iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XII **Month of publication:** December 2021

DOI: https://doi.org/10.22214/ijraset.2021.39420

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 9 Issue XII Dec 2021- Available at www.ijraset.com

Speech Based Emotion Recognition Using Machine Learning

Vaibhav K. P.[1], Parth J. M.[2], Bhavana H. K.[3], Akanksha S. S.[4]
[1, 2, 3, 4]*Computer Engineering, Sinhgad Academy of Engineering Pune*

Abstract: Speech emotion recognition is a trending research topic these days, with its main motive to improve the human-machine interaction. At present, most of the work in this area utilizes extraction of discriminatory features for the purpose of classification of emotions into various categories. Most of the present work involves the utterance of words which is used for lexical analysis for emotion recognition. In our project, a technique is utilized for classifying emotions into Angry',' Calm', 'Fearful', 'Happy', and 'Sad' categories.

I. INTRODUCTION

Human speech, through speech, tone, pitch and many such features of the human vocal system, conveys information and con- text. Speech emotion recognition are widely used in many fields with natural human-computer interaction requirements. Speech emotion recognition, which is characterized as extricating the emotional condition of a speaker from his or her discourse. This sort of recognition is supposed to be used to extract useful semantics of speech recognition systems.

The model of SER includes the discrete speech emotion model and continuous speech emotion model. The discrete emotion model expresses several independent emotions, indicating that a certain speech has a single independent emotion, while the continuous speech emotion means that the emotion is in the emotion space, and every emotion has different strength on each dimension.

Determining the emotional state of humans is an idiosyncratic task and may be used as a standard for any emotion recognition model. It uses various emotion such as anger, disgust, surprise, fear, joy, happiness, neutral and sadness.

II. PROBLEM DEFINITION AND OBJECTIVES

For a dataset containing audio files of different actors, design a Speech Emotion Recognition (SER) system which will analyse audio using required machine learning techniques. Also, create a wave plot of the voice.

A. *Objectives*

1) The main goal of this module is to recognise emotion.
2) To design a system which will extract, characterize and recognize the information of speaker's emotions.
3) To build a model to recognize emotion from speech using the CNN and the RAVDESS, SAVEE, CREMA-D, TESS dataset. Literature survey.

III. USAGE SCENARIO

Usage scenario of the project is where we want to detect the emotion by using voice. The project can be used by directly by audience to analyse the authenticity by providing input of their voice(speech).

A. *User Profiles*

1) *Actors:* The person who will use this system must have essential computer knowledge. The user must read the user manual and apply it.
2) *Software Developers:* People with very great knowledge of programming language, in order to get it and be able to extend project's source code.
3) *Doctors:* The doctor will use this system to track the emotional changes of patients with depression as a basis for disease diagnosis and treatment.

B. *Use-Cases*

Users would use the platform for verification of emotions from popular sources. The client is not required any kind of specialized information to utilize this system, fair required the essential knowledge of computer. User can easily interact with the system as it is available in a popular international language such as English. The admin is responsible for authenticate the user and to show the recognized speech and identifies proper emotion along with gender.

©IJRASET: All Rights are Reserved

IV. PROJECT SCOPE

For a dataset containing audio files of different actors, A Speech Emotion Recognition (SER) system is designed which will analyse human emotion using speech as an input. The system ex- tracts, characterizes and recognizes the information of speaker's emotions for both male and female using the CNN Algorithm and the RAVDESS dataset. The system is limited to English language where real-time input speech data is supported. Also, it is also ex- tended to recognize more than 4 emotions as multiple datasets are being used. Speech emotion recognition could be used as a feedback application in hospitals to take feedback from patients and order the feedbacks into their deserved categories such as anger judgement, or a sad judgement

V. SOFTWARE QUALITY ATTRIBUTES

Our software has many quality attributes that are given below:

1) *Availability:* The system is available for all operating systems.
2) *Reliability:* System reliability will move forward as long as the audio's quality is sweet and the person's voice is clearly capable of being heard. Since the measure and sort of the record to be transferred is constrained, no system crashes will be permitted.
3) *Re-usability:* Certain percentage of the system must be de- signed generically to support re-usability.
4) *Ease of Use:* Since the developed application is a user- oriented project, it should provide simple usage to the user. Therefore, the interface we will prepare will be understand- able and user-oriented.

