International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Depression Detection of Audio, Image and Text Using Machine Learning Technique

Mrs. Shikha Pachouly, Praveenkumar Saw, Pratik Nawale , Gaurav Khaire , Akshay Padwal

Mrs. Shikha Pachouly(Assistant Prof. at AISSMS College of Engineering Pune (M.E IT / PhD Pursuing)
Praveenkumar Saw/Computer Engineering (All India Shri Shivaji Memorial Society's College of Engineering Pune)
Pratik Nawale/Computer Engineering (All India Shri Shivaji Memorial Society's College of Engineering Pune)
Gaurav Khaire/Computer Engineering (All India Shri Shivaji Memorial Society's College of Engineering Pune)
Akshay Padwal/Computer Engineering (All India Shri Shivaji Memorial Society's College of Engineering Pune)

Abstract - Depression is common mental disorder that affect millions of people in world.in past few years, depression detection has gained popularity through its speech. However, many challenges remain which includes feature providing the best discrimination between classes and depression levels in this model we give different type of features for depression detection by using comparative analyses by using this same collection, the test base and speech base system is evaluated and how the system is built we found that performance can be increase by using combination of feature which is drawn from both speech and text. Nowadays there are lot of platforms like Twitter, Facebook, and Instagram for an individual to express their emotions which can be in the form of images, videos, audios and mostly through text. These social media platforms provide huge quantity of user data which can be used for explorative analysis. This textual data is most widely used data analysis which offers a bunch of characteristics and the state of emotion of an individual. So, we will be using Emotion artificial intelligence for emotion detection, which involves the field of data mining. Twitter, Facebook, YouTube are number of social media platforms. Emotions, opinions of majority people are expressed through their social media images, videos and text. Hence due to this presence of social media there is large user data available for explorative analysis. Mainly textual data is most widely used which gives number of characteristics. Hence it became best choice for data analysis, for emotion AI.

Key Words: Depression, Data, Audio, Video, Feature extraction.

1.INTRODUCTION

- **AIM: -**

This project will focus on feature development for depression detection by investigating how to build a detection system that extracts features from Text and speech and detect emotions using facial expressions. This work aims to discover which features provide the best discrimination between depression levels.

- **MOTIVATION: -**

The main motivation of this research is to make the interface which is simpler, interactive with a common person and more flexible. Depression is a common and major mental illness and spreading worldwide, which lead to suicides. Every year around globe, more than 300 million people are estimated to suffer from depression. Generally, depression is diagnosed through face to face by using clinical methods. However, at early stages of depression, 70% of the patients would not consult doctors, which may take their condition to advance stages. The developers and experts have more knowledge to develop such system which can visualize facial, Audio, Text Features.

2.OBJECTIVE

• To Predict mood level or activity based on score with class label.

• To successfully implement the test model based on training set as deep learning approach.

• To execute the proposed system maximum accuracy.

• The main goal of this project is to detect stress of person.

• The aim of the system to detect suicidal intent.

• To detect stress through speech analysis, to use these MFC coefficients is the computation of the mean and the standard deviation of each of them, instead of using them as a single feature which can lead to very large feature sets.

3. LITERATURE SURVEY

1. **Depression Detection Using Machine Learning Techniques on Twitter Data (IEEE 2021).** (*Kuhaneswaran A/L Govindasamy, Naveen Palanichamy*)

Description: -

This research is used to detect the depression of people on twitter data of people by analyzing the data such as comment text etc. Firstly, the data collected from twitter of any platform is gathered. Then that data is fed into two different classifier which is naïve baiyes and NB Tree which is hybrid model. Therefore, we can observe both dataset gives same accuracy level and perform equally.

Pros: -

- Fastest Algorithm and time to process is less easily predict the class
- Multi class prediction Problem Can be solved using this model.

Cons: -

- If your test data is not present or empty then the naïve baiyes model assign probability to zero and then the accuracy of model becomes less and cannot take data into consideration

2. Depression Detection by Analyzing social media Posts of User (2019 IEEE)

(*Nafiz Al Asad, Md. Appel Mahmud Pranto, Sadia Afreen, Md. Maynul Islam)*

Description: -

The paper aims to model that takes a username and perform analysis of the social media posts of the user to determine the levels of critical to depression. This model is trained to classify the depression criteria in six ranges including Considered Normal, Mild, Moderate, Borderline, Severe, Extreme. If a person is above 55% verdict is depressed. The tweets and post from Facebook and twitter is collected and analyzed by the model and label it whether it is depressed or not.

Pros: -

- Detects at various level of prediction and give more priority to particular person.

