

Depression Detection Using Video and Speech

Noor Fathima Arfa¹, Pooja K², Prema³, Shivani Gahlot⁴

**Computer Science Engineering, B.M.S College Of Engineering, Bengaluru*

Abstract - Depression is a common mental health disorder affecting millions of individuals worldwide. Timely detection and intervention are crucial for effective treatment and management. This project aims to develop a system for depression detection using facial emotion and speech analysis, leveraging advancements in computer vision and natural language processing.

The suggested approach employs cutting-edge machine learning methods to examine speech patterns and facial expressions to look for potential depression indicators. The process of facial emotion analysis entails the extraction of face features from still photos or moving pictures in order to identify important emotions including grief, happiness, and neutral expressions. Speech analysis focuses on removing words, categorizing them as positive or negative, and then using that information to analyze depression.

By combining facial emotion and speech analysis, this project aims to develop a non-intrusive and accessible tool for depression detection. The system has the potential to assist mental health professionals in early identification and intervention, improving the overall well-being and quality of life for individuals at risk of depression.

Key Words: Depression, Xception CNN, Naive-Bayes Classifier, HAAR Classifier, NLP, FER Datasets, Google Speech API.

1. INTRODUCTION

1.1 Overview :

This depression Detector is a web-based tool designed to detect potential signs of depression using advanced facial emotion and speech analysis techniques. It aims to complement existing mental health services by providing an easily accessible and preliminary screening tool. The various modules include:

Website Application : The website offers a user-friendly interface where individuals can sign in and upload a video in mp4 format which is used in analysis for depression. It handles user input data and services requests on the user side.

Server Application: The server responds to inquiries from the website application. It executes the machine learning models and delivers the website's output. It also handles data storage for user data. Figure 1 shows an overview of the system.

1.2 Motivation:

Depression is a serious mental health problem that affects people all over the world, and its effects on people's lives and society at large cannot be overstated. Innovative approaches that can support the early identification and treatment of depression are urgently needed.

First of all, conventional approaches to diagnosing depression sometimes rely on subjective self-reporting or drawn-out clinical evaluations, which can lead to under diagnosis or under-treatment. The study provides a non-intrusive and objective method to identify probable indicators of sadness by merging face emotion and voice analysis.

The second critical issue is access to mental health services. Many people encounter obstacles such as a lack of access to qualified professionals, limited resources, or social stigma. The project's web-based design solves this accessibility issue by offering a user-friendly interface that is accessible to people from all backgrounds.

Last but not least, early detection and intervention are necessary to improve the results of mental health for those who are depressed. Early detection of probable depression symptoms enables people to obtain prompt assistance and treatment, improving both their condition's management and quality of life.

2. PROPOSED SYSTEM

A model for detecting depression uses a person's speech and facial expressions to make the diagnosis. FER (Facial Emotion recognition dataset) is the dataset used for facial emotion recognition. Grayscale portraits of faces measuring 48 × 48 pixels make up the data. The faces have been automatically registered so that each face roughly fills the same amount of space in each image and is roughly centered.

The suggested system's key components are as follows:

Video capture : Video of a subject is uploaded by the user after registering and logging in.

Haar feature classifier : It is used so as to detect if there is a face in the uploaded video. Haar-Features are good at detecting edges and lines. This makes it especially effective in face detection.

The Xception CNN model, which was initially created for picture classification tasks, has displayed outstanding performance in recognizing face emotions. The Xception model successfully extracts characteristics from facial photos

and accurately predicts emotions by utilizing its depthwise separable convolutions and transfer learning capabilities.

Naive Bayes is a probabilistic classifier that can be used for positive and negative word recognition. It calculates the conditional probabilities of words belonging to these categories and assigns a label during training which helps in the classification during testing.

Natural Language Processing (NLP): The Natural Language Toolkit (NLTK) is a well-liked open-source Python package for NLP and machine learning. It offers a wide range of resources and tools for text processing, including tokenization, stemming, tagging, parsing, and classification techniques.

