iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56635

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XI Nov 2023- Available at www.ijraset.com

Comprehensive Review and Proposed Model for Multimodal Depression Detection

Deepali M. Ujalambkar[1], Sharayu Rasal[2], Shivani Bhalsakle[3], Sahil Wadhwani[4], Bhushan Shinde[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, All India Shree Shivaji Memorial Society College of Engineering, Pune, India*

Abstract: Depression has become prevalent in today's world and is an ever-increasing concern. Amidst the rising cases of depression witnessed every passing year, it has become vital to identify the signs and symptoms leading to it. This paper reviews the advances made in machine learning to detect these signs by exploring four different learning models- text, facial, audio, and EEG. The accuracy of correct identification is analyzed along with the ease of extraction of data, and three modalities are selected for further research based on the aforementioned factors. Furthermore, the paper proposes a hybrid approach that combines the tools and techniques employed in the previous works by integrating the three modalities. A multimodal system is proposed that overcomes the limitations of the unimodal systems explored in the previous papers.

Keywords: Depression Detection, HRV, Facial, Textual, Multimodal, Depression Symptoms

I. INTRODUCTION

Depression is a common and serious mental illness that affects millions of people worldwide. An average of 3.8% of the global population suffers from depression, with over 700,000 dying from suicide [1]. It affects about 5% of adults and is 50% more common in women. Depression can cause a wide range of symptoms, including sadness, hopelessness, fatigue, difficulty concentrating, changes in appetite, and unnatural sleep patterns. The World Health Organization (WHO) uses the term 'depressive episode' for a period when the person experiences single or multiple symptoms of depression [1]. This episode can be ranked based on severity and frequency of occurrence, to base the study for diagnosis and treatment of the patient. Despite the growing number of antidepressant treatments, it remains a major concern for clinicians owing to its high rates of relapse and resistance, and prolonged duration of treatment. Moreover, most people do not receive treatment. Almost 75% of the depressed population refrain from seeking therapy, due to various factors like the high cost of treatment procedures, lack of awareness, and social stigma around mental health, to name a few [3].

In these times, identifying various symptoms of depression is of prime importance. Extensive research is being carried out to determine the signs and provide accurate diagnosis. While clinical diagnosis is a prevalent and popular methodology, the level of accuracy can be enhanced by integrating Machine Learning techniques for prediction. Machine learning encompasses techniques such as Sentiment Analysis, Emotional Artificial Intelligence, and Image Recognition which can be used to analyze textual data and facial data. This data can be sourced from social media tweets, comments, posts, and pictures, which would prove to be a more reliable strategy to assess the mental state of a person as compared to consciously self-filled reports since it could introduce personalbiases or dishonest responses. These machine learning techniques can be combined to achieve an even higher accuracy. Such a multimodal approach provides several benefits - enhanced accuracy through cross-validation of multiple indicators, improved sensitivity to subtle variations in emotional expressions, and reduced false positives/negatives by considering multiple data sources. Various machine learning algorithms, including Convolutional Neural Networks (CNNs) for image analysis, Support Vector Machines (SVMs), Random Forests, and Neural Networks for text and HRV feature extraction, have shown promising results in individual modality-based depression detection. Combining these modalities through fusion techniques and sophisticated models offers the potential for improved accuracy and sensitivity, paving the way for more effective early detection and intervention in depressive disorders. This holistic approach holds the promise of making a significant positive impact on mental health assessment and treatment. The feasibility of a multimodal depression detection system integrating text, images, and HRV is high. There is a growing body of research demonstrating the effectiveness of multimodal approaches for depression detection. Additionally, multimodal systems can be more robust to noise and variability in individual modalities. Furthermore, the data required for multimodal depression detection is becoming increasingly available. Textual data can be collected from surveys and online communities. Image data can be collected from Health Service Research Centers.

©IJRASET: All Rights are Reserved

HRV data can be collected from wearable devices, such as smartwatches and fitness trackers. The availability of this data makes it possible to develop and deploy multimodal depression detection systems on a large scale.

