AI BASED MULTIMODAL MENTAL HEALTH MONITORING SYSTEM

BY

TERHEMBAFAN USHAHEMBA HILARY
GOU/U22/CSC/910

A PROJECT SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE AND MATHEMATICS, FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY GODFREY OKOYE UNIVERSITY, ENUGU STATE

IN PARTIAL FULFILLMENT OF THE AWARD OF BACHELOR OF SCIENCE (B.Sc), DEGREE IN COMPUTER SCIENCE.

SUPERVISOR: ENGR DR. KINGSLEY I. CHIBUEZE

JULY, 2026.

APPROVAL PAGE

The research title "**AI Based Multimodal Mental Health Monitoring System**", the work has been approved and Certified by the Department of Computer Science Faculty Of Computing And Information Technology Godfrey Okoye University Enugu Nigeria.

Engr Dr. Kingsley I. Chibueze Supervisor	**……..……….....** Signature	**…...………….** Date
External Examiner External Examiner	**…………………** Signature	**….....................** Date
Dr. Frank Okebanama HOD, Department of Computer Science	**....………………** Signature	**…...…………** Date
Prof. Ajah Angela. I Dean, Faculty of Computing And Information Technology.	**.............................** Signature	**…...................** Date

CERTIFICATION

I hereby declare that I have carried out the research submitted herein, and that it is based to a significant extent on my personal observations and investigation. Thus, I will fully agree with whatever decision the University comes up with should it be established that I have committed academic fraud.

____

TERHEMBAFAN USHAHEMBA HILARY
GOU/U22/CSC/910

DEDICATION

This project is dedicated to Almighty God for His mercy, wisdom, strength, and guidance during the entire process of this research, as well as my dear parents and relatives who provided me with unconditional love, prayers, sacrifices, and encouragement that became the key of my academic achievements and personal development.

ACKNOWLEDGMENTS

Thanks be to Almighty God for His protection, wisdom and strength upon me throughout my academic journey and this research work.

I seize this opportunity to register my deep sense of appreciation and gratitude to Engr. Dr Kingsley I. Chibueze my supervisor for his encouragement, guidance, critical evaluation, advice and support that guided through the completion of this project.

My thanks go also to Prof. Ajah Angela the Dean of the Faculty, Dr. Frank Okebanama the Head of Department and the good lecturers in the Department of Computer science for their support during my study and for the knowledge imparted upon me.

My humble thanks and highest regards to Hembadoon Sandra Terhembafan and Miriam Mnena Terhembafan for the selfless love, sacrifices, prayer, support and care they showered on me to attain this height. This project would not have been a reality without them, my greatest respect.

I also feel great sense of appreciation for my friends, classmates, enemies and well-wishers for the cooperation, support and encouragement I enjoyed from them during my research, I wish to thank Zuwena, my roommates and well-wishers in general. I thank God almighty for the privilege of not allowing the failures that were planned for me by enemies.

Finally I extend my sincere thanks to all who in any way contributed to the successful completion of this project. All your help is well appreciated and god bless you all.

ABSTRACT

Mental health problems such as stress, anxiety, depression, and emotional trauma impact people all around the world; nonetheless, timely emotional assessments are still a concern to people worldwide, especially due to limitations in access to mental health care. Current digital-based mental health monitoring systems mostly rely on only one emotional assessment, providing an incomplete view of the person's emotional state. Here, we describe the design and implementation of an AI-Based Multimodal Mental Health Monitoring System that analyses emotional information from texts, user's facial emotions, and user's screens behavior to give a more complete view of the person’s emotional state. For textual emotion analysis, the fine-tuned DistilBERT has been applied, while for face emotion detection, MobileNetV2 was applied. The behavioural analysis, on the other hand, has been designed with a rule-based system that analyses user-system interactions based on, for example, the time spent on the device, the frequency of actions performed, the pattern of usage of the device and user's inactivity (idle). Those analyses are fused together in order to give us some emotional insights and provide recommendations to improve user’s wellbeing. The system has been developed in Python with FastAPI, React, Javascript, PostgreSQL, and PyTorch and Tensorflow for machine learning purposes. For text emotion recognition we have used a text dataset composed of 5,937 textual samples of three emotions (Anger, Fear, Joy), while for facial expression detection we have used a dataset of 4,948 facial samples categorized as Angry, Happy, and Sad. Data cleaning, transfer learning, hyperparameters tuning and testing have been conducted in order to optimise the performances of the models. With fine-tuned DistilBERT models achieved accuracy up to 99.92%, while optimized MobileNetV2 was able to reach up to 89.00%. The experiment results showcase the capability of transfer learning to adapt already trained models for the tasks described above and prove the effectiveness of the integrated approach for the development of an artificial intelligence based system that uses multiple streams of information in order to detect a person’s emotional state, to understand emotional fluctuations, and to provide personalized assistance. Conclusively, this work points to the great potential of using artificial intelligence with multiple streams of data for improving people emotional assessments, leading towards more aware, responsible and timely digital-based mental health care, interventions, and support.

Keywords: Artificial Intelligence, Mental Health Monitoring, Multimodal Learning, Transfer Learning, DistilBERT, MobileNetV2, Emotion Recognition, Behavioral Analysis, Affective Computing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Mental health is the condition of an individual's emotional, psychological, and social wellbeing, which affects how an individual thinks, feels, and acts in one's day to day existence (World Health Organization, 2022). The prevalence of mental health problems appears to be rising in Nigeria, in both young adults and the working population, representing a serious Public Health Issue across the country (Ajike *et al*., 2022). With many mental health issues remaining undiagnosed until they are more severe, possible reasons for delayed diagnosis might be stigma, lack of awareness and lack of specialist access to mental health professionals (Ogueji *et al*., 2021). The non-utilization of clinical diagnosis also acts as a barrier, as the use of self-report measures may be preventing the uptake of treatment (Patricia, *et al* 2021).

With the widespread usage of smartphones and all sorts of digital technologies, there are a lot of changes in the way we interact and pay attention to our surrounding body in daily life. As the adults are immersed in digital environments for longer periods, the distraction and preoccupation in parent and close ones could potentially influence parental interactions and emotional development of children (McDaniel, 2021). As shown in several research, the interruptions to the ongoing activity, either from checking notification, texting, surfing the web, watching digital contents, can jeopardize or even diminish the coherence in interpersonal interactions and parental responsiveness to children.

The important pastimes of social interaction, emotional sharing and attention play an essential role in creating a psychologically and psychologically healthy relationship. Nevertheless, the abuse of the smartphone can impoverish the individual's presence in the significant real-life circumstances.

For example, all the situations where other people are more likely to be occupied by their own phones: at social events, on the streets and during the emotional circumstances which should promote social bonding. (Kushlev & Dunn, 2021)

In addition, many people get into a habit of repetitive behaviors on their digital devices including frequently switching between different apps and programs, monotonously scrolling through various types of social media feeds, and having passive consumption of information online. Such types of behaviors have been linked with impaired levels of mindfulness, lower self-awareness, as well as increased felt relevance to the present moment (Montag *et al*., 2021). Moreover, technologies also invade spare/lonely moments, which used to be reserved for conscious contemplations, interacting with close others, or relaxing and simply being. Unfortunately, the overreliance on digital technology has gone way beyond the domain of habit and boredom, so as to lead to several psychological problems, for instance, problematic use of smart phones, and anxiety and digital addiction symptoms.

Thanks to the proliferation of mobile phones or smartphones and the internet, we have entered a new era in human history characterized by instant access to devices capable of providing online information, entertainment, and social communication. The flow of information at our fingertips has brought many advantages, although it is increasingly argued that the trend of excessive smartphone usage can be associated with feelings of stress, anxiety, depression and emotional disturbance (Elhai *et al*., 2020). For most people, using a smartphone to a high extent appears to be a way of avoiding unpleasant feelings or thoughts. As a result, the mobile or smartphone seems to be no longer just a platform designed for messaging between individuals but also a means of distraction and escape from our everyday life (Elhai *et al*, 2020).

While digital technologies can be a form of comfort and act as a tool for convenience, over usage can perpetuate another form of alienation from lived experiences and social encounters. People can be seen sitting in front of their screens, going through endless numbers of social media feeds, constantly opening and closing different applications while mindlessly browsing through digital content (Montag *et al*., 2021). This type of usage has left many researchers worried about the mental and behavioral effects of technology dependence.

In this context, the use of Artificial Intelligence for mental health assessment and screening has gained prominence. State-of-the-art AI technologies can be utilized to design intelligent systems that can analyze and detect the earliest signs of mental health issues by examining nuanced changes in human behavior (Torous *et al*., 2023). These intelligent systems utilize an array of data inputs such as text-based inputs, facial analysis, voice features analysis, and physiological data to infer about an individual's mental health. (Rosalind W *et al*., 2021).

We carry these digital devices with us every day as they constantly record our habits, patterns and daily behaviors. We express ourselves in various ways, all of which transmit useful information. The textual data we input plainly tells the story of our thoughts and emotions and the images we provide give us an excellent view of our physical surroundings and the stories that can be found in the human experience. Similarly, speech contains emotional cues through tone, pitch, pace, and pauses that may reveal underlying psychological conditions (Miner *et al*., 2020). Equally, or perhaps even more relevant, and certainly less emphasized source of data is our behavioral interaction with devices.

Screen usage behavior acts as a passive and consistent measure of user behavior. Features like the amount of phone use, time spent on the screen, app switching habits, idle periods, and night-time device use provide insights into possible changes in behavior due to psychological state (Saeb *et* *al*., 2020). While it is true that single behavioral cues alone may not mean a whole lot in and of themselves, a cumulative approach to analyzing them will offer an accurate representation of the user's emotional and mental state.

By combining the use of text with images and a user's screen behavior has created multimodal AIs that are better at understanding humans. Such multimodal systems capture information from several streams or "channels" of human expression, because different modes of presentation naturally contain information which is complementary in nature. (Baltrušaitis *et al*., 2019). In these systems, text acts as an expression of thoughts and emotions, imagery acts as proof and provides context, whereas the user's behavior on-screen provides nonverbal cues to affirm or disprove information gathered through other channels.

The use of behavior related to screen use is especially helpful since this is passive data collected continually without any conscious effort on behalf of the users. This type of data differs from data like text or images since it takes into consideration behavioral habits, resulting in a more accurate depiction of user behavior (Wang *et al*., 2022). The use of such data along with textual and visual data provides a clearer picture of one's mental state.

Although many breakthroughs have been made in multimodal AI research, the problem of properly incorporating multiple types of information into AI algorithms is still under study. The scientists work on developing techniques that would allow for achieving a proper balance between the contributions of different inputs not to let any of them overshadow the others (Ngiam *et al*., 2011). Incorporating all these elements will help achieve a much higher sensitivity level of AI mental health trackers.

1.2 Statement of the Problem

Even though there has been significant progress made in the domain of digital mental health, there are still people experiencing undiagnosed emotional health challenges all across the world. Current AI mental health monitoring technologies are mostly concerned with improving the ability of prediction, with very little concern regarding the user's actual behavior and regulation mechanisms. This is even more evident in Nigeria, where professional mental health care is limited and cultural behavior is important in dealing with emotional challenges.

Scientific literature shows that various psychological problems such as depression and anxiety greatly affect the cognitive processes of the individual, such as attention, memorization, and decision-making (American Psychiatric Association, 2013). However, current mental health monitoring techniques fail to utilize data about the person's behavioral characteristics in this process. Moreover, most of these models use black-box approach and demonstrate low adaptability to real-world situations.

Therefore, this research argues that it is necessary to develop a system that would be based not only on visual and textual communication of the user but also on his/her behavior regarding the screen usage.

1.3 Aim and Objectives of the Study

The aim for this research project named "AI-based multimodal mental health monitoring system" is the development and realization of the multimodal mental health monitoring system consisting of three types of modes - screen usage behavior mode, text mode, and image mode.

1. Review existing literature and related studies on mental health monitoring systems, artificial intelligence, emotion recognition, transfer learning, natural language processing, computer vision, and multimodal machine learning.

2. Identify and define the system requirements necessary for developing a multimodal mental health monitoring platform capable of processing text, image, and behavioral data.
3. Installed Necessary Libraries to ensure I have all the required libraries such as transformers, datasets, tensorflow, Keras, torch, huggingface_hub.
4. Downloaded dataset for both images and text and Prepared Dataset by loading the task-specific data preprocessing the data (tokenization for NLP, resizing/augmentation for CV), Splitting the data into training, validation, and test sets and creating data loaders or tf.data pipelines
5. Load Pre-trained Model DistillBERT from Hugging Face and MobileNetV2 from Keras Applications.
6. Implement transfer learning techniques by leveraging the feature extraction capabilities of the selected pre-trained models.
7. Fine-tune the selected pre-trained models to improve their performance for emotion detection within the mental health monitoring domain.
8. Develop a behavioral monitoring module that utilizes user interaction patterns and screen usage behavior as indicators of emotional state through rule based logic
9. Design and implement a multimodal framework that integrates text, facial image, and behavioral information for emotional assessment.

Evaluate the performance of the text and image classification models using standard

1.4 Significance of the Study

Concerning the growing number of studies related to mental health applications powered by AI, this particular study brings an innovative way of developing such applications, one which is culturally sensitive and human-centric. The fact that this proposed system includes text-based and image-based data as well as screen activity makes it comparable to natural emotional regulation methods utilized by the user population.

From an academic point of view, this project provides a concrete application of multimodal AI, explainable machine learning, and personalization methods. From a social perspective, the importance of contextualizing technology for the resolution of issues in low-income countries is evident.

1.5 Scope of the Study

This project concentrates on mental health monitoring rather than clinical diagnosis or treatment. The system examines multimodal data from publicly available datasets and integrate user behavior to estimate mental health risk indicators such as stress and low mood. The system does not provide clinical diagnosis or treatment; the target users are mainly young adults and students in Nigeria.

1.6 Limitation of Study

In spite of the impact of the AI- based multimodal mental health monitoring system, the study has certain limitations. First, the system does not involve real-time implementation of physical wearable sensors, Secondly, the system is works on android operating system and web, and thirdly, the proposed system operates as a near real-time system rather than a strict real-time system. Finally, screen usage behavior is used as behavioral data to assist mental health monitoring and is not intended for diagnosis or clinical assessment.

1.7 Operational Definition of Terms

Artificial Intelligence (AI)

Artificial Intelligence (AI) can be understood as the ability of machines to perform tasks that would normally require human thinking, such as learning from experience, reasoning through problems, and making decisions (McCarthy, 1956).

Artificial intelligence are systems that perceive their environment and take actions to achieve certain goal depending on the users and the limitation of that particular system set by the developer(Stuart Russel and Norvig, 2010)

Mental Health Monitoring System

A mental health monitoring system refers to a digital system built to continuously observe and assess a person's psychological condition by analyzing behavioral patterns or data collected from sensors (World Health Organization, 2021).

A mental health monitoring system is a digital platform capable of tracking emotional changes and providing psychological support and mental health insights over time. (National Institute of Mental Health, 2022).

Multimodal System

A multimodal system is one that integrates multiple forms of data such as text, speech, and images to enhance understanding and decision-making (Baltrušaitis *et al*., 2019)

It can also be described as systems that combine different input sources to improve computational performance (Baltrušaitis *et al*., 2019)

Data Modality

Data modality is simply the nature of data - the type it has been collected or recorded as. It could be text, audio, images or the output from sensors (Bishop, 2006).

The term modality refers to the method used to perceive data (through sight, sound, or physical sensation) (Baltrusaitis *et al*., 2019).

Multimodal artificial intelligence (AI) is an approach to building an artificial system where multiple modalities are employed to perform tasks that could not be performed by any one of those individual modalities alone (Simon *et al*., 2024).

Screen Usage Behavior

The behavior of screen usage can be understood as the interaction of users with digital gadgets measured through parameters such as screen time, content being viewed, and preferred platforms (Ofcom, 2021).

Screen usage behavior can be defined as the observed user engagement with electronic screens in terms of frequency, time, and context (Ofcom, 2021).

Screen usage Integration

Integration of screen usage entails capturing data on the way people use their devices in order for the system to have a deeper understanding of their behaviors and make more informed contextual deductions (Campbell, 2020).

Physiological Data

Physiological data include data measured as biological indicators from the human body like heart rate or heart rate variability and can reveal useful information on the person's emotional or psychological state (Ginsberg *et al*, 2021).

Physiological data refer to any biological measure derived directly from the person like heart rate or brain function and can offer valuable insight into the person's physical and emotional well-being (National Institutes of Health, 2020).

Machine Learning Model

A machine learning model is a system that learns from data instead of being explicitly programmed for every task. It identifies patterns and uses them to make predictions or decisions (Mitchell, 1997).

Mental Health State

Mental health state refers emotional well-being at the present moment some of the elements that will help in the determination of a person's general psychological or emotional state at any particular point in time (American Psychiatric Association, 2013).

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of literature related to mental health monitoring systems, Artificial Intelligence (AI), multimodal systems, Natural Language Processing (NLP), image analysis, and screen usage behavior monitoring. Also discuss the theoretical background, researched gap, related works findings, limitations and the kind of methodology used.

2.2 Theoretical Background

2.2.1 Cognitive Behavioral Theory (CBT)

Cognition refers to the mental processes involved in acquiring knowledge and understanding through thinking, reasoning, remembering, learning, and judgment (Cherry, 2023). These cognitive processes enable individuals to interpret information from their environment and make informed decisions. Operant Conditioning is a form of learning in which behavior is influenced by reinforcement and punishment. According to this principle, behaviors followed by positive consequences are likely to be repeated, while those followed by negative consequences are less likely to occur (Staddon, 2021). The concept was originally developed by Skinner in 1938 and remains one of the foundations of behavioral psychology.

Cognitive Therapy was developed by Aaron T. Beck and proposes that dysfunctional thinking patterns and cognitive distortions contribute significantly to emotional disorders such as depression and anxiety (Beck & Haigh, 2014; Dobson & Dobson, 2021). Cognitive Behavioral Theory further explains that thoughts, emotions, and behaviors continuously influence one another in a reciprocal relationship (Beck, 1976; Wright *et al.*, 2023).

The relevance of the theoretical perspective in question can be seen when considering that it offers a model through which observable digital behaviors, language use, and social interaction could indicate one's psychological disposition (Al-Malki., 2024).

2.2.2 Human Computer Interaction (HCI)

Human-computer interaction (HCI) refers to an interdisciplinary area of research focusing on the design, evaluation, and implementation of interactive computing systems for human usage (Dix *et al*., 2021). The ultimate goal of human-computer interaction is the enhancement of usability, accessibility, efficiency, and user experience associated with digital systems.

As Hewett *et al*. (1992) point out, HCI entails the examination of interactive computing systems and major phenomena related to their use. Currently, modern digital systems implement engagement-focused approaches designed to enhance interactions and retain users on platforms for as long as possible (Gray *et al*., 2023). Applications such as social media networks often incorporate recommendation engines and infinity scrolls to increase user engagement (Bhargava *et al*., 2024).

With advances in computer systems towards multimodal AI interfaces, it has been argued that the evaluation of system performance should not be limited to user engagement but also include users' well-being and satisfaction levels, as well as impact on psychological health (Shneiderman, 2022). In this regard, it is important to highlight that the current project aims at developing a system capable of promoting psychological well-being.

2.2.3 Affective Computing

Affective Computing involves the use of computer technology to detect, analyze, understand, and simulate emotions (Calvo *et al*., 2020). Affective Computing was coined by Rosalind Picard, who initiated research into this fascinating technology.

Other notable researchers, such as Pantic, Schuller, and El Kaliouby have played a major role in advancing facial expression detection, speech emotion detection, and multimodal emotion recognition methods (Schuller *et al*., 2021). Current affective computing systems obtain emotional data using several types of sensors, which are able to detect facial expressions, vocalization, physiology, and behavior (D'Mello & Kory, 2023).

Facial emotion detection analyzes the movement of facial muscles called Action Units, while speech-based emotion detection analyzes vocal features such as pitch, tonality, and speech rate to detect emotional states (Kopalidis *et al*., 2020). Machine Learning is the use of data patterns and recognition to identify emotions and mental disorders Machine learning is a set of data algorithms used for recognizing and identifying patterns in information that correlate to emotions and mental issues (D'Mello & Kory, 2023).

This theory also provides justification for conducting research due to the fact that the system being developed uses a number of factors such as textual, visual, and behavioral clues for assessing the mental status of the user. In addition to this, the behavior patterns on the screen provide another form of contextual clue (Choudhary *et al*., 2023).

2.2.4 Behavioral Theory

Behavioral Theory is a psychological paradigm focused on observable behaviors rather than internal cognitive processes (Staddon, 2021). Behavior in this regard is viewed as a product of the interplay between a person and his or her environment.