3. IMPLEMENTATION

Haar cascade Classifier for Face Detection:

For face detection in this system, we employ the Haar classifier technique. The algorithm permits the face candidate to move on to the next phase of detection when one of these features is discovered. A sub-window, which is a rectangular portion of the original image, is a face candidate. These sub-windows normally have a fixed size (usually 2424 pixels). This Sub-window is frequently scaled to produce faces of various sizes. This window is used by the algorithm to scan the entire image, designating each region as a potential face candidate. The algorithm processes the Haar features of a face candidate in constant time using an integral image. It quickly eliminates non-face candidates through a cascade of phases. There are numerous distinct Haar traits in each step. A Haar feature classifier categorizes each feature. The stage comparator can receive the output that the Haar feature classifiers produce. To evaluate if the stage should be passed, the stage comparator adds up the results of the Haar feature classifiers and compares the result with a stage threshold. The face candidate is declared to be a face if all phases are successfully completed.

Haar Feature Classifier :

In a Haar feature classifier, a feature's value is determined by the rectangle integral. In the Haar feature classifier, each rectangle's area is multiplied by its weight, and the sums are calculated. Various Haar feature classifiers make up a stage. The results of the Haar feature classifier are aggregated at each step and compared to the stage threshold by a stage comparator. Each level can have any number of Haar properties. Depending on the characteristics of the training data, each stage may have a variable number of Haar features.

Haar Features:

A Haar feature is composed of two or three rectangles. The most recent stage of Haar features are examined on candidate faces using facial scanning technology. To determine the value of each Haar feature, the area of each rectangle is multiplied by its appropriate weight before the results are summed.

Xception CNN model:

The Xception architecture uses a series of depth-wise separable convolutional layers, followed by a global average pooling

layer, and finally a fully connected layer. The depthwise separable convolutions consist of two separate operations: first, a depthwise convolution applies a single filter to each input channel, and then a pointwise convolution applies a linear combination of the outputs of the depthwise convolution to generate the final output channels. This approach reduces the number of parameters required to train the network while maintaining good performance.

Overall, Xception has been shown to perform well on a variety of image recognition tasks. Its combination of depth-wise separable convolutions and Inception-style modules makes it a powerful tool for image classification and object detection.

Naive Bayes model :

Naive Bayes model works by learning the probabilities of each feature (word) appearing in each sentiment class (positive, negative, or neutral) and then using these probabilities to make predictions on new, unseen text data.

Modules:

Login module:

This module is in charge of generating the user's account and keeping the system's results and recommendations.

Dashboard module:

Gives users a way to access the depression detection system, which includes a function that lets them take pictures with their laptop's built-in camera and choose which ones to process for the other modules. Additionally, the dashboard section has a questionnaire that users can complete for test analysis.

Face Detection module:

The FER dataset[8] and HAAR feature[6] based cascade classifier are loaded by this module. It accurately recognises the frontal face in an image. Compared to other face detectors, it is real-time and quicker. We employ an OpenCV implementation. Module for expression detection:

This module makes use of the Mini_Xception, 2017 CNN module from Xception.[8] We'll develop a classification CNN model architecture that uses a bounded face's (48*48 pixels) input to forecast the likelihood of seven different emotions in the output layer.

Expression Detection module:

This module uses an Xception CNN module (Mini_Xception,2017).[8] We will train a classification CNN model architecture which takes bounded face (48*48 pixels) as input and predicts probabilities of 7 emotions in the output layer.

Speech analysis module :

The speech analysis module detects depression from the words spoken by the subject of the video. The speech taken from the mp4 video is extracted and is fed into Google speech API which converts the given speech into text. This is then processed by nltk machine learning library

NLTK (Natural Language Toolkit) is a popular open-source library for natural language processing (NLP) and machine learning in Python. It provides a wide range of tools and resources for text processing and analysis, including algorithms

for tokenization, stemming, tagging, parsing, and classification.

After this the Naive Bayes model is used to classify positive and negative words that are spoken by the subject.

Result module:

The depressed is detected only when the video of subject is analyzed on regular intervals. Therefore a popup message is shown to come back if that number of days are not yet completed. A user stats graph is shown to depict the result over a long interval of time so that the user can easily understand. If depression is detected link to the nearby psychologist available is given so that the subject of the video can consult them.

Chat-bot:

Chat-bot is available for user to interact and share their problem to get solution based on their category of problem. The chat-bot is interactive and gives various resources like videos, books, quotes etc so as to tackle depression.

