MENTCARE AI: ENHANCING THE EFFICIENCY OF DIAGNOSING AND MANAGING MENTAL DISORDERS USING MACHINE LEARNING

Project ID: 24-25J-322

Kudathanthirige Kavindi Sathsarani Perera

Bachelor of Science Honors Degree in Information Technology Specializing in Data Science

Department of Computer Science

Sri Lanka Institute of Information Technology

Sri Lanka

August 2024

DIAGNOSING AND MANAGING DEPRESSION THROUGH TEXT AND FACIAL IMAGES

Project ID: 24-25J-322

Project Proposal Report

Kudathanthirige Kavindi Sathsarani Perera

IT21178368

BSc (Hons) Degree in Information Technology specializing in Data Science

Department of Computer Science

Sri Lanka Institute of Information Technology

Sri Lanka

August 2024

DECLARATION

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Student ID	Name	Signature
IT21178368	Perera K K S	

The above candidate is carrying out research for the undergraduate Dissertation under my supervision.

22/08/2024

Signature of the Supervisor

(Ms. Wishalya Tissera)

Date

22/08/2024.

Signature of the Co-Supervisor

(Dr. Kapila Dissanayake)

Date

ACKNOWLEDGEMENT

I would like to extend my sincere gratitude to all the individuals who supported me in continuing my fourth-year research project. First and foremost, I sincerely thank our research project supervisor, Ms. Wishalya Tissera, for her invaluable guidance, encouragement, and support. I am also grateful for the advice and support given by our co-supervisor, Dr. Kapila Dissanayake. I extend my sincere thanks to the CDAP panel for their guidance throughout our project work. Furthermore, I would like to thank our medical domain expert, Dr. Anuradha Baminiwatte, for his coordination and information sharing, which enabled us to begin our study endeavor. I would also like to extend my sincere gratitude for the contributions made by our research project leader, Dilshan G.A.M., and other group members who have supported me throughout the project. Finally, I would like to extend my heartfelt gratitude to my parents and friends for their invaluable support in making this project a success.

ABSTRACT

Depression is a mental health disorder that affects millions of people worldwide, impacting their daily functioning and overall quality of life. In clinical settings, early diagnosis and accurate assessment of depression levels are important for effective treatment and management. This research aims to develop a multi-modal approach combining text and image data to improve the accuracy of depression diagnosis and level identification. I will implement and compare two fusion methods which are Weighted Average and Multi-modal Neural Networks. These models will be trained to assess depression and suggest levels of depression to doctors for further examination. The process will also integrate an attention mechanism and employ Explainable AI techniques to enhance the transparency and interpretability of the model's decisions for clinical experts. The final model outputs will be evaluated by real clinical professionals to ensure their reliability and applicability in healthcare settings. By leveraging this multi-modal approach, I aim to enhance the efficiency and accuracy of depression diagnosis, providing doctors with more comprehensive tools to understand and treat their patients effectively.

Keywords: Depression Detection, Multi-modal Neural Networks, Weighted Average, Explainable AI, Attention Mechanism, Fusion Methods

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviations	Description
ML	Machine Learning
AI	Artificial Intelligence
XAI	Explainable AI
NLP	Natural Language Processing
NN	Neural Network
API	Application Programming Interface
DL	Deep Learning
LSTM	Long Short-Term Memory
AUC	Area under the curve
ROC	Receiver operating characteristics
IDE	Integrated Development Environment
UI	User Interface
UX	User Experience

1. INTRODUCTION

1.1 Background Literature

Mental health disorders, including depression, have become a growing concern globally. As per the World Health Organizations' estimation, 3.8% of the population suffer from Depression. [1] Millions of individuals are impacted globally, and it has a big influence on their everyday life, interpersonal connections, and general well-being. Many people don't receive the proper diagnosis or treatment, even though there are plenty of options available. These obstacles include the stigma attached to mental health issues and the shortcomings of the diagnostic techniques used today.

Conventional techniques for diagnosing depression mostly depend on clinical assessments and self-reported questionnaires, both of which have limitations due to the patient's limited capacity to describe their symptoms and can be subjective. Moreover, these techniques are frequently needed for in-person meetings with mental health specialists, which can be expensive, time-consuming, and unavailable to many. As a result, more effective and scalable techniques for the diagnosis and treatment of depression are required.

New opportunities in the field of mental health have been made possible by recent developments in ML and AI. Artificial intelligence (AI)-driven solutions have demonstrated potential in comprehending people's emotional and mental states, especially in the analysis of text and visual data. Through the use of computer vision techniques for picture analysis and natural language processing (NLP) for text analysis, machine learning models are able to identify precise patterns and traits that may be indicative of depression. By enabling real-time input and analysis through mobile applications, these AI-driven models can improve the accessibility and individualization of mental health care.

With this study, I hope to investigate how integrating textual and visual data sources could improve the precision of depression diagnosis. In order to discover indicators of depression, I will use multimodal methodologies to analyze textual inputs, such as patients' feelings in texts and their photographed facial expressions. I'm going to specifically implement two approaches: an intermediate fusion method that uses a multi-modal neural network to integrate both text and image data types within a single architecture, and a late fusion method that combines the outputs of separate text and image models using a weighted average model. My goal is to find the best method for diagnosing depression and its level by analyzing these models' performances and contrasting their results.

Additionally, this research will incorporate attention mechanisms to enhance the focus of the models on relevant features within the text and images, leading to more accurate predictions. The final output of the system will be a comprehensive analysis provided to mental health professionals, assisting them in diagnosing depression and its levels. By utilizing XAI, I aim to improve the transparency of the model's decisions, ensuring that clinicians can trust and understand the AI's recommendations. The proposed system will be evaluated with input from clinical experts to ensure its practical applicability and effectiveness.