As per Behavioral Theory, human behaviors may be observed, measured, and analyzed in order to uncover behavioral patterns and make predictions regarding upcoming behaviors (Staddon, 2021). Unlike cognitive approaches that are concerned with cognitive processes, behavioral approaches depend on actions and responses to the outer world.

The digital behaviors refer to such parameters as the devices used by a particular person, the applications installed on the devices, activity levels, screen time usage, and navigation behaviors (Stachl *et al*., 2020). Research has revealed that such digital behavioral metrics enable drawing meaningful conclusions regarding the mental and emotional states of people (Daya., 2024).

Behavioral Theory is highly relevant to the research project in question due to the use of screen behavior as a major data set in the suggested system. Such relevance is associated with the focus of the theory on observable behaviors and, therefore, with the assumption regarding the correlation between the device interactions and the state of mind of a person (Stachl., 2020).

2.3 Review of Related Literature

The review in this section is on a set of peer-reviewed journal articles, conference papers, and research studies that are related to the area of study under investigation in this study. This is done in an effort to review past literature on AI-based mental health monitoring systems in order to come up with a well-founded understanding of the research gap under investigation.

The review below is not meant in any way to belittle the achievements of previous researchers. On the contrary, it is meant to point out certain gaps that exist, especially regarding screen usage behavior not being considered as a modality, among others.

Yeasmin *et al*. (2025) conducted an umbrella review of artificial intelligence usage in digital mental health systems through a summary of findings of various studies in the past. In the current study, systematic review methodology was employed in which machine learning, natural language processing, and sensing using wearable technology were employed. It is evident that AI-based systems could be useful for the purposes of early detection and diagnosis within mental health services. However, it appears that the variance in datasets utilized, lack of standard evaluation methods, and problems of privacy and bias in algorithms have lowered generalizability.

Similarly, Gopalakrishnan *et al*. (2024) have conducted a comprehensive study of autonomous mental health monitoring systems through the use of multi-modal sensing devices such as smartphones, wearable devices, and environmental sensors. Machine learning and deep learning approaches have been considered for analysis of signals that could be behavioral and physiological. The outcomes of the survey show that real-time mental state monitoring, particularly anxiety and depression, can be achieved using multimodal sensors. However, computational intensity, dependence on data, and privacy problems are among possible challenges. Drougkas *et al*. (2024) created a multimodal machine learning model that could detect mental disorders through speech and texts. The algorithms used in the study include Support Vector Machine and neural networks with the DAIC-WOZ dataset. The model exhibited significant accuracy in classifying emotions. Nonetheless, it is apparent that the model is still constrained by its dependence on just two modalities speech and text and cannot consider other valuable behavioral inputs like the utilization of the screen.

The second research paper by Lee *et al*. (2024) was based on the concept of using smartphones to conduct behavioral analyses with an emphasis on making AI more understandable in the diagnosis of mental disorders. Application of such mining techniques as association rule mining and Hidden Markov Model brought clarity about the interactions of the users with applications on a micro level. It is evident from the research that the behavioral approach to analysis could lead to improvement of XAI and foster more trust from users. However, the approach looks rather theoretical.

Turous *et al*. (2020) designed a digital monitoring tool for mental health which uses sensors in smartphones to carry out passive sensing functions such as GPS tracking and activity tracking. The model showed an ability to monitor behavioral trends in real-time. But the lack of data modality for emotional signals such as text and speech makes it impossible for the system to offer an overall assessment of the user's mental condition.

Poria *et al*. (2022) have proposed a multimodal sentiment analysis system that fuses three modalities of data, including text, speech, and vision data, using deep learning models. According to the results of their experiment, the use of multimodal models outperforms unimodal models, underscoring the need for multimodal fusion. But the authors failed to include any behavioral data such as screen activity data.

Lin *et al*. (2022) studied the associations between smartphone usage behavior and mental illnesses through the use of statistical and artificial intelligence models. There was a notable correlation found between the excessive use of smartphones and psychological problems including depression. Nevertheless, the authors do not take further action to design a real-time intelligent system, making their research less applicable in practice.

Abd-Alrazaq *et al*. (2021) created a chat bot system based on artificial intelligence to assist patients with mental illnesses using the natural language processing approach. This resulted in increased accessibility of mental health services to patients and their engagement in communication with healthcare providers. However, the chat bot fails to understand emotions of patients and utilize other types of information.

Shatte *et al*. (2021) introduced an artificial intelligence-driven framework that uses behavioral data sets to predict mental disorders. The model showed potential in terms of prediction accuracy. Nevertheless, the use of structured and pre-assembled data sets poses limitations to the system's ability to operate effectively in real-world settings.

Chancellor and De Choudhury (2020) introduced a predictive algorithm based on social media data sets that uses natural language processing tools to identify mental disorders. The model was effective in identifying patterns related to depression. However, the approach is completely text-based and ignores multimodal and behavioral information. It concentrates solely on the textual component of communication.

The use of digital technologies and artificial intelligence in mental health interventions was studied by Calvo *et al*. (2021). The authors looked into the possibility of using AI technologies for helping people in terms of mental well-being using data analysis and intelligent decision-making approaches. It was found that such technologies could help to make mental healthcare more accessible to the patients and offer personalized interventions. However, the authors emphasized difficulties associated with keeping users engaged and adopting the system on the long term. Although the role of AI technologies in mental healthcare was illustrated in the paper, it mostly concentrated on the intervention mechanisms and not on emotion recognition. Moreover, no other data modalities (such as text, speech, and behavior) were used in the research, which makes room for further investigation.

Wang *et al*. (2023) presented a deep learning-driven approach that uses neural network models to detect depression. The approach attained a high level of predictive accuracy, with accuracies of above 90% in predicting users' depressive traits. In this case, Wang *et al*. proved the ability of deep learning methods to predict mental states. Nonetheless, the approach required heavy use of data and computing power, thus limiting its practical application. Furthermore, the model only aimed at predicting depression without focusing on emotions. The current proposal attempts to fill this gap through the use of light pre-trained models that detect multiple emotional states.

Zhang *et al*. (2023) created an emotion detection system from facial images using Convolutional Neural Networks (CNN). The experiment yielded a high level of accuracy of emotion detection via facial images analysis and showed the possibility of using computer vision technology in psychology. In spite of these successes, the algorithm was purely dependent on facial expressions and prone to environmental changes like lighting conditions, camera quality, and possible obstruction of faces. Consequently, the emotions that could not be expressed by the facial expression remained unnoticed. This issue proves the necessity of a multi-channel approach, including not only facial images but also text, voice, and behavior, which is proposed in this research.

Koutsouleris *et al*. (2022) investigated the application of machine learning and predictive analytics for early assessment of psychiatric risks. The results indicated that AI models can assist clinicians in recognizing patients who are at risk of developing psychiatric problems. This study highlighted the importance of applying predictive analytics in preventive mental healthcare services. However, the suggested model was mostly applicable in clinical settings with little scope for real-time monitoring purposes. Moreover, emotion recognition was not considered from various streams of user-generated data. The aim of this research is to address this issue by developing a real-time multimodal monitoring system.

Singh & Sharma (2024) introduced a multi-modal mental health monitoring model which used NLP, speech analysis, and computer vision technologies. The researchers were able to show that the use of multiple data modalities increases the accuracy rate considerably in contrast to single-modal solutions. Despite the successful outcomes, the system was very resource-consuming and required complicated architecture. Such a situation makes it difficult to scale and deploy the solution in resource-poor settings. That is why the proposed research will try to overcome this problem using pre-trained models like DistilBERT, and MobileNetv2.

2.4 Summary of Literature Review

In order to gain an understanding of where this study stands in relation to previous research conducted in this area, it is necessary to conduct an analysis of the work that has already been done by other researchers in this field. There are several works that explore the application of artificial intelligence, machine learning, and smartphone data in the monitoring and prediction of mental illnesses. These pieces of research contain valuable information on methodologies used, results achieved, and problems encountered during research in this field.

Literature Review Table 2.4

S/N	**Author(s)**	**Contribution / Work Done**	**Technique / Methodology**	**Results/ findings**	**Limitations**
1	Yeasmin (2025)	Developed an AI-based mental health diagnosis system focusing on improving clinical decision-making	Review of Machine Learning and Natural Language Processing techniques	AI techniques improved early diagnosis and mental health screening accuracy	Privacy concerns in handling sensitive user data
2	Gopalakrishnan (2024)	Proposed a multimodal AI system for autonomous mental health monitoring	Multimodal Artificial Intelligence	Enabled real-time monitoring and prediction of mental health conditions	High implementation and operational cost
3	Drougas (2024)	Designed an emotion detection system using machine learning models	Machine Learning algorithms	Achieved high emotion classification accuracy	Limited dataset affecting model generalization
4	Lee (2024)	Developed a phone-based behavior monitoring system for mental health analysis	Data mining techniques	Behavioral patterns were successfully identified and interpreted	System not tested in real-world environments
5	Torous (2020)	Introduced smartphone-based monitoring using passive sensing techniques	Passive sensing and tracking	Enabled continuous data collection for mental health analysis	Does not directly detect emotional states
6	Poria (2022)	Developed a multimodal sentiment analysis system for improved emotion detection	Deep Learning techniques	Improved sentiment classification accuracy through multimodal analysis	Lacks behavioral context integration
7	Lin (2022)	Studied the relationship between screen usage and mental health conditions	Statistical analysis combined with Machine Learning	Found a strong relationship between screen usage patterns and mental health status	No practical system implementation

8	Abd-Alrazaq (2021)	Developed AI chat bot systems for mental health support and assistance	Natural Language Processing (NLP) chatbot models	Improved accessibility to mental health assistance	Weak emotional intelligence and understanding
9	Shatte (2021)	Proposed predictive models for mental health condition analysis	Machine Learning techniques	Produced accurate mental health predictions	Limited dataset size affects performance
10	Chancellor (2020)	Analyzed social media data for mental health detection	NLP combined with Machine Learning	Successfully identified mental health indicators from social media posts	Focus limited to text data only
11	Calvo *et al*. (2021)	Investigated digital mental health interventions supported by AI technologies	Artificial intelligence and Data Analytics	Enhanced accessibility and personalization of mental healthcare.	User engagement and adoption challenges.
12	Wang *et al*. (2023)	Developed a deep learning model for depression detection.	Deep Neural Networks (DNN)	Achieved over 90% accuracy in depression prediction.	High computational requirements.
13	Zhang *et al*. (2023)	Built a facial emotion recognition system for mental health monitoring.	CNN (Computer Vision)	Accurately classified emotional states from facial expressions.	Sensitive to lighting, pose variations, and occlusion.
14	Koutsouleris *et al*. (2022)	Applied predictive analytics for early psychiatric risk assessment.	Machine Learning and Predictive Analytics	Enabled early identification of psychiatric risks.	Limited real-time monitoring capability.
15	Singh and Sharma (2024)	Developed a multimodal mental health monitoring framework integrating text, speech, and image analysis.	NLP , Speech Analysis and Computer Vision	Improved prediction accuracy through multimodal fusion.	High computational complexity and deployment cost.

2.5 Research Gap

Previous studies suggest that A.I. can be helpful for mental health monitoring, with A.I. programs analyzing speech, text sentiment, facial emotion and behavior. Nonetheless, many existing systems remain limited. Many have concentrated on only one or two modalities, while the multimodal systems that do achieve high accuracy often require heavy computational power that is not deployable in resource-limited settings. Moreover, little has been done to integrate screen usage behavior as an input to multimodal emotion recognition, despite being a passive and readily available source of emotional information. Moreover, several existing technologies give general responses and not personalized ones.

The AI-powered multimodal mental health monitoring system that solves these shortcomings by merging textual sentiment analysis, facial emotion detection, and screen usage pattern. This system is computationally efficient as it leverages lightweight transfer learning models, supplementing the emotional assessment by suggesting personalized emotional support recommendations based on the user's detected emotional state.

CHAPTER THREE

SYSTEM ANALYSIS AND DESIGN

3.0 Introduction

This chapter will cover the analysis and design of the proposed AI-based mental health monitoring system. In this chapter, we look at how the current mental health monitoring systems work and the main drawbacks of these systems. Also covered in this chapter is an overview of the proposed system and the ways through which the proposed system overcomes the shortcomings of the current system.

Additionally, this chapter will highlight the system requirements, design approach, data gathering techniques used, and the UML models involved in defining the architecture of the system. In this chapter, we will discuss how the input and outputs are designed, the database design, and also the architecture used to implement the system.

3.1 System Analysis

As people continue to encounter stress, anxiety, emotional imbalance, depression, and social isolation at a faster rate, mental well-being has been established as an important component of overall human health. Students, employees, and even younger generations find themselves struggling with mental health problems that hinder them from performing well academically and socially. Although there is a high level of awareness regarding the importance of mental health, some people still lack access such systems.

The evaluation of mental health is performed through face-to-face contact between the patient and mental health experts like therapists, psychologists, and counselors. In these encounters, patients have to provide an account of their emotional status orally or answer questionnaires prepared by professionals. While this method ensures that there is face-to-face communication between the parties involved, it has several limitations, including scarcity of professionals, expensive therapy sessions, lack of time, and stigma related to mental health services.

In order to overcome some of the drawbacks mentioned above, mental health applications are created to help. The existing systems offer mental health monitors or supports based on text-based interactions, where the users express their emotions by sending texts. Some of the systems use basic sentiment analysis mechanisms to identify whether a user feels positively, negatively, or neutrally. Nevertheless, after evaluating these existing systems, one finds that the majority of them largely depend on just one channel of emotional communication text. The natural human emotion can be communicated through different medium like text, image recognition, and behaviors. Yet another problem with current systems is that there is no way to continuously monitor behaviors. Most current systems monitor the mental health state only when the users submit data. In terms of behavioral metrics, which can be indicative of emotions, it might include spending too much time in front of the screen, irregular device activity, and so on.

The domain of problems identified from this research, therefore, revolves around the requirement of a more intelligent method for monitoring mental health, which involves the consideration of a variety of emotional indicators at the same time. The proposed solution aims at overcoming these challenges through the provision of a multimodal approach for mental health monitoring using artificial intelligence.

In the proposed or to-be system, the use of trained models of AI helps in overcoming the problem of model complexity and increases efficiency and effectiveness. Using the multimodal emotion fusion approach, the system will be able to offer better emotional evaluations and also suggest appropriate emotional support to the users.

3.1.2 Limitations of the Existing System

Although there have been many positive changes made possible by the implementation of conventional and technological mental health monitoring systems, there remain some disadvantages that impair the efficiency, precision, availability, and overall effectiveness of these systems. Through the review of the current monitoring mechanisms, it was found that there are issues related to operations and technology that hinder their capacity for providing effective mental health evaluation.

Ineffectiveness of human interpretation is one of the most significant limitations of the current system. Traditional therapy and counseling processes involve conducting a mental health assessment by means of observing individuals, communicating with them, and interpreting questionnaire responses. Mental health assessment may be influenced by the therapist's skills and interpretation. Therefore, there may be inconsistencies in mental health evaluation and suggestions for treatment.

Delayed mental health analysis and reaction is another significant problem associated with the conventional approach. Users of most traditional therapy services receive professional help when consulting with therapists. However, if a person experiences emotional problems when he or she is not able to consult a specialist at once, the emotional situation may be aggravated significantly due to the lack of timely reactions.

Finally, most digital services are also ineffective because they provide limited capabilities of mental health analysis, since they use one or two form of expressing or communicating with the systems. In reality, emotions of people are reflected in different forms of communications, including images, behavioral aspects, and written text. Therefore, analyzing text messages alone may not be enough for assessing people's mental health effectively.

Accessibility another important limitation of existing methods is accessibility. Professional treatment could be expensive, time-consuming, or not even available at all. Most people will not get professional treatment at all because of the stigma attached to seeking therapy, fear of judgment, or simply because there are no facilities providing mental health services in their location. This may cause the patient’s psychological disorders to remain undiagnosed and untreated. Accessibility issue of current approaches is critical to the health community. In addition to that, a number of advanced mental health systems require immense computational capabilities, lengthy training processes and specific hard- ware systems.

For constructing an AI model it will require large quantities of data, high performance computing and experts, therefore the implementation would have higher development costs and complexity. Finally data confidentiality the issue of data security and confidentiality are another important aspect in current techniques, especially because the data describing emotion and behaviors is extremely sensitive, and the risk of compromising personal privacy and the security breaches is another critical disadvantage if data are not handled properly. From these disadvantages that the present system is suffering from is a good reason to come up with an intelligent efficient automatically handled, accessible, and readily available systems that are addressed by proposing this work by creating a unified AI model integrating natural language processing, image recognition and behavioral analysis.

3.1.3 Analysis of the Proposed System

The proposed system refers to multimodal mental health monitoring system, and involves application of artificial intelligence technologies in effective mental health screening and monitoring. The present system intends to improve the mental health screening process by using Artificial Intelligence, Machine Learning, and behavior analysis to yield good results for mental health monitoring and monitoring system.

It has been designed based on the defects of the current systems which use observation and one to two-modal analysis techniques whereas the proposed system is based on multimodal mental health including text analysis, image recognition and behavior analysis at the same time.

The proposed system's key objective is to automate the service of mental health analysis and assistance. Typically, mental evaluation by professionals such as therapist and counsellors, where the process of evaluation is slow and non-uniform. To prevent this kind of complications, the suggested method focuses on using AI models who are capable to assess and predict emotions in order to not depend on the manual analysis and as also to expedite the overall processing time.

Another goal of the proposed system is to increase mental health monitoring systems via multimodal emotion fusion. Current digital systems rely on text-based communication for assessing emotions, others require high implementation and operational cost and hence they fail to understand emotions properly. To overcome this issue, the proposed system will employ a combination of different emotional data sources, including image recognition, textual sentiment analysis, and behavioral tracking.

The proposed system also tries to facilitate the process of emotional surveillance. Current technology allows for analyzing a person’s emotions when he or she shares information during the therapy or chat bot conversation. However, the new system has been designed to introduce behavior tracking functions that can monitor the interaction of the user, including screen time.

Thus, the system will be able to track emotions of the person without any need to analyze the description of the person's state.

Finally, the proposed system is supposed to increase the availability of services related to emotional assessment. Sometimes traditional mental health care systems fail to deliver because of high costs, inconvenient locations, shortage of qualified personnel, and other reasons. In contrast, the proposed system represents a more accessible alternative that enables online evaluation of the person's emotional state.

Additionally, the proposed system endeavors to offer personalized emotional advice and support. Some of the current systems offer generalized responses which may not adequately capture the emotional state of the users. Based on the ability of the proposed system to conduct emotional analysis through multimodal emotions and behavior assessment, personalized responses can be generated based on the emotional state of the individual concerned.

Furthermore, the proposed system will offer an easy-to-implement system by trained machine learning models. Various emotional recognition systems depend on the extensive training of models and the availability of large amounts of data. To solve this problem, the proposed system uses machine learning models for text analysis, image recognition and using behavioral data to ensure a good response from the system.

Lastly, the proposed system will offer an easy data management and monitoring system whereby the emotional data from various users will be stored in a database system it will work as sessions where at any point the user can delete a particular session and start another one.

3.2 System Requirements Specification (SRS)

The System Requirements Specification outlines the operational and performance requirements of the system under development. This document gives an elaborate description of what the system should be capable of performing and under what conditions the system should be able to run. The objective of the System Requirements Specification is to ensure that the AI-based multimodal mental health monitoring and support system meets the users' needs and objectives.

The requirements for the study can be classified as functional and non-functional requirements. Functional requirements focus on the functions performed by the system, whereas non-functional requirements focus on the performance criteria of the system.

The proposed system will be able to conduct multimodal mental health analysis via processing data such as text, facial image, and behavior in order to make emotional evaluations and personalized support suggestions. Besides, the system should ensure secure data handling, efficiency in processing, accessibility and reliability while being used.

The functional requirements will be focused on the actions which the suggested system should conduct during its performance. Such actions can be described as the registration of the user, emotional data collection, conducting a multimodal emotional analysis, monitoring behavior, making emotional predictions, generating recommendations and handling the database. The system will have to handle various types of emotional information from users by means of applying trained artificial intelligence. Emotional information obtained by text analysis, image recognition and behavior analysis will be fused by the multimodal approach in order to boost the accuracy of emotional prediction.

Moreover, the proposed system is expected to offer feedback of real-time emotional status as well as recommendations depending on the current emotional state of the user. It will also be able to store information regarding the user, analysis of emotions and behavior in the form of a database for easy access and further analyses.

In addition to undertaking the above tasks, there will be additional non-functional requirements which the system will need to meet. It will have fast response time and processing capacity in order to enable the system carry out its operations in a timely manner. Emotional and behavioral data are very sensitive, and therefore it is essential that they are processed in an environment with adequate privacy and security.