II. LITERATURE REVIEW

The Literature Survey in this paper is divided into two sub-sections based on the different types of approaches used. The first sub-section talks about research based on a singular modality or the unimodal approach, while the second section is about a multimodal or hybrid approach.

A. Unimodal Approach

Prabhu et al. [4] proposes a new method for detecting depression using EEG signals. The method uses a deep learning model called EEGNet to extract features from the EEG signals and classify the subject as depressed or not depressed. The method is sensitive to the quality of the EEG data. If the EEG data is noisy or incomplete, the model may not be able to make accurate predictions. However, the researchers believe that the method has the potential to be a valuable tool for detecting depression and monitoring treatment progress. Tazawa et al. [6] proposes a new method for detecting and assessing depression using a multimodal wristband-type wearable device. The device collects data on physical activity, sleep, heart rate, skin temperature, and ultraviolet light exposure. A machine learning model is then used to analyze this data and identify patterns that are associated with depression. The authors evaluated their method on a dataset of 45 depressed patients and 41 healthy controls. The method achieved an accuracy of 76% for screening depression and a correlation coefficient of 0.61 for assessing severity. However, demographic data (e.g., age, sex, etc.) was not matched between patients and healthy controls. Therefore, differences in the demographic data may have had some impact on the study results.

Sangwon Byun et al. [7] HRV data was collected from 37 MDD patients and 41 healthy subjects and Support Vector Machine (SVM) was used for feature extraction. The HRV features were statistically compared between the control and MDD groups using the Mann –Whitney U test. On executing the model, it showed that Non-linear HRV performs better than the Linear HRV model. SVM-RFE outperformed the statistical filter as a feature selection method, but the differences in performance measures between the two methods were not substantial.

Kim NH et al.'s model [5] used natural language processing to categorize phrases made by social media users according to the nine symptoms of depression included in the Patient Health Questionnaire-9 (PHQ-9). The model's output was used to determine the user's level of depression. One potential drawback of the suggested model is that it classifies users as either depressed or not.

Guo et al. [16] The approach used a combination of 2D and 3D facial expressions to extract features that are relevant to depression. These features were then used to train a deep neural network to classify individuals as depressed or not depressed. The authors evaluated their approach on a dataset of 100 individuals, including 50 individuals with depression and 50 healthy controls. The approach achieved an accuracy of 96.2%, which is significantly higher than the accuracy of previous methods for depression recognition using facial expressions. One shortcoming of the approach mentioned in the paper is that it required a large dataset of facial expressions from individuals with depression and healthy controls. This can be difficult to collect, especially for 3D facial expressions.

B. Multimodal Approach

A hybrid method that uses both facial characteristics and EEG was proposed by Danniel Shazmeer Bin Abdul Hamid et al. [8] for the detection of depression. The system classified the person as depressed or not depressed by first extracting characteristics from the EEG and face data, and then using a bidirectional long short-term memory (BiLSTM) network. Additionally, the scientists demonstrated that the BiLSTM network outperformed other deep learning techniques including convolutional neural networks (CNNs) and recurrent neural networks (RNNs) by comparing their suggested system to these techniques. Nevertheless, the technique does not investigate alternate approaches, instead concentrating on a particular model architecture (BiLSTM).

MDD patients can be identified using physiological indicators, as established by C. Á. Casado et al. [9]. It used signals from photoplethysmography (PPG) to compute the HRV rate. Using a light source and a photodetector, photoplethysmography is a low-cost optical technology that measures changes in blood volume at the skin's surface. They examine how various depression levels affect the blood volume pulse (BVP) signal's physiological response. The findings indicate that it is a useful tool for diagnosing depression and that it may be used in conjunction with other visual data modalities to enhance its capabilities. It is difficult to separate the effects of these tasks on depression prediction because confounding factors were introduced when data from two independent tasks (Freeform and Northwind) were combined and added to the AVEC2014 dataset.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Sahana Prabhu et al. [10] proposed an approach for depression detection using a combination of textual, speech, and facial features. Using a Long Short-Term Memory (LSTM) network, the method first extracts features from each modality and then combines the data to determine if the person is depressed or not. The 100-person dataset, which consisted of 50 depressed people and 50 healthy controls, was used by the authors to assess their methodology. The approach achieved an accuracy of 90.2%, which is significantly higher than the accuracy of previous methods for depression detection using a single modality. On the other hand, ISED could make labeling the facial expressions quite difficult and utilizing the IMEOCAP dataset relatively requires extensive testing.

