IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning-Based Depression Classification Model

Himanshu Garud, Harsimran Singh Dhillon

Student, Student
Computer Engineering,
K. K. Wagh Institute of Engineering Education & Research, Nashik, India

Abstract: Depression is a serious mental condition that affects people of all ages and genders. Depression is caused by work culture, demanding living, emotional imbalance, family troubles, and social life. As a result, according to the World Health Organization's predictions, depression will become the second greatest cause of sickness (WHO). Despite the availability of well-trained physicians, medical and psychological treatments for depression, individuals and families are reluctant to speak out or contact doctors about the disease for a variety of social reasons. The diagnosis of depression disorder entails multiple interviews with patients and their families, clinical analysis, and questionnaires, all of which take time and require well-trained physicians. The detection of depression is automated using a machine learning approach. Extraction of acoustic and facial features becomes more efficient and accurate with the use of improved machine learning techniques such as the MLP algorithm and the DeepFace library.

Index Terms **- Speech processing, PHQ-9, Machine Learning, DeepFace, MLP Classifier**

1.INTRODUCTION

More than 264 million people of all ages suffer from depression around the world. Depression is a leading cause of disability worldwide and contributes significantly to the global disease burden [4]. Depression affects one out of every 15 adults in a particular area, with women having double the risk of males. People regard depression as a taboo subject, and they frequently dismiss it as a result of their sadness. This has an impact on their lifestyle and may result in a variety of health problems.

People frequently mix depression with sadness, which delays diagnosis and has a significant impact on their health. The depression classification model is a method for predicting the severity of depression in people using images, audio, and questionnaire data.

The goals are to develop a system that uses machine learning and deep neural network models to predict the degree of depression in people and to automate the diagnosis process to reduce the time necessary for diagnosis.

2.LITERATURE SURVEY

Depression is a severe mental condition that affects people of all ages and genders. Depression affects one in every 15 adults in a given area and risk in women is twice than men [4]. Symptoms include: change in food and habits, loss or gain in weight, loss in concentration, anxiety, hopelessness, feeling of less use, heart diseases, inflammation, sexual health problems, sleep disorder, etc [4]. Humans are social animals that have been trained to show only their strengths and hide their weaknesses, so they are afraid to discuss these issues publicly. The diagnosis of depression necessitates one's openness and honesty. The app's goal is to identify depression in an individual at any time and in any location. This application will use diagnostic and evaluation tools to diagnose depression using machine learning approaches, such as interview style assessment, automatic detection using speech, and face feature extraction from video. This application can detect depression, and the user could then be directed to a qualified and experienced physician.

• The PHQ-8 as a measure of current depression in the general population(2014)[3]:

The Patient Health Questionnaire (PHQ) is a self-administered questionnaire for patients. The PHQ evaluates eight diagnoses, which are divided into threshold disorders (disorders that correspond to specific diagnoses: panic disorder, other anxiety disorder, and bulimia nervosa) and sub-threshold disorders (disorders whose criteria include fewer symptoms than are required for any specific diagnoses: other depressive disorder, probable alcohol abuse/dependence, somatoform, and binge eating disorder).

• Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression(2019)[7]:

The PHQ-9 is the 9 depression module questions. Major depression is diagnosed if 5 or more symptoms out of the 9 depressive symptoms criteria have been present at least “more than half the days” in the past 2 weeks, and 1 of the symptoms is depressed mood or anhedonia [2]. Other depression is diagnosed when two, three, or four depressive symptoms have been present for at least "more than half the days" in the previous two weeks, with depressed mood or anhedonia being one of the symptoms [2]. If you have one of the nine symptom criteria ("thoughts that you would be better off dead or hurting yourself in some way"), it counts if you

have it for any length of time. Before reaching a definitive diagnosis, the doctor is expected to rule out physical causes of sadness, typical grieving, and a history of a manic episode, much like in the original PRIME-MD. Since each of the 9 items can be scored from 0 (not at all) to 3 (almost), the PHQ-9 score can vary from 0 to 27 as a severity measure. "How difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?" was added to the end of the diagnostic section of the PHQ-9 questionnaire, asking patients who marked off any problems on the questionnaire.

