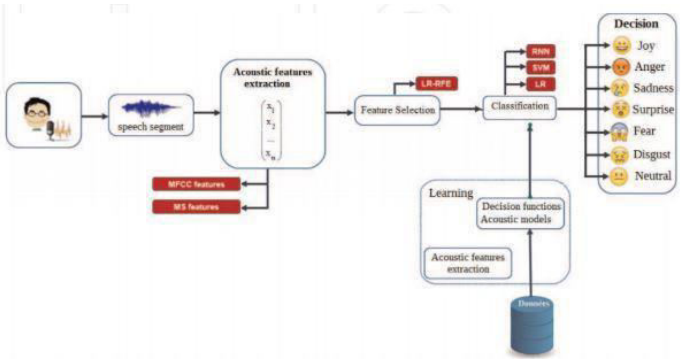


Fig: System Architecture Diagram

The Above system architecture diagram consist of different modules like:

- Speech Segment
- Feature Extraction
- Feature Selection
- Classification
- Learning
- Database
- Emotion Decision

- Speech Segment

Speech segmentation is the process of identifying th e boundaries between words and syllables, in spoken languages . It applies both to the mental processes used by humans, and to artificial processes of natural language processing . This speech segment is pass to the Acoustic Feature Extraction.

- Feature Extraction

It is a process of reduction by which a set of raw data is reduced to more manageable groups for processing. After extracting the features it further passed to the feature selection. A characteristic of these large data sets is a large number of variables that require a lot of computing resources to process . It mainly consist of two methods such as:

- MFCC Features
- MS Features

a. MFCC Features

The MFCC feature extraction technique basically contains windowing the signal, applying the DFT, taking the log of the magnitude, and then warping the frequencies on a Mel scale, followed by applying the inverse DCT.

b. MS Features

It is nothing but the Modulation Spectral Feature extraction from the speech segment.

- Feature Selection

In machine learning, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features for use in model construction. It can be done by using LR-RFE.

- Classification

It is also called as the categorization. grouping of data according to the given criteria is nothing but the classification. this classification will be done by using 3 algorithms such as follows:

- RNN Algorithm
- SVM Algorithm
- LR Algorithm

- Database

where all the speech segments and their extracted features were stored for emotion recognition. this data is pass to the learning module for further processing.

- Learning

In This module Acoustic features extraction is done and Decision functions Acoustic Models are designed. Then these learning's are passed to the classification module to make decision of speakers emotion.

- Emotion Decision

This is the last module where the speakers emotion is recognized were the speaker is Joyful, Angry, Sad, Surprised, Feared, Disgust and Neutral etc.

V. CONCLUSION

The research over this idea fetches us the knowledge, that this techniques is yet play the vital role in medical and technical field. Voice and face detection will play a wise role in upcoming equipment and systems . The authentication processes also highly lay their concern over this recognition formula. This idea may gain its assert over the vast field of computer science and other related branches by collecting lots of data on the predictable form and lay its root firm to become the indispensable one of the future world. The reason for evolution of all these technique is just for the advancement and reduction to time that assist people.

VI. REFFERENCES

1. B. W. Schuler, **"Speech emotion recognit**ion: Two decades in a nutshell, **benchmarks, and ongoing trends,"** Communications of the ACM, vol. 61, no. 5, pp. 90–99, 2018

2. M. S. Husain and G. Muhammad, **"Emotion recognition** using deep learning approach from audio–visual emotional big **data," Information** Fusion, vol. 49, pp. 69–78, 2019.

3. Livingstone SR, Russo FA (2018) The Ryerson Audiovisual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal express ions in North American English. PLoS ONE 13(5): e0196391. https://doi.org/10.1371/journal.pone.0196391.

4. **Practical Cryptography, "Mel Frequen**cy Cepstral **Coefficients (MFCC) tutorial."** Internet: http://practicalcryptography.com/miscellaneous/machi ne-learning/guide-Mel-frequency-cepstral-coefficients -mfccs/, [Feb.27, 2019]

5. **Sunil Ray, Analytics Vidhya, "Under**standing Support Vector Machine **algorithms from examples." Internet:** https://www.analyticsvidhya.com/blog/2017/09/ understating support-vector-machine-example-code/, Sept.13, 2017 [Mar.10, 2019].

6. K.V. Krishna Kishore, P. Krishna Satish, **"Emotion** Recognition in speech **Using MFCC and Wavelet Features",** 3rd IEEE International Advance Computing Conference (IACC), 2013