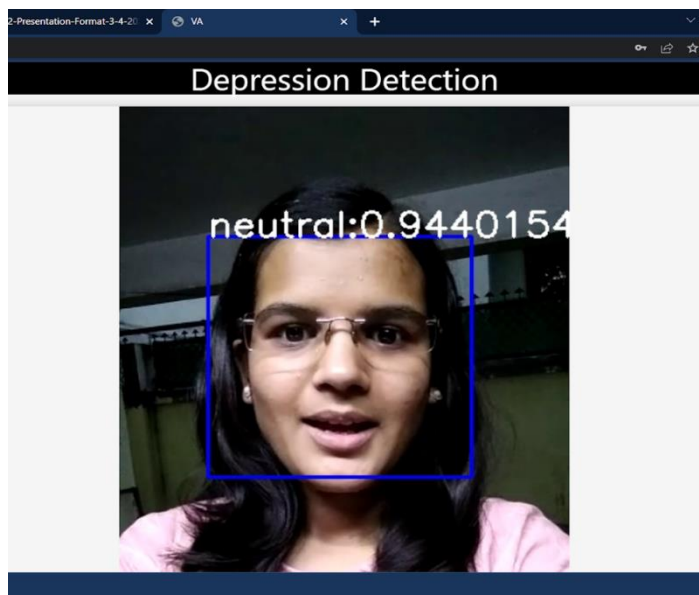


Fig -1: Facial Emotion Recognition

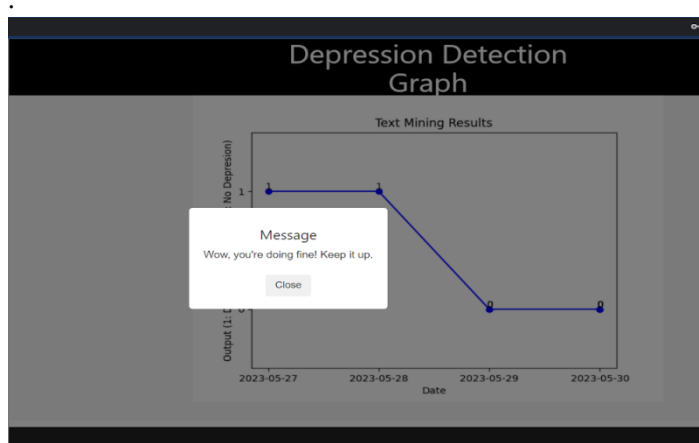


Fig - 2: Result

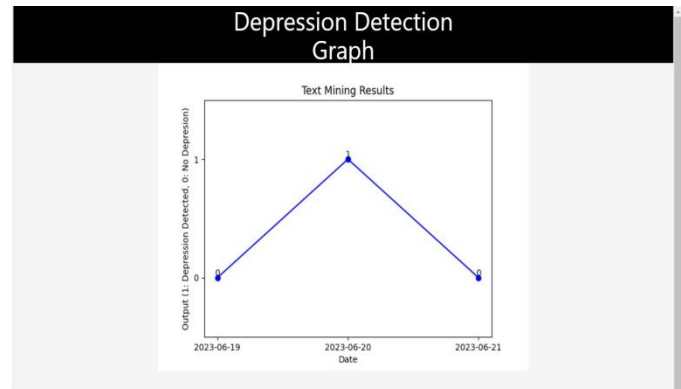


Fig - 3 : Statistics of detection(Graph)