• Tracking depression severity from audio and video based on speech articulatory coordination(2019)[2]:

We integrate face features and acoustic features to obtain optimum accuracy. These extracted features are then cross-checked against PHQ-9 questionnaire scores, and the final result is evaluated. Using the COVAREP toolkit, various acoustic features are retrieved and fused to improve classification effectiveness. Principal Component Analysis (PCA) is used to choose features. Different speech kinds and emotions are used to diagnose depression using machine learning methods such as K Nearest Neighbors(KNN), Gaussian Mixture Model(GMM), and Support Vector Machine(SVM). Facial traits, expressions, and posture are derived from video and image data. In speech, feature engineering entails frame-by-frame feature extraction, statistical measure calculation, feature fusion, and feature selection based on correlation. COVAREP is being used to extract 73 baseline Low Level Descriptors (LLD), including Prosodic(2), Voice quality(8), and Spectral(63) features [6].

3. RESEARCH METHODOLOGY

3.1 Data and Sources of Data

The datasets used are REVDESS(Ryerson Audio-Visual Database of Emotional Speech and Song) and TESS(Toronto emotional speech set). The dataset of REVDESS consists of about 24.8 GB of song and speech, which is accessible from Zenodo. It features 24 trained actors—12 male and 12 female—who each deliver two lexically related lines with a neutral North American accent. The sample speech includes 7 expressions which are calm, happy, sad, angry, fearful, surprise, and disgust, and the song contains calm, happy, sad, angry, and fearful emotions[10].

TESS has 2800 audio files in total. Two actresses (26 and 64 years old) recited a set of 200 target words in the carrier phrase "Say the word _," and recordings of the set expressing each of the seven emotions (anger, disgust, fear, happiness, pleasant surprise, sorrow, and neutral) were made. The format of the audio file is a WAV format[11].

For this project the dataset is built using 5252 samples from the above mentioned two dataset[9].

3.2 System Architecture

When signing up, the user must provide information such as their name, email address, gender, and age. Following registration, the user must complete the PHQ-9 Questionnaire, which consists of nine questions based on symptoms experienced in the previous 14 days. The replies to these nine questions are fed into a rule-based feature categorization engine.

The user's facial and audio samples are collected in a model in which the user must read a paragraph after granting access to their microphone and camera. Every 3 seconds, the camera would capture a frame, which would be saved in the database. This procedure is critical for determining the user's feelings. The mic also collects an audio sample in addition to the facial features. The audio sample would last approximately 45 seconds. The pre-processing/feature extraction block receives the collected user data and performs the necessary normalization and standardization. The pre-processing phase is essential since it improves the accuracy of feature extraction. The image and audio models both require various features. These categorization features are retrieved individually for the image and audio models.

The collected characteristics are fed into the DeepFace and MLP classifier models, which classify the severity of the user's depression. The MLP classifier is used to categorize the user's acoustic sentiment from audio samples, while the DeepFace model is used to classify the user's facial mood. This model assigns the result to one of the five severity levels of depression. This will assist the user in determining the severity of their depression so that they can receive appropriate treatment.

