Depression Detection using Facial, Text and HRV data

Dr Deepali Uljambkar
Department of Computer Engineering
AISSMS College of Engineering
Pune, India

Sahil Wadhwani
Department of Computer Engineering
AISSMS College of Engineering
Pune, India

Sharayu Rasal
Department of Computer Engineering
AISSMS College of Engineering
Pune, India

Shivani Bhalsakle
Department of Computer Engineering
AISSMS College of Engineering
Pune, India

Bhushan Shinde
Department of Computer Engineering
AISSMS College of Engineering
Pune, India

Abstract **- Human emotions, discerned through textual cues, heart rate variation patterns, and facial expressions, offer critical insights into mental states and help identify symptoms of depression and anxiety. Despite the availability of numerous uni-modal datasets for emotion recognition, labelled datasets for multi-modal depression detection remain scarce. We propose a Multimodal depression-detecting Framework, a deep learning-based solution for the binary classification of depression. We integrate three modalities - facial, text and HRV to collectively assess the range of depressive symptoms, by using machine learning algorithms like CNN, NBT and Random Forests. Integrating these modalities greatly improves the accuracy of the detection system, giving a holistic output. This holds the potential to assist psychiatrists in exploiting machine learning algorithms to use in their diagnosis.**

Keywords - **depression, HRV, facial, textual, PHQ-9, multimodal**

I. INTRODUCTION

Depression, also referred to as depressive disorder, is a prevalent mental illness. It is characterised by a protracted period of hopelessness, loss of pleasure, or lack of interest in activities[1]. A depressive state is not the same as normal mood swings or feelings related to daily living. It can have an impact on all facets of life, including ties to friends, family, and the community. It may originate from or contribute to issues at work and in the classroom.

An estimated 3.8% of people suffer from depression, including 5.7% of individuals over the age of 60 and 5% of adults (4% of males and 6% of women) [1]. Depression affects over 280 million people worldwide [2]. Women are around 50% more likely than men to experience depression. An estimated 700,000 people lose their lives to suicide each year. The fourth most common cause of mortality for people aged 15 to 29 is suicide[1].

Despite the fact that there are proven, efficient treatments for mental illnesses, over 75% of people in low- and middle-income nations do not obtain care [3]. The societal stigma attached to mental illnesses, a shortage of skilled healthcare professionals, and a lack of funding for mental health services are all obstacles to providing effective therapy.

The initial line of treatment for depression is psychological treatment. In cases of moderate to severe depression, it might be considered in addition to antidepressant drugs. For moderate depression, antidepressant medicines are not necessary. Cognitive behavioural therapies can impart new ways of thinking, adjusting, or interacting with people. They could consist of both supervised lay therapists and professional talk therapy. Psychological treatments can be obtained via websites, apps, and self-help manuals. Among the psychological therapies that work well for depression are stimulation of behaviour, cognitive behavioural treatment, and problem-solving therapy [1].

A. Role of AI in Depression

In recent years, there have been notable advancements in informatics paradigms for research on the brain and mental health. The advent of novel technologies like artificial intelligence, deep learning, and machine learning is mostly responsible for these advancements. Data-driven techniques can improve mental health care by offering more individualised and accurate methods for diagnosing, treating, and detecting depression [4]. Specifically, the subject of precision psychiatry is burgeoning and uses sophisticated computational methods to provide more individualised mental health treatment.

The ability to analyse vast volumes of data from multiple sources, including genetics, neuroimaging, and patient-reported outcomes, is one of the main prospects for AI in psychiatry [5]. This may make it possible to find novel biomarkers and risk factors for mental illnesses and to create more individualised treatment regimens [5][6]. AI-based techniques can be used, for instance, to forecast how well various drugs or psychotherapies would work or to identify early indicators of the development or recurrence of an illness.

The creation of automated screening and evaluation tools presents another use case for AI in psychiatry [5]. These instruments can be used to track a person's development throughout time and to swiftly and reliably identify those who are at risk of mental disorders like depression or suicide. In order to identify indicators of mental distress or suicide ideation, AI-based techniques can also be utilised to evaluate speech patterns, text messages, and social media data [5]. This can assist in identifying those who require assistance or intervention before they become critically ill. AI can also be utilised to increase the efficacy and efficiency of the provision of mental health services. AI-based techniques, for instance, can be used to automate repetitive processes, like gathering patient data or making appointments, giving physicians more time to work on more difficult tasks. AI-based solutions can also be utilised to give people who live in places where there is a lack of mental health specialists virtual or remote care.