A healthcare monitoring system based on ontologies and bidirectional long short-term memory (Bi-LSTM) was proposed by Farman Ali, Shaker El-Sappagh, et al. [11] to accurately evaluate huge data in healthcare and increase classification accuracy. Large volumes of the most beneficial healthcare data were retrieved by the framework from a variety of sources, including social media, wearable sensors, cell phones, and medical records. Furthermore, the extracted data is stored in a big data cloud repository, and MapReduce is used to handle and process structured and unstructured data intelligently. Word2vec, a neural network-based word embedding model, has been utilized to represent textual data related to healthcare with semantic meaning. Furthermore, the Word2vec model was integrated with certain domain ontologies. Principal component analysis (PCA) and information gain (IG) were used in a number of experiments with the ontology and LSTM-based models, and the outcomes were compared to the corresponding reference models. SentiWordNet lexicons alone, however, are insufficient to accurately comprehend textual input for the purpose of predicting a patient's mental health and potential pharmacological side effects.

Vandana et al. [12] proposed a hybrid model for depression detection using deep learning. The authors compare their model to several previous methods, including machine learning models trained on text and audio data separately. Their hybrid model outperforms all the other methods, which suggests that there is value in combining text and audio features for depression detection. The authors evaluated their model on a dataset of 47 subjects, including 26 subjects with depression and 21 healthy controls. The model achieved an accuracy of 98% for detecting depression, which is significantly higher than the accuracy of previous methods. The authors also investigate the impact of different feature extraction methods on the performance of their model. They find that using a combination of deep learning and statistical features achieves the best results. The authors also discussed the potential limitations of their study, including the small dataset size and the lack of external validation. They also acknowledge that their model may not be generalizable to all populations. More research is needed to validate the model on larger datasets and to investigate its performance in clinical settings.

III. PROPOSED METHODOLOGY

This architecture model is divided into four stages - Data Collection, Data Preparation, Model Training, and Model Testing. The data is collected and tested in three formats - textual, facial, and HRV, each of which is elaborated in detail below.