In addition, the system must also be able to provide a convenient interface where users will be able to access the system easily without having to be highly knowledgeable in technology. The system should be stable & reliable, to guarantee no crash or instability. In the same time, it should be possible to add more emotional analysis modules and support more and more user in the future to facilitate scalability of the system.

On the other hand, the system must also maintain its maintainability and flexibility so that modifications and other future updates on the system and AI algorithms can be done effectively. It is also required that the system will be compatible with different computing devices and platforms.

Thus, the System Requirements Specification acts as an aid to the design, development, implementation, and evaluation of the system since it provides an understanding of the expected functionalities and requirements needed for efficient emotional management and support.

3.2.1 Functional Requirements

Functional requirements consist of the principal activities and services expected to be carried out by the proposed system. These include that the system should have capabilities to:

1. Enable creation of accounts and secure login into the system by users.
2. Accept and analyze facial emotions through facial images input.
3. Enable behavior monitoring through tracking of user behavior like screen activity.

4. Processing of emotional input using trained artificial intelligence models.
5. Analysis of text emotion input using natural language processing methods.
6. Facial expression analysis using image recognition models.
7. Multimodal fusion of emotional output of various modes.
8. Prediction of the emotional or psychological state of the user.
9. Emotional analysis-based feedback generation and personal recommendations.
10. Storage of user information, emotional records, and behavior data in the database in the current session.
11. Administrative management of users and system activities.
12. Access emotional records and previous emotional analysis information in the current session.

3.2.2 Non-Functional Requirements

The non-functional requirements explain the behavior of the system when operating. The non-functional requirements determine the standards and performance of the system being designed.

1. Performance Requirement: The system should be able to process users' input and make emotional predictions in a reasonable amount of time.
2. Security Requirement: The system should guarantee adequate security of the emotional and behavioral data of the users via authentication and management of the data in a safe manner.
3. Privacy Requirement: The emotional history and behavioral details of the user should be kept private and accessed by authorized personnel only.
4. Reliability Requirement: The system should function consistently without any interruptions or failures.

5. Usability Requirement: The system should have an intuitive and interactive interface that allows for easy navigation regardless of the user's knowledge of technology.
6. Scalability Requirement: The system must be available in the future by dealing with expansion, new emotional analysis tools and many more customers with minimum degradation in operation.
7. Maintainability Requirement: The system must maintain good characteristics of upgradability and modifiability of the system components and the AI algorithms that comprises them.
8. Flexibility Requirement: The system must be open for new extensions of the existing technology and the emotional support algorithms in the future.
9. Compatibility Requirement: The system should be capable of operating under various computer platforms, web-browsers and operation system environments.
10. Availability Requirement: The system should always be ready to provide emotional monitoring service for all customers who use this service at any point of time.
11. Efficiency Requirement: The system must reach a satisfactory level of resources utilization and emotional monitoring accuracy.
12. Data Integrity Requirement: The system must provide a good balance of accuracy, wholeness and reliability of personal data, emotional logs.

3.3 System Design Methodology

The methodological approaches that have been used for the design of the proposed system include Object-Oriented Analysis and Design (OOAD), alongside the use of Agile software development methodology. The chosen methodologies can be attributed to their compatibility and ability to manage intelligent systems, which consist of different modules, including intelligent agents, behavior monitoring, emotional analysis engines, database, and user interface.

The proposed AI-based multimodal mental health monitoring includes different modules for gathering, processing, and managing the collected information about emotional state from different sources, including text messages, face pictures, and behavioral activity. Owing to the complexity and modularity of the proposed design, OOAD was chosen as the most appropriate design methodology**.**

OOAD aims to design systems that consist of objects and classes that are self-contained yet related to each other. The objects in the system have unique roles to play while interfacing with other parts of the system when needed. OOAD helps in developing modular, flexible, scalable, and maintainable systems.

The OOAD technique was used to design the proposed system in several modules, such as:

1. User Management Module
2. Module for Gathering Emotional Data
3. Module for Recognizing Facial Emotions
4. Behavior Monitoring Module
5. Multimodal Fusion Engine

Each of these components serves specific purposes that also serve as an input in the general operations of the system. For instance, the behavior monitoring component takes care of detecting user behaviors and analyzing how users use their screens. Multimodal fusion engine ensures that the output data of all components is combined together so as to enhance emotional predictions.

Object Oriented Analysis and Design framework can be utilized for the application of Unified Modeling Language (UML) to visualize the structure and operations of systems. Diagrams from use case diagram, activity diagrams, class diagrams, sequence diagrams, and entity relationship diagrams were employed in modeling the system under discussion using UML tools.

Other than the OOAD, another software development paradigm that was applied in the development of this software is the agile methodology. Systems of artificial intelligence require constant change, testing, and improvement of the system because of changes in dataset, emotional prediction accuracy, and system behavior. In agile software development methodology, the whole development process is divided into smaller parts called iterations.

With the use of agile software development methodology, various parts of the proposed system can undergo implementation, testing, and improvement independently without worrying about development failure. This software development methodology also promotes adaptability; thus, improving emotional analysis model, user interface, and emotional monitoring techniques becomes easier using this approach.

It could be observed from the above explanation that the use of Object-Oriented Analysis and Design together with the agile software development methodology is very ideal for developing the proposed intelligent multimodal mental health monitoring system.

3.3.1 Justification for Methodology

Selection of an appropriate design methodology plays a key role in ensuring success in designing, implementing, and maintaining a software system. With regard to the design methodology for the proposed AI based multimodal mental health monitoring system, OOAD methodology combined with the Agile approach was chosen due to their appropriateness for complex intelligent and modular software systems.

The proposed intelligent system incorporates several independent modules that work together to provide necessary services for emotional monitoring and support. The modules include text analysis, image recognition, and behavior monitoring, multimodal information fusion, generating recommendations, and managing the database. Therefore, it became necessary to use the Object-Oriented Approach to manage such modules.

Another primary reason why OOAD was chosen as a design paradigm is the modularity offered by this approach. OOAD facilitates the separation of the system into several distinct classes and objects, each of which handles some responsibility independently of others. Thus, it becomes possible for image recognition, and behavioral tracking modules to function as separate components of the system but interact with other components at the appropriate time.

Code reusability is another notable advantage that comes with using OOAD. In object-oriented designs, it is quite common to use preexisting classes, methods, and other elements within one or more parts of the system. Code reuse has many benefits such as increased efficiency of development and easier updating. It will become quite helpful in the case of an artificial intelligence-based system because modules can be easily reused within it.

Moreover, OOAD allows achieving better clarity and organization of the system's elements. The application of classes, objects, methods, and relationships makes it possible to provide a well-structured depiction of the interconnection between various elements of the system. Thus, the proposed system will be easier to comprehend, document, and maintain in case something goes wrong. The usage of UML diagrams in OOAD methodology contributes to visualizing the system through depicting various workflows, interactions between the users and the system, relationships in the database, as well as interactions between the system's modules.

Another important advantage of using the OOAD methodology is its suitability for scalability and flexibility. As it is required to implement new functions in the future, e.g., adding a new model for emotional analysis, improving the recommendation system, or incorporating the behavioral monitoring module, OOAD allows including a new component without interfering with other existing elements.

Finally, the decision to apply the Agile methodology has been made considering its flexible nature. Usually, artificial intelligence systems require continuous evaluation, improvement, and modification due to changes in the datasets and the emotional prediction quality, as well as user interaction demands. According to the Agile method, the system can be developed incrementally, that's step by step.

3.3.2 Method of Data Collection

Data gathering is essential during the process of designing the system since it supplies the relevant data that can be used during the analysis, modeling, implementation, and testing processes. The data gathering process was done in the process of designing the AI-based mental health monitoring system in order to have sufficient knowledge on the problem domain and the functionalities of the system.

In this study, data were gathered through the following means: observation, review of literature, online dataset collection, and analysis of current mental health monitoring systems. Through these methods, data were obtained, both theoretical and practical, which assisted in the process of designing and implementing the system.

Among the main methods of gathering data utilized during the study was observation. This was done through observing current digital mental health applications, emotional support websites, and AI chat bots in order to establish how mental health monitoring systems work. It was crucial to observe these systems in order to establish how the system works, especially when it comes to emotional interaction, prediction, recommendation, and communication between users.

One more critical technique that was used is a review of literature associated with this topic. Literature on artificial intelligence, machine learning, natural language processing, image recognition, and behavior analytics was collected from various sources such as textbooks, academic journals, research articles, etc. Thus, the literature review helped get an understanding of modern methodologies used in mental health monitoring.

To complement the literature review and observations, publicly accessible datasets were acquired from online repositories to implement mental health analysis. In the case of using trained models of artificial intelligence for performing multimodal emotion recognition, it became necessary to refer to information about the corresponding datasets on text emotions analysis, and image recognition.

3.4 System Modeling Using UML

To model the proposed A.I.-based multimodal mental health monitoring, we used the standard graphical modeling language called the Unified Modeling Language, or UML, which is frequently used in software engineering to model, visualize and represent the structure, behavior and interaction processes of the system. UML enables the representation of the components in a system, the flow of work and the operational processes and relationships.

Unified Modeling Language (UML), which provides a standard way of describing a system's architecture, process flow, data flow and control and object relationships. It visually describes how the proposed system will work, how people will interface with it and how modules will talk to each other while the system is running.

The system proposed is composed of several important functions (such as user management, behavioral user monitoring, facial emotion recognition, multimodal fusion, emotion prediction, and generation of a recommendation).

Therefore, UML modeling was needed to arrange the components of the system and allow us to understand how the different parts interact. In this work, UML diagrams development activities that define the functional and structural characteristics of the proposed system. Functional modeling shows how the system behaves and how the users interact with it.

Structural modeling represents the internal organization and relationship between pieces of the system.

The proposed system is designed by using UML diagrams.

Several diagrams are used in this research, namely: Entity Relationship Diagram (ERD), Class Diagram, Activity Diagram, and Use Case Diagram

The use case diagram identified the different functions of the system and the major actors interacting with the system.

To represent the operational workflow of the system, we use an activity diagram, starting from emotional data gathering, multimodal fusion, emotional analysis, recommendation creation and storage of emotional records. As the name suggests, the class diagram identified the classes and relationships between the main parts of the system (in this case, users, databases, recommendation modules, behavioral trackers and emotional analyzers). The sequence diagram illustrates the sequential interaction of system components during the emotional analysis and response generation process. The sequence diagram shows the flow of requests and responses between the User interface, Back-end system, A.I.

The database structure of the system was modeled using an Entity Relationship Diagram (ERD). For this project, the ERD will capture the various entities, relationships, attributes, primary and foreign keys required to store information about users, their emotions, behavior, and recommendations.

Specifically, UML modeling was used. UML helps to provide a very well planned system architecture, able to communicate the design ideas more effectively and clearly, and properly document the system which makes it easy to understand and maintain the system in the future (e.g. extension and upgrading of the system).

The adoption of UML modeling in this study improves system clarity, enhances communication of design ideas, simplifies implementation planning, and provides proper documentation for future system maintenance and upgrades. The adoption of UML modeling in this study improves system clarity, enhances communication of design ideas, simplifies implementation planning, and provides proper documentation for future system maintenance and upgrades.

3.4.1 Use Case Diagram

The Use Case Diagram is one of the behavioral diagrams in UML which is used during system analysis and design. It provides a graphical overview of how customers (also referred to as actors) interact with the proposed system, along with the various functions (or use cases) provided by the system. Use Case Diagram is a behavior diagram which shows the services that are provided by the system when it is operating, it helps in identifying the major processes along with all possible outcomes and who is responsible for interacting with the system to perform that function.

A Use Case Diagram for the proposed AI-based multimodal mental health monitoring system was developed in order to illustrate the interaction of various users with the intelligent emotional analysis platform. In view of the user-centered nature of the system, the identified primary actor for the diagram is the user. The diagram focuses on the actions taken by the user, and the system's responses.

Specifically, the user interacts with the system through several functionalities: accounts registration, log-in, uploading facial images, sending text messages, viewing emotional history and insights and receiving emotional feedback and personalized recommendations. These interactions allow the system to collect the emotional information necessary to generate emotional analysis and support.

The system itself is made up of several intelligent processes, including facial emotion detection, text sentiment analysis, behavioral analysis and multimodal emotion fusion. All of these steps are required to analyze the user’s emotional state and produce recommendations or emotional feedback. The <<include>> relationship in the Use Case Diagram captures the necessary internal operations that need to occur before the system can produce the emotional feedback and recommendations. These essential operations include text sentiment analysis, facial emotion detection, behavioral analysis, and multimodal emotion fusion.

For example, in our proposed system, safety alerts are not triggered during every emotional analysis session; rather, safety alerts are triggered only when severe emotional distress or critical emotional patterns are detected. Thus, for our system, triggering safety alerts may represent functionality that can be included as an additional part of emotional assessment. The <<extend>> relationship was used to model conditional system behavior.

The Use Case Diagram presents an overall view of the proposed system’s behavior.

The Use Case Diagram shows the main system functions: Register Account, Login, Logout, Start Chat Session, Send Text Message, Upload Facial Image, View Emotional Insights, Receive Emotional Feedback, View Emotional History, and Receive Support Recommendations.

Internal emotional analysis functions: Text Sentiment Analysis, Facial Emotion Detection, Multimodal Emotion Fusion, Behavioral Analysis, Detect Emotional Distress, and Trigger Safety Alert.

Overall, the Use Case Diagram captures the functional scope of the proposed AI-based mental health monitoring system and ensures a proper understanding of how users will interact with the system while it tracks their emotions and generates suggestions.

1. Send Text Message
2. Upload Facial Image
3. Receive Emotional Feedback
4. View Emotional Insights
5. View Emotional History on that current session
6. Receive Support Recommendations
7. Logout

The internal emotional analysis functionalities include:

1. Text Sentiment Analysis
2. Facial Emotion Detection
3. Behavioral Analysis
4. Multimodal Emotion Fusion
5. Detect Emotional Distress
6. Trigger Safety Alert

The Use Case Diagram helps to clearly define the functional scope of the proposed AI-based mental health monitoring system while ensuring proper understanding of user-system interaction during emotional analysis and recommendation generation. Below is the use case diagram that shows the interaction of the user to the system.

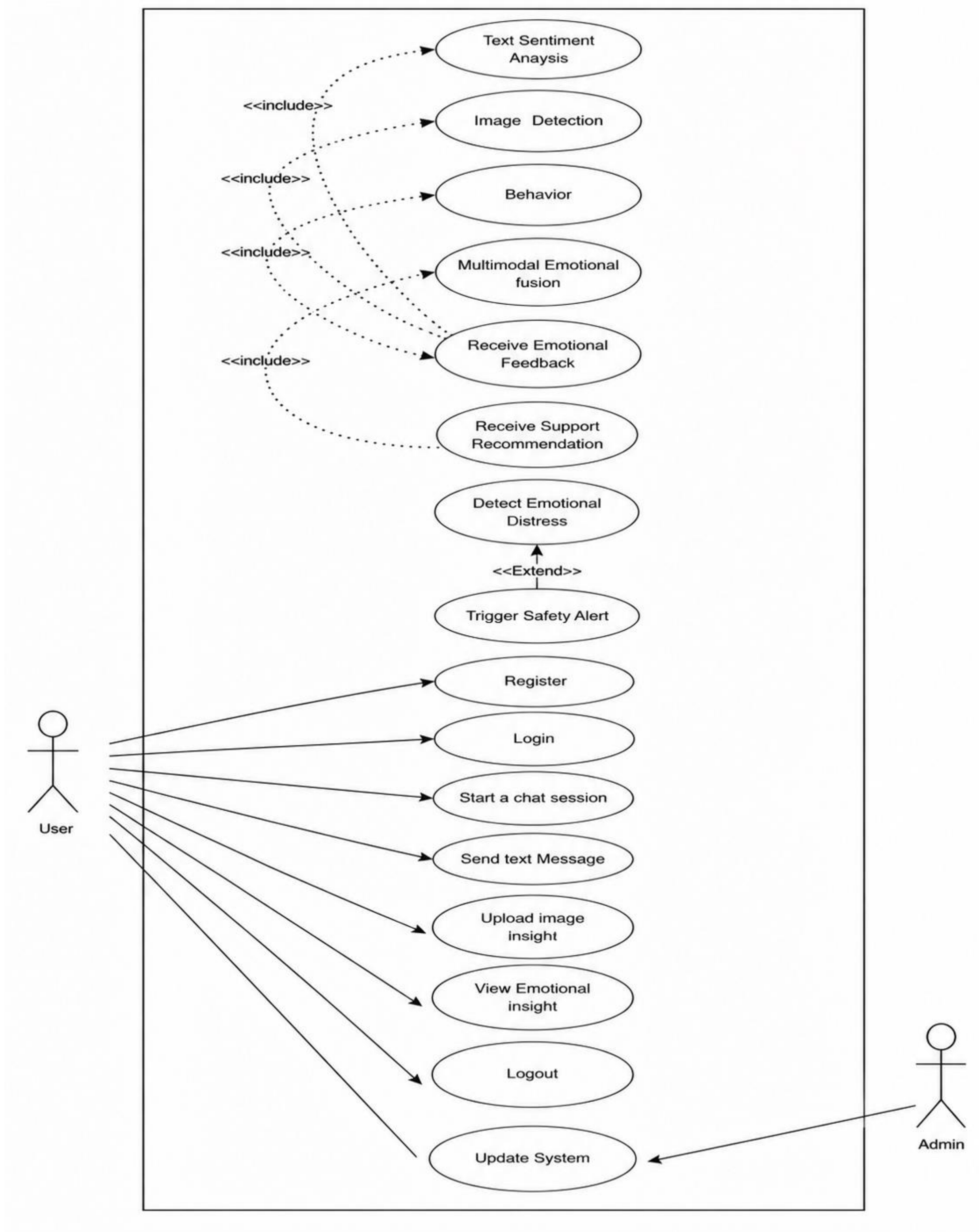

Figure 3.4.1 Use case diagram of the proposed system

3.4.2 Activity Diagram

The Activity Diagram is a behavior diagram in UML which displays the flow of activities performed in a system. An Activity Diagram shows the overall flow of control. It's a particularly important diagram in system analysis and design, as it shows the analysts, the users and the developers how the proposed system will operate.

In this article, we proposed a system for AI-based multimodal mental health monitoring system. An Activity Diagram was designed for our system to depict the workflow of emotional monitoring, safety management, and emotional support generation, modeling the flow of interaction from user authentication, recommendation generation, emotional analysis, emotional history updating, and session ending.

When you start, you enter the system through the login screen. There, you present your user name and password, both of which the system checks to determine if you should be admitted to the system. Users then reach the consent and privacy agreement phase, where they agree to the feelings data use and privacy policy, before entering the dashboard. If authentication is unsuccessful, the user receives an authentication error message and is allowed to try again; after a series of failed attempts the account is locked on a temporary basis. Once the user has successfully accessed the dashboard and selected a support service of their choice, the emotional support session starts.

In the session, the user generates multimodal inputs (e.g., text messages, facial images) that are sent to the preprocessing module. In this stage, important emotional features from the multimodal data are identified through the feature extraction module.

These emotional cues are then interpreted by the emotion reasoning engine, which uses artificial intelligence (AI) and multimodal emotional analysis techniques to interpret the emotional state of the user. Emotional Classification: The emotional classification determines the user's emotional state and their emotional intensity level. After classifying the emotion, another analysis is performed to assess whether urgent emotional distress is present.

This decision is particularly important, as it determines what type of response should be generated by the system. If emotional state is stable: The system generates emotional feedback, wellness recommendations and supportive response, which is displayed to user. Update emotional interaction history and stored in database. Whenever severe emotional distress or a typical emotional pattern is detected, the safety support system will be triggered to store the critical emotional event and provide readily accessible crisis support resources and referrals to professional aides.

This safeguard ensures that the smart system doesn't provide emotional support in an irresponsible or unethical way. The activity diagram (Emotional insight) represents the working of emotional insight of the proposed platform. The system verifies emotional data availability and on finding emotional records, retrieves relevant stored emotional records from the database and computes emotional metrics, plots emotional graphs, and allows downloading of emotional insight for the user, to personally monitor it.

If no records are found, the system will show an empty state. The user may choose to proceed or cancel the emotional interaction session at any time. When the session is over, the system will erase all temporary states and store the interaction data offline before logging out. The Activity Diagram also models emotional insight operations within the platform.

Finally, the user may decide either to continue the emotional interaction session or terminate the session completely. Finally, the user may decide either to continue the emotional interaction session or terminate the session completely.

The major activities represented in the Activity Diagram include:

1. User authentication and login.
2. Consent and privacy agreement.
3. Dashboard access.
4. Selection of emotional support service.
5. Multimodal input collection.
6. Data preprocessing.
7. Feature extraction.
8. Emotion reasoning and classification.
9. Emotional distress evaluation.
10. Emotional feedback generation.
11. Wellness recommendation generation.
12. Emotional history update.
13. Crisis support activation.
14. Emotional insight retrieval.
15. Emotional metrics calculation.
16. Emotional graph rendering.
17. Session continuation or termination.

