MENTAL HEALTH AND DEPRESSION DETECTION

NOOR FATHIMA ARFA
Undergraduate Student
B.M.S. College of Engineering
Bengaluru, India
noorfathima.cs19@bmsce.ac.in

Dr. Prasad G R
Professor
B.M.S. College of Engineering
Bengaluru, India
prasad.cse@bmsce.ac.in

POOJA K
Undergraduate Student
B.M.S. College of Engineering
Bengaluru, India
poojak.cs19@bmsce.ac.in

PREMA
Undergraduate Student
B.M.S. College of Engineering
Bengaluru, India
prema.cs19@bmsce.ac.in

SHIVANI GAHLOT
Undergraduate Student
B.M.S. College of Engineering
Bengaluru, India
shivani.cs19@bmsce.ac.in

Abstract- In many humanities-related professions, understanding a person's interior emotional condition is crucial. Emotions are a crucial part of our existence because of the profound influence they have on people's physical and mental wellbeing. Depression is a prevalent mental illness. The sluggish, hesitant, monotonous voice that is a notable trait of sad people may change as affective sensing technology advances with a focus on acoustic qualities.The system's goal is to use voice and facial expressions to classify emotions and depressive states. The system's goal is to use voice cues to classify emotions and depressive states. The system will employ

Haar based classifier and CNN model for facial emotion detection. The model for predicting depression based on the attributes is 3DCNN. The content of the speech itself is another aspect of speech that can be utilized to gauge a person's emotional state. It contains the crucial data that sentiment analysis can utilize to anticipate depression. You can convert speech to text using the Google Speech API. After feature extraction, the acquired

text can be trained and tested using an SVM model.

Keywords : Depression,3D CNN, SVM, DAIC-WOZ, MFCC feature extraction

I. INTRODUCTION

More than 264 million individuals worldwide suffer from depression, making it a widespread ailment.

Depression is a mental health disease that is primarily distinguished by a consistent lack of enthusiasm for activities or a pessimistic attitude. Early identification is necessary for quick intervention, which could stop the illness from getting worse. Despite the fact that known, effective treatments exist for such mental illnesses, between 76% and 85% of people in low- and middle-income countries do not receive care for their disease.

This absence of therapy is a result of a lack of resources and skilled medical workers (besides the obvious social stigma attached to these conditions). The largest challenge, though, is probably inaccurate assessment. Patients' depression is frequently misdiagnosed.

The majority of laboratory tests are unhelpful for identifying depression. In reality, the primary method for diagnosing depression still involves speaking with the patient. To test for depression, doctors ask a set of common questions. However, a diagnosis of clinical depression is frequently difficult to make because it can manifest itself in so many different ways. Depression only seldom and clearly manifests itself as a result. Setting up objective metrics becomes crucial during diagnosis as a result.

Speech is one of these objective criteria. Speech is appealing because it can be easily, cheaply, remotely, and non-intrusively measured. However, it has a great deal of diversity and is highly communicative.Prosodic speech problems in depressed people include monotonous pitch and loudness, repetitive pitch inflections and stress patterns, and diminished loudness variance. The likelihood of an incorrect diagnosis and social stigma could both be reduced by simply implementing this diagnostic technique in a wearable gadget or mobile phone application.

II. PROBLEM STATEMENT

Undetected mental problems can be very harmful in the long run. Because it can affect everyday living of an individual. It can also lead to physical side effects and tendencies for self harm.

The depression detection system's objectives are to identify depression and to make appropriate suggestions for modifying one's lifestyle.By the use of input parameters as speech , non verbal and verbal parts can both be used on models 3D CNN and SVM respectively so as to detect depression . This can help the user to keep track of their mental health and maintain a mentally healthy lifestyle.