VI. UML STATE TRANSITION DIAGRAM

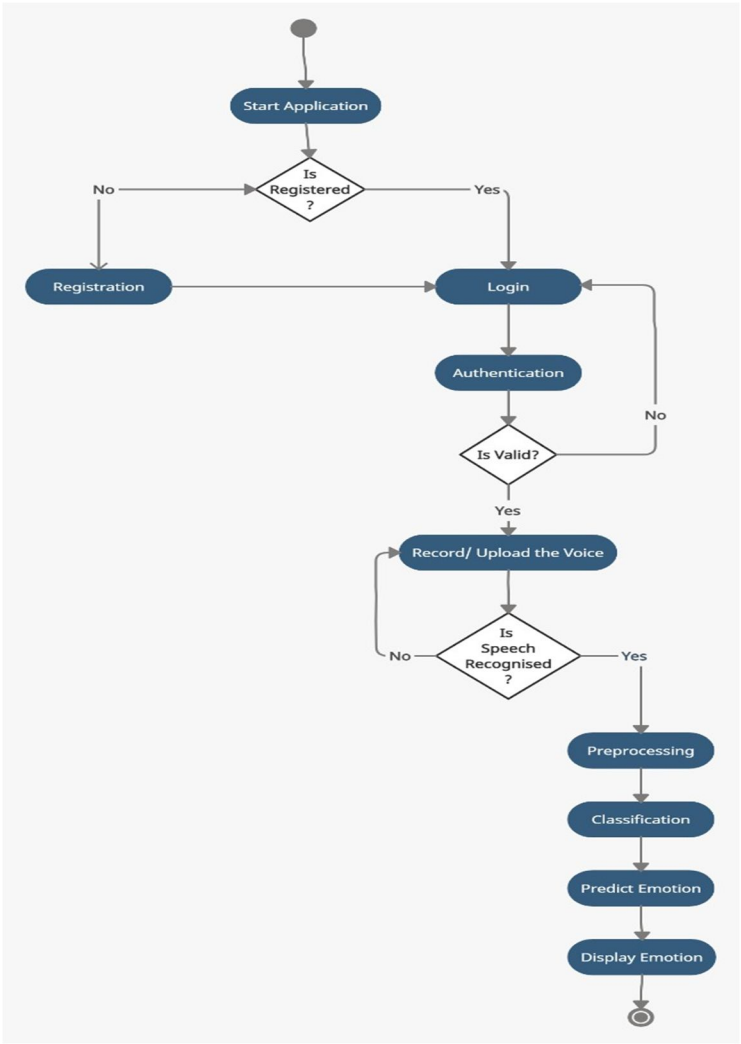


Fig. 1 Design for emotion recognition software.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 9 Issue XII Dec 2021- Available at www.ijraset.com

VII. RESULT(OUTCOME)

In the proposed work, a procedure has been presented to evaluate or predict the gender and emotions through their speech. The evaluation or prediction of Gender and emotions are done through Convolutional Neural Network by plotting the waveform & spectrogram. In order to recognize emotion from speech, 12162 samples are given as an input and a CNN model is built which is used to name the emotion. By using the 4 datasets (RAVDESS, SAVEE, CREAMA-D, TESS) in our project for the proposed model, over- all accuracy is calculated using only one feature that is MFCC from the speech. Initially the accuracy is calculated for each emotion along with gender then the accuracy is found to be 44% only, but when the accuracy for each emotion is calculated then the accuracy becomes 50%. The final accuracy of model is calculated based on the gender distribution which is found to be 78.52%.