With the integration of AI into depression diagnosis, I hope to make mental health care more accessible, personalized, and efficient, ultimately improving outcomes for individuals suffering from depression.

1.2 Literature Review

This paper explores the application of machine learning models to enhance the detection of depression using a tabular dataset. The dataset consists of 10 columns, including a target variable, and captures various patient responses to general questions aimed at identifying depressive symptoms. The study focuses on predicting whether a patient is experiencing depression based on their answers to these questions. By leveraging machine learning techniques, the authors demonstrate the potential for automating depression detection, which could lead to faster and more efficient assessments. However, the research is limited to the detection of depression through tabular dataset and does not extend to further analysis or treatment suggestions. The findings emphasize the importance of using structured data for improving diagnostic accuracy in mental health. [2]

In their study on depressive and non-depressive tweet classification, the authors proposed a sequential deep learning model for depression detection using only text-based data from Twitter. The model comprises three layers: an embedded layer, a 1D-convolutional layer, and an LSTM layer. The study compared the performance of this deep learning model with traditional machine learning models, including Naïve Bayes, KNN, and Random Forest. The sequential deep learning model demonstrated superior performance, achieving an accuracy of 98.47%, surpassing the other models tested. The results highlight the efficacy of the deep learning approach in effectively distinguishing between depressive and non-depressive tweets. [3]

This study addresses the challenge of identifying depression from social media data through machine learning, focusing on improving classification efficiency by integrating feature selection techniques. Depression, particularly major depression, manifests through symptoms such as prolonged low mood and diminished interest in previously enjoyable activities. The research highlights the severe societal impact of depression and explores two machine learning classifiers that predict depressive states based on a combination of demographic and psychological factors. A key aspect of the study is the use of Recursive Feature Elimination (RFE) with linear regression to optimize the feature set, thereby enhancing the model's performance by identifying the most relevant characteristics within the dataset. This approach aims to strike a balance between model accuracy and computational efficiency, offering a novel contribution to the domain of mental health analysis through social media data. [4]

The article demonstrates the development of a system designed to identify and quantify depression levels using visual data, specifically facial expressions. The paper focuses on utilizing machine learning models, particularly Convolutional Neural Networks (CNNs), to analyze visual inputs and detect depressive symptoms. The research is grounded in the premise that non-verbal cues, such as facial expressions, can serve as reliable indicators of an individual's mental health status. The system's architecture leverages pre-trained models, such as VGG19, to capture features from facial images, which are then classified into various depression levels. The paper emphasizes the importance of early detection of depression, highlighting the potential for automated systems to assist clinicians in diagnosing mental health conditions more efficiently. Moreover, the study also discusses the integration of such systems into real-world clinical settings, considering factors like accuracy, reliability, and interpretability of the results. Through experimental evaluations, the researchers demonstrate the efficacy of their approach, showing promising results in the automatic detection of depression levels. [5]

The article delves into the utilization of visual facial cues for identifying and diagnosing depression. The research underscores the importance of facial expressions as non-verbal indicators of emotional and mental health. The study reviews various machine learning techniques, focusing primarily on deep learning approaches like Convolutional Neural Networks (CNNs), which are employed to analyze facial features and classify depressive symptoms. The paper also explores the integration of pre-trained models to extract relevant features from facial images, enhancing the accuracy and reliability of depression detection systems. Furthermore, the research highlights the challenges faced in real-world applications, such as balancing detection accuracy and computational efficiency, as well as addressing issues related to data variability across different demographic groups. The study concludes by emphasizing the potential of integrating these visual-based models into clinical settings to assist healthcare professionals in diagnosing depression, making the process more efficient and scalable. [6]

The paper consists of presents a novel approach to detecting depression levels by leveraging both spatial and temporal features from multimodal data sources. The authors utilize a combination of facial expressions and textual data to create a robust model that captures the nuanced indicators of depression over time. By integrating multimodal inputs, including video frames and speech transcripts, the proposed system enhances the accuracy of depression detection. The research employs a hybrid architecture combining Convolutional Neural Networks (CNNs) for spatial data and Long Short-Term Memory (LSTM) networks for temporal data, effectively addressing the dynamic nature of emotional expressions. Additionally, the study explores the role of feature fusion techniques to merge information from different modalities, thereby improving prediction outcomes. The paper concludes by highlighting the potential of this multimodal approach for real-world applications in mental health diagnostics, offering a promising avenue for early intervention and personalized treatment plans. [7]

1.3 Research Gap

Millions of individuals worldwide suffer from depression, a common mental health illness that makes early diagnosis and treatment extremely difficult. There is still a study gap in the thorough application of multi-modal techniques to diagnose depression by merging various data inputs, such as text, facial expressions, and speech, despite advances in machine learning and artificial intelligence for mental health applications.

A large number of previous research works have concentrated on single-modal methods that employ facial, audio, or textual data to diagnose depression. Although these techniques show promise, they frequently fall short of the depth and accuracy that come from multi-modal analysis. Moreover, rather than offering a comprehensive examination of depression severity levels, the majority of recent research has focused on categorization tasks for determining if an individual is depressed. This restricts these models' practical application in actual clinical settings.

The suggested remedy, "Mentcare AI," seeks to bridge this research gap by employing multi-modal neural network techniques to incorporate text and visual data for more accurate depression diagnosis. When XAI is combined with these models, it offers a clearer grasp of the reasoning behind diagnosis, giving medical professionals more insight into the prediction process. This improves the system's accuracy while also boosting healthcare specialists' confidence in it.

By comparing the proposed solution with existing models, "Mentcare AI" shows promise in advancing the field of AI-driven mental health applications, offering a more holistic and explainable approach to depression diagnosis.