Cons: -

- Accuracy of model is less
- The limitation of language in this model

3. Recognition of Audio Depression Based on Convolutional Neural Network and Generative Antagonism Network Model (2020 IEEE)

(*Zhiyong wang, longxi chen , lifeng wang , and guangqiang diao****)***

Description: -

Using convolution neutral network and generative antagonism network model recognise audio depression. Data pre-processing is done then removed long term mute segment from dataset and merge the rest audio in new audio file. Then Mel-scale frequency cepstral coefficient (MFCC), spectral entropy, etc features are extracted using audio difference normalisation algorithm.

Pros: -

- in field of mental healthcare audio depression recognise has maximum importance.
- Values obtained from RSME and MAE gives 5% more accuracy than comparison algorithm.

Cons: -

- there should be long term muted audio cause decrease in accuracy
- CNN has several layers then the training process takes a lot of time when there isn't good GPU of your computer.

4. The Detection of Depression Using Multimodal Models Based on Text and Voice Quality Features **(2021 IEEE)**
(Hanadi Solieman, Evgenii A. Pustozerov)

Description: -

This paper aims to make multimodal models using text and voice to predict depression. In this study we create text analysis model using natural language processing techniques and create voice quality analysis model to analyze voice. The model analysis text and made its best performance with an F1-score equal to 0.8/0.5 for non-

depressed and depressed individuals, while the voice quality model scored 0.76 /0.38.

Pros: -

- it is easy to implement
- natural language processing system provides answers to the question in natural language.

Cons: -

- accuracy is low
- natural language processing system does not have a user interface

5. A Depression Recognition Method for College Students Using Deep Integrated Support Vector Algorithm. (2021 IEEE)

(*Yan ding, xuemei chen, qiming fu , and shan Zhong)*

Description: -

Due to the increase in the population the pressure on students in order to perform well while dealing with the difficult competition around them has affected the mental health of the students. The social networking sites like Weibo, QQ and WeChat have provided these students with platforms where they can interact with their friends and express their emotions and relieve their stress. The data which these social network platforms contain can be used for detecting the mental health of an individuals. This paper analyses the text of Sina Weibo from college students to detect the level of depression among them. This text info is converted into input data from which features are extracted using deep neural networks which are used in the machine learning model to predict depression among college students. A DISVM (Deep integrated support vector machine) algorithm is used to classify input data and find the results.

Pros: -

- requires less memory to process
- If Multiple dimensional datasets are given then this algorithm is efficient

Cons: -

- does not perform well for datasets with more noise i.e., target classes are overlapping.

6. Diagnosis of Depressive Disorder Model on Facial Expression Based on Fast R-CNN. (2022 MDP)
(Young-Shin Lee 1 and Won-Hyung Park)

Description: -
In this study, diagnosis of depressive disorder model on facial expression based on fast R-CNN Is used to detect depression from facial expression. This application provide user with the facial expression recognition feature that are used to measure the level of depression of the user. On the basis of the depression level the application would recommend the types of treatment best suited for the user.

Pros: -

- depression detection using facial expression recognition is to provide a self- detection tool for user to identify user depress or not.

Cons: -

- R-CNN detection is slow process.
- Long time for processing which does not work properly with large and big life datasets

7. Natural language processing methods for acoustic and landmark event-based features in speech-based depression Detection. (2019 IEEE)

(*Zhaocheng Huang, Julien Epps, Dale Joachim, Vidhyasaharan Sethu)*

Description: -

This paper framework model is created for analysis of acoustic and landmark event from speech to determining depression. Acoustic and landmark words are extracted from speech using mfcc and landmark extraction respectively. These extracted words are tokenised using n-gram algorithm. These tokens are extracted using BoW, W2V, LDA and these extracted tokens are reached to SVM algorithm that determines depression or non-depression.

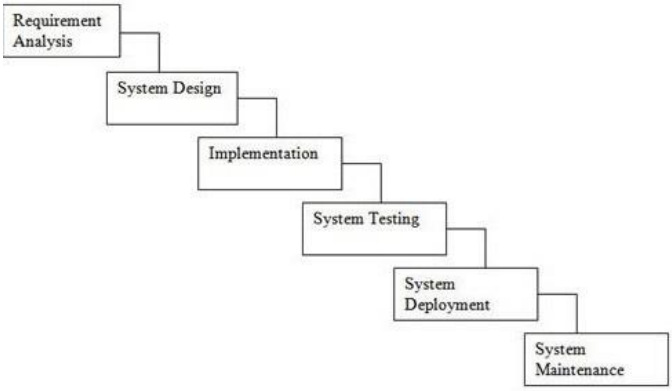
Pros: -

- this framework works for drastically different datasets (i.e., clean vs noisy for DAIC-WOZ and SH2-FS)
- different types of speech events are combined to analyse the depression

