DEPRESSION DETECTION USING FACE, TEXT AND AUDIO USING MACHINE LEARNING

Article · November 2022

CITATIONS
0

READS
205

7 authors, including:

Mrs.Kavita Sawant
Bharati Vidyapeeth's College of Engineering for women Pune
12 PUBLICATIONS **2** CITATIONS

SEE PROFILE

JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DEPRESSION DETECTION USING FACE, TEXT AND AUDIO USING MACHINE LEARNING

[1]K.S.Sawant [2]P.D.Kale, [3]S.P.Deshamukh,

Professor of BVCOEW

[4]Snehal Avhad, [5]Gitanjali Gadakh [6]Pooja Shinde, [7]Pooja Jagtap

BE Student at BVCOEW,

Department of Computer Engineering,

Bharati Vidyapeeth College of Engineering for Women, Pune, India

Abstract : Depression is a serious illness that affects millions of people globally. From child to senior citizens are facing depression. Major area is occupied by adults, college going students and teenagers also. In recent years, the task of automatic depression detection from speech has gained popularity. However, several challenges remain, including which features provide the best discrimination between classes or depression levels. We provide a comparative analyses of various features for depression detection. Using the same corpus, we evaluate how a system built on text-based , audio- based and speech-based system. We find that a combination of features drawn from both speech and text lead to the best system performance.
By doing a survey we have find most efficient algorithms for detection purpose. We have used CNN (Convolutional Neural Network) for Face images training, for Face recognition we have used Harr Cascade Algorithm. To detect depression using Text ,we have used SVM(Support Vector Machine) Algorithm. Lastly for Audio input, we have used MFCC for speech recognition.

Keywords **- Machine Learning, Face detection, Image preprocessing, segmentation, extraction, CNN, SVM, MFCC, Depression detection**

I. INTRODUCTION

Depression is a serious illness that affects millions of people globally. From child to senior citizens are facing depression. Major area is occupied by adults, college going students and teenagers also. In recent years, the task of automatic depression detection from speech has gained popularity. However, several challenges remain, including which features provide the best discrimination between classes or depression levels. We provide a comparative analyses of various features for depression detection. Using the same corpus, we evaluate how a system built on text-based and speech-based sys- tem. We find that a combination of features drawn from both speech and text lead to the best system performance. Nowadays people tend to express their emotions, opinions and disclose their daily lives through a variety of social media platforms like Twitter, Facebook and Instagram. This expression can be through images, videos, audios and mainly through text. Due to the widespread presence and reach of these social media platforms, there is a plethora of user data available for under taking explorative analysis. Textual data being the most widely used form of communication offers a bunch of characteristics which makes it the best choice for doing data analysis, for emotion AI. So, we are also using Emotion artificial intelligence field of ongoing research in emotion detection, specifically in the field of data mining. Depression detection from images alone, mainly de- pends on a clear and proper definition of a depressed face. The facial expression of a depressed face is slightly different from that of sadness. A depressed face expression has the same characteristics of a sad expression, such as the upward slanted eyebrows etc. but the main difference is that there is no major frown involved. Also a sad face may have eyes lowered looking downward showing the helpless, dejected mood. On the contrary a depressed person can put forth a face devoid of depression.

II. Motivation

Main motivation of this depression detection system is to detect depression symptoms in early stage of depression, so that person will get alert and take necessary action, consult doctor in early stage itself, so that later stages will be avoided, and last stage that is Suicide will get avoid. We want to create happy people with mental peace and avoid and detect depression. The main motivation

