iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50255

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 11 Issue IV Apr 2023- Available at www.ijraset.com

Speech Based Emotion Recognition Using Machine Learning

Makineedi Rajababu[1], P. Abhinav[2], N. K Subhash[3], M. Ajay Chowdary[4]
[1, 2, 3, 4]*Department of Information Technology Aditya Engineering College, Surampalem, India*

Abstract: ***Speech-based emotion recognition is a developing field that has attracted a lot of interest lately. We suggest a machine learning method for identifying emotions from speech samples in this article. From the speech examples, we extract acoustic features that we then use to train and test a variety of machine learning methods, such as decision trees, support vector machines, and neural networks. Using a dataset of speech samples with emotional labels that is publicly accessible, we assess how well these models perform. The experimental findings demonstrate that, with an accuracy of 87%, the neural network model beats the other models. Applications for the suggested method include human-computer interaction, education, and the diagnosis of mental illnesses. Overall, this article makes a contribution to the improvement of speech-based emotion recognition systems.***

I. INTRODUCTION

Speech-based emotion recognition is an active area of research in the field of human-computer interaction. Recognizing emotions from speech is important for a range of applications, including mental health diagnosis, education, and entertainment. Many studies have been conducted on thistopic, but there is still a need for more accurate and reliable emotion recognition systems. Machine learning has emerged as a promising approach for speech-based emotion recognition due to its ability to learn patterns from data and adapt to new situations.

We suggest a machine learning method for identifying emotions from speech samples in this article. Then we extract acoustic features from the speech samples and use these features to train and evaluate several machine learning algorithms. We evaluate the performance of these models using a publicly available dataset of speech samples labeled with emotions. We also compare our approach with existing methods in the literature.

The major contributions of this paper are the experimental assessment of this approach on a dataset that is freely available, as well as the proposal of a novel machine learning approach for speech-based emotion recognition. Our findings demonstrate that the proposed method works better in terms of accuracy than current approaches. The proposed approach has the potential to be used in various applications, and can contribute to the development of more accurate and reliable speech-based emotion recognition systems.

II. LITERATURE REVIEW

[1] Sukanya Anil Kulakarne, Feature extraction and classifier training are required for emotion recognition from audio signals. In order to effectively train the classifier model to distinguish a certain emotion, the feature vector is made up of audio signal components that characterise speaker-specific properties including tone, pitch, and energy. The acted voice corpus of both male and female speakers from the English. Language open-source dataset RAVDESS was manually separated into training and testing. Mel-frequency cepstral coefficients, which we have derived from the audio samples in the training dataset, are used to represent speaker vocal tract information. With genuine speech recordings, we also put feature extraction into practice. We measured the energy and MFCC coefficients of audio samples representing various emotions, including neutral, rage, fear, and melancholy. The classifier model receives these extracted feature vectors. The extraction process for the test dataset will be followed by the classifier's determination of the underlying emotion in the test audio. K Vaibhav.P [4] The earliest speech recognition system, created in 1952 by Davis at Bell Laboratories in the US, could identify male voiced numbers from 0 to 9. The obstacles associated with speech processing, such as continuous speech recognition and emotion recognition, were too great for researchers to overcome. Not just through facial expressions, but also through speech, emotions can be understood. Every human's speech is accompanied by an emotion. Emotions are crucial because they let a person comprehend their feelings. Speech reveals a person's feelings, such as happiness, sadness, etc. [5] Elsevier B. V.

©IJRASET: All Rights are Reserved

But, behind facial cues, audio cues are the most frequently used information to determine an individual's emotional state. We merged all the techniques into a single input vector in order to increase the identification rate. Because these techniques are more frequently employed in speech recognition and have high recognition rates, we decided to use the coefficients MFCC, ZCR, and TEO in our investigation. And secondly, we suggested using an auto-encoder to minimise the input vector dimensions in order to optimise our system. Support vector machines were employed (SVM).