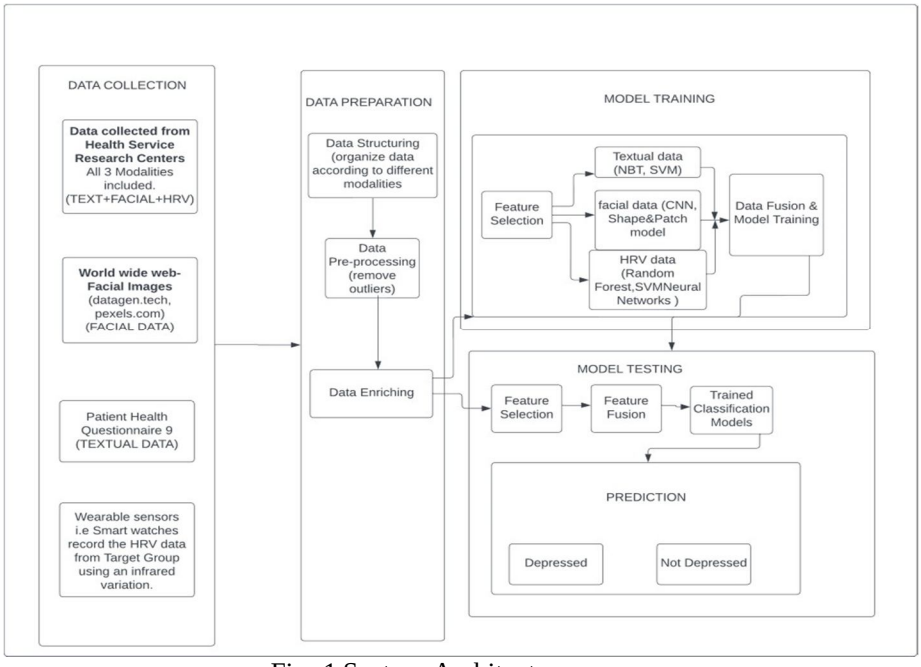


Fig. 1 System Architecture

A. *Pixel Data*

Pixel data can be used to identify patterns in facial expressions that are associated with depression. Pixel data in depression detection typically refers to the use of image analysis techniques to identify potential signs of depression from visual data, such as photographs or videos. This approach is often used in conjunction with facial expression analysis, body language recognition, and other visual cues to assess an individual's emotional state. Analyzing facial expressions captured in images or videos can provide valuable information about an individual's emotional state. Algorithms can be trained to detect patterns associated with depressive symptoms, such as sadness, hopelessness, or lack of emotional expressiveness.

1) *Convolutional Neural Network (CNN):* CNNs can be used to analyze facial expressions in images or videos to detect signs of depression. Features like furrowed brows, downturned mouth corners, or limited eye contact can be indicative of depressive symptoms. CNNs can be trained to recognize these subtle facial cues. CNNs can be used in combination with other data sources, such as audio analysis (voice and speech patterns) and textual analysis (written or spoken language). Multimodal approaches can provide a more holistic assessment of an individual's mental health.

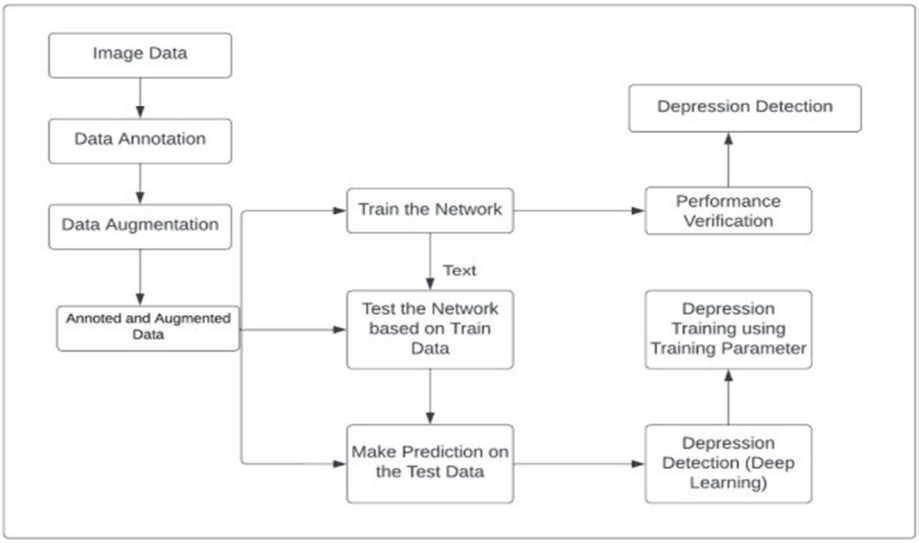


Fig. 2 Convolutional Neural Network

In the CNN block diagram, the facial image is first preprocessed. This may involve tasks such as resizing the image, normalizing the image, and converting the image to grayscale. The preprocessed image is then fed into the CNN. The Convolutional Neural Network model is comprised of a series of inter-connected layers, inserting a layer of added complexity to increase the precision of the prediction model. Scaling reduces the size of the image, while the connected layers categorize the image.