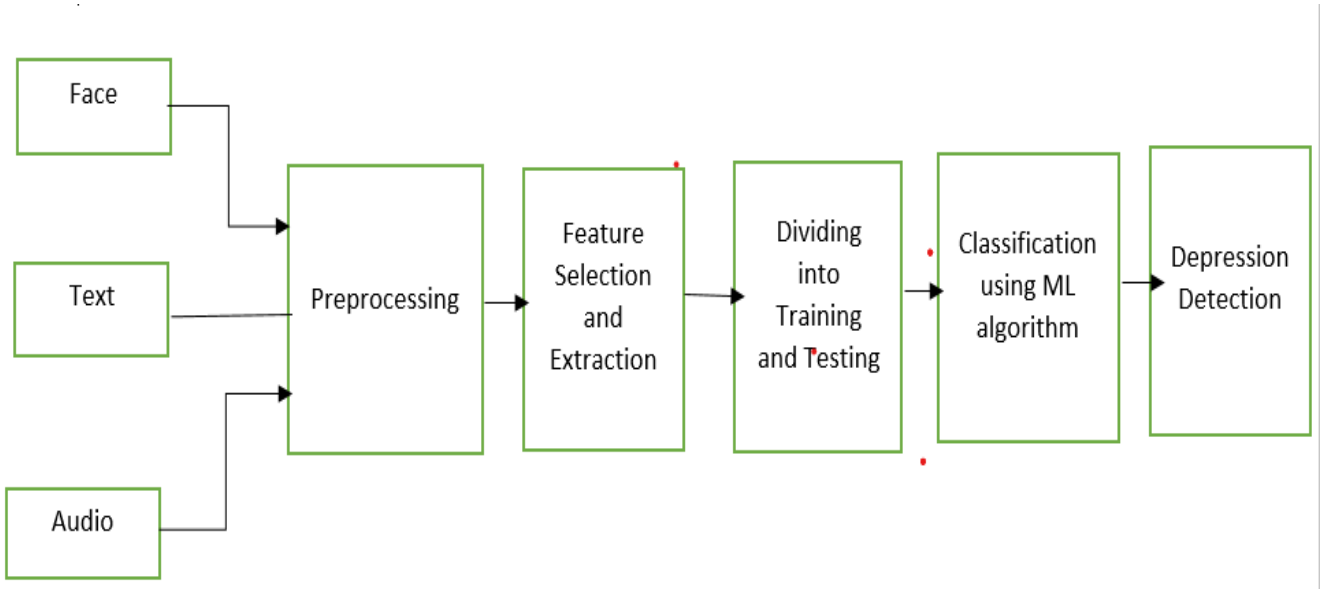
of such research is to make the man-machine interface more flexible and more easy for the user. Depression is a common mental illness and a leading cause of disability worldwide, which may cause suicides. Globally, more than 300 million people are estimated to suffer from depression every year. Generally, depression is diagnosed through face-to-face clinical depression criteria. However,at early stages of depression, 70percent of the patients would not consult doctors, which may take their condition to advance stages. .Human experts will have privileged knowledge that codes the facial, text and audio features.