Device/ Application	Depression Detection	Multi-modal Approach	Model Explainability	Attention Mechanism	Depression Level Detection	Real-time Data Integration
Research [2]	Yes	No	No	No	No	No
Research [3]	Yes	No	No	No	No	No
Research [4]	Yes	No	No	Limited	No	No
Research [5]	Yes	Limited	No	Yes	No	No
Research [6]	Yes	No	No	No	No	Yes
Research [7]	Yes	Yes	No	Yes	No	Yes
Proposed Solution	Yes	Yes	Yes	Yes	Yes	Yes

Table 1 : Comparison between existing methods and the proposed tool

2. RESEARCH PROBLEM

Depression is a prevalent mental health disorder that affects millions of people worldwide, contributing to substantial emotional, social, and economic burdens. It is characterized by persistent sadness, loss of interest, and a range of cognitive and physical symptoms, which can severely impair an individual's ability to function in daily life. Despite its widespread impact, depression often goes undiagnosed or is misdiagnosed, leading to inadequate treatment and prolonged suffering.

Conventional approaches to depression diagnosis mostly rely on self-reported questionnaires and clinical interviews, which are highly subjective by nature and can differ greatly depending on the experience of the therapist. In situations where patients are unable or unable to describe their symptoms, these approaches may fall short of fully capturing the complexity of the illness. Furthermore, long appointment wait times are a common consequence of the rising demand for mental health care, which postpones necessary interventions.

Recent technological developments, especially in the area of AI, offer a chance to improve the precision and effectiveness of depression diagnosis. But the majority of current AI-driven methods concentrate on single-modal data, including text-based analysis or facial emotion detection, which might not adequately represent the multifaceted character of depression. It is also challenging for physicians to trust and use AI technologies in practice because of the "black-box" character of many of these models, which raises questions about the transparency and interpretability of the diagnostic process.

This research seeks to address these challenges by developing and evaluating two advanced multi-modal approaches: a Weighted Average method and a Multi-modal Neural Network. These approaches integrate text and facial emotion data to improve the identification and classification of depression levels in patients. By leveraging a mobile application for real-time data collection, the model will provide personalized and timely insights into a patient’s mental health. With the integration of data from many sources, the suggested model seeks to offer a more thorough and precise diagnosis. Additionally, the model's predictions will be visible and comprehensible thanks to the integration of explainable AI approaches, enabling doctors to comprehend the underlying logic and make defensible conclusions. By addressing the shortcomings of single-modal AI models and conventional diagnostic techniques, this strategy hopes to improve patient outcomes for the identification and treatment of depression.

3. RESEARCH OBJECTIVES

3.1 Main Objective

The main goal of this research is to develop and evaluate a comprehensive system that employs two multi-modal techniques which are Weighted Average and Multi-modal Neural Networks to enhance the efficiency and accuracy of depression diagnosis. By integrating two essential modalities, facial emotion recognition and text-based analysis these methods aim to provide a deeper understanding of a patient's mental state. This multi-modal approach captures the complexity of depression by simultaneously analyzing verbal and non-verbal cues, in contrast to traditional methods that often rely on subjective and potentially unreliable self-reported questionnaires and clinical interviews.

A key contribution of this study is its ability to categorize depression severity rather than just identifying the presence of depression. By evaluating varying levels of depression, this approach provides more detailed insights for individualized treatment planning. This is especially important in clinical settings, where a timely and precise diagnosis can significantly impact the treatment plan and the patient's overall outcomes.

Incorporating XAI techniques, this research aims not only to improve diagnostic accuracy but also to enhance the transparency and interpretability of the system's predictions. Explainable AI will allow clinicians to understand not only the diagnosis but also the reasoning behind it, increasing confidence in the system and encouraging its adoption in real-world medical settings. By highlighting which components of the text and facial expressions contributed most to the depression level assessment, the approach becomes a valuable tool in the decision-making process.

Ultimately, this study seeks not only to develop these two multi-modal approaches but also to strictly compare their performance to determine which is most effective. The goal is to identify the best approach for assisting mental health practitioners in making well-informed, evidence-based decisions regarding patient care. By bridging the gap between traditional diagnostic procedures and the capabilities of modern technology, this research aims to significantly improve the overall quality of mental health care.

3.2 Sub Objectives

To achieve the main objective of developing a multi-modal neural network for depression level identification, the following sub-objectives have been outlined:

Sub Objective 1: Develop a Weighted Average Model for Depression Diagnosis

The first sub-objective is to develop a Weighted Average model that combines the outputs from text-based analysis and facial emotion recognition. This model will assign appropriate weights to each modality, calculating a weighted average that reflects the relative importance of verbal and non-verbal cues in diagnosing depression. This approach aims to enhance the overall diagnostic accuracy by leveraging the strengths of both data types.

Sub Objective 2: Design and Implement a Multi-modal Neural Network

The second sub-objective is to design and implement a Multi-modal Neural Network that integrates both text and image data within a unified architecture. This neural network will be trained to learn the combined features of verbal and non-verbal cues, allowing for a more strongly matches analysis of the patient's mental state. The network will be capable of not only detecting depression but also categorizing the severity of the condition.

Sub Objective 3: Compare the Performance of the Two Models

The third sub-objective involves a comprehensive comparison of the Weighted Average model and the Multi-modal Neural Network. The performance of each model will be evaluated based on various metrics, such as accuracy, precision, recall, and F1-score. This comparison will identify which approach provides the most reliable and accurate diagnosis of depression, considering both the detection and severity classification aspects.

Sub Objective 4: Integrate Explainable AI Techniques

The fourth sub-objective is to incorporate Explainable AI techniques into both models. This will involve developing methods that make the decision-making process of the models transparent and interpretable. By highlighting the key features and data points that influence the depression diagnosis, these techniques will provide clinicians with insights into how the models arrive at their conclusions, thereby increasing trust and facilitating the adoption of the system in clinical practice.