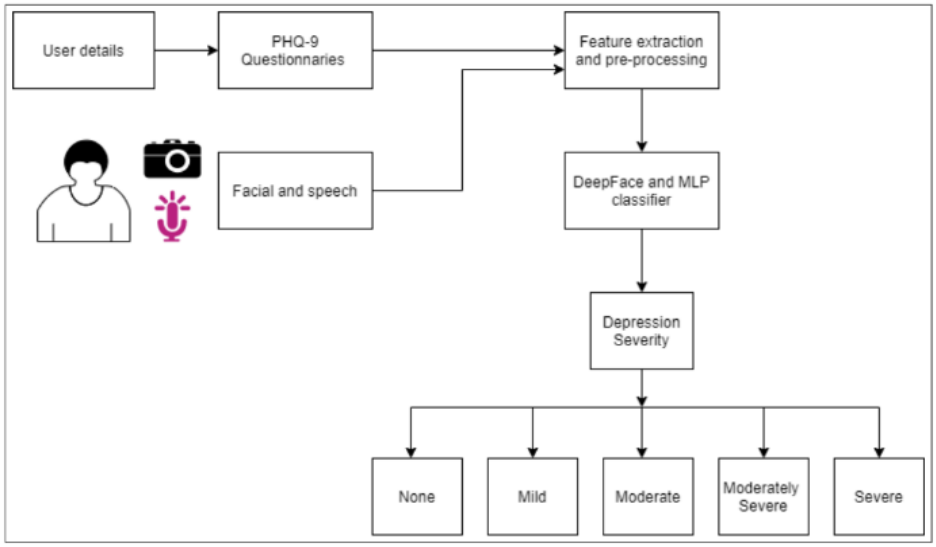

Figure: Architecture diagram

3.3 UML Diagram

The system's UML diagram is shown below. To access the PHQ-9 questionnaire, the actor must input proper login credentials. After completing the PHQ-9 questionnaire, the user is redirected to the facial and speech model, which collects the user's facial and speech samples. The features would be extracted, and the severity would be classified using the model.

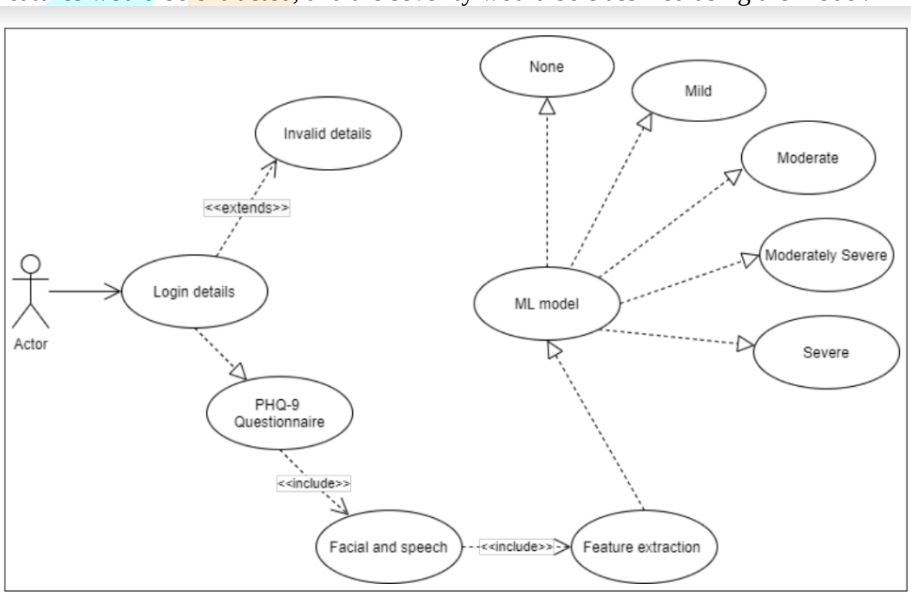

Figure: UML diagram

3.4 Overview of project modules

- **Questionnaire**

  This module is a rule based learning model. In this module, the user must answer a series of questions created by professionals and based on the user's previous 14 days of experience. The rule-based learning algorithm then categorizes an individual's depression severity. This model has an 88 percent specificity rate and an 89 percent sensitivity rate

- **Facial**

  In this module the system collects the facial features of the user. Required preprocessing steps are applied to remove noise and disturbance from the collected samples. These samples are then given to the applied DeepFace algorithm which classifies the facial features and gives the dominant facial
  feature as output out of all the samples collected

- **Acoustic**

  This is the system's final module, in which the user must read a paragraph and the system will collect the user's audio samples. The user's voice is captured using pyaudio, and the relevant features are extracted with the MLP classifier to classify the user's speech. The final output is generated by mapping the categorized speech output with the face expression classification and the questionnaire classification

3.4.1 Model for MLP Classifier

MLP classifier is used for classifying the speech emotion classification. The dataset used for classification are RAVDESS and TESS. The speech samples are collected dynamically using pyaudio. The features of the speech samples were extracted using the librosa library. The functions defined for extraction of speech features are : MFCC, MEL, CHROMA. The dataset is loaded and is split into 75 and 25 percent for training and testing the model respectively. MLP Classifier is initialized and trained. Collected speech samples are tested for classification using this trained model.