Using the RML database, our system is assessed. Ankitha Mathew Chinnu [3] One of the most often used study areas is speech processing. There are numerous researchers working on various speech processing systems all over the world. In 1920, the company Radio Rex produced a celluloid toy that marked the beginning of speech processing. This toy was the first speech recognition device that relied on the 500 Hz acoustic energy that the vowel "Rex" releases. The earliest speech recognition system, created in 1952 by Davis at Bell Laboratories in the US, could identify male voiced numbers from 0 to 9. The obstacles associated with speech processing, such as continuous speech recognition and emotion recognition, were too great for researchers to overcome. [2] S . Padmaja Karthik, In recent times, the significance of understanding human speech emotions has grown in order to enhance the effectiveness and naturalness of human-machine interactions. The difficulty in differentiating performed and natural emotions makes it a very difficult task to recognise human emotions. In order to correctly determine emotions, experiments have been done to extract the spectral and prosodic elementsWeprovided an explanation of the classification of emotions based on calculated bytes utilising human speaking utterance. How to categorise the gender using estimated pitch from human voice was explained by Chiu Ying Lay et al. Acoustic cues from speech can be extracted in order to identify emotions and classify them, according to Chang-Hyun Park et al. Nobuo Sato et al. provided information on the MFCC technique. Their primary goal was to use MFCC on human speech and accurately classify emotions with over 67% accuracy. In an effort to improve accuracy, Yixiong Pan et al. applied Support Vector Machines (SVM) to the problem of emotion classification. With more than 60% accuracy, Support vector machines in neural networks were used by Keshi Dai et al. to recognise emotions. The implementation of speech-based emotion recognition utilising machine learning and deep learning concepts has been the subject of numerous articles. Humans have a wide range of heterogeneity in their capacity to identify emotion. It's crucial to remember that there are many sources of "ground truth," or details about what the real emotion is, when learning about automated emotion recognition. Take into account that we are trying to determine Alex's emotions.

"What would most people say that Alex is feeling?" is one source. The "truth" in this case may not be what Alex feels, but it may be what the majority of people would assume Alex thinks. For instance, Alex might appear pleased even when he's truly feeling depressed, but most people will mistake it for happiness. Even if an automated technique does not truly represent Alex's feelings, it may be regarded accurate if it produces results that are comparable to those of a group of observers. You can also find out the "truth" by asking Alex how he really feels.

This works if Alex is conscious of his internal state, is interested in conveying it to you, and is able to express it precisely in words or numbers. Yet, some people with alexithymia lack a strong awareness of their internal emotions or are unable to express them clearly through words and numbers. . In general, determining what emotion is actually present can be difficult, depend on the criteria that are chosen, and typically require retaining a certain amount of uncertainty. Due to this, we decided to examine the effectiveness of three alternative classifiers in this instance. Both regression and classification issues can be solved using the machine learning approach known as multivariate linear regression classification (MLR) [6]Gaurav Sahu , We used Machnine Learning models like Random forest,gradient boosting,Support Vector Machnies and Multinomial Naïve Bayes, Logistic Regression models to extract the emotion from the audio.

III. PROPOSED METHODOLOGY

In this proposed work, an audio file is given as an input. Here we trained a machine learning model which this, firstly it converts the .mp3 format files into the .wav format files. Using the mp32wav.py file. Where this will convert the .wav format. And Pydub library is used for converting the audio.

Created the embeddings(OpenI3) for both training and test data set using pretrained audio models with size of 512.

1) Libraries used: openl3, soundfile

2) There are different arguments present in openl3 which I changed to obtain different embeddings.

3) Best embeddings are when [input_repr=”m1l256”, hop_size=0.5, content_type=”env”]

For given audio files embeddings are of shape (N,512) . N depends on duration of audio.

I converted each embedding into N 512D embeddings and given same label to all those which I later used for training.

A. Training

KNN classifier with standard-scaler is used to train embeddings Libraries used – sklearn.K Fold Cross validation with different scaling methods, splits are also experimented.
we I experimented with different CNN models created embeddings from pretrained