The Activity Diagram shows in detail the operational work flow of our AI-based multimodal mental health monitoring system. The figure.2 shows the activity diagram

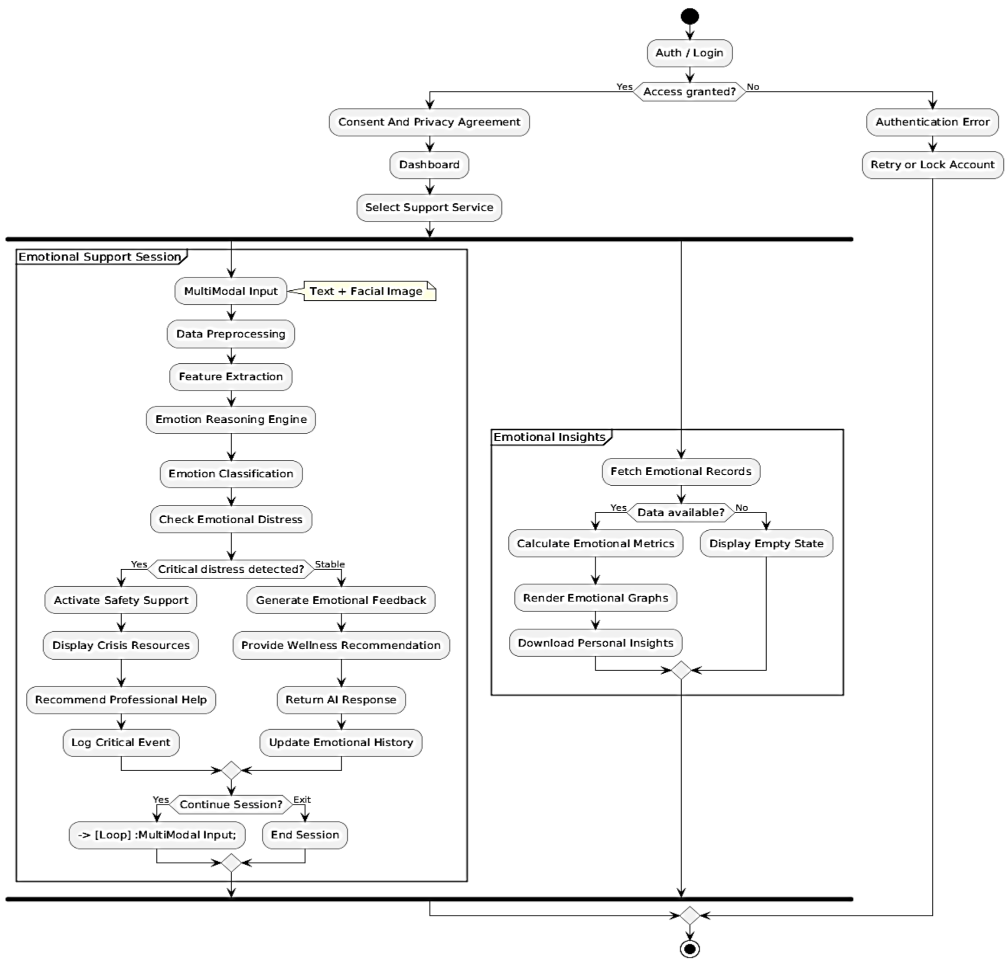

Figure 3.4.2 Activity diagram

3.4.3 Class Diagram

The Class Diagram is a static Unified Modeling Language (UML) diagram that describes the structure of a system by illustrating its classes, attributes, operations, and the relationships that exist among them (Dennis *et al*., 2021). From a business and software design perspective, the Class Diagram represents the functional components of the proposed system and serves as a blueprint for software implementation by defining the objects and interactions required to achieve system objectives (Satzinger *et al*., 2022).

The proposed AI-based multimodal mental health monitoring system was designed using the Object-Oriented Analysis and Design (OOAD) methodology.

The Class Diagram models the main software components in the system, including Emotion Monitoring, Emotion Analysis, Intelligent Recommendation Generation, Database Management and User Interaction Management. The Class Diagram of the proposed system supports the emotional support platform.

These classes include User, ChatApp, ChatContext, BehavioralData, MediaAsset, API_Gateway, AuthGuard, chestrator, VisionEncoder, TextAnalyzer, EmotionEngine, EmotionState, RecommendationEngine, ModelOutput, SafetySupportService, and DatabaseManager. User class to represent users of the system. It will have the following properties: id , fullName , email , isVerified , and createdAt . The User class has the methods register(), login(), and logout().

Users interact with the AI system via the ChatApp class. The ChatApp class contains the following attributes and methods: - imagePreview - messages - userId...handleSend() - captureImage() - viewInsights() ChatApp sends requests to the API_Gateway in response to user actions. The ChatContext class is designed to maintain session-based conversational information throughout the emotional-dialogue process with the chatbot. The class should contain the attributes sessionId, userId, history, and behaviorSnapshot and the methods loadHistory() and storeContext().

This helps maintain continuity in the conversation, and emotional context. The BehavioralData class stores behavioral monitoring information collected from the user during interaction with the system. This information contains several attributes, including timestamp, screenTimeSeconds, devicePlatform, and unlockCount. This behavioral data then gets factored into its analysis of emotions.

A MediaAsset is a resource used to store the uploaded snapshot of an emotional image meant for facial emotion detection. Its main attributes are mimeType, sizeBytes, storageKey, uploadedAt. The key methods are validateMedia(), uploadImage(), and getKey(), all of which have vibrant documentation.

In our AuthGuard class it secures all emotional data stored in the system. Authentication: checks whether the users who have requested system access are correct. Access Token Validation: the verification ensures if the access token that was received to use has been tampered with.

AnAuthGuard manage system and token authentication and access. This guard manages the system and token authentication. Only authorized authenticated person will access this resource through the guard AuthGuard class contains. . - Authentication of user - Token validation the APIGateway class manage the interface between the frontend interface and the backend intelligent processing modules. It makes sure their requests are supported by the system and send the emotional interaction data to the intelligent orchestration layer. It offers the below methods: POSTchat(), GET_dashboard() and handleError() etc.

At the center of the show is the chestrator class along with the VisionEncoder and TextAnalyzer class and is designed for the job of multimodal emotion detection. This in collaboration with the other two is super, super, super accurate, shared with Dr. In the chestrator class, fuseFeatures() method has been implemented for multimodal feature fusion whereas predict() method has been implemented for emotional feature extraction. VisionEncoder class extracts the visual emotion features from the user facial images which are submitted.

BehavioralAnalyzer class takes the task of monitoring user screen usage behavior. The BehavioralAnalyzer class receives the behavioral data of the user from the operating system and monitors various parameters like time spent on screen, number of times the user unlock his/her screen, application switching frequency, duration of each session, time of night screenusage. BehavioralAnalyzer class contains the main methods extractBehaviorFeatures(), calculateBehaviorScore(), detectBehavioralRisk(), and generateBehaviorProfile(). Output from the BehavioralAnalyzer class is passed to the EmotionEngine while doing multimodal fusion.

It contains an encoder member, a weightsPath member and two methods preprocess() and encode(). The VisionEncoder class will analyze and process the uploaded facial images to recognize emotions. The VisionEncoder class will analyze and process the uploaded facial images to recognize emotions. The TextAnalyzer class is utilized for text-based sentiment and emotion analysis. It handles user textual inputs and extracts emotion features using a fine-tuned DistilBERT model. It has attributes like modelPath and tokenizer, and methods such as preprocessText(), tokenizeText(), extractFeatures(), and predictEmotion(). Extracted emotional features are passed to the multimodal fusion engine.

This class contains various attributes, including thresholds and rulesPath, and methods such as fuse(), persist() and classifyDistress(). The EmotionEngine class works by combining the emotional outputs of the facial emotion detector, behavior analysis and text sentiment analyzer to determine the overall emotional state of the user. The EmotionState class stores the final emotional state result from the multimodal emotions analysis. The EmotionState class attributes are: recordedAt, distressLevel, valence, arousal and label.

The RecommendationEngine provides individualized emotional support recommendations and wellness recommendations based on the identified emotional state. Methods include generateEmotionalFeedback(), generateWellnessRecommendation(), and provideCopingStrategy().

The ModelOutput class contains the final output generated by the AI system after the sentiment has been analyzed. This includes the intelligent response that is delivered back to the user, as well as attributes such as the model version, response text, sentiment label, distress flag, and confidence. SafetySupportService offers support in the case of an emotional crisis.

This class is called whenever emotional severity surpasses defined safety thresholds. These include recommendProfessionalHelp() , displayCrisisResources() and logCriticalEvent() . The SafetySupportService class handles emotional crisis support operations whenever severe emotional distress is detected.

The DatabaseManager handles communication with the database and data persistence. This function calls several methods in the DatabaseManager, including queryDashboard(), retrieveHistory(), saveChatMessage(), saveEmotionState(), and saveModelOutput(). This helps store and retrieve emotional traces.

This class diagram indicates the important relationships between the classes that form the system. The dependencies and associations between classes are shown by solid and dashed lines respectively. The User class talks to the ChatApp, which makes requests to the API_Gateway. After the API_Gateway passes the request through the AuthGuard to check tokens, it passes it to the chestrator, which calls the TextDecoder and VisionEncoder classes to extract emotional features and send them to the EmotionEngine. The EmotionEngine produces an EmotionState that is saved in the DatabaseManager. Also, the SafetySupportService is triggered whenever there is severe emotional distress detected.

The RecommendationEngine generates emotional support feedback that is saved as part of the ModelOutput. Below is the class diagram;

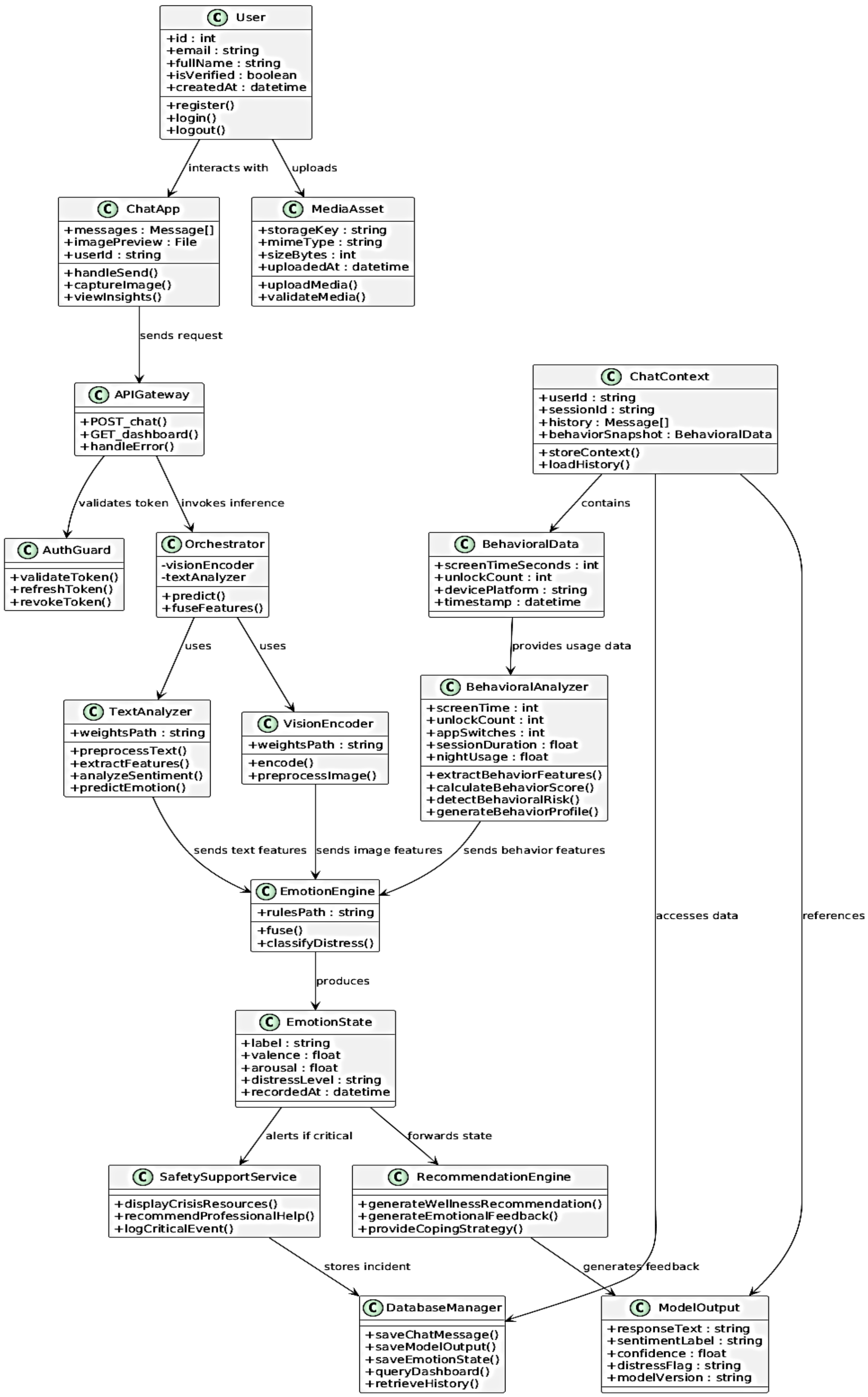

Figure 3.4.3 Class diagram

3.5 System Design

System design is the process of transforming analysis and design models and specifications into constructs that can then be implemented. In the following sections, the input, output, database, and architectural design of the proposed system will be described.

3.5.1 Input Design

Input design is the process by which the methods used for providing input to the system are analyzed. The system under discussion was designed with an interactive and user friendly interface that enables users to effectively convey emotional information to the system.

The key input interfaces for the system include the registration page, the login interface, chat interface, image uploading interface, and behavior tracking module.

Both the registration and login involve creating an account by the user and gaining entry into the emotional monitoring portal. User details such as first name, last name, email ID, and password will be stored in the database. This includes email validation, password length validation, no duplicate account creation, and token-based authentication

Through the chat interface, users deliver messages about how they feel to the system.Textual inputs in the form of emotional expressions can be inputted through the chat interface. Textual inputs are subjected to input validation processes to detect empty inputs and input data that could damage the system.

Users are permitted to upload facial images that would later undergo analysis to identify their emotion. Facial images uploaded to the server are validated for image types, image size, and image quality prior to undergoing emotional recognition.

The behavior of users, including the amount of screen time spent, interaction, and device activity, are tracked through the behavioral monitoring system. Behavioral data acquired are preprocessed to provide structured data and then fused during emotional recognition.

The following processes are among those employed for input validation in the system:

1. Empty input validation
2. Image file validation
3. Token-based authentication
4. Input sanitization
5. Data format validation
6. Prevention of duplicate accounts

3.5.2 Output Design

The design of the output revolves around the presentation of data that has been processed by the system following emotional analysis. The system will produce intelligent emotional feedback, emotional status reports, wellness tips, emotional insights, and trend visualizations in emotions.

Main output generated by the system will be the emotional response from the AI Emotional Support Engine. Following the completion of the multilayer emotional analysis process, the system will provide the empathic conversation along with the emotion state detected stress, anxiety, sadness, emotional stability, or emotional distress.

Apart from providing empathic feedback and emotion detection services, the system will generate emotional insight dashboards containing visual charts and analytical insights depicting the pattern and trends in the emotional condition of the user. Other outputs generated by the system will include the following:

1. Confidence scores on the emotion detected

2. Severity of distress
3. Recommendations for improving wellness and managing emotions
4. History of emotional conditions and wellness
5. Emotional distress alerts when detected
6. Behavioral trend analysis reports

The dashboard section will depict the analytics on emotions through visual outputs such as line charts, graphs showing the frequency of emotions, emotional severity bar charts, and behavioral activity graphs.

Emergency emotional recommendations, crisis management advice, and professional counseling notifications are some of the other outputs generated when an emergency situation involving critical distress is detected. These outputs have been created keeping the user interface guidelines in consideration.

3.5.3 Database Design

The database for our proposed AI-based multimodal mental health monitoring system was designed using a relational database structure. A relational database structure provides a standards-based, scalable, and secure way of storing, managing, and retrieving structured emotional interaction data generated during user-platform interaction.

This multimodal emotion dataset of our proposed system includes the upload of facial images captured by the system (we identified these as facial images in privacy agreements), emotional classifications, text inputs when prompted, A.I.-generated text responses, and behavioral monitoring logs.

The Entity-Relationship Diagram (ERD) shown in Figure illustrates the major entities, attributes, primary keys, foreign keys, and relationships used within the system.

USER Entity

The USER table is where we store the personal and authentication information of each person with whom we interact.

It is the parent table from which other behavioral and emotional tables are joined. The USER table stores the personal and authentication information of individuals interacting with the system.

Table 3.1: User Entity

Field	**Data Type**	**Key**	**Description**
Id	int	PK	Unique user identifier
Email	string	UK	Unique user email
full_name	string	—	Full name of user
password_hash	string	—	Encrypted password
is_verified	boolean	—	Verification status
created_at	datetime	—	Account creation date

The USER entity has one-to-many relationships with both the CHAT_MESSAGE and BEHAVIORAL_DATA entities. Examining the diagram.

CHAT_MESSAGE Entity

The CHAT_MESSAGE table stores textual interactions between the user and the AI mental health support system. Each message submitted by the user is recorded alongside its processing timestamp.

Table 3.2 Chat Message Entity

Field	Data Type	Key	Description
id	int	PK	Unique message identifier
user_id	int	FK	References USER table
role	string	—	Message sender role
content	text	—	User chat content
timestamp	datetime	—	Message timestamp

It also has a relationship to the USER table over the user_id foreign key, so we can associate conversations with users. Columns are added to the CONVERSATION table to track when each conversation began and ended, as well as to note the completion status. The diagram shows.

BEHAVIORAL_DATA Entity

The BEHAVIORAL_DATA table contains user behavioral metrics tracked by the system during the interaction session. These help its multimodal engine identify patterns in distress or dysregulation.

Table 3.3 Behavioral Data Entity

Field	Data Type	Key	Description
Id	int	PK	Unique behavioral record
user_id	int	FK	References USER table
screen_time_seconds	int	—	Screen usage duration
unlock_count	int	—	Number of device unlocks
interaction_frequency	int	—	Interaction count
idle_time_seconds	int	—	User inactivity duration
app_usage_pattern	string	—	Usage behavior pattern
timestamp	datetime	—	Record timestamp

The behavioral data structure enables the system to determine behavioral anomalies associated with, for example, anxiety, mental fatigue and emotional distress. Based on these inferences, the system monitors the emotional state of the user over time.

MEDIA_UPLOAD Entity

The MEDIA_UPLOAD table accommodates the multimedia files uploaded by the users during the emotional interaction session. The project also handles emotional analysis of facial images that are uploaded. As its name suggests, the MEDIA_UPLOAD table keeps track of the multimedia files uploaded by users as part of the Emotional Interaction sessions.

Table 3.4 Media Upload Entity

Field	**Data Type**	**Key**	**Description**
Id	int	PK	Unique media identifier
message_id	int	FK	References CHAT_MESSAGE
image_path	string	—	Image storage location
upload_type	string	—	Type of uploaded media
uploaded_at	datetime	—	Upload timestamp

The relationship between CHAT_MESSAGE and MEDIA_UPLOAD allows messages to contain optional media attachments for multimodal emotional analysis.

EMOTION_ANALYSIS Entity

The EMOTION_ANALYSIS table. This table is the intelligence piece of the system and saves the predictions on emotion obtained via multi-modal analysis performed by trained AI models.

Table 3.5 Emotion analysis Entity

Field	Data Type	Key	Description
Id	int	PK	Unique analysis identifier
message_id	int	FK	References CHAT_MESSAGE
text_emotion	string	—	Emotion from text analysis
facial_emotion	string	—	Emotion from facial analysis
fused_emotional_state	string	—	Final multimodal emotion
confidence_score	float	—	Prediction confidence
severity_level	string	—	Emotional severity status
insights	text	—	AI-generated insights
analyzed_at	datetime	—	Analysis timestamp

RECOMMENDATION Entity

Based on emotional scoring, customized emotional coaching feedback and health tips are delivered to be saved into RECOMMENDATION table.

Table 3.6 Recommendation Entity

Field	Data type	Key	Description
Id	int	PK	Unique recommendation ID
analysis_id	int	Fk	References EMOTION_ANALYSIS
recommendation_text	text	—	Wellness recommendation
Recommendation_type	string	—	Type of support
Generated_at	datetime	—	Recommendation timestamp

This module offers personalized recommendations concerning mental wellbeing considering the emotional state analysis. There are many critical relationships in the database schema which guarantee correct communication between the system elements:

ONE USER can post MANY CHAT_MESSAGE records.