III. LITERATURE SURVEY

A. *Key Terminologies*

1. **Acoustic Features** - They are any acoustic characteristics of a speech sound that may be captured and studied, such as its fundamental frequency or formant structure.
2. **Feature extraction** -It is the process of transforming unprocessed data into numerical features that can be further processed while retaining the details of the original data set.
3. **MFCC** - The MFCC technique aims to extract information from the audio stream that can be applied to voice recognition of phones.
4. **3D CNN** - Simply defined, a 3D CNN accepts either a series of 2D frames or a 3D volume as input (such as slices from a CT scan). For learning representations for volumetric data, 3D CNNs are a useful paradigm.
5. **API** - A group of definitions and protocols called an application programming interface, or API, are used to develop and integrate application software.
6. **SVM** - Support vector machines (SVMs) are a subset of deep learning algorithms that carry out supervised learning tasks such as data group classification or regression.
7. **RNN** - RNNs, a subset of artificial neural networks that use sequential or time-series data, are referred to as recurrent neural networks (RNNs).
8. **LSTM** - Deep learning and artificial intelligence both use long short-term memory (LSTM), a kind of artificial neural network.
9. **DAIC-WOZ** - 189 subjects, both male and female, who underwent psychological distress evaluation are featured in audio-visual interviews in the database known as the Distress Analysis Interview Corpus (DAIC-WOZ).

B. *Existing Systems*

There are numerous mechanisms in place for identifying depression. This could involve identification by contacting a psychiatrist, as well as machine learning-based methods such using social media language as a basis, eeg, eye movement, and eye blink rate. Many people fear to speak about themselves when consulting a psychiatrist, it is not cost-effective, and The complexity of each mental condition makes it challenging for psychiatrists to recognise the presence of mental illness in a patient, making it difficult to provide the patient with the necessary care before it's too late. Social media can only be used to identify a specific person's depression. Eye movement, EEG, facial expression, and eye blink rate when using a text-based technique are less accurate than speech recognition.

Since sadness is not always transient and might linger for a very long time, using facial recognition requires sensors like cameras, which makes it unusable. Although everyone needs an EEG system to identify sadness, the

EEG is not user-friendly despite being a powerful instrument. In order to increase the precision and reliability of the depression detection process, hybrid systems incorporate several methodologies, such as utilizing both speech and text approaches.

According to David William et al[1] .'s discussion on the text-based detection of depression, BiLSTM + Attention method outperformed methods like RNN and SVM for determining the appropriate context from a word in a long sentence, but this method still requires optimization like summarizer configuration or hyperparameter tuning to further improve the performance of this model.

Hanai et al. [2] reported a depression detection model based on audio and text transcription sequences. They experimented with a regularized logistic regression model (context-free and weighted) and a long short-term memory (LSTM) model (using the sequences of responses, and without context of the questions). Their findings showed that context-free modeling based on text characteristics outperformed audio features when classifying for a binary outcome (depressed vs. non-depressed). However, auditory components provided a more accurate prediction of the multi-class depression score. As expected, the multi-modal model performed the best.

The whole range of depression scores and the length of all audio recordings were used by Ying Yang et al [3].Depression score was the dependent variable, and prosodic characteristics were the independent factors. The method of supervised learning was applied. With the use of the Transcriber programme and CMU Sphinx III, each pair of recordings was manually transcribed before being force-aligned. More voice features must be employed in this situation to boost accuracy because the vocal features now being used are sparse.

Using deep learning, Akshay Doke et al.[4] were able to determine the Twitter user's stress level and attempt to offer the user a way out of the difficult circumstance they were in. They observed a strong correlation between language used in NLP (Natural Language Processing) and LSTM and depression. The research revealed that terms associated with anxiousness, melancholy, anxiety, rage, and suicide thoughts were included in linguistic forecasts for depression.

Shephali Santosh Nikam et al. [5] used natural language processing (NLP) to analyze the user's speech to determine whether he was stressed, furious, depressed, apprehensive, etc. The corresponding output is displayed using a recommender system. The user's emotions will be predicted by the web application AI Therapist. It is intended to analyze users' mental health difficulties and make recommendations regarding their issue. The user will behave as though he is speaking with a real person while communicating with a chatbot. The limit is that the responses provided by the created AI chatbot may not be tailored.

Speech emotions were divided up into stages with appropriate feature subsets by Gintautus et al. [6]. For the composition of feature subsets at various levels of classification, they used the most effective feature selection criterion. Experimental evaluation of the multi-level structure of categorization and features in two-, three-, and four-emotion recognition tasks was carried out with comparisons to standard feature combination strategies.