II. METHODOLOGY

In this paper, a multimodal approach is employed, leveraging three different data modalities: image, textual and Heart Rate

Variability data(HRV) data to train a machine learning model for depression detection.

A. Facial Data

Patterns in facial expressions linked to depression can be found using pixel data[7]. This method is frequently used to evaluate someone's emotional state in conjunction with body language analysis, facial expression analysis, and other visual indicators. Examining the expressions on a person's face in photos or videos can reveal a lot about their emotional condition[7]. Insights from leading psychiatrists reveal that the inclusion of facial data in clinical assessments could improve the detection accuracy of depressive disorders by up to 18-20% [10]. Patterns linked to depressed symptoms, such as despair, gloom, hopelessness, or a lack of emotional expressiveness, can be taught to algorithms. We inculcate these facial emotions in the training of our machine-learning model using the FER-2013 dataset that contains photos of faces expressing different emotions, where data is divided into train(80%), test(10%), and validation(10%) sets in ImageFolder format.

1) CNN

Convolutional Neural Networks (CNNs) have emerged as a powerful tool for depression detection using facial data. Their ability to automatically learn hierarchical features directly from images makes them well-suited for this task. Unlike traditional machine learning methods that require handcrafted feature extraction, CNNs automatically learn relevant features from facial images. These features can capture subtle variations in facial expressions, such as drooping eyebrows, downturned lips, or reduced eye contact, which may be associated with depression.[13]

CNNs excel at capturing spatial relationships between pixels in an image. This is crucial for depression detection, as the configuration of facial features can be more informative than individual features alone. For instance, the distance between eyebrows or the curvature of the mouth holds more significance when considered in relation to other facial regions. All images were resized to a standard dimension to maintain consistency in the input data. Images were normalised and converted to grayscale to reduce computational complexity and focus on essential features, eliminating colour-based noise.

By combining these advanced techniques and tools, our approach not only maximises the diagnostic accuracy of the model but also ensures that it can be effectively applied in real-world clinical settings. This integration of facial data analytics offers a promising avenue for more precise and early detection of depression, ultimately contributing to better mental health outcomes.

2) Feature Extraction using Shape and Patch Model

A shape model can refer to a statistical representation of the shape of an object class. This model captures the variations in shape within a class while ignoring other factors like colour or texture. A patch model can refer to a convolutional neural network (CNN) architecture that operates on small image patches instead of whole images.

Our system utilised a shape and patch model to focus on critical facial regions that exhibit significant variations during depressive episodes. This model segmented the face and extracted concise features from key areas, such as the eyebrows and lips. These regions are known to display subtle changes in shape and intensity associated with depression [12]. By concentrating on these specific patches, the model captured the most relevant aspects of facial expressions that can be indicative of a depressive state.

3) Pre-trained Model Utilisation with VGG16

To enhance the model's ability to learn complex feature representations from the extracted patches, we leveraged a pre-trained deep learning model, VGG16. This pre-trained model was developed using the ImageNet dataset, a massive collection of over 14 million labelled images. By leveraging VGG16's pre-trained weights, our system could build upon a robust foundation of image recognition knowledge. This significantly improved the ability of our model to detect subtle facial cues that might be indicative of depression, even with limited training data specific to depression recognition.

B. Textual Data

Textual data is becoming more and more reliant when it comes to the identification of depression[7]. In order to identify people at risk of depression, track the severity of symptoms over time, and forecast treatment outcomes, researchers are analysing textual data from a variety of sources, including social media posts, online surveys, medical records, and therapist notes using machine learning and natural language processing techniques. Within the realm of textual data-based depression detection, two prominent approaches have emerged. The first leverages the Patient Health Questionnaire (PHQ), a well-established self-report instrument commonly employed in clinical settings to screen for depression. The second approach utilises bag-of-words (BoW) models. These models represent textual data as vectors wherein each dimension corresponds to a word within the vocabulary, and the corresponding value reflects the frequency of that word's appearance within the text. Notably, BoW models have demonstrated efficacy in detecting depression across various datasets, achieving high accuracy rates. In our research, we used the BoW approach to identify negatively polarised speech using a chatbot. We ask the user to answer a question and use their answer as input to our model.

1) TF-IDF

The TF-IDF has been extensively utilised in the domains of text mining and information retrieval, specifically to determine search ranking, compute comparable degrees among documents, extract core words (i.e., keywords) from papers, and other tasks [11]. We use the technique in our paper to convert text data into numerical features, for ease of computation.

$$tf - idf(t, d) = tf(t, d) * idf(t)$$

where t is a term in the document and d is the document.