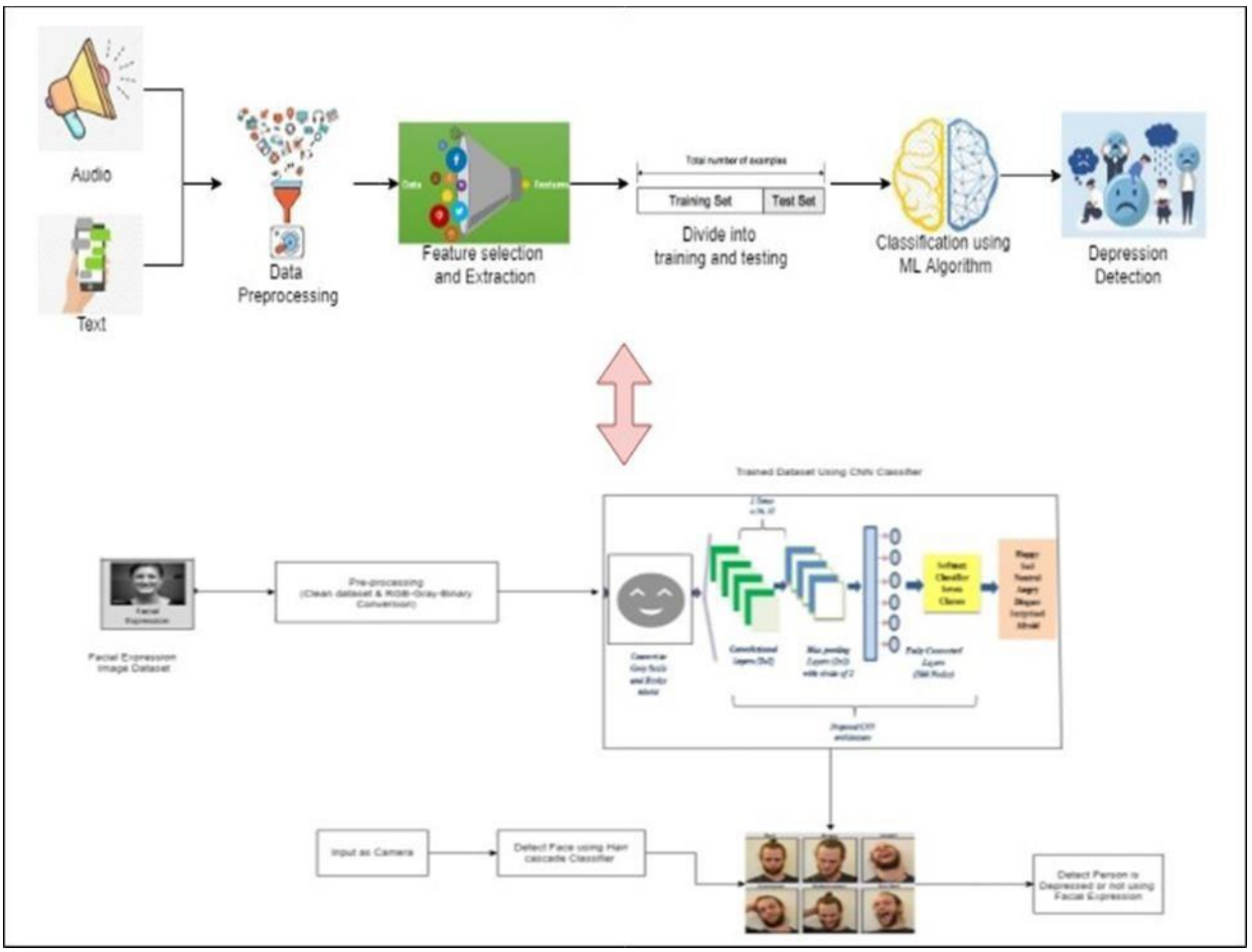
III . Literature Survey

SR. NO.	Paper	Year	Author	Description
1	Depression Detection Using Machine Learning Techniques on Twitter Data	2021 IEEE	Kuhaneswaran A/L Govindasamy, Naveen Palanichamy	The proposed research work aims to detect the depression of the user by their data, which is shared on social media. The Twitter data is then fed into two different types of classifiers, which are Naïve Bayes and a hybrid model, NBTree. The results will be compared based on the highest accuracy value to determine the best algorithm to detect depression.The results shows both algorithm perform equally by proving same accuracy level.
2	Depression Detection by Analyzing Social Media Posts of User	2019 IEEE	Nafiz Al Asad, Md. Appel Mahmud Pranto, Sadia Afreen, Md. Maynul Islam	This paper proposed a model that takes a username and analyzes the social media posts of the user to determine the levels of vulnerability to depression. The machine learning model is trained to classify the depression criteria in six ranges (Considered Normal, Mild, Moderate, Borderline, Severe, Extreme). The verdict is depressed when the percentage is above borderline (above 55%). The collected tweets and the facebook posts are analyzed by the model and labeled the user as depressed or non-depressed.
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	This paper proposes an audio depression recognition method based on convolution neural network and generative antagonism network model. First of all, preprocess the data set, remove the long-term mute segments in the data set, and splice the rest into a new audio file. Then, the features of speech signal,such as Mel-scale Frequency Cepstral Coefficients (MFCCs), short-term energy and spectral entropy, are extracted based on audio difference normalization algorithm.