Albino Nogueiras et al. [7] use RAMSES, the voice recognition system of the UPC, to provide the first method for emotion recognition. A popular type of voice recognition technology, hidden semi-continuous Markov models, provide the foundation of the method. It is debated whether to employ low level features and how to create the recognition system. Results for speaker-dependent emotion recognition are shown using the Spanish corpus of the INTERFACE Emotional Speech Synthesis Database.

According to Bhanusree Yalamanchili et al.[8], the "DAIC-WOZ" data set used in this work has class imbalances and is accessible through the AVEC 2016 challenge. The disparity in availability between "depressed" and "not depressed" audio samples is the cause of the class imbalance. A biased machine learning model results from this. SMOTE analysis is employed to solve the class imbalance issue. The outputs of different categorization methods including SVM, Random Forest, and Logistic Regression are compared. SVM has performed well among all the other classifiers following SMOTE examination.

In order to model the characteristics of depression, He, L, Cao, C, and colleagues[9] employed the raw and spectrogram DCNN. They also recommended employing joint tuning layers, which integrate the raw and spectrogram DCNN, to improve the capacity to recognise depressions. Their methodology outperforms previous audio-based approaches for depression recognition, according to experimental findings on the AVEC2013 and AVEC2014 depression datasets.

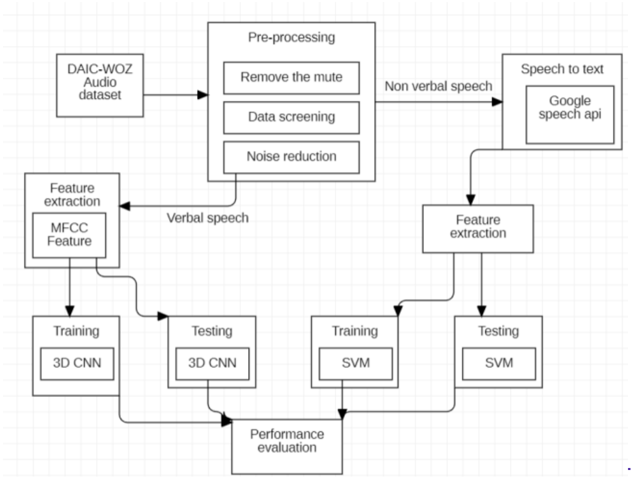
Cohn et al. [10], a pioneer in the field of emotional computing, conducted research on depression analysis in which he combined the visual and aural characteristics with behavioral considerations, which are closely related to mental diseases. Their findings suggest that it is possible to create a system that automatically recognises depression, which will benefit clinical theory and practice. According to research by Yang et al. into participant audio feature fluctuations, switching pauses and F0 together can predict depression scores to a limited extent.

According to Xingchen Ma et al.[11], a cutting-edge deep neural network called DepAudioNet is recommended for ADD. This hierarchical structure produces a complete audio representation by employing CNN to capture the short- and middle-term temporal and spectral correlations and LSTM to extract the long-term correlations. In this situation, the unequal sample distribution is balanced using a random sampling strategy. The evaluation of the DAIC-WOZ used for the AVEC 2016 competition reveals that the suggested tactic is successful.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

IV. PROPOSED SYSTEM

A. High Level Design



V. CONCLUSION

In conclusion, depression is one of the main diseases that endanger people's mental health. The existing traditional diagnosis methods have significant limitations, thus a strategy of objective evaluation of depression based on intelligent technology needs to be created to aid in early detection and patient treatment. Since atypical speech patterns in people with depression are somewhat correlated with their mental state, it is desirable to use speech acoustic qualities as objective indicators for the diagnosis of depression. Depression, which is swiftly becoming an epidemic disease, commonly affects people from all socioeconomic levels, racial groups, and countries.This can be solved by remote diagnosis of depression from speech at a person's ease. The model discussed does just that using the machine learning models. This can be implemented in smartwatches and in an application for people to use.

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] Gintautus , 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.