B. *Textual Data*

Textual data is playing an increasingly important role in depression detection. Researchers are using machine learning and natural language processing techniques to analyze textual data from a variety of sources, such as social media posts, online surveys, medical records, and therapist notes, to identify individuals at risk of depression, monitor the severity of depression symptoms over time, and predict treatment outcomes. One promising approach to textual data-based depression detection is PHQ is a self-report questionnaire that is commonly used to screen for depression in clinical settings. It consists of nine questions that assess the severity of depression symptoms over the past two weeks. Another promising approach to textual data-based depression detection is to use bag of words (BoW) models. BoW models represent textual data as vectors of word counts, where each dimension in the vector represents a word in the vocabulary and the value of the dimension is the number of times that word appears in the text. BoW models are effective at detecting depression with high accuracy using a variety of datasets.

1) *Naive Bayes Classifier:* Naive Bayes (NBT) is a probabilistic classifier that assumes that the features of a text are independent of each other. For depression detection, NBT can be used to classify text based on the frequency of words that are associated with depression. In the context of text classification, the Naive Bayes theorem is used to classify text documents into different categories. The algorithm works by calculating the probability that a word belongs to each category and then assigning the word to the category with the highest probability. The classifier is fast to train and can be very accurate, even with training datasets.

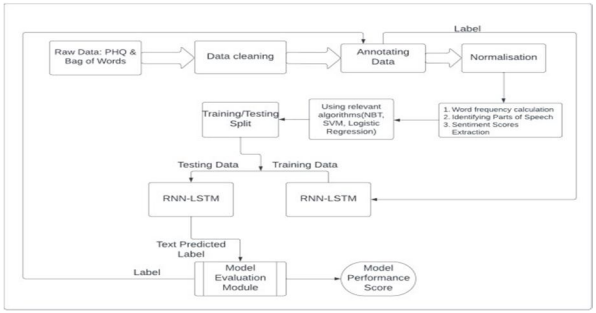


Fig. 3 Naive Bayes Classifier

NBT classifiers are effective for text classification, especially when used in conjunction with other techniques such as feature selection and regularization. Our approach to using the Naive Bayes theorem (NBT) for text classification in multimodal depression detection is to first extract features from the text data. These features can be words, n-grams (phrases of two or more words), or other types of features that are considered to be relevant for depression detection. Once the features have been extracted, they can be input to a Naive Bayes classifier. The Naive Bayes classifier will then calculate the posterior probabilities of the different classes (depressed and healthy), given the features. The posterior probability of a class is the probability of that class, given the evidence. In this case, the evidence is the features extracted from the text data. The Naive Bayes classifier will then output the class with the highest posterior probability as the prediction. The textual data would first need to be converted into a numerical representation, predictions from the Naive Bayes classifier are then combined with the predictions from the other modalities (images and HRV data) to get a more accurate prediction of whether a person is depressed or not.

C. HRV Data

The tim interval between each heartbeat is known as heart rate variability or HRV technology. This variation is caused by the autonomic nerve system (ANS) of the body, which also regulates our blood pressure, breathing, digestion, and heart rate automatically. The Electroencephalogram (EEG) is a crucial method for researching brain activity and yields valuable information about alterations in mental state. There was a strong link between HRV and EEG, which suggests that emotional disorders may be associated with alterations in the nervous system. Wearable sensors like Smart watches can record the HRV using an infrared variation. They are easily obtainable and comfortable to wear. We aim to utilize the HRV readings of our participants to record the fluctuations in their stress levels. Given the dearth of accessible biomarkers in the field of psychiatry, wearable sensors provide an efficient and non-invasive way to collect HRV data, which can be used to further study the biological changes in a person suffering from depression.

1) *Random Forests:* Random Forest is another popular machine learning algorithm used in depression detection models that use Heart Rate Variability (HRV) data. One such study [13] used Random Forest to construct a depression detection model based on HRV parameters. The study used a dataset of 60 participants, including 30 patients with depression and 30 healthy controls. The HRV parameters were extracted from the electrocardiogram (ECG) signals of the participants. The study used a Random Forest classifier to classify the participants into two groups: depressed and non-depressed. The results showed that the Random Forest classifier achieved an accuracy of 90% in classifying the participants into the two groups [13].