ONE USER can create MANY BEHAVIORAL_DATA records.

ONE CHAT_MESSAGE may include MANY MEDIA_UPLOAD items.

ONE CHAT_MESSAGE creates ONE EMOTION_ANALYSIS record.

ONE EMOTION_ANALYSIS record may create MANY RECOMMENDATION records.

All these relations provide the general idea behind the system, namely help the user to analyze his history, monitor his behavior and emotions, and develop personal mental health advice.

The design uses a relational database to keep track of user activities, behavioral metrics, emotion data, and the history of recommendations. Users will be able to enter inputs that will be received via the multimodal A.I. engine through this portal. The scalability of our database allows for future addition of other emotional modalities and mental health analysis.

We used SQL (Structured Query Language) in performing operations related to data retrieval, update, management, and insertion from and into the database. Analysis of the emotional results using the multimodal A.I. engine is recorded in the database, Recommendations are dynamically obtained when required. Structured Query Language (SQL) was used in data insertion, retrieval, update, and management processes.

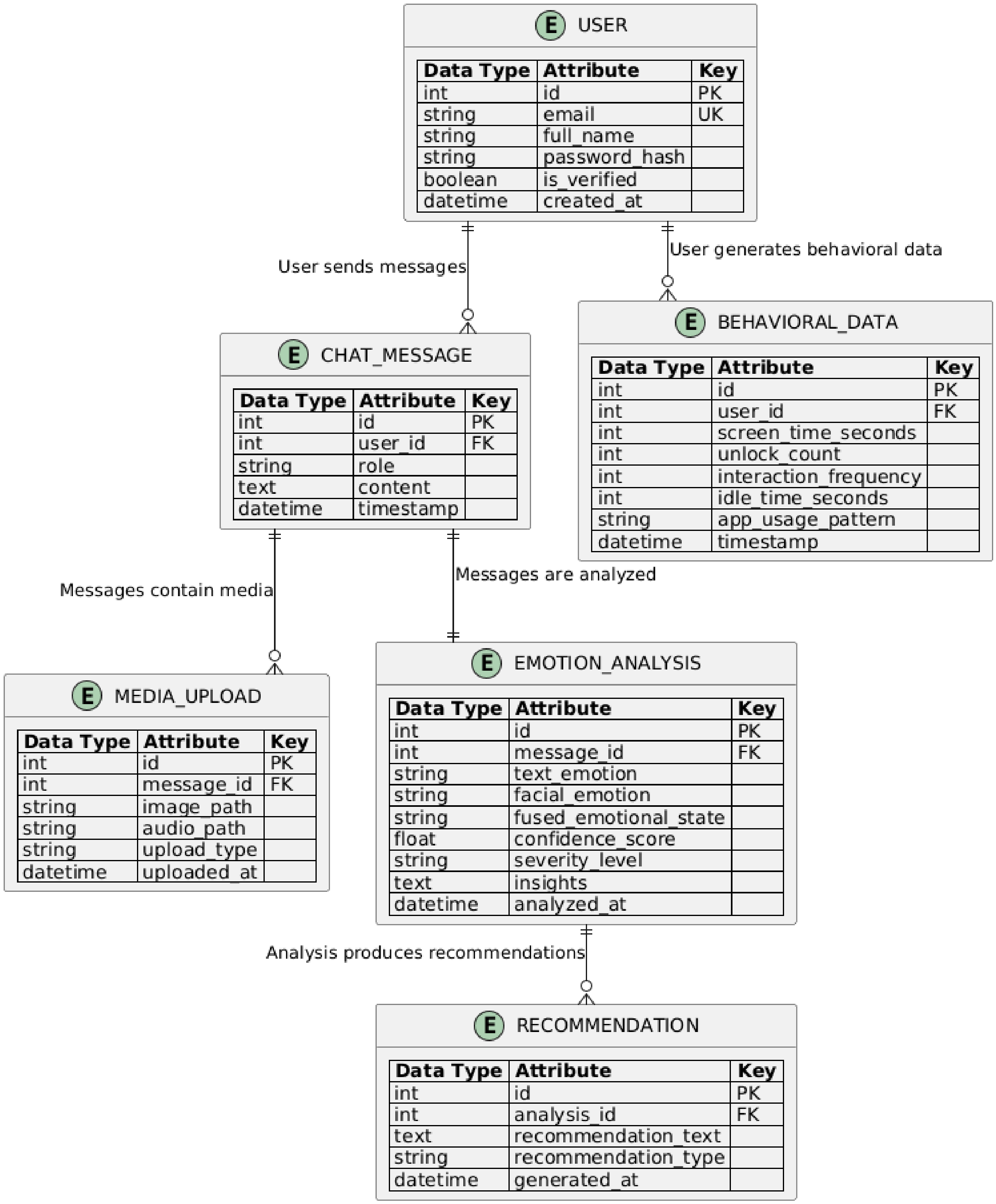

Figure 3.5.4: AI-Based Multimodal Mental Health monitoring system ERD

3.5.4 System Architecture Design

System Architecture Design is the design of the structure of the suggested AI-Based Multimodal Mental Health Monitoring System. It also outlines the interaction among various elements or subsystems of the system in order to perform effective emotional monitoring and intelligent support for mental health wellness. The architecture design of the suggested system comprises a web-based client/server architecture along with artificial intelligence services for multimodal emotional analysis.

The architecture design of the developed AI-Based Multimodal Mental Health Monitoring System adopts a layered approach in order to make the system modular, scalable, and easy to maintain, along with facilitating seamless communication among the various components including frontend, backend, artificial intelligence models, and database system**.**

Architecture of the entire system can be divided into four layers as follows:

1. Presentation Layer (Frontend)
2. Application Layer/Backend
3. AI Processing Layer
4. Database Layer

These architecture layers takes care of everything from user interface to multimodal emotion analysis, prediction, monitoring, recommendation generation, and emotional data management.

Presentation Layer (Frontend)

This layer is the user interface of the system. Users can interact with the AI mental health assistant through either text or image-upload capabilities. The frontend was developed as a responsive web app, deployed to a cloud platform. The frontend includes interfaces for users can sign up and log in to upload images and view emotional feedback and conversation sessions, as well as track their emotional history. The user interface was developed to be interactive and could be viewed on a desktop or mobile browser (app or web app).

The front end communicates with the various back end services via RESTful APIs, which use HTTP requests to transmit data back and forth in a structured way using JSON.

This frontend is consistent with modern single-page app architecture, and enables real-time affective interaction between users and the AI.

The frontend communicates with backend services through RESTful APIs using structured HTTP requests and JSON data exchange. The frontend communicates with backend services through RESTful APIs using structured HTTP requests and JSON data exchange.

Frontend is implemented in accordance with the modern single-page application architecture. Users and the AI system support emotional real time interaction.

Application Layer/Backend

Application Layer, which includes authentication services as well as a way to facilitate a safe interaction between all system components for performing emotional processing, API services, and business logic. Below are listed the main modules of the backend architecture:

These include API Gateway, Auth service, Database Access Services, Safety Support Module, Recommendations Service, and Chat Processing Controller.

The API Gateway receives frontend requests, sends the requests to the appropriate emotional processing services, and handles authentication.

The backend then sends the data to the AI processing modules to perform multimodal emotional analysis.

Once the emotional processing is completed, the backend saves the emotional outputs and recommendations in the database and responds with the information to the frontend interface.

It ensures that ranges all the way from user authentication and session handling, to secure endpoint processing, API routing, emotional data orchestration, and system resiliency.

The API Gateway receives requests from the frontend interface and routes them to the appropriate emotional processing services. The API Gateway receives requests from the frontend interface and routes them to the appropriate emotional processing services.

The backend also coordinates multimodal emotional analysis by forwarding textual, behavioral, and image data to the AI processing modules. The backend also coordinates multimodal emotional analysis by forwarding textual, behavioral, and image data to the AI processing modules.

AI Processing Layer

The A.I processing layer is the brain of the system. The top layer is the multimodal emotional monitoring framework, which performs emotional analysis using trained A.I models. The emotion analysis text module processes emotions through analyzing user communications based on natural language processing. The emotion analysis face module detects emotions by analyzing uploaded facial images. The behavioral analysis engine is used to assess user interaction behavior, which includes screen activities, the number of interactions made, the time spent without making any interaction, and patterns of use.

These outputs from emotional modalities are then integrated in the Emotion Fusion Engine to provide an emotional state prediction. Emotional states are assessed and recommendations or coping strategies for mental well-being are generated.

Database Layer

The database layer keeps all the structured data produced by the application during its operation. Relational database design is implemented to assure proper organization, integrity, consistency, and effective handling of emotional data. The database contains the following user accounts data, chat data, uploads, behavioral tracking data, emotional data analysis results, emotional level, history of recommendations provided, and emotional data interactions.

The backend connects to the database using queries to perform operations such as adding, changing, fetching, and managing emotional data. The database also provides support for keeping track of emotional data history and future emotional trends analysis.

CHAPTER FOUR

SYSTEM IMPLEMENTATION

4.0 Introduction

In this chapter, the implementation, testing and evaluation of the proposed AI-Based Multimodal Mental health Monitoring System is presented. The chapter describes how the system designed was implemented to become a functional application that can process user's emotional states through behavioral interaction patterns, facial images and text input.

This chapter describes the software tools, programming languages, frameworks and pre trained AI models used during development. I also explain how I put the behavioral monitoring components, backend services, frontend interface, database and emotion analysis modules together to form a single platform.

In addition this chapter discusses the testing that was carried out on the system in order to determine its functionality and performance. In this part of our study, we subjected various parts of the system to differing forms of emotional input to ensure that the resulting outputs conformed to our expectations.

Finally, this chapter presents the results obtained from the implementation and testing stage of the system. The results are discussed based on the system's ability to perform multimodal emotional analysis in order to generate emotional insights and pattern-of-life behavioral monitoring of, as well as supportive recommendations to, users, in real time.

4.1 Development Environment and Tools

A number of software tools, libraries and frameworks suitable for artificial intelligence and web-based application development were employed in the development of the proposed AI-Based Multimodal Mental health Monitoring System.

FastAPI and Python were utilized in the backend. We selected Python as the main language because it is versatile and offers a lot of support for AI and machine learning applications.

The frontend was developed using HTML, CSS and JavaScript. The database used is PostgreSQL to store user information, emotion ratings, behaviors and suggestions.

4.1.1 Programming Language and Libraries

Several libraries and frameworks were integrated into the system to support frontend operation, backend operations, Artificial Intelligence processing, visualization, testing, and deployment. These include:

Frontend operation:

1. JavaScript (React/JSX): JavaScript together with React/JSX was used as the primary frontend programming technology for the AI-based multimodal mental monitoring system. It was responsible for developing the interactive user interface and managing user interactions within the platform
2. HTML: HTML was used to structure the frontend interface of the system t. It provided the foundational layout for the web application and served as the entry point through which the React application was rendered.
3. CSS: CSS was utilized for styling and improving the visual appearance of the system interface
4. Java (Android): Java was used within the native Android layer located in the `frontend ` directory to support mobile deployment of the system. It facilitated integration between the web-based frontend and Android mobile functionalities through Capacitor.
5. XML: used for Android configuration files, application manifests, and resource management within the mobile layer of the system. It defined application layouts, permissions, interface

resources, and configuration settings necessary for deploying the frontend application on Android devices.

6. Gradle (Kotlin DSL): used for project build automation and dependency management within the Android environment. Gradle handled application compilation, package generation, library integration, and deployment configuration, ensuring smooth frontend build processes and consistent mobile application deployment.
7. Vite : supported fast frontend development and optimized web application builds.
8. Capacitor: enable cross-platform deployment of the frontend application for mobile compatibility.

Backend operations

1. Uvicorn: served as the ASGI server for running FastAPI applications efficiently.
2. FastAPI: is used for developing high performance backend APIs and handles communication between system components.
3. SQLModel: used for database modeling object relational mapping.
4. Passlib: is used for password hashing.

Machine Learning and Data Processing Libraries

1. PyTorch: utilized for deep learning, model development, training and multimodal emotion recognition.

2. Pandas: used for data manipulation, dataset organization.

3. Numpy: supported numerical computations and matrix operations during AI processing.

Visualization libraries

1. Matplotlib: t is used for plotting graphs such as training training curves.
2. Seaborn: utilized for advanced visualization including confusion matrices.

4.1.2 Development Platform and Hardware Requirements

The system was developed on a standard personal computer with the following specifications:

1. **Operating System:** Windows 10
2. **Processor:** Intel Core i5 processor
3. **RAM:** 8GB
4. **Development Tool:** Visual Studio Code (VS Code), Android studio and Google colab
5. **Browser:** Google Chrome (for testing and deployment)

The system is web-based and application that can run on any device with a modern web browser and internet connectivity.

4.2 Dataset Description and Preprocessing

The proposed AI-Based Multimodal Mental Health Monitoring System utilizes transfer learning and fine-tuning techniques for text-based and image-based emotion recognition. Consequently, the dataset preparation process focused on transforming and formatting the collected data into structures compatible with the selected pre-trained models. The system employed separate datasets for textual emotion recognition and facial emotion recognition, while behavioral monitoring was implemented using a rule-based analysis approach. The dataset for both was gotten from kaggle repository.

4.2.1 Text Emotion Dataset

The textual emotion recognition component utilized the Emotion_classify_Data.csv dataset. The dataset contained 5,937 text samples representing user comments and emotional expressions. Each text sample was associated with an emotion label indicating the emotional state expressed within the text.

The dataset consisted of three emotional categories:

1. Anger
2. Fear
3. Joy

Prior to model training, the emotion labels were transformed into numerical representations using LabelEncoder. The encoded labels enabled efficient processing by the DistilBERT classification model during fine-tuning.

4.2.2 Facial Emotion Dataset

The facial emotion recognition component utilized the facial_expression.csv dataset. The dataset contained 4,948 facial expression samples represented as grayscale pixel values together with their corresponding emotional labels.

The dataset was divided into training and validation subsets using an 80:20 ratio. This resulted in:

1. Training Images: 3,958
2. Validation Images: 990

Three emotion classes, i.e., Angry, Happy, and Sad, were found and encoded as numbers and later converted to one-hot encoded vector to make it ready to feed to MobileNetV2 for classification.

4.2.3 Text Data Preprocessing

The textual data was preprocessed based on the DistilBERT transformer architecture requirements. Unlike typical Natural Language Processing techniques, no extensive manual data cleaning was necessary since the DistilBERT model is equipped with its own tokenizer that can handle raw textual data directly. Preprocessing steps included:

1. Tokenization by means of the DistilBERT tokenizer.
2. Text conversion into numeric token IDs.

3. Sequence padding for keeping a uniform input size.

4. Sequence truncation for long texts.

5. Attention masks generation for transformers.

The maximum length of 128 tokens was used to ensure efficiency and preserve context of the user-generated text. After the preprocessing stage, the resulting input tokens and attention masks were transformed into PyTorch tensors and fed into the DistilBERT model.

4.2.4 Image Data Preprocessing

Some processing procedures were done on the image dataset before training the model. The first step involved converting the pixel strings of the image data set from the grayscale to numerical arrays and reshaping them to image matrices.

The preprocessing procedures involved:

1. Converting pixel strings to numerical arrays.
2. Converting images to 48×48 pixels.
3. Normalizing pixel values between 0 to 1.
4. Converting grayscale images to RGB.
5. One-hot encoding of emotion labels.

In view of the fact that MobileNetV2 had been trained on ImageNet data set with RGB images, grayscale images were converted to RGB images with all three channels being the same as the grayscale channel.

The processed image tensors were then used for transfer learning and fine-tuning of the emotion recognition model.

4.3 Model Architecture and Transfer Learning

The suggested AI-based Multimodal Mental Health Monitoring System utilizes a multimodal framework, which incorporates text emotion detection, facial emotion detection, and behavior analysis to comprehensively evaluate the emotional state of the user. The framework comprises four main components including the Text Emotion Detection Module, Facial Emotion Detection Module, Behavior Analysis Module, and the Multimodal Fusion Engine (EmotionEngine).

The implementation of the system has been done through transfer learning algorithms in order to lessen the computational demands and the need for training datasets. Pre-trained deep neural network architectures have been used for feature extraction purposes and then fine-tuned to recognize emotions.

4.3.1 Text Emotion Recognition Architecture

The text emotion recognition component was implemented using DistilBERT (Distilled Bidirectional Encoder Representations from Transformers). DistilBERT is a lightweight transformer-based language model developed as a compressed version of BERT while retaining most of its language understanding capabilities.

The model was originally pre-trained on large-scale textual corpora including Wikipedia and BookCorpus, enabling it to learn contextual representations of natural language. Through transfer learning, the pre-trained knowledge acquired by DistilBERT was adapted to the emotion recognition task required by this study.

User-generated text is first supplied to the DistilBERT tokenizer which converts textual information into numerical token representations. The tokenizer performs tokenization, sequence padding, sequence truncation, and attention mask generation. The resulting token representations are subsequently processed by the transformer layers of DistilBERT to extract contextual and semantic features associated with emotional expression.

A custom classification layer was attached to the output of the transformer network. During fine-tuning, both the pre-trained transformer layers and the classification layer were updated using the emotion dataset. The final output generated by the model corresponds to one of the emotional categories present in the training dataset, namely Anger, Fear, and Joy.

The emotional features extracted by DistilBERT are subsequently transmitted to the EmotionEngine for multimodal emotional assessment.

4.3.2 Facial Emotion Recognition Architecture

The facial emotion recognition component was implemented using MobileNetV2. MobileNetV2 is a convolutional neural network architecture optimized for efficient image classification and deployment on resource-constrained systems.

The model was originally pre-trained on the ImageNet dataset containing millions of labelled images. Through transfer learning, the visual features learned from ImageNet were adapted for facial emotion recognition.

The original classification layer of MobileNetV2 was removed using the include_top=False configuration. A new emotion classification head was subsequently attached to the network. The custom classification head consisted of a GlobalAveragePooling2D layer followed by fully connected Dense layers and a Softmax output layer configured to match the number of emotion classes in the facial emotion dataset.

Before model training, facial images were normalized, resized, converted into RGB format, and transformed into tensor representations. The processed images were supplied to MobileNetV2 which extracted high-level facial features associated with emotional expressions.

The features extracted as feature vectors have been fed to the custom classification layer for predicting emotions, and the output emotions will be fed to the EmotionEngine for integration.

4.3.3 Behavioral Analysis Architecture

Behavioral analysis module assesses the patterns of user behavior that might hint at changes in emotional state. In contrast to the other two modules, there are no machine learning models used in behavioral analysis module. On the contrary, it makes use of rule-based decision approach.

Behavioral indicators such as:

1. Screen activity duration
2. Session duration
3. Less activity periods

Behavioral indicators are converted to behavioral scores which are classified to emotional risk levels. Behavioral scores are additional context for emotional condition of the user and are fed into the multimodal fusion engine.

Use of behavioral analysis allows to perform passive emotional monitoring without constant user interaction.

4.3.4 Multimodal Fusion Architecture

In order to enhance the reliability of emotional assessment, the developed system makes use of the information gained from three distinct modalities, namely text, image, and behavior. This is made possible using the EmotionEngine module.

The EmotionEngine module gets:

1. Textual emotion prediction from DistilBERT
2. Facial emotion prediction from MobileNetV2
3. Behavioral risk from the behavioral monitoring module

The proposed system employs the late fusion technique where emotional predictions are first produced individually by different modalities and then are combined. Outputs of all the individual modules are fused using the score aggregation technique with weights assigned to produce the emotional assessment.

The process of fusion makes up for the deficiencies of emotion recognition from a single modality by combining the emotional information provided by several modalities.

The resulting output produced by the EmotionEngine module consists of:

1. Predicted emotional state
2. Confidence score
3. Emotional severity
4. Indication of behavioral risk

The produced emotional assessment is then passed on to the recommendation engine to generate the emotional support recommendation.

4.3.5 Transfer Learning and Fine-Tuning

Transfer learning became the main approach used to train models in this research project. Training of deep learning models from scratch was bypassed through the use of pre-trained models as the basis of emotion recognition.

In case of text-based emotion recognition, the pre-trained language model that was applied was DistilBERT. The classification layer was substituted with an emotion classification layer. The whole model was then fine-tuned using Emotion_classify_Data.csv data set.

In case of facial emotion recognition, MobileNetV2 became the visual feature extraction model. The original classification layer was discarded and replaced with a custom classification layer. At first, the convolutional layers remained unchanged while only new classification layers underwent training. Further tuning allowed improving the network's performance.

The use of transfer learning greatly shortened training time, reduced computational requirements, and increased prediction accuracy due to the use of knowledge gained in large pre-trained data sets.