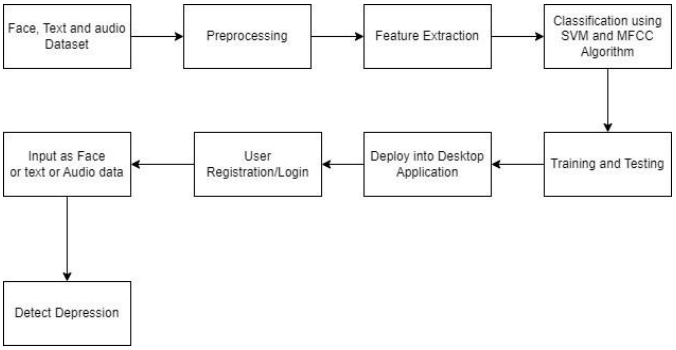
Cons: -

- it takes the short term depressional characteristics during emotional speech rather than long term characteristics associated with depression

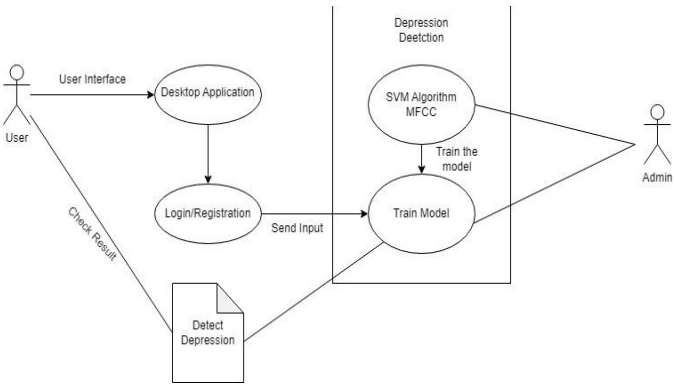
3.ANALYSIS MODEL :-



4.DFD DIAGRAM :-



5.USE CASE DIAGRAM:-



ALGORITHMS AND MODEL USED:-

- **CNN:-**

In deep learning for image recognition and tasks that involve the processing of pixel data one common network architecture is use that is CNN.

There are many types of neural networks in deep learning, but for identifying and recognizing objects, CNNs are the network architecture of choice

. CNN has many layers each layer used to detect different features of image. After each layer the output get better and more detailed. In the lower layers, the filters can start as simple features.

At each layer the complexity to check and identify features get increased Thus, the output of each convolved image -- the partially recognized image after each layer -- becomes the input for the next layer. In the last layer, which is an FC layer, the CNN recognizes the image or the object it represents.

Images goes through q set of these filters using convolution. As each filter detect certain features from the image, it does its work and passes on its output to the filter in the next layer. Each layer learns to identify different features and the operations end up being repeated for dozens, hundreds or even thousands of layers. Finally, all the image data progressing through the CNN's multiple layers allow the CNN to identify the entire object.

- **SVM:-**

Support Vector Machine or SVM is a Supervised learning algorithm which can be used to solve both Classification and regression problems.

In SVM the main aim is to create the best line or decision boundary which is helpful in segregating n-dimensional space into classes which would be helpful to classify new data point in the future. SVM is primarily used for binary classification but it can also be used for multi-class classification.

SVM use kernel functions to transform the data into a higher-dimensional space where classes can be easily separated. This allows SVMs to handle more complex problems which would be difficult to solve using simple linear separation.

SVMs have applications in image classification, bioinformatics and text classification.

- **MFCC (MEL FREQUENCY CEPSTRAL COEFFICIENT):-**

MFCC is used to study audio and understand various feature of audio.

It is used as an internal representation of audio.

In MFCC follow the following steps to convert audio to MFCC

1.get audio in time domain format

2.convert this audio in a periodogram using help of fast Fourier transform

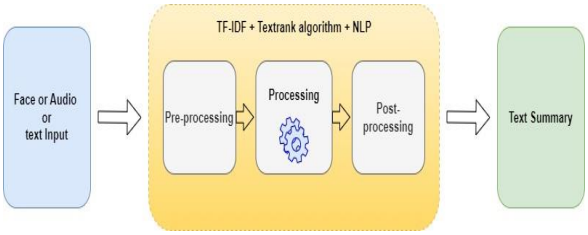
3.next step to convert periodogram into spectrogram

4.after that to perform hamming window to prevent spectral leakage.