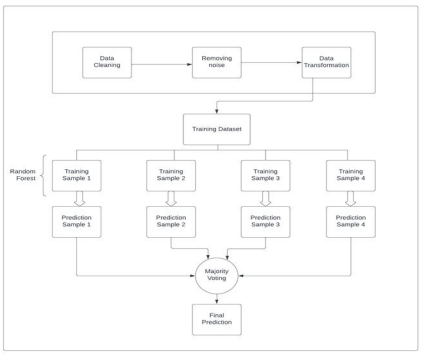


Fig. 4 Random Forest

In the random forest block diagram, the HRV data is first preprocessed. This may involve tasks such as cleaning the data, removing noise, and transforming the data into a format that is suitable for machine learning. The preprocessed data is then fed into the random forest. The random forest consists of a number of decision trees. Each decision tree is trained on a subset of the data. The predictions from the individual decision trees are then combined to produce a final prediction.

2) *Neural Networks:* Neural Networks are a popular machine learning algorithm used in depression detection models that use Heart Rate Variability (HRV) data. Based on two distinct types of deep belief network (DBN) models, one such study [14] suggested a novel method for identifying possible depression risk. The created dataset comes from multiple healthcare service centers. Another large dataset from a Norwegian youth-oriented public online information channel is used to test the methodology. Recurrent neural networks (RNNs) based on long short-term memory (LSTM) were employed in the study to recognize texts that described the participants' subjectively perceived symptoms of depression [15]. In general, Neural Networks are trained on a dataset of HRV parameters extracted from electrocardiogram (ECG) signals of participants. The trained model can then classify participants into two groups: depressed and non-depressed. The accuracy of the model depends on the quality and quantity of the data used for training.

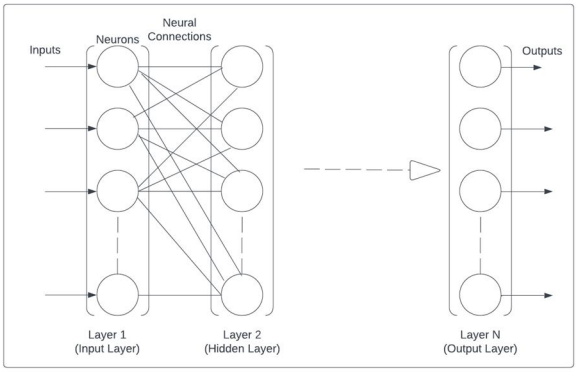


Fig. 5 Neural Network

In the neural network block diagram, the HRV data is first preprocessed. This may involve tasks such as cleaning the data, removing noise, and transforming the data into a format that is suitable for machine learning. The preprocessed data is then fed into the neural network. The neural network consists of a number of layers of neurons. Each neuron is connected to a number of other neurons in the previous layer. The weights of the connections between neurons are adjusted during training. The trained neural network can then be used to make predictions on new data.