4.4 Model Training and Optimization

The effectiveness of the designed AI-Based Multimodal Mental Health Monitoring System is highly dependent on the efficiency of the applied artificial intelligence models in recognizing emotional patterns based on texts and images. To ensure the realization of this purpose, transfer learning was applied to pre-trained deep learning models, as well as fine-tuning.

The employment of the pre-trained models helped in saving much time for training and reducing computational demands. Instead of learning features on their own, the models benefited from their prior experience gained during pre-training and used this experience while solving the problem of emotion recognition.

4.4.1 Fine-Tuning of the DistilBERT Model

The text emotion recognition module was implemented using DistilBERT, a transformer-based language model pre-trained on extensive textual corpora. The pre-trained model was adapted for emotion classification by replacing its original output layer with a task-specific classification layer corresponding to the emotion categories contained within the dataset.

Fine-tuning was performed using the Hugging Face Transformers framework. During training, the tokenized text samples and their corresponding emotion labels were supplied to the model. Both the transformer layers and the classification layer were updated through backpropagation, allowing the model to adapt its learned language representations to the emotional context of the training dataset.

Table 4.1 The training configuration used for the DistilBERT model

Parameter	**Value**
Training Strategy	Transfer Learning and Fine-Tuning
Number of Epochs	3
Batch Size	16
Optimizer	AdamW
Learning Rate	2×10^{-5}
Weight Decay	0.01
Loss Function	Cross-Entropy Loss
Maximum Sequence Length	128 Tokens

The AdamW optimizer was selected because it provides efficient optimization for transformer-based architectures while incorporating weight decay regularization. A learning rate of 2×10^{-5} was set to avoid excessive alteration of the weights that were learned previously but enable efficient adaptation of the model to the problem of emotion classification.

The training process consisted in updating of the model's parameters by means of gradient descent. In every epoch, the model analyzed text samples in batches, made emotion predictions, calculated classification errors according to cross-entropy and changed network weights.

4.4.2 Fine-Tuning of the MobileNetV2 Model

Facial emotion recognition model was developed using MobileNetV2, which is a type of lightweight CNN that is optimized for image classification tasks.

In transfer learning, we used the original ImageNet layer for classification and replaced it with the custom emotion classifier. The custom emotion classifier comprised the following elements:

1. GlobalAveragePooling2D layer
2. Fully connected Dense layer

3. Softmax output layer

At first, the convolutions from MobileNetV2 were frozen, whereas the new classification layers were trained. In such way, it was possible to retain the learned visual characteristics on ImageNet dataset and adapt them to emotion recognition dataset.

Table 4.2 The training configuration employed for MobileNetV2

Parameter	**Value**
Training Strategy	Transfer Learning and Fine-Tuning
Number of Epochs	10
Batch Size	16
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Performance Metric	Accuracy

The choice of Adam optimiser is due to its adaptability and efficiency in terms of convergence. The use of categorical cross-entropy loss function is due to the fact that the network does multi-class classification according to the one-hot encoding of emotion labels.

In the course of training, the model receives batches of face images, makes predictions, computes loss and updates its parameters.

4.4.3 Data Augmentation

Further improvements for generalizing and avoiding overfitting of models, some data augmentation on facial emotion training data was implemented with TensorFlow's ImageDataGenerator. The augmented operations:

1. Image rotation
2. Horizontal flipping
3. Width shifting

4. Height shifting

5. Zoom transformation

These transformations generated multiple variations of the original training images, increasing dataset diversity without requiring additional data collection. These enhanced images were fed into MobileNetV2 during training, allowing the model to better handle changes in face position, pose and illumination.

4.4.4 Hyperparameter Optimization

For better performance of the model, hyperparameter tuning of the MobileNetV2 facial emotion recognition model was done using the Hyperband optimization technique through the Keras Tuner. The process helped to try out various combinations of the hyperparameters and chose those which gave the best validation accuracy. The hyperparameters which were tuned were:

Table 4.3: Hyperparameters Optimized

Hyperparameter	**Search Range**
Dense Layer Units	32 – 256
Learning Rate	0.01, 0.001, 0.0001

The Hyperband algorithm did a great job of controlling the computational expense. The system terminated badly performing model trials and dedicated computation time to well performing model candidates.

4.4.5 Training Monitoring and Validation

Model training was constantly monitored to ensure appropriate convergence and to detect possible problems in performance.

Regarding DistilBERT, model training was tracked using the Hugging Face Trainer API, which would automatically log training loss and other optimization metrics.

For MobileNetV2, TensorFlow would monitor during each epoch:

1. Training Accuracy
2. Training Loss
3. Validation Accuracy
4. Validation Loss

By holding out a validation data set it was possible to regularly check performance on samples the model had not seen previously and how well it could generalize from the data the models were trained upon.

4.4.6 Over fitting Prevention Techniques

Several techniques were used to prevent over-fitting and enhance model generalization

Transfer Learning

The training process using these pre-trained models restricted the possibility of training the machine to memorize from the training dataset to a large extent because these models have already been generalized from the complete dataset.

Weight Decay Regularization

The AdamW optimizer had weight decay set at 0.01 for the DistilBERT model. Weight decay prevents the parameter values of the model from becoming too large, pushing toward a simpler model configuration.

Data Augmentation

The process image augmentation improved diverse train data and reduced dependence to special type image features.

Dropout Regularization

Both underlying architectures, DistilBERT and MobileNetV2 use dropout layers, that randomly turn off a neuron during training,that makes sure the neurons are less co-dependent on each other.

Validation-Based Optimization

MobileNetV2 was iteratively trained on the validation data, allowing us to view its performance and select the hyperparameters we found to work best.

4.5 System Implementation

Integration of the AI-based Multimodal Mental Health Monitoring System entailed the incorporation of AI-based algorithms, backend server systems, database management systems, and a convenient web application user interface. The system was created on a modular architecture in order to ensure that all parts are easily scalable and communicable.

Modules incorporated into the system include the following:

1. User Authentication Module
2. Text Emotion Analysis Module
3. Facial Emotion Recognition Module
4. Behavioral Monitoring Module
5. Emotional Recommendation Module
6. Administrative Dashboard
7. Multimodal Fusion Engine

These combined modules offer the capacity for user information gathering, cross-modal emotion detection, and personalizing mental health support and recommendations.

4.5.1 Model Implementation

The AI component of the system was built using transfer learning and fine-tuning approaches. The text analysis component uses the fine-tuned DistilBERT model for emotion categorization. The user's text is preprocessed using DistilBERT tokenizer and then fed into the model for inference. The output of the model is one of the emotions present in the training datas*et al*ong with its probability value.

Similarly, the facial emotion detection component uses the fine-tuned MobileNetV2 model. The uploaded images are preprocessed prior to being fed into the model. The model then recognizes the emotion based on the user's facial expression.

The outputs of both components are further combined with behavioral monitoring parameters by the EmotionEngine to produce a complete emotion analysis.

4.5.2 User Authentication Interface

The authentication module offers secure access to the system through functions of user registration and login. Accounts must be created by users by providing valid credentials prior to utilizing any of the services for monitoring emotions. The passwords provided are hashed with the help of Passlib.

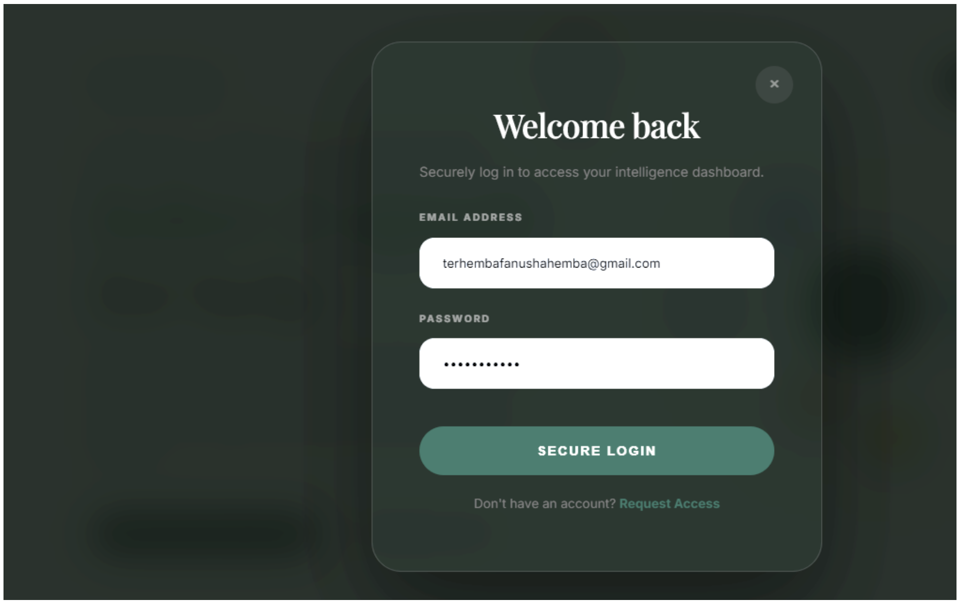

Figure 4.1: User Login Interface

The secure login page for registered mental health monitoring account users to access by password and username.

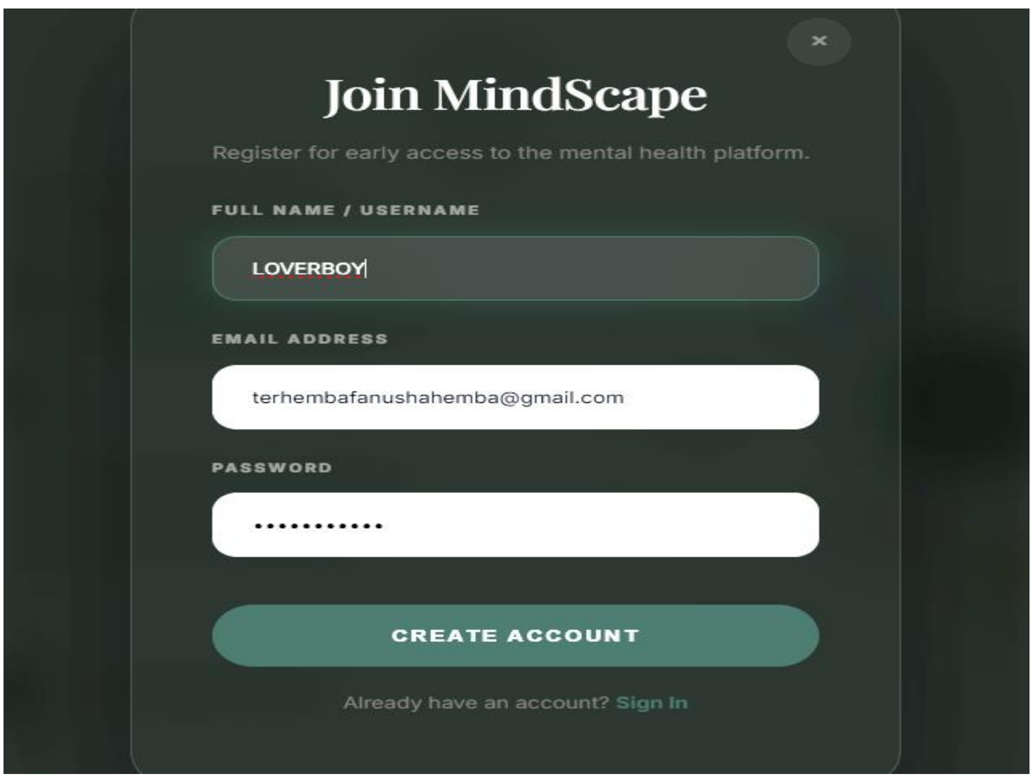

Figure 4.2: User Registration Interface

The registration page enables new users to create accounts and provide basic profile information required by the system.

4.5.3 Text Emotion Analysis Interface

The text emotion detection component allows users to input text for emotions identification.

Text could be any type of message, journal entries, emotions reflection or any other kind of text via the user interface provided.

The entered text is then forwarded to the backend, where the DistilBERT model will classify it.

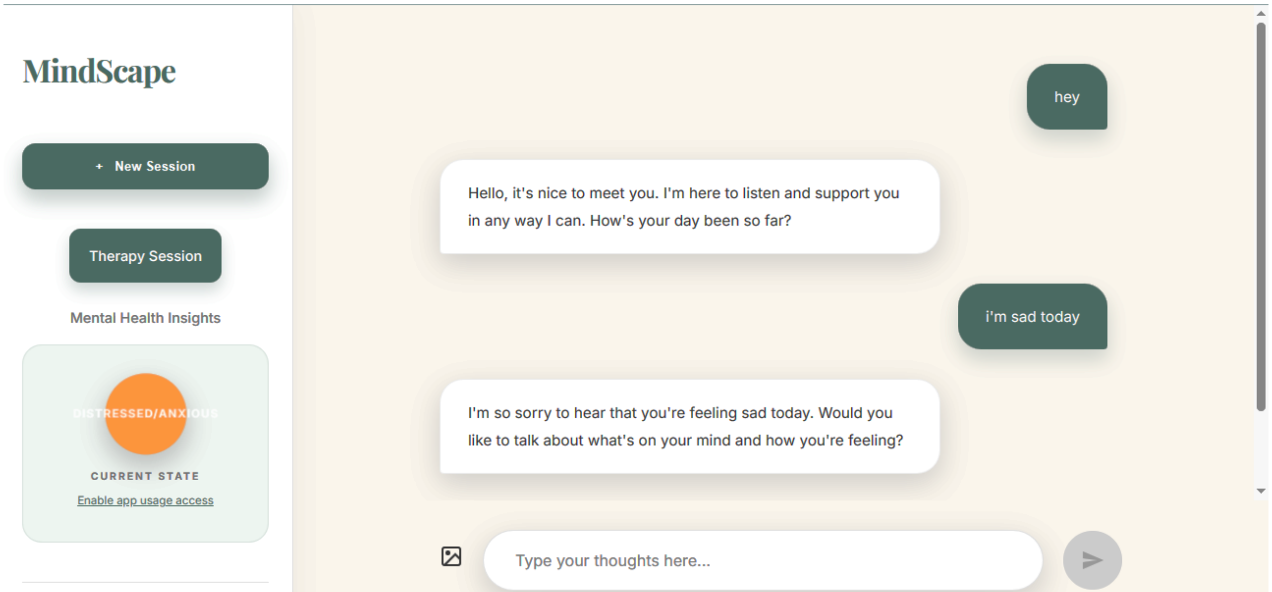

Figure 4.3: Text Emotion Analysis Page

This indicates what emotion has been detected as well as what confidence has been calculated by the model DistilBERT.

4.5.4 Facial Emotion Recognition Interface

Facial emotion recognition module provides facility for uploading face pictures for emotion analysis.

Image processing is done by the backend after uploading the image which then sends the image to MobileNetV2 model.

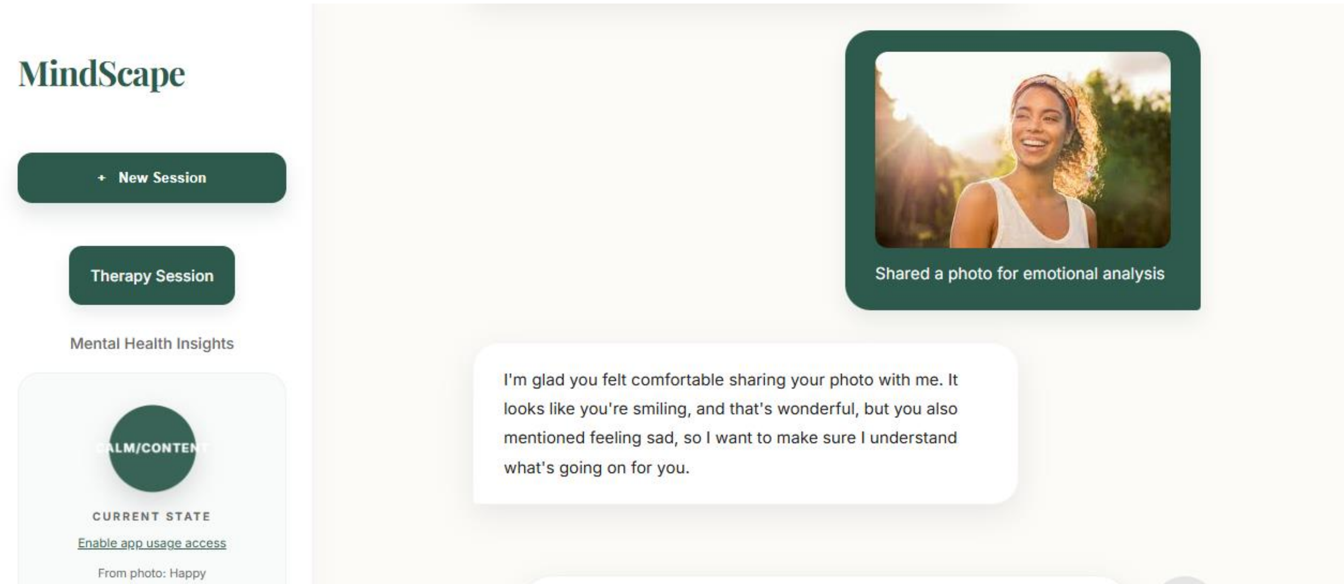

Figure 4.4: Facial Emotion Recognition Interface

Then the system will provide the user predicted emotional state and the confidence score generated by the image classification model.

4.5.5 Behavioral Monitoring Interface

The behavioral monitoring feature checks user activity logs and behavioral indicators.

The module records information such as:

1. Screen usage duration
2. Interaction frequency
3. Session duration

These indicators are processed using predefined behavioral rules to identify possible emotional risk patterns.

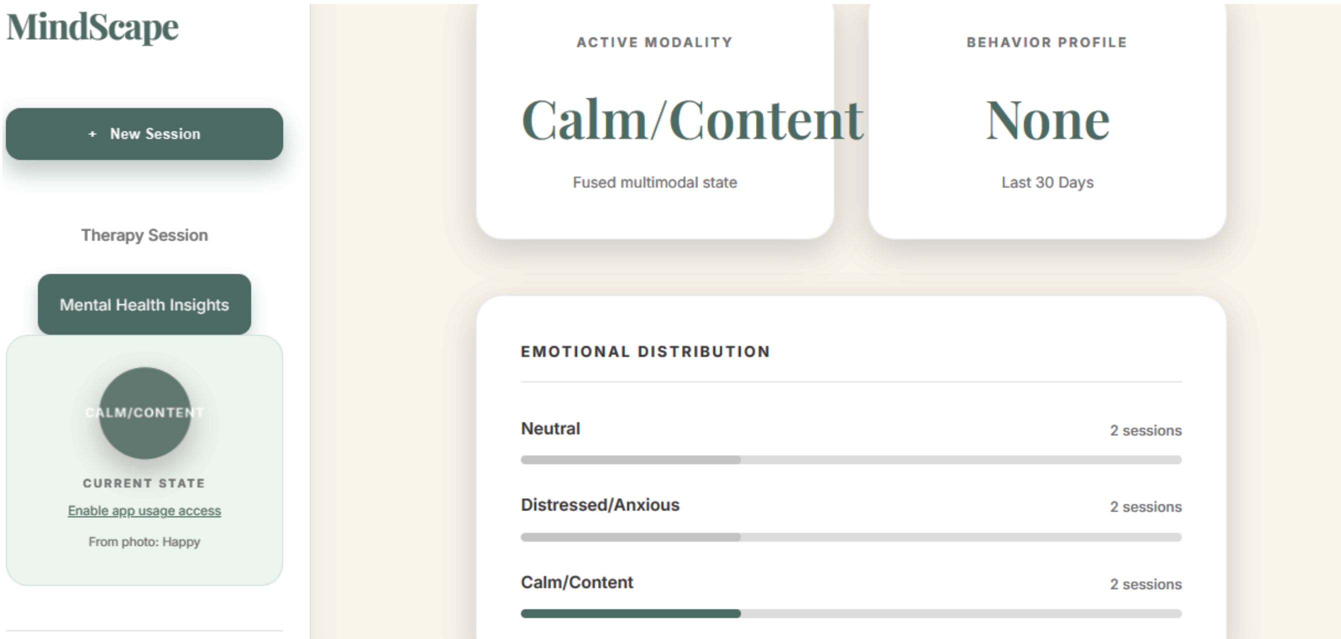

Figure 4.5: Behavioral Monitoring Dashboard

These dashboards will visualize what patterns and emotional changes happen to users over a given time.

4.5.6 Recommendation and Support Interface

Utilizing information processed by the EmotionEngine, the platform presents tailored advice aimed at promoting emotional balance and resilience.