B. Data visualization

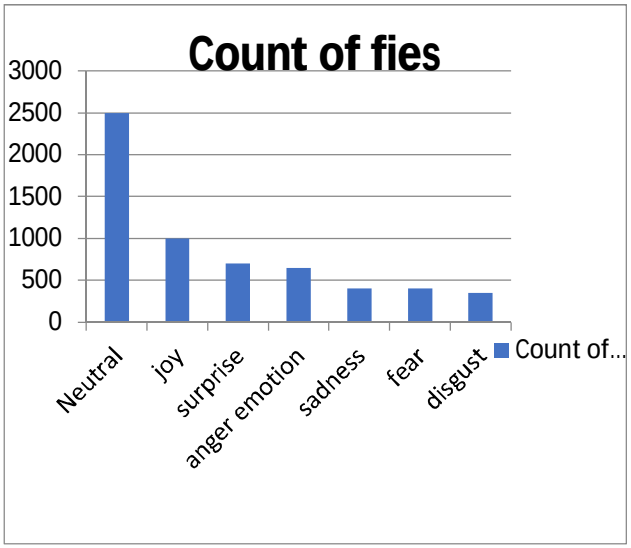


Fig.1 Before Preprocessing

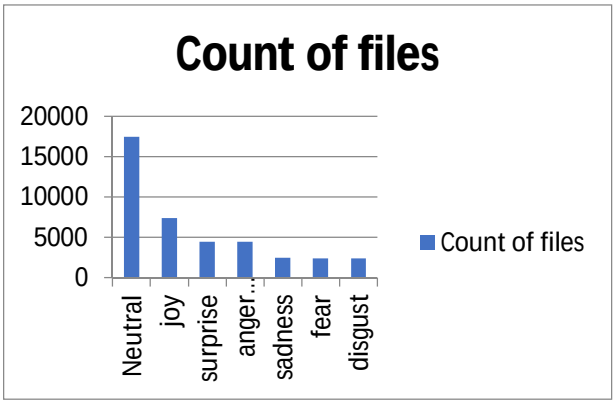


Fig.2 After Pre-Processing models like VGGISH, edg13. But open13 with KNN classifier gave good results.

IV. CONCLUSION

We suggested a machine learning method for identifying emotions from speech samples in this article. Using an openly accessible dataset of speech samples that had been emotionally labeled, we tested the effectiveness of these models. Our findings

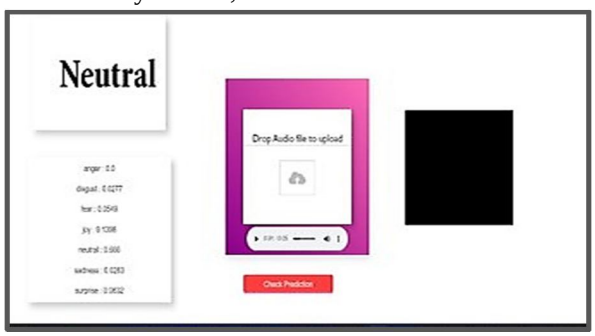


Fig.3 Result1

The main contribution of this paper is the development of a new machine learning approach for speech-based emotion recognition that can be used in various applications, including mental health diagnosis, education, and entertainment. Our approach has the potential to improve the accuracy and reliability of speech-based emotion recognition systems, which is important for these applications.In conclusion, our proposed approach shows promising results for recognizing emotions from speech samples using machine learning.