SR. NO.	Paper	Year	Author	Description
7	Depression detection using emotional artificial intelligence and machine learning: A closer review	2022	Manju Lata Joshi a,⇑ , Nehal Kanoongo b	This study analyses how facial expressions, images, emotional chat_x0002_bots and texts on social media platforms can be effectual in detecting one's emotions and then depression. Naive-Bayes, Support Vector Machines (SVM), Long Term Short Memory (LSTM) – Radial Neural Networks (RNN), Logistic Regression, Linear Support Vector, etc. are the various ML techniques used to recognize emotions from text processing; Artificial Neural Network (ANN) is used for feature extraction and classifications of images to detect emotions through facial expressions. This paper aims to provide the survey of various AI and ML techniques which help in the detection and analysis of emotion and hence depression along with their related research issues.
8	Deep Learning for Depression Detection from Textual Data	2022	Amna Amanat 1 , Muhammad Rizwan 1,*, Abdul Rehman Javed 2 , Maha Abdelhaq 3 , Raed Alsaqour 4,*, Sharnil Pandya 5 and Mueen Uddin 6	This paper proposes a productive model by implementing the Long-Short Term Memory (LSTM) model, consisting of two hidden layers and large bias with Recurrent Neural Network (RNN) with two dense layers, to predict depression from text, which can be beneficial in protecting individuals from mental disorders and suicidal affairs.We train RNN on textual data to identify depression from text, semantics, and written content. The proposed framework achieves 99.0% accuracy, higher than its counterpart, frequency-based deeplearning models, whereas the false positive rate is reduced
9	Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions	2022	Shumaila Aleem 1 , Noor ul Huda 1 , Rashid Amin 1,* , Samina Khalid 2 , Sultan S. Alshamrani 3 and Abdullah Alshehri	This review paper enlists different machine learning algorithms used to detect and diagnose depression. The ML-based depression detection algorithms are categorized into three classes, classification, deep learning, and ensemble. A general model for depression diag_x0002_nosis involving data extraction, pre-processing, training ML classifier, detection classification, and performance evaluation is presented. Moreover, it presents an overview to identify the objectives and limitations of different research studies presented in the domain of depression detection. Furthermore, it discussed future research possibilities in the field of depression diagnosis.

IV . Proposed Architecture





Explaination of System Architecture

Input –Audio, Speech and text (Data Set): The first step of the system is to input an audio speech andtext

Pre-Processing: The second phase of the system deals with quality enhancement of the input signals of the audio speech and text. It may include silence removal, Preemphasis, noise removal, windowing and unwanted pauses, etc.

Feature Extraction: Feature extraction involves the analysis of the speech signal and text. The speech signal contains large number hidden information which reflects the emotional characteristics. It is considered as an important phase of the system as extraction of relevant and significant features heavily impact on the final recognition. Some of the featuresextracted by various researchers are MFCC (Mel-Frequency Cepstral Coefficients), LFPC(Log Frequency Power Coefficients), pitch, energy, and voice quality.

Classification: The fourth step is the main step of the system in which the audio speech and text is classified into different emotions based on the features extracted from the audio speech using CNN classifier. With the helpof the features extracted, the audio speech is classified into different emotions. Then detect person audio is stress or not.

Depression detection through emtions: A depressed face expression has the same characteristics of a Sad expression, such as the upward slanted eyebrows etc. but the main difference is that there is no major frown involved. Also a sad face may have eyes lowered looking downward showing the helpless, dejected mood. On the contrary a depressed person can put forth a face devoid of depression. This depicts a case of concealed expression of depression, i.e. the depressed face may not be a sad face, and instead the personmay put forth a happy face to conceal depression. Individual person are classified as neutral or negative,based on a curated word-list to detect depression tendencies.

V. Algorithms Used

1. Convolutional Neural Network: Convolutional Neural Networks specialized for ap- plications in image and video recognition. CNN is mainly used in image analysis tasks like Image recognition, Object detection Segmentation.

There are Four types of layers in Convolutional Neural Networks:

1) ConvolutionalLayer: In a typical neural network each input neuron is connected to the next hiddenlayer. In CNN, only a small region of the input layer neurons connect to the neuronhidden layer.

2) Pooling Layer: The pooling layer is used to reduce the dimensionality of the fea-ture map. There will be multiple activation pooling layers inside the hidden layer ofthe CNN.

3) Flatten: - Flattening is converting the data into a 1-dimensional array for inputtingit to the next layer. We flatten the output of the convolutional layers to create a singlelong feature vector.

4) Fully-Connected layer: Fully Connected Layers form the last few layers in the network. The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer.