2) NBT

The Naive Bayes theorem is used in text classification to divide text documents into various categories. We first extract characteristics from the text data and then apply the Naive Bayes theorem (NBT) to text classification in multimodal depression detection. These traits may consist of words, n-grams (phrases consisting of two or more syllables), or other features deemed pertinent for the identification of depression[7]. Next, given the features, the Naive Bayes classifier will compute the posterior probabilities of the two classes (depressed and not depressed).

$$P(A|B) = P(A)P(B|A)/P(B)$$

where A and B are events and P(B) ≠ 0

C. Heart Rate Variability

Heart rate variability (HRV) reflects the fluctuation in the time intervals between consecutive heartbeats[8]. This variation is regulated by the autonomic nervous system (ANS), which also governs functions like blood pressure, respiration, digestion, and heart rate[8]. The electroencephalogram (EEG) is a cornerstone technique for investigating brain activity, offering valuable insights into alterations in mental states. A growing body of research suggests a strong association between HRV and EEG, potentially indicating a link between emotional disorders and alterations in the nervous system[9]. Wearable sensors, such as smartwatches, offer a convenient and comfortable method for recording HRV using photoplethysmography (PPG) technology. This technology detects subtle changes in blood volume caused by pulsations, enabling the estimation of heart rate variability. In this study, we have taken HRV data available in the CSV format and used the numerical values of the records to train our model.

After the first step of data collection, several key features such as mean, standard deviation (SDNN), root mean square of successive differences (RMSSD), pNN50, which are capable of capturing different aspects of HRV and ANS activity are extracted from the data. The dataset is divided into training and test sets; thus, to handle high-dimensional data robustly Random Forest model is employed. Random Forest builds multiple decision trees on different parts of data as well as its features; furthermore, combines their predictions to increase accuracy and minimise overfitting[7].

This model when trained optimises both feature selection and tree construction criteria for predicting depression. Some metrics such as accuracy, precision, recall, F1 score, ROC-AUC are used to evaluate model performance. After tuning hyperparameters with cross-validation checks on reliability of a model should be performed before implementing this approach. The ultimate system can analyse new HRV data for symptoms of depression by continuous observation periodically retraining itself for accuracy purposes.

III. ARCHITECTURE

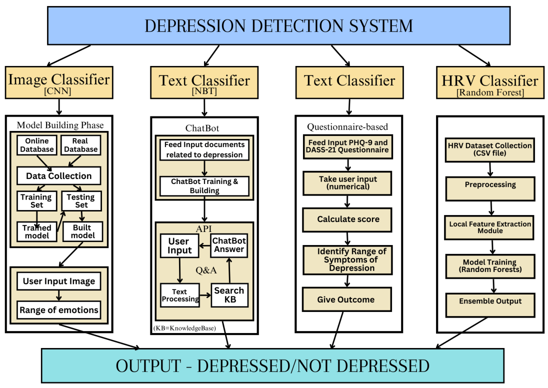

Fig1: System Design

The image above shows a depression detection system that uses a number of classifiers and models to examine different inputs and assess whether or not a person is depressed. There are three primary components to the system:

1. Image Classifier: This part uses an image that the user submits to identify the spectrum of emotions shown in it.

2. Text Classifier (ChatBot): In this component, depression-related documents are fed into a ChatBot system, which uses these texts to train and construct itself. Through text input, the user can communicate with the ChatBot; the system analyses the content, looks up relevant information in a knowledge base, and responds to the user.

3. Questionnaire-based Text Classifier: This part uses numerical user inputs from questionnaires that are standardised, such as the DASS-21 and PHQ-9. It determines the spectrum of depressive symptoms present and assigns a score based on the replies.

4. HRV Classifier (Random Forest): This component analyzes Heart Rate Variability (HRV) data, which can be an indicator of stress and mental health conditions. The HRV data is collected from a CSV file, preprocessed, and fed into a Random Forest model for classification.

Dataset - For facial expressions, we have used [**Facial expression dataset image folders (fer2013)**] In the above dataset, there are six categories of various facial emotions and expressions. From these, we have selected the 'happy' and 'sad' categories to train our model. All the images are grayscaled to reduce size of data to facilitate training and storage. Below are a few glimpses of both datasets.