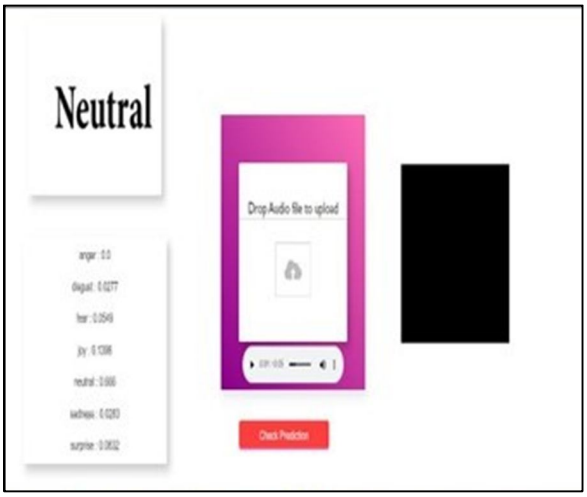
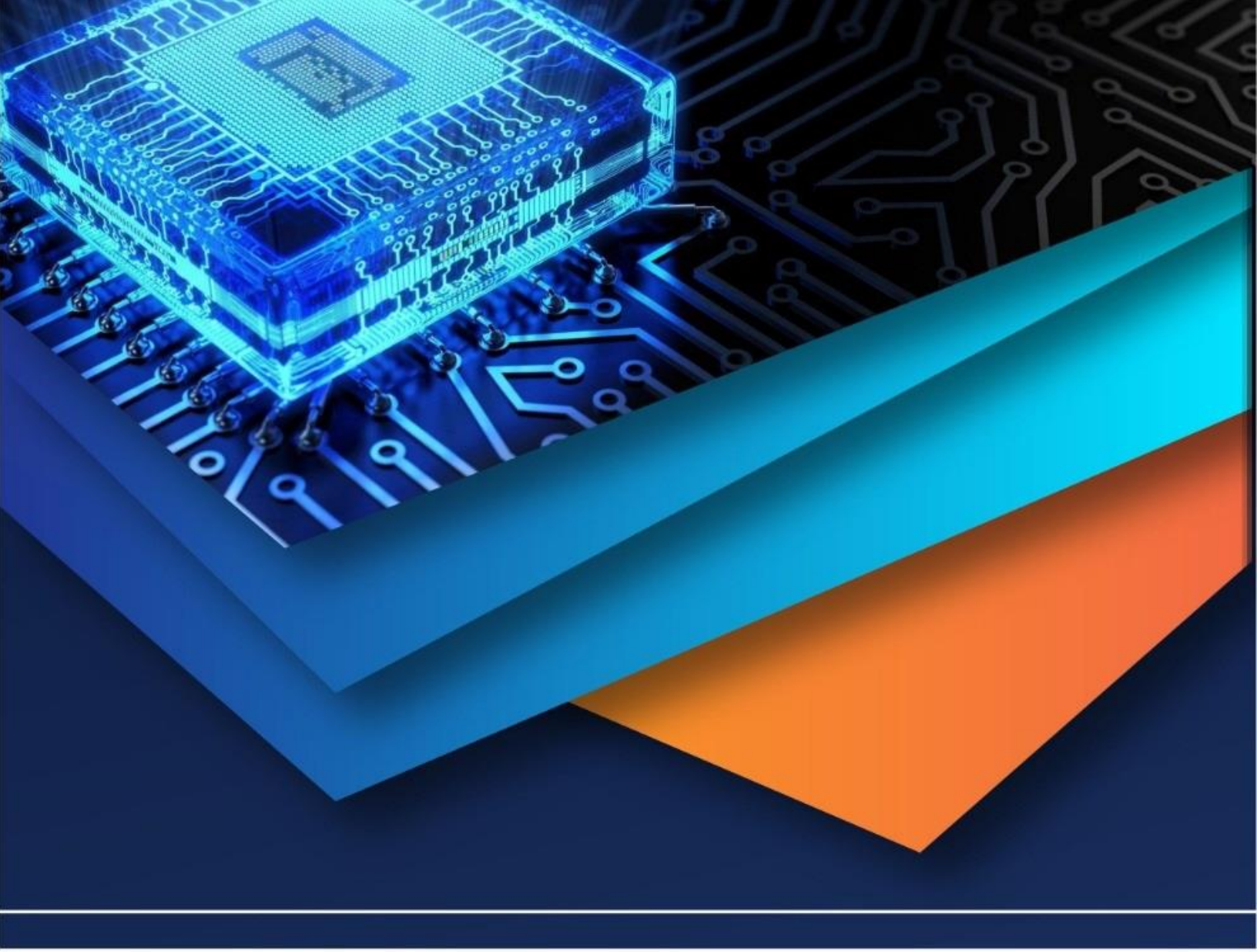


Fig.4 Result1

REFERENCES

[1] Sukanya anil Kulkarni , “Speech based Emotion Recognition machine Learning “ March 2019
[2] Mahalakshmi Selvaraj, R.Bhuva, S.Padmaja Karthik ,”Human Speech emotion recognition “ February 2016
[3] Amitha Khan K H, AnikithaChinnu Mathew, Ansu Raju, Navya Lekshmi M, Raveena R Maranagttu, Rani Saratha R ,”Speech Emotion Recognition Using Machnine Learning” 2021
[4] Vaibhav K.P,Parth J.M, Bhavana H.K, Akanksha S.S, ”Speech Based Emotion Recognition Using Machnie Learning” , 2021
[5] Elsevier B.V , “Speech Emotion Recognition with Deep Learning” , Procedia Computer ,2020
[6] Gaurav Saahu ,”Multimodal Speech Emotion Recognition and Ambiguity Resolution”,2019



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)