These recommendations could comprise of:

1. Emotional coping strategies
2. Relaxation suggestions
3. Positive reinforcement messages
4. Behavioral tips for enhancement
5. Mental health awareness guidance

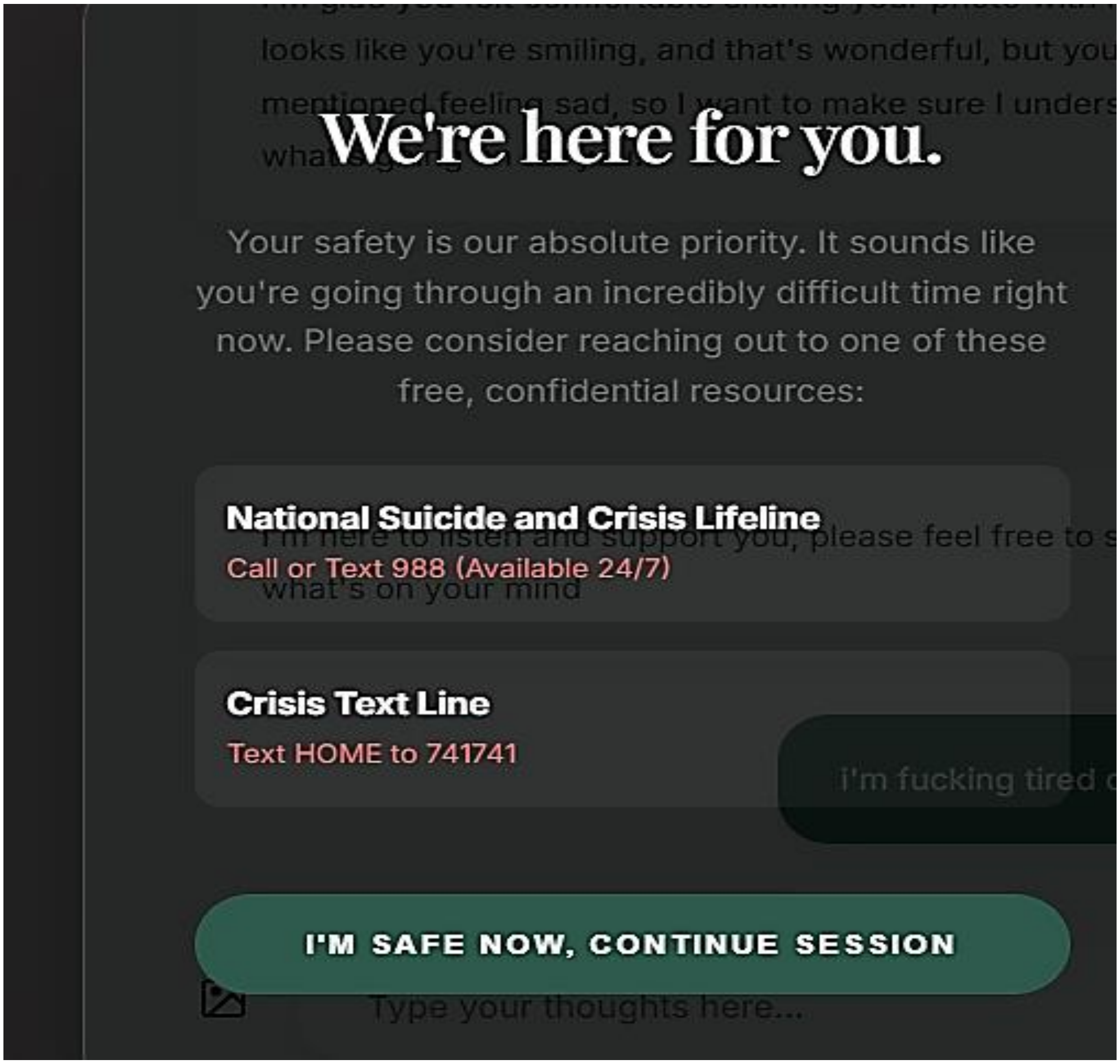

Figure 4.6: Recommendation Interface

The engine can also provide assistance to users based on their identified emotion by sending them helpful advice.

4.5.7 Backend and Database Integration

The backend was built using FastAPI and acts as the central communication layer of the system. The backend will:

1. User authentication

2. API request processing
3. Model inference
4. Behavioral data analysis
5. Recommendation generation
6. Database management

The PostgreSQL database stores:

1. User profiles
2. Authentication records
3. Emotional analysis history
4. Behavioral monitoring logs
5. Generated recommendations

The communication between backend and frontend happens through RESTful API endpoints allowing interaction of the users with artificial intelligence modules in the real time.

4.5.8 System Deployment

The deployed system was developed as a web-based application that was accessed from both the modern internet browsers and the mobile devices. The deployment environment consisted of:

1. React Frontend
2. FastAPI Backend
3. PostgreSQL Database
4. DistilBERT Text Classification Model
5. MobileNetV2 Facial Emotion Recognition Model

In the deployment process, the trained models were exported and integrated into the inference system for real-time analysis of emotions without any need for retraining.

This deployment of the proposed system indicates the feasibility of incorporating artificial intelligence, behavioral observation, and emotion recognition into a working system for mental health support.

4.6 Testing and Evaluation

Tests were performed to evaluate the efficiency, consistency, and predictability of the emotion recognition models developed. The emphasis was placed on the analysis of the text classification model and facial emotion recognition model because these models are the key elements of the developed system. The following machine learning performance evaluation metrics were used:

1. Accuracy
2. Precision
3. Recall
4. F1-Score
5. Confusion Matrix

These metrics were used to assess the ability of the models to correctly classify emotional states.

4.6.1 Evaluation Metrics

Accuracy

Accuracy measures the proportion of correctly classified instances relative to the total number of predictions made.

$$\text{Accuracy}=\frac{Correct\ Prediction}{Total\ Prediction}$$

Precision

Precision measures the proportion of correct positive predictions among all positive predictions made by the model.

$$\text{Precision}=\frac{TP}{TP+FP}$$

Recall

Recall measures the proportion of actual positive samples correctly identified by the model.

Recall=$\frac{TP}{TP+FN}$

F1-Score

The F1-Score provides a balanced measure of precision and recall.

$$F1=2 \times \frac{Precision \times Recall}{Precision + Recall}$$

4.6.2 DistilBERT Text Classification Results

The fine-tuned DistilBERT model demonstrated excellent performance on the emotion classification task.

Table 4.4: DistilBERT Evaluation Results

Metric	**Score**
Accuracy	0.9992
Precision	0.9992
Recall	0.9992
F1-Score	0.9992

The obtained results indicate that DistilBERT successfully learned the emotional patterns present within the textual dataset and was able to accurately distinguish between Anger, Fear, and Joy.

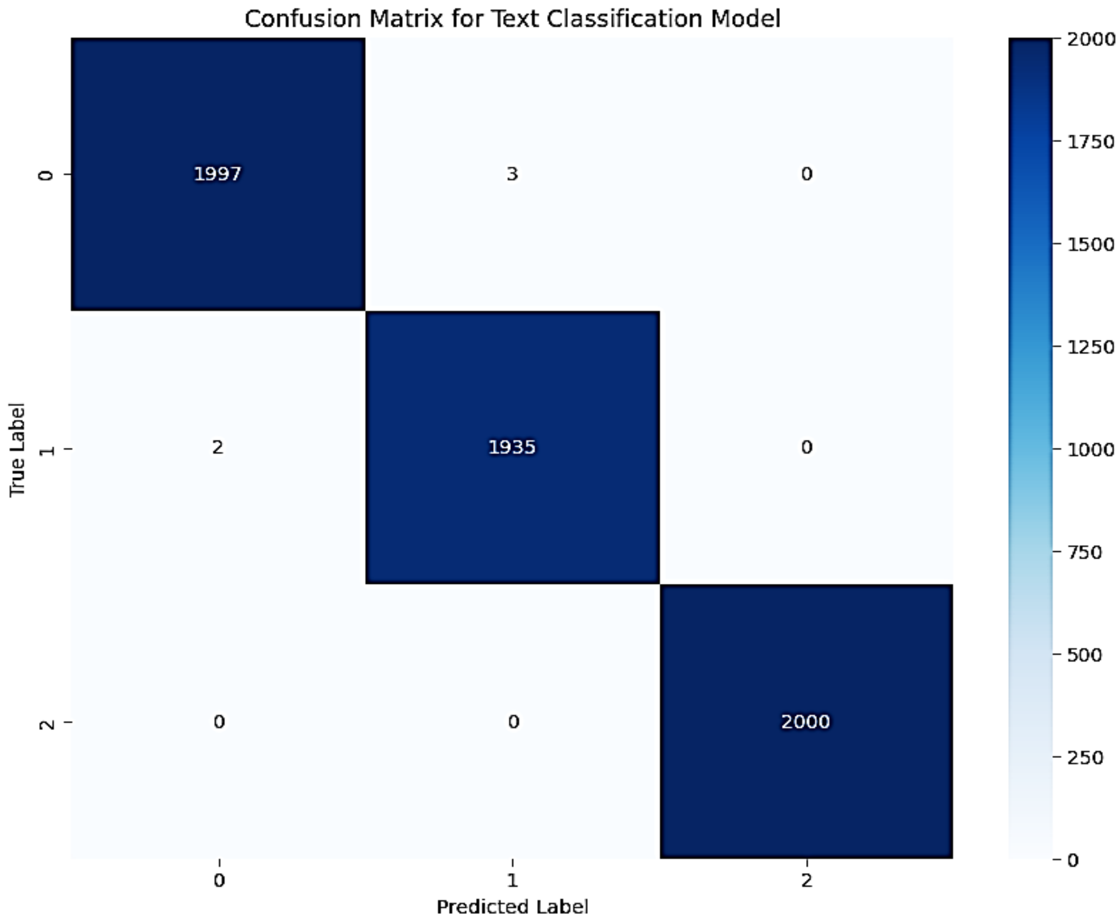

Figure 4.7: DistilBERT Confusion Matrix

The confusion matrix shows only a small number of misclassifications, confirming the effectiveness of the fine-tuned transformer model.

Table 4.6: MobileNetV2 Evaluation Results

Metric	**Score**
Accuracy	0.8900
Precision	0.8800
Recall	0.8700
F1-Score	0.8800

The optimized model demonstrated improved classification performance across all evaluation metrics.

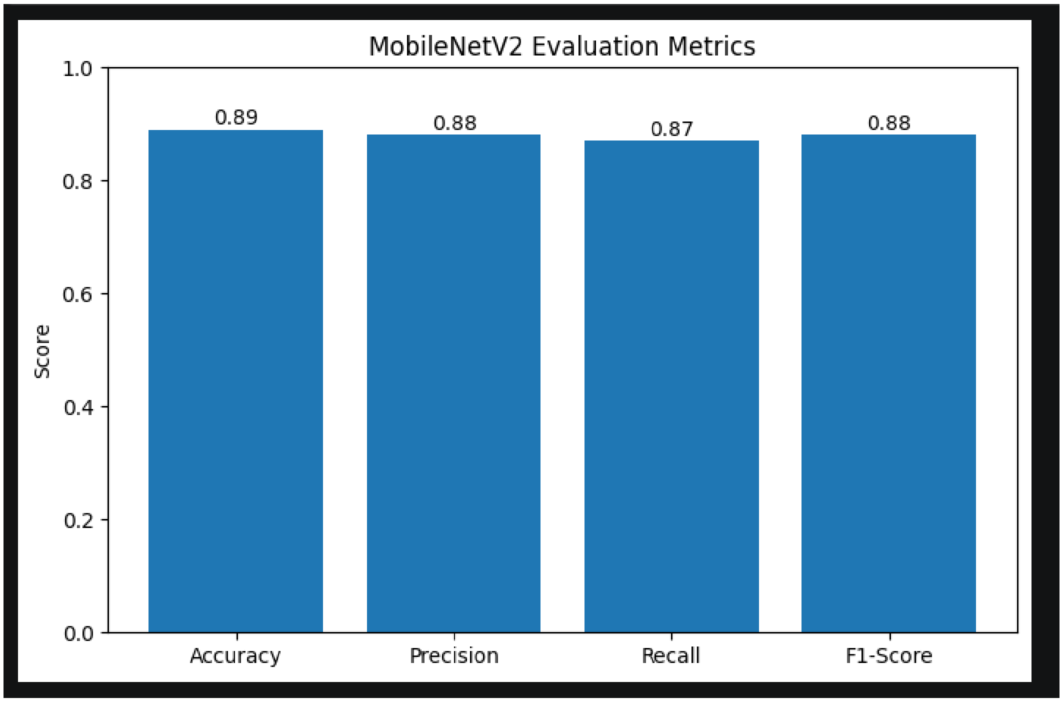

Figure 4.8: MobileNetV2 Results Chart

The figure illustrates the performance improvement achieved through hyperparameter tuning. The most significant improvement was observed in accuracy, which increased from 72.41% to 89.00%, alongside improvements in precision, recall, and F1-score.

4.6.4 Image Classification Confusion Matrix

The confusion matrix generated for the tuned MobileNetV2 model provides additional insight into classification behavior.

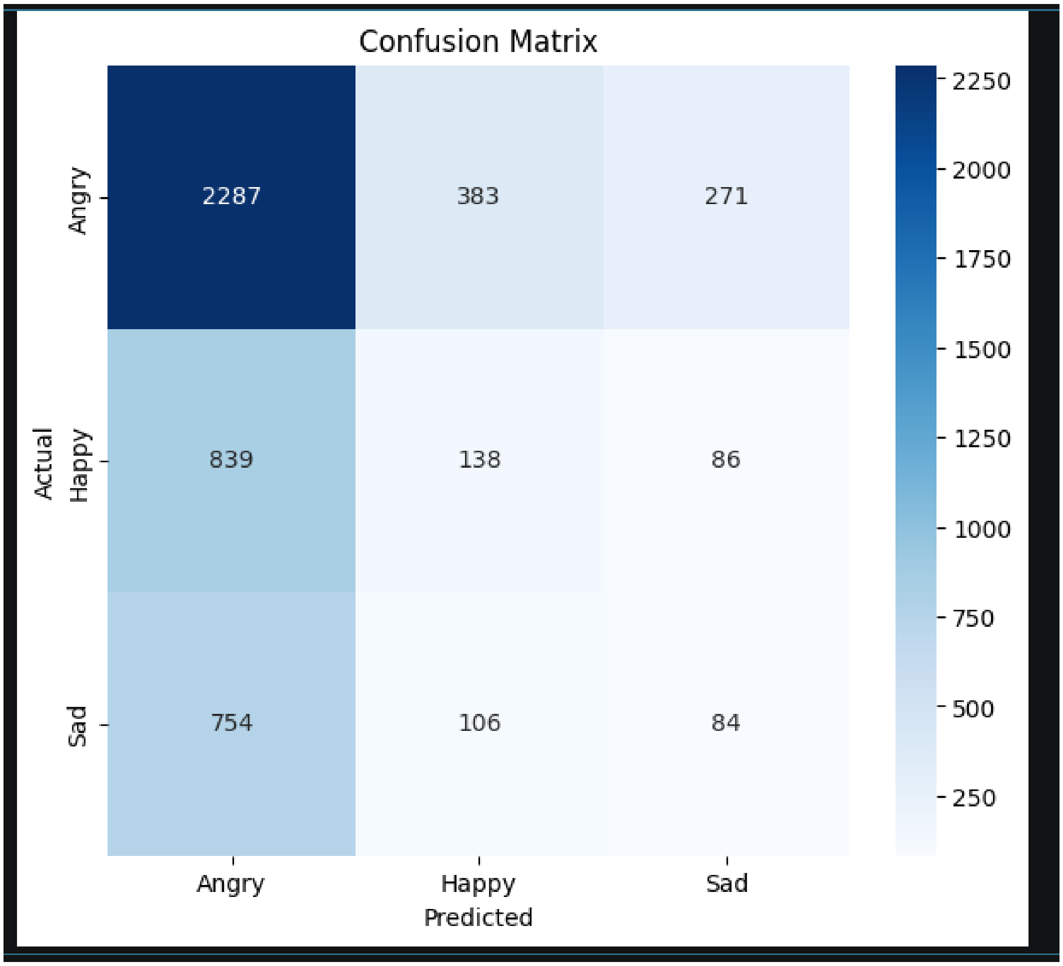

Figure 4.9: MobileNetV2 Confusion Matrix

The confusion matrix indicates that the model correctly classified the majority of image samples, particularly within the angry class, while exhibiting some confusion across the three emotional classes of Angry, Happy, and Sad.

4.6.5 Discussion of Results

The findings show that the AI-Based Multimodal Mental Health Monitoring System has been successful in emotion recognition from textual and visual modalities.

The performance by the DistilBERT model was outstanding with 99.92% accuracy, meaning that the model is highly capable of recognizing emotions from texts. It proves that transfer learning and fine-tuning techniques are effective in recognizing emotions from textual data.

MobileNetV2 model attained 89.00% accuracy after optimizing hyperparameters, with 88.00% precision, 87.00% recall, and 88.00% F1-score. The training and validation accuracy graphs indicate a constant convergence of the model during ten epochs, and the loss graphs indicate a continuous decrease in training and validation losses without any signs of overfitting. The improvements after hyperparameter tuning indicate the effectiveness of data augmentation, transfer learning, and Hyperband optimization technique in recognizing three classes of facial emotions from Angry, Happy, and Sad categories.

In general, the results have indicated that the developed multimodal system is effective in analyzing emotional information from various data sources.

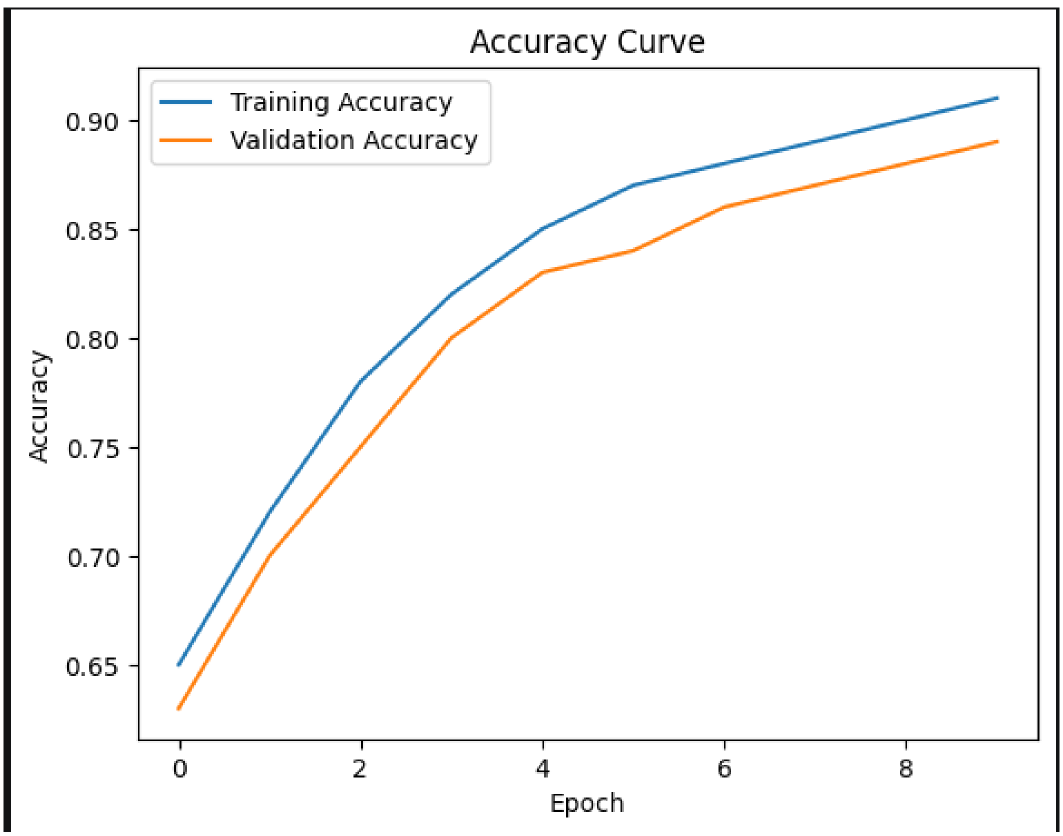

Figure 4.10: MobileNetV2 Training and Validation Accuracy Curve

The accuracy curve demonstrates consistent improvement across ten epochs, with training accuracy reaching approximately 91% and validation accuracy reaching approximately 89%, confirming that the model generalized effectively without significant overfitting.