Sub Objective 3: Develop a Mobile Application for User Input Collection

The third sub-objective is to develop a mobile application that will engage with users to gather real-time data on their current moods and emotions through text and facial expressions. By asking users how they feel at the moment and so on, the facial images will capture each and every response of patient. The application will collect essential input data, which will be processed by the Weighted Average and Multi-modal Neural Network models for depression diagnosis.

Sub Objective 6: Provide Progress and Current Situation Reports to Doctors

The sixth sub-objective is to develop a system that regularly updates doctors on the progress and current situation of patients based on the analysis performed by the models. This system will generate detailed reports that include both the current depression severity and trends over time, offering clinicians a comprehensive view of the patient's mental health status. These reports will assist doctors in making informed decisions about treatment adjustments and interventions.

Sub Objective 7: Determine the Optimal Model for Clinical Use

The final sub-objective is to determine which of the two models is most suitable for clinical implementation. Based on the comparative analysis and clinical evaluation, the study will conclude by recommending the optimal approach for diagnosing depression. This recommendation will consider both the diagnostic performance and the ease of interpretation, ensuring that the chosen model can be effectively integrated into routine mental health care.

4. METHODOLOGY

To develop an effective system for diagnosing and managing depression through text and facial images, I will follow a systematic methodology encompassing data collection, model development, and evaluation phases.

The first stage focuses on data collection, where three distinct datasets will be utilized. The first text-based dataset, comprising peoples' comments and messages labeled as 'depressed' or 'not depressed,' and the second text-based dataset is consisting of messages labeled as 'suicidal' or 'non suicidal,' will be used one of these datasets to train the text analysis model to identify depression indicators from written communication with future needs. In parallel, the image-based dataset, consisting of facial expressions categorized into seven emotions which are sad, surprised, neutral, happy, fearful, disgust, and angry, will train the image analysis model to recognize emotional states from facial images.

Type	Number of tweets
Depressed	3832
Not Depressed	3441

Table 2 : Text Dataset 1

Type	Number of tweets
Suicidal	116034
No Suicidal	116013

Table 3 : Text Dataset 2

Emotion	Angry	Disgusted	Fearful	Happy	Neutral	Sad	Surprised
No of Train set	3995	436	4097	7215	4965	4830	3171
No of Test set	958	111	1024	1774	1233	1247	831

Table 4 : Image Dataset

Following data collection, preprocessing steps will be conducted to prepare the data for model training. Text preprocessing will involve cleaning and tokenizing the data, removing noise, and normalizing the input using techniques like stop-word removal, stemming, and lemmatization. For the image data, preprocessing will include resizing, normalizing, and augmenting the facial images to ensure consistency and enhance model robustness. Basically, extract all the features need to be train the models.

The core of the system development lies in building two models: the text analysis model and the image analysis model. The text analysis model will employ Natural Language Processing (NLP) techniques, specifically using the BERT model, to analyze the text data for signs of depression. In contrast, the image analysis model will utilize Convolutional Neural Networks (CNNs) with architectures such as VGG19 or ResNet to classify facial expressions into emotional categories.

These models will work in collaboration to detect depression by analyzing both textual and visual cues.

To combine the predictions from these models, two multi-modal fusion approaches will be implemented. The first approach, the Weighted Average Method, involves generating predictions from both models and combining them using a weighted average to derive a final depression score. The weights will be determined based on the performance and relevance of each model. (w_{text} and w_{image} are the weights for the text and image predictions, respectively)

$$\textbf{\textit{Final Score}} = (w_{text} \times \textbf{\textit{Text Prediction}}) + (w_{image} \times \textbf{\textit{Image Prediction}})$$

The second, more advanced approach, is a Multi-Modal Neural Network that integrates text and image features into a single model with distinct branches for each modality. These branches will be fused at a later stage to produce a unified prediction, with an attention mechanism incorporated to dynamically focus on the most relevant features from both modalities, enhancing the model's performance.

An attention mechanism will play an important role in the multi-modal neural network by helping the model focus on significant aspects of both text and image data, thereby improving the accuracy of depression detection. This mechanism will assign different attention weights to various parts of the input data, allowing the model to better understand and integrate information from both modalities.

To ensure the system's transparency and trustworthiness, Explainable AI techniques will be implemented. These techniques will generate explanations for the model’s classifications, such as highlighting important features or decision pathways that led to a particular depression level assessment. By providing these explanations, clinicians will gain a better understanding of the rationale behind the model’s predictions, enabling them to make more informed decisions in the diagnosis and treatment of patients.

The models and fusion methods will be evaluated using a combination of performance metrics and clinical expert feedback. Metrics such as accuracy, precision, recall, F1 score, and AUC-ROC will assess the performance of the models, while clinical experts will review the model outputs to validate the accuracy and relevance of the predictions. Feedback from these experts will be used to refine the models and enhance their clinical applicability.

Upon successful evaluation, the models will be integrated into a mobile application designed for real-time mood and emotion capture. This application will allow patients to input text and images, which will be analyzed by the models to provide depression level assessments. The results will be summarized and presented to doctors, assisting them in diagnosing and managing depression more effectively.

In conclusion, this methodology aims to create a comprehensive system for depression diagnosis that leverages both textual and visual data. By combining multi-modal approaches with advanced techniques like attention mechanisms and Explainable AI, the system strives to offer a robust and clinically useful tool for mental health professionals. Through iterative development and evaluation, this approach ensures that the system will be effective and reliable in real-world scenarios.