3.4.2 Model for DeepFace

DeepFace is used for classifying the facial features of the user. DeepFace in default configuration compares faces with the VGG-Face model[8]. Cosine similarity metric is used for finding out the similarity or differences. OpenCV is used to capture the camera frames at run time. Haarcascade frontal face is used preprocessing the collected sample data. DeepFace analyses the test sample data and classifies the output.

4. RESULTS AND DISCUSSION

4.1 Outcomes

- PHQ-9 questionnaire classifies the severity of depression with a sensitivity of 89 percent
- DeepFace classifies emotion of the person with an accuracy of around 84 percent
- MLP Classifier classifies emotion using acoustic features with an average accuracy of 85 percent

4.2 Results of DeepFace classifier

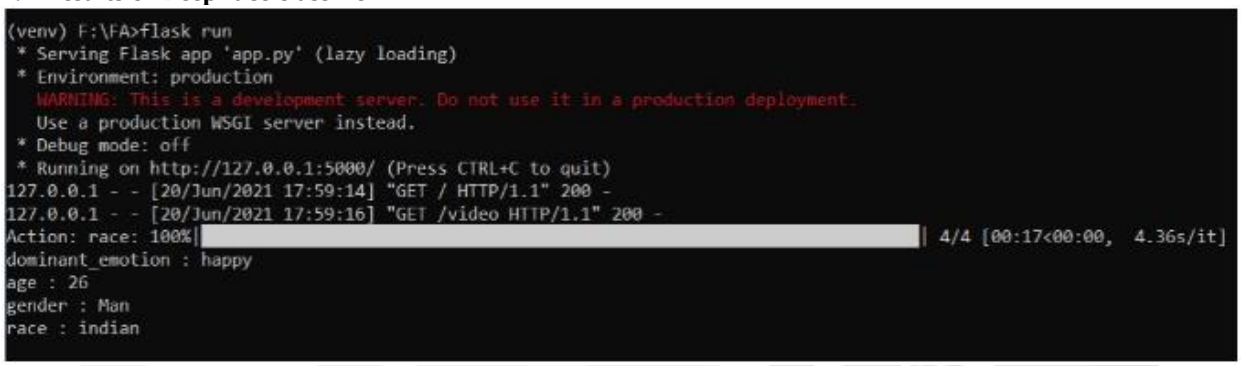

Figure: Result of DeepFace classifier

Above image displays the emotion, age, gender and race of the user. This model consists of 4 iterations and is based on real time classification, capturing pictures from the video and then using it to detect the emotion of the user. This whole classification process is achieved in around 3-5 seconds per iteration.