2. Mel-Frequency Cepstral Coefficients

MFCC are popular features extracted from speech signals for use in recognition tasks. In the source-filter model of speech, MFCC are understood to represent the filter (vocal tract). The frequency response of the vocal tract is relatively smooth, whereas the source of voiced speech can be modeled as an impulse train. The MFCCfeature extraction technique basically includes 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. MFCCs are commonly used as features in speech recognition systems, such as the systems which can automatically recognize numbers spoken into a telephone. MFCCs are also increasingly finding uses in music information retrieval applications such as genre classification, audio similarity measures, etc.

3. Support Vector Machine

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algo- rithm builds a model that assigns new exmples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scal- ing exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

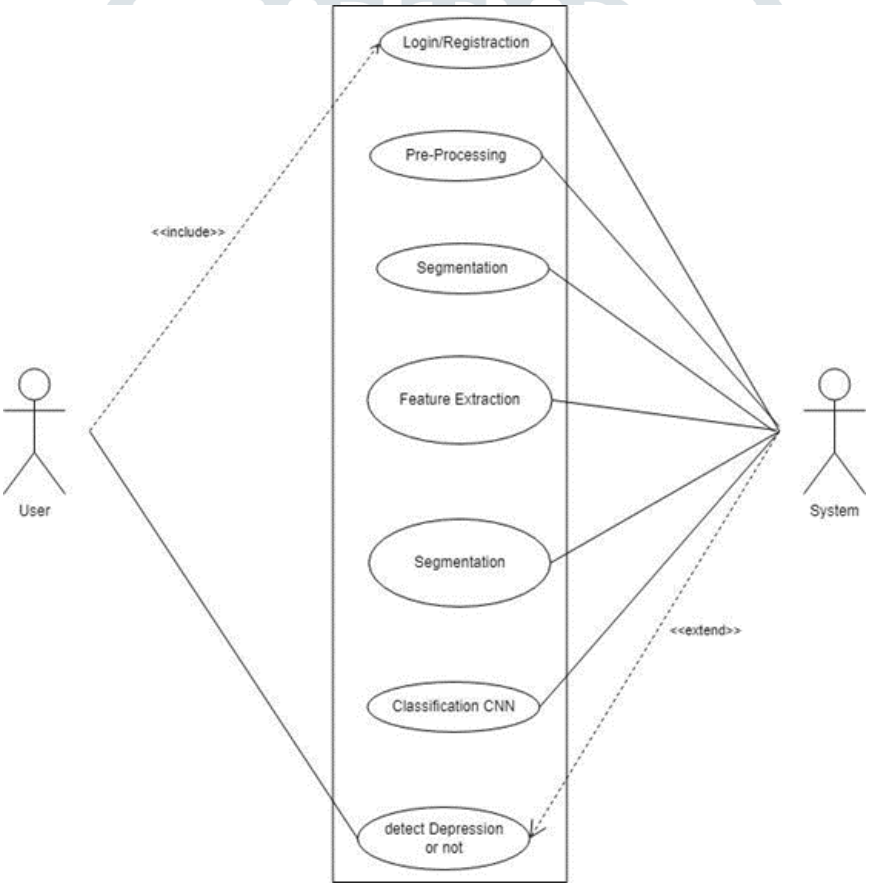
Two types of SVM

Linear SVM: Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

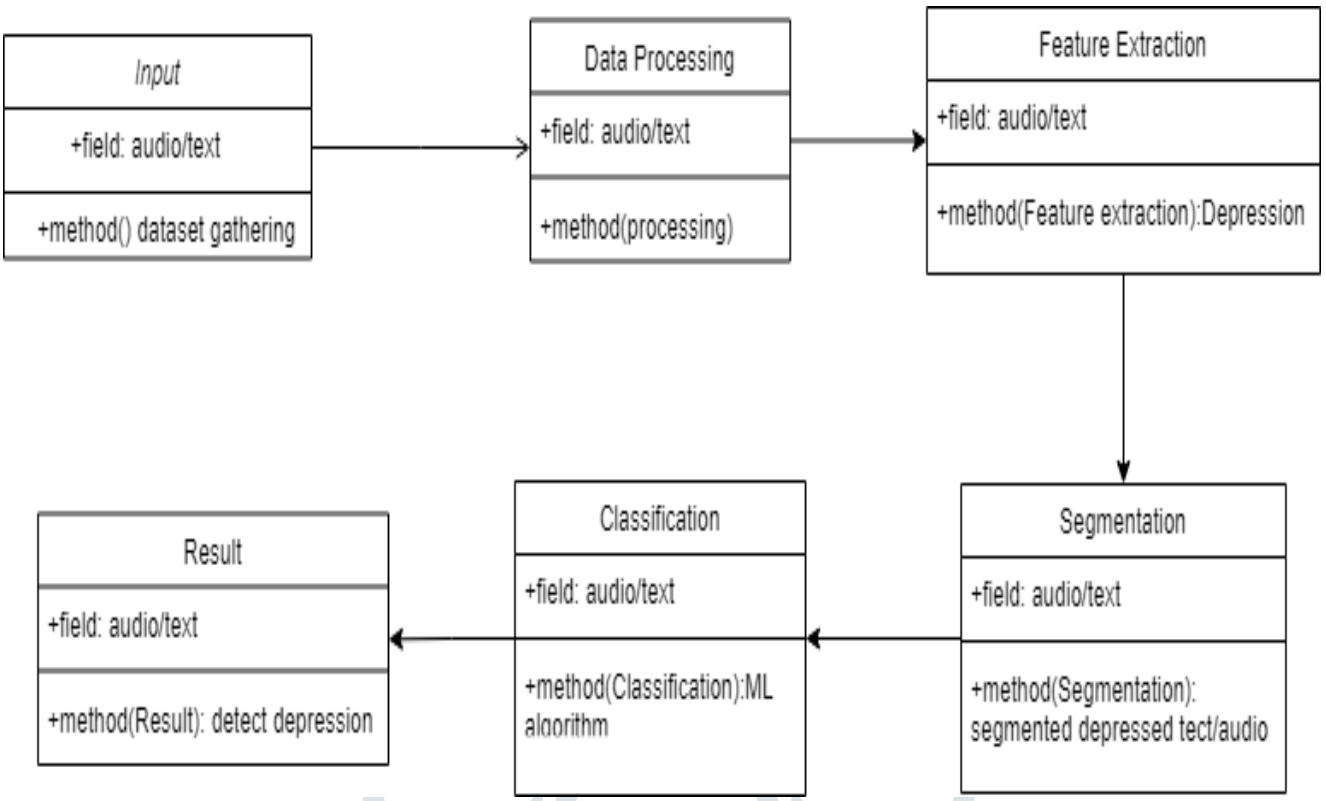
Non-linear SVM: Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier

VI. Diagrams

UML Diagram



Class Diagram



VII. Conclusion