4.1 System Architecture Diagram

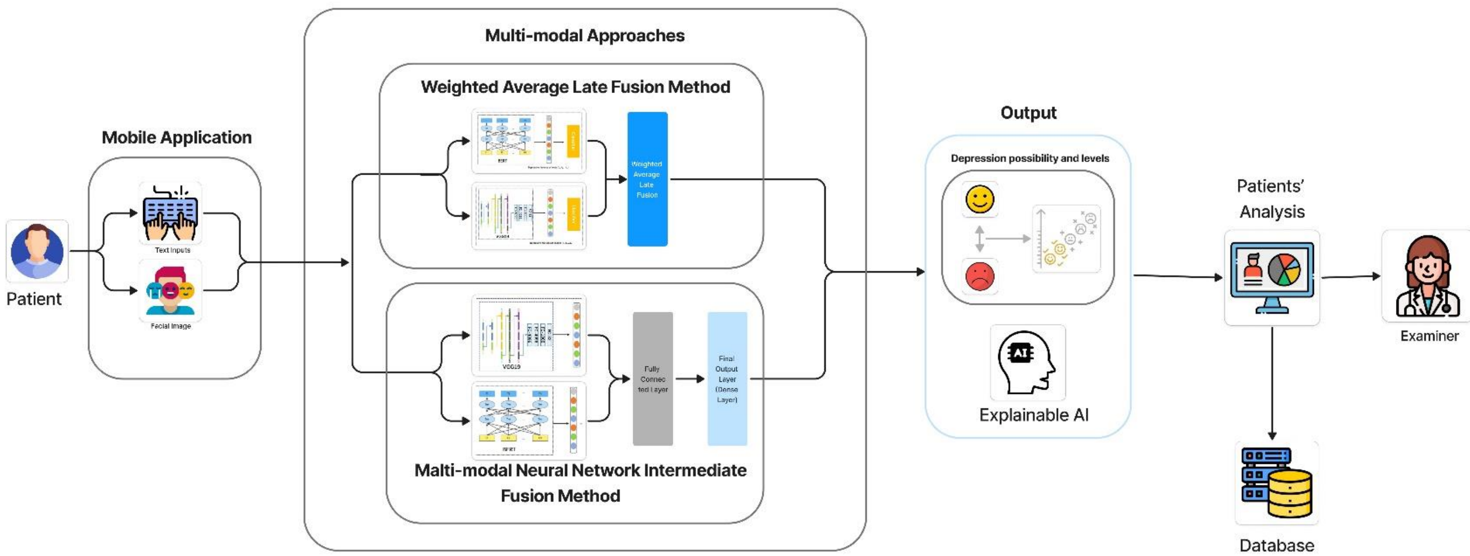

Figure 1 : Individual Component System Overview Diagram

The system is designed to assist doctors in diagnosing and managing depression through a multi-modal approach that combines both textual and visual data. Patients interact with the system via a mobile application, where they input text and facial images. The system uses a pre-trained BERT model to analyze text data for signs of depression and a Convolutional Neural Network (CNN) model, such as VGG19 or ResNet, to classify emotional states from facial images.

The internal process is structured into several stages, starting with preprocessing of the text and image data, followed by separate model predictions for each modality. The predictions are then combined using two fusion methods: a weighted average and a multi-modal neural network that integrates both text and image features. An attention mechanism is incorporated into the multi-modal neural network to dynamically focus on the most relevant features.

Once the analysis is complete, the system provides a depression level assessment, which is summarized and made available to doctors through a user-friendly interface. The system also incorporates Explainable AI techniques to enhance transparency by providing explanations of the model's predictions, helping doctors understand the underlying reasoning. Additionally, clinical experts will review the model's outputs to ensure that the predictions are accurate and relevant for real-world clinical applications. The system will continue to monitor patient data, allowing doctors to track progress and adjust treatment plans as needed.

4.2 Requirements

4.2.1 Functional Requirements

- The application must capture and process both text and image inputs from patients through a mobile interface.
- The system must analyze text inputs using the BERT model to detect depression indicators in written communication.
- The system must analyze facial images using a CNN model (e.g., VGG19 or ResNet) to classify emotional states.
- The application must fuse the predictions from text and image models using both a weighted average and a multi-modal neural network.
- The system must provide a depression level assessment based on the combined analysis of text and image inputs.
- The application must present the analysis results to doctors through a user-friendly dashboard, summarizing key insights and depression levels.
- The system must generate explainable AI outputs to help doctors understand the reasoning behind the model's predictions.
- The application must allow doctors to track patient progress over time by storing and comparing historical assessments.
- The system must support real-time processing and response to patient inputs for immediate depression analysis.

4.2.2 Non-Functional Requirements

- **Reliability:** The system must be highly reliable as it assists in diagnosing mental health conditions. To ensure reliability, the models' outputs will be validated through extensive testing and expert review by clinical professionals.
- **Security:** The system will handle sensitive patient data, including text input and facial images. Therefore, strong security measures must be implemented, such as encryption, secure user authentication.
- **Availability:** The system should be available 24/7 to allow doctors to access patient data and provide timely diagnoses without disruptions. The system's availability will be verified through rigorous uptime testing before deployment.
- **Usability:** The application must be designed to be intuitive and easy to navigate, allowing doctors to access patient data and analysis results efficiently. The user interface should

include clear navigation, a well-organized dashboard, and responsive design to enhance usability.

- **Scalability:** The system must be scalable to handle an increasing number of users, patients, and data inputs as it grows. Cloud infrastructure, such as AWS services, should be utilized to ensure that the system can scale efficiently with demand.
- **Performance:** The system must deliver fast and responsive performance, ensuring minimal latency in data processing, model inference, and dashboard loading times. Regular performance testing will be conducted to ensure the system meets these standards.
- **Maintainability:** The system should be modular and well-documented to ensure ease of maintenance and updates. Developers should be able to easily modify and upgrade components of the system without disrupting overall functionality.