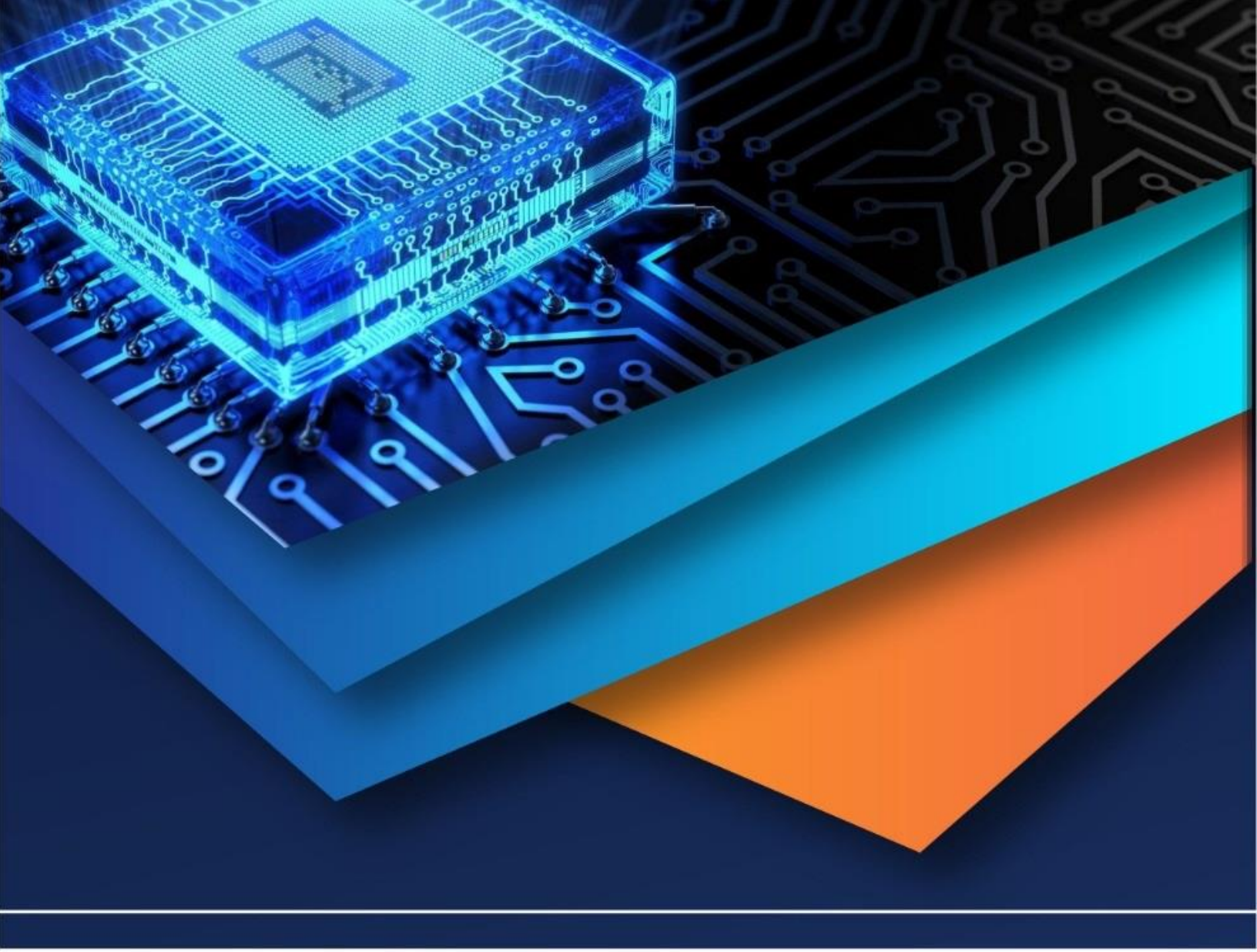
IV. CONCLUSION

In recent years, the development of multi-modal depression detection systems has garnered significant attention in the field of mental health research and technology. This innovative approach leverages a combination of text, image, and Heart Rate Variability (HRV) data, aiming to provide a more comprehensive understanding of an individual's mental well-being. Various machine learning algorithms, including Convolutional Neural Networks (CNNs) for image analysis, Support Vector Machines (SVMs), Random Forests, and Neural Networks for text and HRV feature extraction, have shown promising results in individual modality-based depression detection. Additionally, shape and patch models have been explored for image analysis. Combining these modalities through fusion techniques and sophisticated models offers the potential for improved accuracy and sensitivity, paving the way for more effective early detection and intervention in depressive disorders. This holistic approach holds the promise of making a significant positive impact on mental health assessment and treatment. Moreover, the multi-modal depression detection system not only advances the state of the art in mental health assessment but also addresses the limitations of unimodal systems. Text-based analyses may capture semantic content and sentiment, while image analysis can provide insights into facial expressions and body language, and HRV data can reflect physiological stress levels. By integrating these diverse modalities, the system can mitigate the challenges associated with self-reporting biases, enabling a more objective and holistic evaluation of an individual's mental health status. Furthermore, the incorporation of machine learning and deep learning techniques allows for the automatic extraction of intricate patterns and relationships within the data, ultimately enhancing the accuracy of depression detection. As a result, this project contributes to the ongoing efforts to develop more reliable and comprehensive tools for mental health professionals and individuals seeking support, thereby promoting early intervention and improved overall well-being.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/depression/?gclid=CjwKCAjw1t2pBhAFEiwA_-A-NN-a-LwiiOj_YkXRYqDiySHfxtMIGhReUiIswN2DXtZNbyMvS19nRxoCBjMQAvD_BwE
[2] Albert, U., Tomasetti, C., Marra, C., Neviani, F., Pirani, A., Taddeo, D., Zanetti, O., & Maina, G. (2023). Treating depression in clinical practice: New insights on the multidisciplinary use of trazodone. Frontiers in Psychiatry, 14, 1207621. https://doi.org/10.3389/fpsyt.2023.1207621