5.after that we again perform fast Fourier transform to convert amplitude into frequency

MFCC is used for speech recognition and is used in music information retrieval

6.SYSTEM ARCHITECTURE :-



7. FEASIBILITY SCOPE: -

- The functionality analysis is a crucial reason for system experiments for general purposes.
- In earlier phases of Software integration, it is important to check whether or not the environment project is effective.

1. Technical Feasibility: -

Software requirements:

Programming Language: Python

IDE: Jupyter, Spyder

Hardware requirements:

Processor: Intel i5 and above

Operating System: Windows 10

Ram: 8 GB and above

2. Economic Feasibility: -

• Economic feasibility estimates projected costs to understand if they are within the projected budget or if the project has a reasonable return on investment

• only needs to be determined if feasible. project cost will be within target budget or investment profit.

3. Operational Feasibility: -

• Operational Feasibility tests the company's readiness to support the plan.

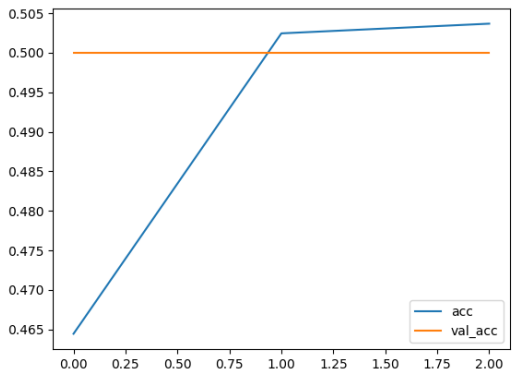
• This is probably the most difficult feature to extend.

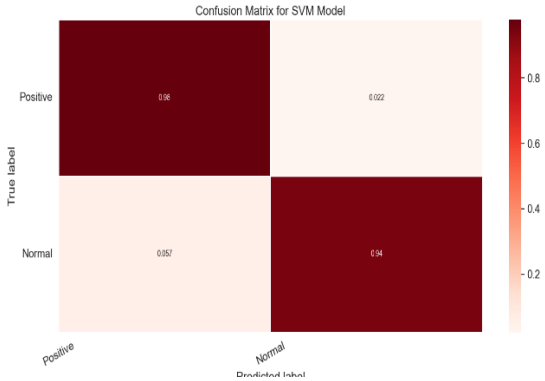
4. Time Feasibility: -

• Similar to the financial implications, it is important to take a rough guess of the business plan to determine if it is feasible within the critical time frame for the overall system project.

• The period is as follows. It is already specified in project task and the Gant map.

RESULT GRAPH:-





8. CONCLUSIONS

That in system we will detect person is depressed or not using Audio and text dataset. The Deep learning technique is used CNN algorithm and Machine Learning used SVM. With the help of These CNN algorithm and SVM algorithm to detect depressed person or not. We are using Harr cascade algorithm to detect facial expression. The accuracy of Depression detection using Face is 86%. The accuracy of Depression detection using Audio is 89% while as the accuracy using Text is 88% approximately.

9.ACKNOWLEDGEMENT

We would like to thank Mrs. Shikha Pancholy, our Project Guide-in-charge for their support and guidance in completing our review on the topic (Depression detection using Machine Learning Technique). It was a great learning experience. I would like to take this opportunity to express my gratitude to all of my group members Pratik Nawale, Akshay Padwal, Gaurav Khaire. The project review would not have been successful without their cooperation and inputs.

REFERENCES

- [1] Cohn J.; Kruez T.; Matthew, I.; Yang, Y.; Nguyen, M.; Padilla, M., Zhou, F.; Torre, F. Detecting Depression from Facial Actions and Vocal Prosody. ACII 2009 978-1-4244- 4799-2/09.
- [2] Hasan, M.; Rundensteiner, E.; Agu, E. EMOTEX: Detecting emotions in Twitter messages. Presented at the 2014 ASE Bigdata Socialcom Cybersecurity Conference, Stanford University.
- [3] Hadjipavlou, G.; Hernandez, C.; Ogrodniczuk, J. Psychotherapy in Contemporary Psychiatric Practice. Can J Psychiatry. 2015 60(6): 294300.
- [4] Wang, X.; Zhang, C.; Ji1, Y.; Sun,L.; Wu, L.; Bao, Z. 2013. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. 201213. In Proceedings of Springer-Verlag Berlin Heidelberg 2013
- [5] Paul, M., J.; Dredze, M. 2011. You are What You Tweet: Analyzing Twitter for Public Health. In Proceedings of ICWSM 2011.
- [6] Nadeem, M.; Horn, M.; Glen Coppersmith, G.; Sen, S. Identifying depression on Twitter. arXiv preprint 2016 arXiv:1607.07384.
- [7] Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems 26 NIPS 2013