4.3 Commercialization of the Product

Mental health disorders, affect a significant portion of the global population, making this a large and critical market to address. Traditional methods of diagnosing and managing mental disorders involve in-person consultations, which can be time-consuming and costly for patients. Our solution aims to bridge this gap by providing a convenient and cost-effective tool for diagnosing and managing mental disorders through a mobile application and web application. This approach not only reduces the challenges associated with traditional healthcare methods but also offers continuous monitoring and personalized insights.

We plan to introduce multiple subscription plans for different user segments:

- **Monthly Subscription:** Rs. 500
- **Annual Subscription:** Rs. 4,500

Our Target Market is General Practitioners of the MOH centers, Institutes who learn about mental disorders, General population who wish to take self-mental care. We will offer the initial mental disorders diagnosis feature free of charge, allowing users to assess their mental health at no cost. However, to access more detailed reports, continuous monitoring, and personalized treatment plans, users will need to subscribe to one of the available plans. This model ensures accessibility while also generating revenue to sustain and enhance the platform. By offering flexible pricing and targeting a wide range of users, including partnerships with community clinics, we aim to create a scalable and impactful solution in the mental health space, addressing a critical global need.

5. SOFTWARE SPECIFICATIONS

5.1 Facilities

- A high-performance computing system equipped with GPU capabilities to efficiently train and test machine learning models for the analysis and classification of multi-modal data (text and images) in real-time.

5.2 Personal Support

- Supervisors with expertise in machine learning, natural language processing (NLP), and image processing to guide and provide feedback throughout the research, implementation, and evaluation phases.
- Clinical experts, such as psychologists and psychiatrists and general practitioners, evaluate the effectiveness and clinical relevance of the proposed multi-modal approach for depression diagnosis and management.

Purpose	Tools and Technologies
Model Building	Python 3.x (TensorFlow, PyTorch)
Mobile App Development	React Native
Web App Development	React, Flask
Data Storage	MySQL
IDEs	Google Colab, Intellij IDEA, JupyterLab, VS Code
Version Control	GitHub
Image Processing	CNN, VGG19, OpenCV
Data Processing	NLTK, Scikit-learn, Tokenization, Stopword Removal
NLP Tools	Hugging Face (BERT)

Table 5: Tools and Technologies

6. BUDGET

Component	Est. Amount in USD	Est. Amount in LKR
Data Collection through open sources	6.77	2000.00
Charges for Tools Used for Research (Cloud Services, Grammarly, etc.)	31.37	10,000.00
Cloud Platforms (AWS, GCP, OpenAI key)	40.00	13,000.00
Total	96.14	**25,000.00**

Table 6 : Budget Allocation

7. CONCLUSION

The proposed research on depression diagnosis using multi-modal machine learning techniques, embodied in the "Mentcare AI" system, has the potential to make a significant impact in mental health care. By integrating both textual and visual data through advanced models and fusion methods, Mentcare AI aims to offer a more comprehensive and accurate assessment of depression. This system not only enhances diagnostic precision through the use of attention mechanisms and explainable AI but also provides actionable insights for clinicians. The real-time analysis of text and facial expressions will facilitate early detection and personalized treatment planning, potentially improving patient outcomes. Overall, Mentcare AI represents a forward-looking approach in mental health diagnostics, promising to advance the field and offer considerable benefits to individuals experiencing depression.

REFERENCES

[1] W. H. Organization, "Depressive disoders (Depression)," 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[2] U. P. a. Y. K. K. Shah, "Machine Learning-Based Approaches for Early Prediction of Depression," *International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE),* p. 7, 2024.

[3] Anupam Agrawal, Sharmistha Dey, Gopal Chandra Jana, "Depressive and Non-depressive Tweets Classification using a Sequential Deep Learning Model," *International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC),* p. 6, 2023.

[4] Manish kumar Jha, Rakesh Ranjan, Gireesh Kumar Dixit, Krishna Kumar, "An Efficient Machine Learning Classification with Feature Selection Techniques for Depression Detection from Social Media," *International Conference on Communication, Security and Artificial Intelligence (ICCSAI),* p. 6, 2023.

[5] Akshada Mulay, Anagha Dhekne,Rasi Wani, Shivani Kadam, Pranjali Deshpande, Pritish Deshpande, "Automatic Depression Level Detection Through Visual Input," *Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4),* p. 4, 2020.

[6] Sana A. Nasser, Ivan A. Hashim, Wisam H. Ali, "A review on depression detection and diagnoses based on visual facial cues," *3rd International Conference on Engineering Technology and its Applications (IICETA),* 2020.

[7] Mingyue Niu , Student Member, IEEE, Jianhua Tao , Senior Member, IEEE, Bin Liu , Member, IEEE,, "Multimodal Spatiotemporal Representation for Automatic Depression Level Detection," p. 14, 2020.

[8] Abu Bakar Siddiqur Rahmana, Hoang-Thang Ta, Lotfollah Najjar, Azad Azadmanesh, Ali Saffet Gönül, "DepressionEmo: A novel dataset for multilabel classification of depression emotions," p. 18, 2024.

[9] A.Pampouchidou∗§, K.Marias∗, M.Tsiknakis∗†, P.Simos‡ and F.Yang§ F.Meriaudeau§, "Designing a Framework for Assisting Depression Severity Assessment from Facial Image Analysis," *E International Conference on Signal and Image Processing Applications (ICSIPA),* p. 6, 2015.

[10] Kuan Ee Brian Ooi, Lu-Shih Alex Low, Margaret Lech and Nicholas Allen*, "PREDICTION OF CLINICAL DEPRESSION IN ADOLESCENTS USING FACIAL IMAGE ANALYSIS," p. 4.