4.3 Results of MLP classifier

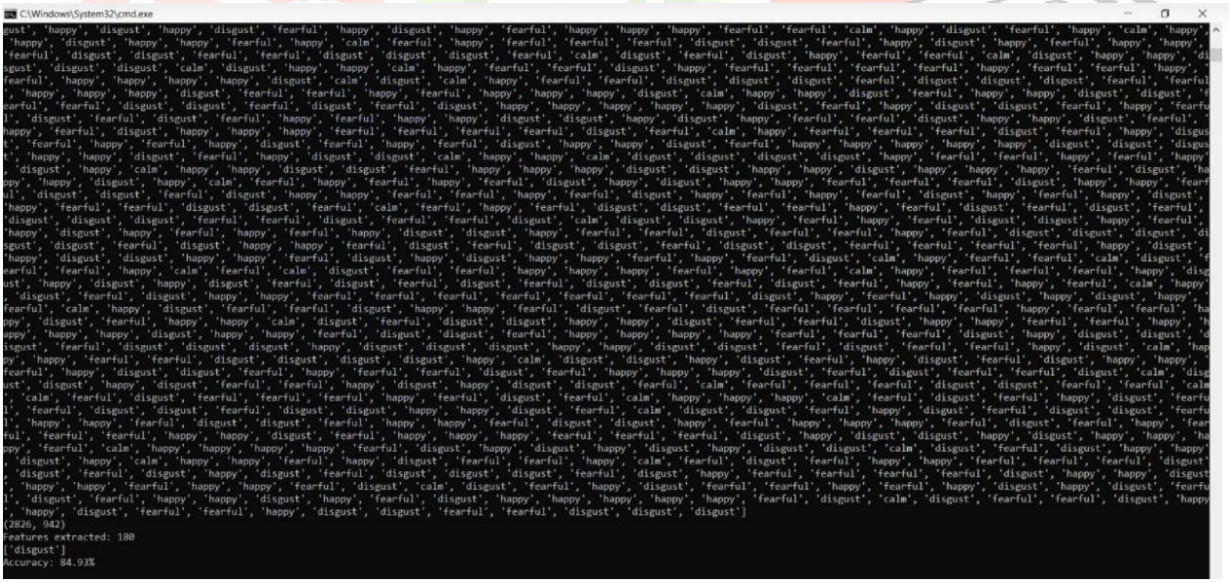

Figure: Result of MLP classifier

Above result provides the features which were extracted from the audio clip provided in run time by the user. This classifier was able to predict the emotion as disgust with an accuracy of 84.9 percent. These results along with face feature classification help us to predict the mood and emotion of the user with higher precision.

4.4 Conclusion

Depression is a major problem faced by many people in today's world. More than 264 million people from all ages suffer from depression. Unlike sentiments it is not visible in external appearance, so many of the people don't even know that they are suffering from depression. One of the major cause of suicides occurring is depression, so detecting depression in people is very important. This system will detect depression with higher accuracy, so people could do the treatment as soon as possible without more affecting their health. Depression reduces productivity so many companies could use this project to keep a track on their employees health condition and could help them get the treatment on time.

This system is divided into 2 parts, one is the PHQ-9 questionnaire and the other one is audio and visual samples collected while reading a paragraph. These 2 parts are very important to attain higher accuracy in determining the output. Output with less accuracy could create a false alarm so for higher accuracy these 2 steps should be done properly. With a good internet connectivity and good quality of camera and microphone, this application would be able to run efficiently with less lag possible.

REFERENCES

[1] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri "A review of depression and suicide risk assessment using speech analysis," Speech Communication, vol. 71, pp. 10–49, 2015

[2] J. R. Williamson, D. Young, A. A. Nierenberg, J. Niemi, B. S. Helfer, and T. F. Quatieri, "Tracking depression severity from audio and video based on speech articulatory coordination," Computer Speech & Language, vol. 55, pp. 40–56, 2019.

[3] K. Kroenke, T. W. Strine, R. L. Spitzer, J. B. Williams, J. T. Berry, and A. H. Mokdad, "The phq-8 as a measure of current depression in the general population," Journal of affective disorders, vol. 114, no. 1-3, pp. 163–173, 2014.

[4] https://www.who.int/news-room/fact-sheets/detail/depression

[5] https://www.kaggle.com/ashishbansal23/emotion-recognition

[6] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "Co-varep-a collaborative voice analysis repository for speech technolo- gies," 2014 ieee international conference on acoustics, speech and signal processing (icassp). IEEE, 2014, pp.960–964.

[7] Brooke Levis, doctoral student1, Andrea Benedetti, associate professor2, Brett D Thombs "Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression" Published 09 April 2019

[8] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato from Facebook AI Research Menlo Park, CA, USA and Lior Wolf from Tel Aviv University Tel Aviv, Israel, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification".

[9] Abhay Gupta, Aditya Karmokar, Khadija Mohamad Haneefa, Chennaboina Hemantha Lakshmi and Shivani Goel, "Identification of emotions from speech using Deep Learning".

[10] Livingstone SR, Russo FA (2018) The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PLoS ONE 13(5): e0196391. https://doi.org/10.1371/journal.pone.0196391.

[11]Toronto emotional speech set (TESS) (https://tspace.library.utoronto.ca/handle/1807/24487)