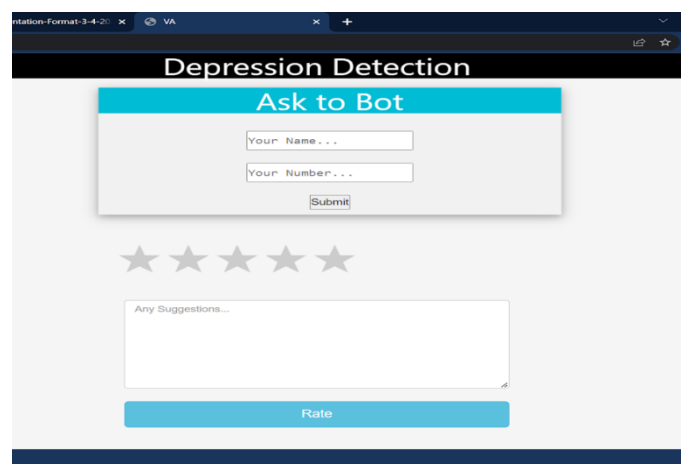


Fig - 4 : Chatbot Support

4. CONCLUSION

4.1 Conclusion

In conclusion, one of the major illnesses that jeopardize people's mental health is depression. A method of objective evaluation of depression based on intelligent technology has to be created in order to aid in early detection and patient treatment because the current traditional diagnosis methods have significant drawbacks. As objective indicators for the diagnosis of depression, speech and facial expressions are preferred since these traits of those who are depressed are in part related to their mental state. Depression is a common illness that affects people of all financial backgrounds, racial backgrounds, and nationalities. It is soon becoming an epidemic condition. Convenient remote diagnosis of depression can alleviate this problem. To do this, the presented model makes use of machine learning techniques.

4.2 Future Enhancements

The project can be further enhanced by individualized intervention strategies based on the individual's detected depression symptoms, through which the system can provide tailored recommendations. We can also foster partnerships with mental health professionals to integrate the system into their practises and improve the diagnosis and treatment processes.

REFERENCES

- [1]David William,"Text-based Depression Detection on Social Media Posts: A Systematic Literature Review",5th International Conference on Computer Science and Computational Intelligence 2020.
- [2] Hanai, Tuka & Ghassemi, Mohammad & Glass, James. (2018). "Detecting Depression with Audio/Text Sequence Modeling of Interviews".1716-1720.10.21437/Interspeech.2018-2522
- [3] Y. Yang, C. Fairbairn and J. F. Cohn, "Detecting Depression Severity from Vocal Prosody," in IEEE Transactions on Affective Computing, vol. 4, no. 2, pp. 142-150, April-June 2013, doi: 10.1109/T-AFFC.2012.38.
- [4] Akshay Doke, Shantanu Pawar, Shridhar Kengar, Sonali Pakhmode,"Analysis Of Text For Depression Detection Using Deep Learning ",International Research Journal of Modernization in Engineering Technology and Science,vol.4,March 2022.
- [5] Shephali Santosh Nikam, Aishwarya Vijay Patil, Gauri Shashikant Patil and Sharvari Pramod Patil
, "AI Therapist Using Natural Language Processing" in https://www.ijresm.com/Vol.3_2020/Vol3_Iss2_February20/IJRESM_V3_I2_187.pdf
- [6] Gintautas , TAMULEVIČIUS, Tatjana LIOGIENĖ, “Low-Order Multi-Level Features for Speech Emotion Recognition”, Baltic J.Modern Computing, Vol. 3(2015), No.4, 234-247
- [7] Albino Nogueiras,Asuncion Moreno, Antonio Bonafonte,Jose B.Marino, “Speech emotion recognition using hidden Markov models”, EUROSPEECH 2001 Scandinavia,7th European Conference on Speech Communication and Technology,2nd INTERSPEECH Event, Aalborg, Denmark September 3-7, 2001
- [8]Bhanushree Yalamanchili,Nikhil Sai Kota,Maruthi Saketh Abbaraju,“Real-time Acoustic based Depression Detection using Machine Learning Techniques”,2020 International Conference on Emerging Trends in Information Technology and Engineering.
- [9]He, L., Cao, C., “Automated Depression Analysis Using Convolutional Neural Networks from Speech”, Journal of Biomedical Informatics (2018).
- [10]A. Jan, H. Meng, Y. F. B. A. Gaus and F. Zhang, "Artificial Intelligent System for Automatic Depression Level Analysis Through Visual and Vocal Expressions," in IEEE Transactions on Cognitive and Developmental Systems, vol. 10, no. 3, pp. 668-680, Sept. 2018.
- [11]Xingchen Ma,Hongyu Yang,Qiang Chen,Di Huang,Yauhong Wang,"DepAudioNet: An Efficient Deep Model for Audio based Depression Classification " in IEEE Transactions on Audio, Speech and Language Processing, 22(10):1533–1545, Oct. 2020.