APPENDICES

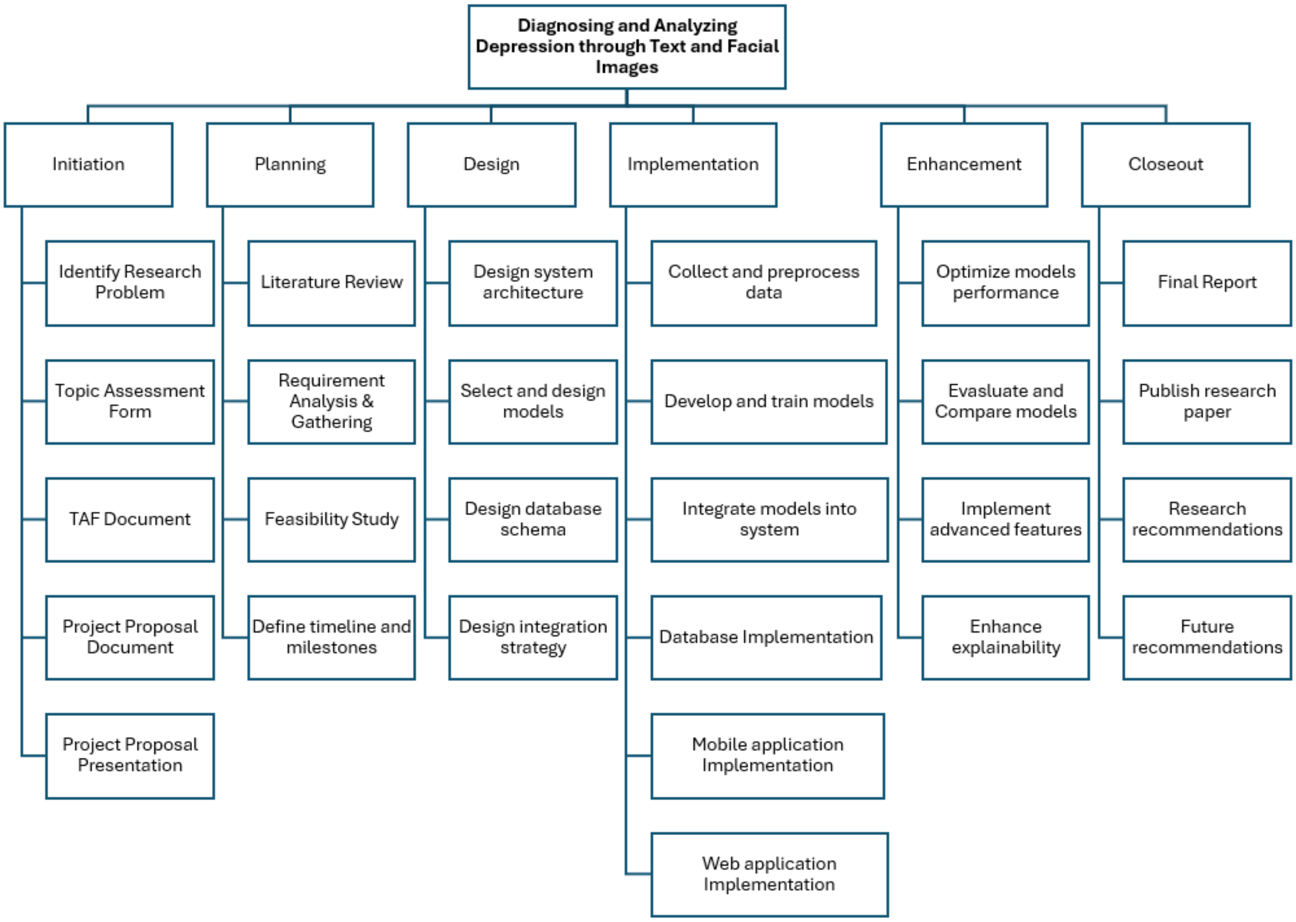

Figure 2 : Work Breakdown Structure

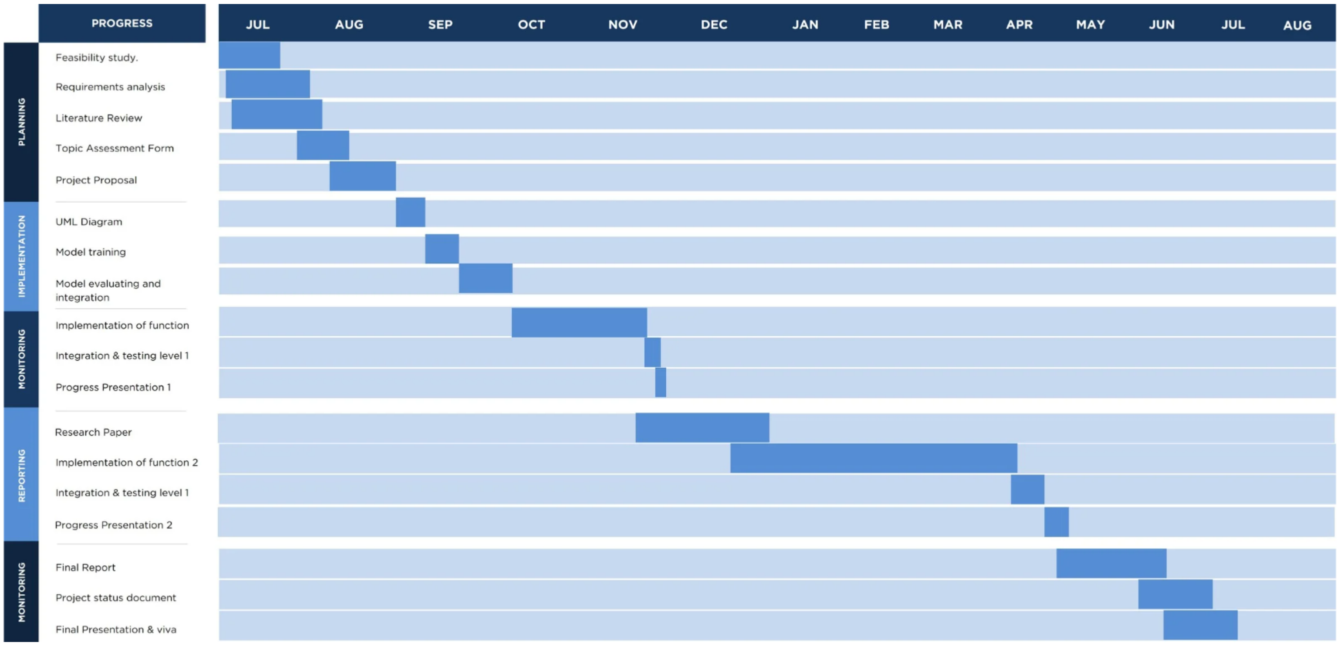

Figure 3 : Gantt Chart