VIII. CONCLUSION

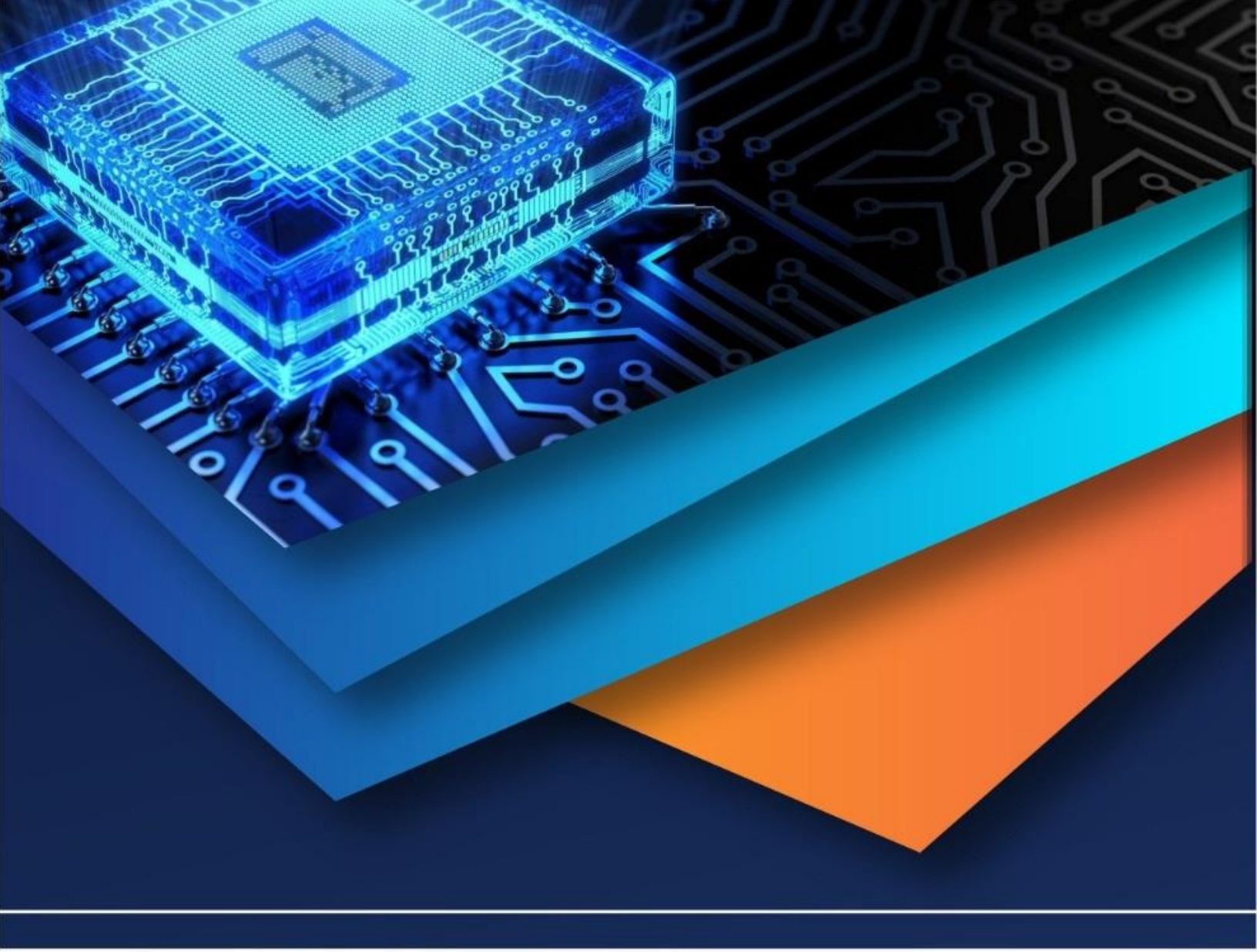
In this paper, procedure for recognition of emotion from speech has been presented. The recognition of emotion is done using CNN algorithm, MFCC and dataset. The performance of a system thoroughly depends upon the types of pre-processing techniques the dataset has been through. And selection of techniques must be done requiring prior knowledge.

However, the scope of the system is limited to English language but its future work scope can be extended. Modifications may change the system into a professional software which can be used to detect emotions in Hospital Feedback system, psychological applications and also for Mood-synchronising music players.

REFERENCES

[1] Girija Deshmukh, Apurva Gaonkar, Gauri Golwalkar, Sukanya Kulkarni, “Speech based Emotion Recognition using Machine Learning”, IEEE, Mar. 2019.
[2] J. Umamaheswari, A. Akila,” An Enhanced Human Speech Emotion Recognition Using Hybrid of PRNN and KNN”, IEEE, Feb 2019.
[3] Michael Neumann, Ngoc Thang Vu, ”Improving Speech Emotion Recognition with Unsupervised Representation Learning on Unlabelled Speech”, IEEE, May 2019.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)