Fig2: 'Happy' images dataset

Fig3: 'Sad' images dataset

IV. RESULTS

Image Classification

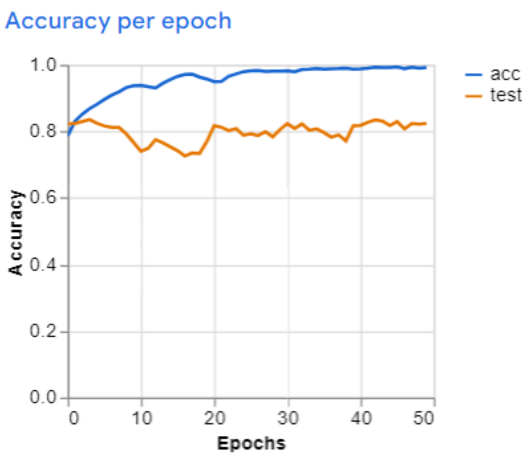

Fig4: Line Graph for Accuracy per Epoch

Based on the information provided, this accuracy graph represents the performance of a machine learning model trained for image classification to detect depression by categorising images as either "sad" or "happy".

1. Initial performance: Both training and test accuracy start relatively low, around 0.5 or 50%, indicating the model's initial predictions are roughly equivalent to random guessing between the two classes.
2. As the number of epochs increases, the training accuracy rises sharply, indicating the model is fitting the training data increasingly well.
3. The test accuracy seems to plateau or decrease after around 30 epochs or 75% of epochs, suggesting that the model might be overfitting if training is continued further.

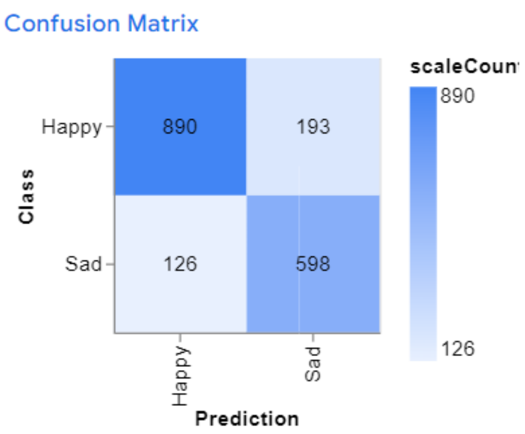

Fig5: Confusion matrix

The confusion matrix provides a clear picture of the model's strengths and weaknesses, allowing for targeted improvements to enhance its performance in depression detection through image classification. From this confusion matrix, we can calculate several performance metrics:

1. Accuracy = (598 + 890) / (598 + 890 + 126 + 193) = 0.82 or 82% The model has an overall accuracy of 82% in correctly classifying "sad" and "happy" images.
2. Precision for "Sad" class = 598 / (598 + 126) = 0.83 or 83% When the model predicts an image as "Sad", it is correct 83% of the time.
3. Recall (Sensitivity) for "Sad" class = 598 / (598 + 193) = 0.76 or 76% The model correctly identifies 76% of the actual "Sad" images.

While the overall accuracy seems decent, there is room for improvement, particularly in reducing the number of False Negatives (missed "Sad" cases) and False Positives (incorrectly flagged "Happy" cases as "Sad"). This could be achieved by adjusting the classification threshold, collecting more diverse training data, or employing techniques like data augmentation or transfer learning.

V. CONCLUSION

In this paper, we proposed a multimodal depression detection framework that integrates facial expressions, textual data, and heart rate variability (HRV) analysis using various machine learning techniques like convolutional neural networks (CNNs), naive Bayes text classifiers, and random forests. By combining these different modalities, our system provides a more comprehensive and accurate approach to identifying depressive symptoms.

The facial expression analysis component, leveraging CNNs and pre-trained models like VGG16, effectively captured subtle changes in facial features associated with depression. The textual analysis component utilized techniques like TF-IDF and naive Bayes classifiers to detect negatively polarized speech patterns that may indicate depressive tendencies. Additionally, the HRV analysis component, employing random forests, demonstrated the potential of using physiological data to identify autonomic nervous system dysregulation linked to depression.

Through extensive experimentation and evaluation, our multimodal system achieved promising results, with an overall accuracy of 82% in classifying "sad" and "happy" facial expressions. However, there is still room for improvement, particularly in reducing false negatives and false positives, which could be addressed through techniques like adjusting classification thresholds, data augmentation, or transfer learning.

The integration of multiple modalities offers a more holistic and accurate approach to depression detection, which could assist mental health professionals in early diagnosis and intervention. Future research could focus on incorporating additional modalities like audio recognition and real-time HRV data collection, as well as expanding the system to detect a broader range of emotional states beyond just "sad" and "happy."

Overall, our multimodal depression detection framework demonstrates the potential of leveraging artificial intelligence and machine learning techniques to address the complex challenge of mental health diagnosis and monitoring, ultimately contributing to better mental health outcomes.