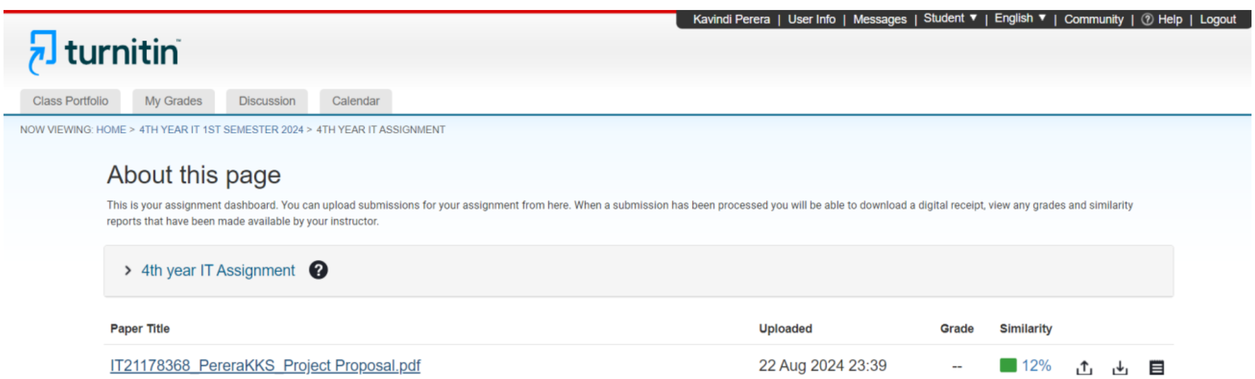

Figure 4 : Plagiarism Value

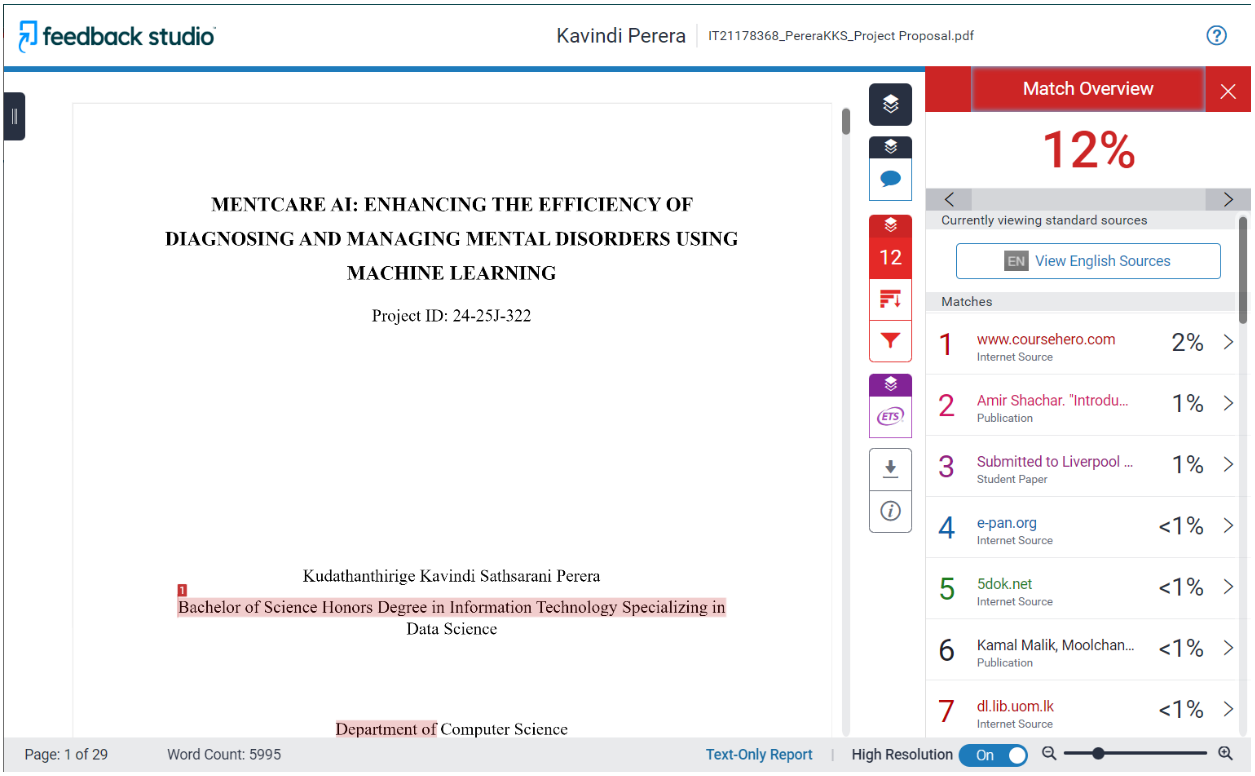

Figure 5 : Plagiarism Results Details

MENTCARE AI

Figure 6 : Applications' Logo