[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] Liu B, Chang H, Peng K, Wang X. An End-to-End Depression Recognition Method Based on EEGNet. Front Psychiatry. 2022 Mar 11;13:864393. doi: 10.3389/fpsyt.2022.864393. PMID: 35360138; PMCID: PMC8963113
[5] Kim NH, Kim JM, Park DM, Ji SR, Kim JW. Analysis of depression in social media texts through the Patient Health Questionnaire-9 and natural language processing. Digit Health. 2022 Jul 17;8:20552076221114204. doi: 10.1177/20552076221114204. PMID: 35874865; PMCID: PMC9297458
[6] Tazawa, Y., Liang, K.C., Yoshimura, M., Kitazawa, M., Kaise, Y., Takamiya, A., Kishi, A., Horigome, T., Mitsukura, Y., Mimura, M. and Kishimoto, T., 2020. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. Heliyon, 6(2)
[7] Sangwon Byun, Ah Young Kim, Eun Hye Jang, Seunghwan Kim, Kwan Woo Choi, Han Young Yu, Hong Jin Jeon, Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol, Computers in Biology and Medicine, Volume 112, 2019, 103381, ISSN 0010-4825
[8] Danniel Shazmeer Bin Abdul Hamid, S.B. Goyal, Pradeep Bedi, Integration of Deep Learning for Improved Diagnosis of Depression using EEG and Facial Features, Materials Today: Proceedings, Volume 80, Part 3, 2023, Pages 1965-1969, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.05.659
[9] C. Á. Casado, M. L. Cañellas and M. B. López, "Depression Recognition using Remote Photoplethysmography from Facial Videos," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2023.3238641
[10] Prabhu, S., Mittal, H., Varagani, R. et al. Harnessing emotions for depression detection. Pattern Anal Applic 25, 537–547how(2022). https://doi.org/10.1007/s10044-021-01020-9
[11] Farman Ali, Shaker El-Sappagh, S.M. Riazul Islam, Amjad Ali, Muhammad Attique, Muhammad Imran, Kyung-Sup Kwak, An intelligent healthcare monitoring framework using wearable sensors and social networking data, Future Generation Computer Systems, Volume 114, 2021, Pages 23-43, ISSN 0167-739X, https://doi.org/10.1016/j.future.2020.07.047
[12] Vandana, Nikhil Marriwala, Deepti Chaudhary, A hybrid model for depression detection using deep learning, Measurement: Sensors, Volume 25, 2023, 100587, ISSN 2665-9174
[13] 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
[14] https://doi.org/10.3389/fnins.2021.609760
[15] Uddin, M.Z., Dysthe, K.K., Følstad, A. et al. Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Comput & Applic 34, 721–744 (2022). https://doi.org/10.1007/s00521-021-06426-4
[16] Guo, W., Yang, H., Liu, Z., Xu, Y. and Hu, B., 2021. Deep neural networks for depression recognition based on 2d and 3d facial expressions under emotional stimulus tasks. Frontiers in neuroscience, 15, p.609760



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)