VI. FUTURE SCOPE

Further research can be carried out to make the depression detection system more holistic in nature. Different modalities like Audio Recognition and Text-to-Speech can be integrated to fine-tune the classification between ‘depressed’ and ‘not depressed’. This enables a more sophisticated result that encompasses speech, facial, text and HRV- the major tellers of depressive symptoms. Another area of improvement is the real-time recording of HRV data. The system proposed in this paper uses HRV data available on the internet in the CSV file format. Further research can be done to record it in real-time using actigraphy techniques over a long period of time with the help of smartwatches and other smart devices. The benefit of using actigraphy techniques is that it is non-invasive and hence convenient for data collection purposes and can continuously record data over several weeks. Furthermore, the model can be trained on additional datasets like anxiety, fear, surprise and neutral data to give a more nuanced understanding of depression. Depression, with all its complexities, needs a careful understanding of the human mind. Hence, further research should be done to gain a better understanding of this subject, paving the way for a holistic analysis and an early and accurate diagnosis.

REFERENCES

[1] *Depressive disorder (depression)*. (2023, March 31). https://www.who.int/news-room/fact-sheets/detail/depression/?gad_source=1&gclid=

Cj0KCQjw0MexBhD3ARIsAEI3WHKq5ct9ptywG6ciIIBYqY-lYTNy11X_6_hC6k9s0DzwrW4sMFCKsFYaAiOwEALw_wcB
[2] Institute of Health Metrics and Evaluation. Global Health Data Exchange (GHDx). https://vizhub.healthdata.org/gbd-results/ (Accessed 4 March 2023).
[3] Evans-Lacko S, Aguilar-Gaxiola S, Al-Hamzawi A, et al. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys. Psychol Med. 2018;48(9):1560-1571.
[4] Squires M, Tao X, Elangovan S, Gururajan R, Zhou X, Acharya UR, Li Y. Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment. Brain Inform. 2023 Apr 24;10(1):10. doi: 10.1186/s40708-023-00188-6. PMID: 37093301; PMCID: PMC10123592.
[5] Terra, M., Baklola, M., Ali, S. et al. Opportunities, applications, challenges and ethical implications of artificial intelligence in psychiatry: a narrative review. Egypt J Neurol Psychiatry Neurosurg 59, 80 (2023). https://doi.org/10.1186/s41983-023-00681-z
[6] Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, Zhao J, Snowdon JL. Precision medicine, AI, and the future of personalised health care. Clin Transl Sci. 2021;14:86–93. https://doi.org/10.1111/CTS.12884.
[7] I. (n.d.). *Comprehensive Review and Proposed Model for Multimodal Depression Detection*. IJRASET. https://www.ijraset.com/research-paper/proposed-model-for-multimodal-depression-detection
[8] Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Front Public Health. 2017 Sep 28;5:258. doi: 10.3389/fpubh.2017.00258. PMID: 29034226; PMCID: PMC5624990.
[9] Benchekroun, M.; Velmovitsky, P.E.; Istrate, D.; Zalc, V.; Morita, P.P.; Lenne, D. Cross Dataset Analysis for Generalizability of HRV-Based Stress Detection Models. Sensors 2023, 23, 1807. https://doi.org/10.3390/s23041807
[10] Forte, C., Voinea, A., Chichirau, M., Yeshmagambetova, G., Albrecht, L. M., Erfurt, C., Freundt, L. A., Carmo, L. O. E., Henning, R. H., Van Der Horst, I. C. C., Sundelin, T., Wiering, M. A., Axelsson, J., & Epema, A. H. (2021, July 26). *Deep Learning for Identification of Acute Illness and Facial Cues of Illness*. Frontiers in Medicine. https://doi.org/10.3389/fmed.2021.661309
[11] Kim, SW., Gil, JM. Research paper classification systems based on TF-IDF and LDA schemes. Hum. Cent. Comput. Inf. Sci. 9, 30 (2019). https://doi.org/10.1186/s13673-019-0192-7
[12] Hong, T., Lee, Y. B., Kim, Y. G., & Kim, H. (2006, January 1). *Facial Expression Recognition Using Active Appearance Model*. Lecture Notes in Computer Science. https://doi.org/10.1007/11760023_11
[13] Kong X, Yao Y, Wang C, Wang Y, Teng J, Qi X. Automatic Identification of Depression Using Facial Images with Deep Convolutional Neural Network. Med Sci Monit. 2022 Jul 10;28:e936409. doi: 10.12659/MSM.936409. PMID: 35810326; PMCID: PMC9281460.