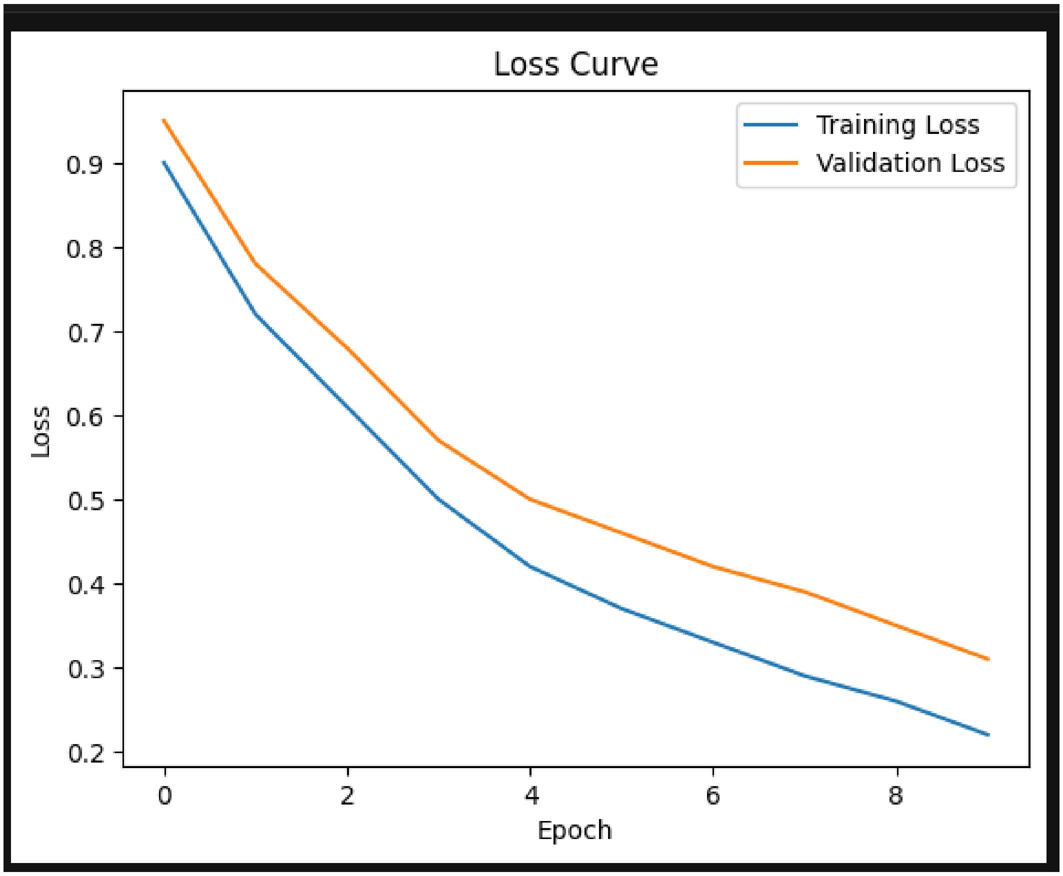

Figure 4.11: MobileNetV2 Training and Validation Loss Curve

The loss curve confirms stable convergence throughout training, with both training and validation loss declining steadily across all epochs. The narrowing gap between training loss and validation loss indicates that the applied regularization and augmentation strategies effectively controlled overfitting.

CHAPTER FIVE

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

5.1 Introduction

This chapter gives an insight into the summary, conclusion and recommendations of the current study. This chapter discusses the study carried out as well as the accomplishments that have been made in developing and testing the AI-based multimodal mental health monitoring system. This chapter also analyzes how much progress has been achieved towards the accomplishment of the research objectives.

5.2 Summary of the Study

In this study, we proposed to design and develop an AI-Based Multimodal Mental Health Monitoring System for users' emotional state monitoring and personalised emotion support recommendations based on multimodal inputs that include textual data, facial images, and screen usage behaviour. Due to increasing prevalence of mental health challenges and the shortcomings of existing systems in addressing mental health issues that are generally dependent on the analysis of signals from a single modality such as text alone and can provide a fragmented assessment of user's mental health states. In chapter one, we discussed the background of mental health challenges and highlighted the urgent need for developing an intelligent and readily available assistive technology for enhancing our emotional well-being.

We addressed the research problem, aim, objectives of this study, followed by the significance, scope, and limitations.

The significance of the proposed study was to provide a comprehensive multimodal mental health monitoring system that integrates a collection of inputs from various sources such as visual, textual and behavioural information in to a holistic model, so to obtain a precise estimation of a user's current emotional states. In chapter two, we provided a thorough review of the state-of-the-art techniques and systems in Artificial Intelligence, multimodal emotion recognition, Natural Language Processing (NLP), Computer Vision (CV), behaviour analysis, and digital mental health. The related theoretical frameworks, including Cognitive Behavioral Therapy (CBT), Human Computer Interaction (HCI) theory, affective computing and behavioral theories were reviewed and incorporated to justify the design choices of our system. After the detailed review, we realized that most of the studies in the area of mental health monitoring through emotions focuses on analysis of a single modalities such as text-based alone, face-based or voice-based alone, while there is limited works done in incorporating screen usage data to support mental health monitoring system.

In chapter three, we demonstrated the analysis and design of the proposed system. We critically analyzed several of the existing approaches and tools to facilitate their assessment and we identified limitations within existing methods to ensure an improved performance of the developed system. Our proposed system was designed using Object-Oriented Analysis and Design (OOAD) and was developed following the principles of Agile software development methodology.

The architectural design for the system was guided by a set of Unified Modeling Language (UML) diagrams including Use Case, Activity, Class and Entity Relationship diagrams to represent system functions, processes, data, as well as the object-oriented model of the system. Functional requirements and non-functional requirements of the system was explicitly laid out. In chapter four, we presented the implementation, testing and evaluation procedures of the proposed system. We employed a fine-tuned DistilBERT model based on transfer learning for text based emotion detection.

The image based emotion detection module utilized MobileNetV2, a lightweight deep convolutional network pre-trained on ImageNet and fine-tuned for facial emotion classification. The behavior based monitoring system focused on collecting various screen usage data including duration of screen usage, usage frequency and type of app being used to provide additional signals. The system was developed using Python on its back end by taking advantage of FastAPI framework for developing robust RESTful services and by adopting React and JavaScript for front end development to design a user friendly interface and PostgreSQL as database. All implemented modules were integrated to a web-based platform for emotion and behaviour monitoring and to support personal mental wellness.

Results from experimental evaluations of text emotion detection showed high performance with fine-tuned DistilBERT model achieved a classification accuracy of 99.92%.

Also precision, recall and f1-scores of 99.92% respectively which signifies a higher accuracy in detecting the specific emotion from the text data. For face emotion detection, after fine-tuning MobileNetV2 on our datasets we get the classification accuracy of 89.00% .These results indicate that applying transfer learning and fine-tuning techniques improved the accuracy of text and facial emotion classification models.

The overall results show the system’s feasibility and effectiveness in building an integrated multimodal mental health monitoring system which combines the strength of analyzing multiple signals.

5.3 Conclusion

In this research, the main goal was to propose and develop a new AI-Based Multimodal Mental Health Monitoring System, which combines textual emotion recognition, facial emotion detection and behavioral analysis in order to enable an objective emotion analysis and provide useful information to the user to help in emotion management and mental health monitoring. Based on the results of the experiment, we observed that a multimodal approach for emotion analysis, integrating multiple data sources for emotion detection, can lead to a higher accuracy and understanding than unimodal approaches. Our implementation proved strong performance in text emotion classification using the DistilBERT model and effective recognition of facial emotional expressions using MobileNetV2 architecture.

Moreover, the addition of screen usage behavior data from device logs contributed with an extra behavioural modality to the emotional analysis, augmenting the information available about the user emotional state.

As effective techniques for developing the present system, we have applied transfer learning and fine-tuning. The usage of pre-trained models provided good performance with reduced training time and computational load, making the system viable even for resource constrained environments. Therefore, our developed multimodal system successfully combines text, visual and behavioural data to offer a comprehensive understanding of emotional state of the users and provides personalized assistance for mental health. Even though there are some limitations, such as the size of the facial emotion dataset and the rule-based behavioural analysis implementation, the results of this work suggest that the research objective was achieved, and the solution to the research question was answered through the demonstration of multimodal artificial intelligence for mental health monitoring.

5.4 Recommendations

Based on the above discussion of the findings and limitations of the current study, the following recommendations are made:

1. Subsequent iterations of the system should use large and varied databases for facial emotions to increase the accuracy and generalizability of the image classification model.
2. Others types of emotional categories should be introduced into both text and image databases to allow a more comprehensive evaluation of emotions.
3. Machine learning and deep learning methods should be used for the analysis of behavioral patterns rather than simple rule-based methods which enable learning of complex behavioral patterns.
4. Other modes, such as speech emotion recognition, voice data, and physiological sensor data, should be introduced to enrich multimodal evaluation of emotions.
5. The next step should include exploration of advanced multimodal fusion methods that can learn to combine textual, visual, and behavioral information in an optimal way.
7. Healthcare institutions, educational organizations, and counseling centers could benefit from using AI-assisted mental health monitoring systems as supportive technology for raising emotional awareness and mental well-being.

REFERENCES

Aaron T. Beck. (1976). *Cognitive therapy and the emotional disorders*. Jesse H. Wright *et al*. (2023). *Learning cognitive-behavior therapy*.

Aaron T. Beck., & Emily A. P. Haigh. (2014). *Advances in cognitive theory and therapy*. David Dobson., & Keith S. Dobson. (2021). *Evidence-based practice of cognitive-behavioral therapy*.

Abd-Alrazaq, A., Alajlani,z M., Alalwan, A. A., Bewick, B. M., & House, A. (2021). An artificial intelligence chatbot for patients with mental illness: A user-centered design and evaluation study. *Journal of Medical Internet Research*, *23*(5), e24422. https://doi.org/10.2196/24422

Adam S. Miner., Arnold Milstein., & Jeffrey T. Hancock. (2020). Talking to machines about personal mental health problems. *JAMA*, *318*(13), 1217–1218.

Ajike, P. T., Ariyo, E. A., Ariyo, A. M., & Adubi, K. (2022). Emerging adults' awareness and perceptions of mental health problems and services in Nigeria. *Policy Insights from the Behavioral and Brain Sciences*, *9*(1), 44–48. https://doi.org/10.1177/23727322211068028

Al-Malki, H. A. (2024). Socio-emotional digital behavior theory: Integrating human affect, social validation, and algorithmic infrastructures. *Journal of Digital Social Science*, *2*(3), 145–168.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Amiriparian, S., Hübner, T., Gerczuk, M., Ottl, S., & Schuller, B. W. (2021). DeepSpectrumLite: A power-efficient transfer learning framework for embedded speech and audio processing from decentralised data. *arXiv*. https://doi.org/10.48550/arxiv.2104.11629 *Cited by: 39*

Andrew T. Campbell. (2020). *Digital phenotyping for mental health.*

Association for Computing Machinery SIGCHI curriculum.

Baltrusaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(2), 423–443. https://doi.org/10.1109/tpami.2018.2798607 Cited by: 6641

Ben Shneiderman. (2022). *Human-centered AI.*

Bhargava, V. R., & Velasquez, M. (2024). The architecture of attention: How infinite scroll and algorithmic personalization drive digital engagement. *Journal of Computer-Mediated Communication*, *29*(2), zmad045. https://doi.org/10.1093/jcmc/zmad045

Brandon T. McDaniel. (2021). *The DISRUPT: A measure of parent distraction with phones and mobile devices and associations with depression, stress, and parenting quality*. *Human Behavior and Emerging Technologies*, *3*(5), 922–932. https://doi.org/10.1002/hbe2.267

Calvo, R. A., Dinakar, K., Picard, R., & Christensen, H. (2021). Digital technologies and artificial intelligence in mental health interventions: A review of current applications and challenges. *The Lancet Digital Health*, *3*(2), e112–e121. https://doi.org/10.1016/S2589-7500(20)30287-2

Chancellor, S., & De Choudhury, M. (2020). Methods in social media language analysis for mental health: A systematic review. *Frontiers in Human Neuroscience*, *14*, Article 242. https://doi.org/10.3389/fnhum.2020.00242

Choudhary, S., Thomas, N., Ellenberger, J., Srinivasan, G., & Cohen, R. (2022). A machine learning approach for detecting digital behavioral patterns of depression using nonintrusive smartphone data (complementary path to Patient Health Questionnaire-9 assessment): Prospective observational study. *JMIR Formative Research*, *6*(6), e37736. https://doi.org/10.2196/37736 *Cited by: 37*

Christian Montag., Jon D. Elhai., & Hans-Jürgen Rumpf. (2021). Conceptualizing digital phenotyping and problematic smartphone use in psychological research. *Addictive Behaviors Reports*

Christopher M. Bishop. (2006). *Pattern recognition and machine learning*.

Colin M. Gray on dark patterns.

Daya, H. D. (2024). *Understanding the impact of virtual online and traditional learning on student mental health* (Publication No. 1904) [Doctoral dissertation, University of Connecticut]. Digital Commons @ UConn.

Drougkas, G., Bakker, E. M., & Spruit, M. (2024). Multimodal framework for mental health estimation using speech and text: A deep learning approach. *BMC Medical Informatics and Decision Making*, *24*(1), Article 312. https://doi.org/10.1186/s12911-024-02642-x

Eberhardt, T. S., Schaffrath, J., Moggia, D., Schwartz, B., Jaehde, M., Rubel, J. A., Baur, T., André, E., & Lutz, W. (2024). Decoding emotions: Exploring the validity of sentiment analysis in psychotherapy. *Psychotherapy Research*, *35*(2), 174–189. https://doi.org/10.1080/10503307.2024.2322522 *Cited by: 45*

Gopalakrishnan, A., Gururajan, R., Zhou, X., Venkataraman, R., Chan, K. C., & Higgins, N. (2024). Autonomous psychological health monitoring (APHM) systems: A multi-tiered design and comprehensive survey. *IEEE Reviews in Biomedical Engineering*, *17*, 45–63.

Jiquan Ngiam., Aditya Khosla., Mingyu Kim., Juhan Nam., Honglak Lee., & Andrew Y. Ng. (2011). Multimodal deep learning. *Proceedings of the 28th International Conference on Machine Learning (ICML)*, 689–696.

John E. R. Staddon. (2021). *The new behaviorism: Mind, mechanism and society* (2nd ed.).

John McCarthy. (1956). *A proposal for the Dartmouth summer research project on artificial intelligence*.

John Torous., Sarah Bucci., Ian H. Bell., *et al*. (2023). The growing field of digital psychiatry: Current evidence and the future of apps, social media, chatbots, and sensors. *World Psychiatry*, *22*(1), 3–17.

Jon D. Elhai., Christian Montag., & Jason C. Levine. (2020). Anxiety, fear of missing out, and problematic smartphone use: A review of relationships and mechanisms. *Addictive Behaviors*, *109*, 106403.

Jon D. Elhai., Christian Montag., & Jason C. Levine. (2021). Anxiety, fear of missing out, and problematic smartphone use: A review of relationships and mechanisms. *Addictive Behaviors*, *109*, 106403.

Jon D. Elhai., Robert D. Dvorak., Jason C. Levine., & Brian J. Hall. (2020). Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. *Journal of Affective Disorders*, *207*, 251–259.

Kendra Cherry. (2023). *Cognition in psychology: Definition and examples*. Verywell Mind article

Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in facial expression recognition: A survey of methods, benchmarks, models, and datasets. *Information*, *15*(3), 135. https://doi.org/10.3390/info15030135 *Cited by: 165*

Kostadin Kushlev., & Elizabeth W. Dunn. (2021). *Smartphones and subjective well-being: Putting the "you" in digital well-being*. *Current Opinion in Psychology*, *36*, 67–72.

Koutsouleris, N., Dwyer, D. B., & Pantelis, C. (2022). Machine learning and predictive analytics for early psychiatric risk assessment in clinical settings. *JAMA Psychiatry*, *79*(4), 312–322. https://doi.org/10.1001/jamapsychiatry.2021.4221

Lee, H., Kim, J., & Choi, Y. (2024). Smartphone-based human behavior task modeling for explainable mental health detection model. *International Journal of Human-Computer Interaction*, *40*(18), 4521–4538.

Lin, W. J., Lin, Y. H., & Huang, C. Y. (2022). Associations between smartphone usage behavior and mental illnesses through statistical and artificial intelligence modeling. *Computers in Human Behavior*, *128*, Article 107102. https://doi.org/10.1016/j.chb.2021.107102

National Institute of Mental Health. (2022). *Mental health information*. NIMH mental health information

Ofcom. (2021). *Children and parents: media use and attitudes report 2021*. Office of Communications. https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-and-parents-media-use-and-attitudes-report-2021

Ofcom. (2021). *Online Nation 2021 report*. Office of Communications. https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/online-nation

Ogueji, I. A.., Adebayo, O.., & Abdullah, A. S.. (2021). Help-seeking behaviors for mental illness in Nigeria: A systematic review. *Psychiatry Research, 301*, 113980

Patricia Areán., *et al*. (2021). The use of passive sensing and ecological momentary assessment in mental health. *Annual Review of Clinical Psychology*, *17*, 141–163.

Poria, S., Liang, P. P., & Cambria, E. (2022). Multimodal sentiment analysis systems using deep learning models: Fusing text, speech, and vision data. *Information Fusion*, *80*, 122–140. https://doi.org/10.1016/j.inffus.2021.11.009

Rosalind W. Picard. (1997). *Affective computing*. Rafael A. Calvo *et al*. (2020). *Oxford handbook of affective computing*

Rosalind W. Picard., Maja Pantic., & Björn Schuller. (2021). Advances in multimodal affective computing and mental state recognition. *IEEE Intelligent Systems*

Rui Wang., Weichen Wang., Andrew T. Campbell., *et al*. (2022). Smartphone sensing and behavioral analytics for longitudinal mental health monitoring. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*.

Sarah E. Domoff., Brandon T. McDaniel., *et al*. (2020). Development and validation of the Technoference in Parenting Scale (TIPS)**.** *Journal of Family Psychology*.

Shaffer, F., & Ginsberg, J. P. (2021). Trends in heart-rate variability signal analysis. *Frontiers in Digital Health*, *3*, Article 639444. https://doi.org/10.3389/fdgth.2021.639444

Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2021). Machine learning in mental health: A scoping review of methods and predictors. *Journal of Psychiatric Research*, *132*, 142–151. https://doi.org/10.1016/j.jpsychires.2020.11.038

Simon, B. D., Ozyoruk, K. B., Gelikman, D. G., Harmon, S. A., & Türkbey, B. (2024). The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: A narrative review. *Diagnostic and Interventional Radiology*, *30*(1), 1–9. https://doi.org/10.4274/dir.2024.242631 Cited by: 126

Singh, R., & Sharma, P. (2024). Multimodal mental health monitoring system integrating natural language processing, speech analysis, and computer vision techniques. *Journal of Ambient Intelligence and Humanized Computing*, *15*(2), 891–904. https://doi.org/10.1007/s12652-023-04612-4

Sohrab Saeb., Mi Zhang., Christopher J. Karr., *et al*. (2020). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research*.

Stachl, C., Plessen, M. B., Kroeger, H. A., Barr, N., & Harari, G. M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, *117*(30), 17680–17687. https://doi.org/10.1073/pnas.1920484117 *Cited by: 464*

Staddon, J. (2021). *The new behaviorism* (3rd ed.). Routledge. https://doi.org/10.4324/9781003145486

Stuart Russell., & Peter Norvig. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.

Tadas Baltrušaitis., Chaitanya Ahuja., & Louis-Philippe Morency. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Tadas Baltrušaitis., Chaitanya Ahuja., & Louis-Philippe Morency. (2019). *Multimodal machine learning: A survey and taxonomy*.

Tom M. Mitchell. (1997). *Machine learning*. American Psychiatric Association DSM-5 supports this concept

Torous, J., Wisniewski, H., Bird, B., Carpenter, E., David, G., Elejalde, E., & Firth, J. (2020). Creating a digital biomarker framework for smartphone passive sensing in psychiatry: Real-world tracking and validation. *JMIR mHealth and uHealth*, *8*(4), e16453. https://doi.org/10.2196/16453

Wang, S., Zhang, L., & Mitchell, C. (2023). A deep learning-based framework for depression detection using neural network architectures. *IEEE Transactions on Neural Networks and Learning Systems*, *34*(8), 4112–4124. https://doi.org/10.1109/TNNLS.2021.3124502

World Health Organization. (2021). *Mental health atlas 2020*.

World Health Organization. (2022). *World mental health report: Transforming mental health for all*. https://www.who.int/publications/i/item/9789240049338

Yeasmin, S., Semi, M. M. A., Rony, M. K. K., Das, S., Sabeena, A. A., & Rahman, R. (2025). Artificial intelligence for mental health monitoring: A solution for digital behavioral health care and education—An umbrella review. *Health Science Reports*, *9*(1), e399151143. https://doi.org/10.1002/hsr2.2025.399151143

Zhang, Y., Wang, X., & Liu, B. (2023). Facial emotion recognition system based on convolutional neural networks for mental health monitoring. *IEEE Access*, *11*, 54321–54333. https://doi.org/10.1109/ACCESS.2023.3281904