In this project we have detected whether the person is depressed or not using Face, Text and Audio. We have used CNN (Convolutional Neural Network) for Face images training, for Face recognition we have used Harr Cascade Algorithm. To detect depression using Text ,we have used SVM(Support Vector Machine) Algorithm. Lastly for Audio input, we have used MFCC for speech recognition. The accuracy of Depression detection using Face is 88%. The accuracy of Depression detection using Audio is 89% while as the accuracy using Text is 87%.

References

[1] Kuhaneswaran A/L Govindasamy & Naveen Palanichamy , Depression Detection Using Machine Learning Techniques on Twitter Data. In 2021 Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021) ©2021 IEEE

[2] Nafiz Al Asad , Md. Appel Mahmud Pranto ,Sadia Afreen & Md. Maynul Islam. Depression Detection by Analyzing Social Media Posts of User. In 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems(SPICSCON) IEEE

[3] zhiyong wang , longxi chen , lifeng wang , & guangqiang diao. Recognition of Audio Depression Based on Convolutional Neural Network and Generative Antagonism Network Model. Digital Object Identifier 10.1109/ACCESS.2020.2998532 IEEE

[4] Hanadi Solieman & Evgenii A. Pustozerov . The Detection of Depression Using Multimodal Models Based on Text and Voice Quality Features. In 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)

[5] Yan Ding , Xuemei Chen , Qiming Fu & Shan Zhong. A Depression Recognition Method For College Students Using Deep Integrated Support Vector Algorithm. Published In 2020 IEEE

[6] Young-Shin Lee 1 and Won-Hyung Park 2, Diagnosis of Depressive Disorder Model on Facial Expression Based on Fast R-CNN, Diagnostics 2022, 12, 317.