Grenze International Journal of Engineering and Technology, June Issue

AI Agent for Mental Wellness Monitoring

Rahul Nasare[1], Saniya Luhure[2], Sejal Shahare[3], Vaishali Dharmik[4], Sanika Shelgaonkar[5] and Harsh Naidu[6]
[1-6]Information Technology, Y.C.C.E, Nagpur, India
Email: rahulnasare28@gmail.com, saniyaluhure3@gmail.com, shahare.sejal05@gmail.com,
vaishalidharmik439@gmail.com, sanikashelgaonkar@gmail.com, naiduharsh45@gmail.com

***Abstract*— With our society becoming more and more technologically advanced, we are witnessing increased prevalence of psychological disorders like stress, anxiety and mood swings being experienced among different demographic groups. A number of people are reluctant to seek help either because of the social biases or fear of being criticized or lack of the necessary resources. Suggesting a new high-tech solution named AI Agent to Mental Health Tracking, it provides an advanced and yet sensitive approach to managing mental health concerns that are responsive to the increasing mental health concerns. This application combines machine learning algorithms, speech processing programs, interactive chatbots, and sophisticated AI techniques into one tool to assess the mental stability of a person based on the written messages and body language. The webcam records real-time information that is then examined using an advanced CNN that can recognize slight expression of faces; advanced Transformers like DistilBERT are used to examine the textual input and determine the sentiments and emotions of the user as accurately as possible. Besides, this platform has an emotion-monitoring interface that displays individual psychological tendencies over a certain period in time, helping one expand the level of self-exploration and take an active step to improve mental health. The AIbased platforms are an inclusive online community in which one can express emotions with no constraints and no limitations are placed on them, and with technology and human understanding being integrated to create a seamless system that would give a person immediate response, personalized guidance, and encourage a conversation. Considering all the factors, this framework would go a long way in helping people to gain access to cheaply provided psychological support in a constantly changing society.**

***Keywords*— Artificial Intelligence, Emotion Recognition, NLP, Visual Computing, Conversational Agent, Neural Networks, Face Emotion Identification, Sentiment Assessment, Mood Monitoring, Emotional Assistance, Health Surveillance.**

I. INTRODUCTION

Mental health is one of the most pressing issues in today's fast-paced and technology-driven world. The effects of rapid urbanization, academic competition, work pressure, and exposure to digital platforms have led to growing levels of stress, anxiety, depression, and emotional exhaustion among people. Although there has been a growing awareness about mental health in recent years, people still feel reluctant to seek professional help because of social stigma, cost, or lack of immediate access to mental health resources. As a result, there is a growing need for discreet and continuous support systems that can help people monitor and manage their emotional health.

Grenze ID: 01.GIJET.12.2.170_24

Artificial Intelligence (AI) has been identified as a promising technology that has the capability of understanding and interpreting human emotions through advancements in deep learning, Natural Language Processing (NLP), and computer vision.
Through facial recognition and text analysis, AI has the capability of recognizing human emotions with a high degree of accuracy.

II. Literature Survey

A. Literature Survey

The growing number of people suffering from mental health issues, specifically depression and anxiety, is a massive concern for public health worldwide. The first reason it has become such a massive concern is the problem of people seeking solutions.
This is because of social stigma, the absence of qualified doctors and therapists, and the high costs involved in conventional therapy. As a result, millions of people are left to fend for themselves without the treatment that could help them shake off their condition. This is a very recent study on how AI is predicting mental health and helping with treatment.
In [1], Shaik et al. proposed a Multi-Agent Deep Reinforcement Learning framework for the continuous monitoring of physiological parameters like heart rate, body temperature, and respiration rate to identify early symptoms of stress and depression. The proposed framework uses multiple intelligent agents working together to analyze the minute variations in physiological parameters.
In [2], Saleem et al. proposed a Multi-Agent System (MAS) using the Belief-Desire-Intention (BDI) reasoning model for diagnosing mental illnesses like anxiety and depression. The BDI model is a simulation of human cognitive reasoning, which increases the interpretability and psychological fit of the diagnosis process. The proposed system had an accuracy rate of up to 95%, proving its efficiency in dealing with the complex and nonlinear emotional patterns. The study clearly states that the inclusion of cognitive reasoning in AI systems increases the transparency, adaptability, and relevance of the system, making it very appropriate for intelligent mental wellness monitoring agents.
In [3], the authors Irfan et al. attempted to apply Machine Learning and Natural Language Processing (NLP) methods for the analysis of user-generated text data, such as diaries, journals, and electronic notes, for emotional and psychological analysis. By detecting linguistic features associated with stress levels, mood swings, and emotional instability, the system facilitated continuous and non- intrusive monitoring. Through sentiment and semantic analysis, the system provided personalized emotional feedback and early warnings for distress patterns. This reference proves the efficacy of text-based behavioral modeling, which is an important module of AI agents for mental wellness monitoring.
In [4], Barath et al. analyzed the use of conversational AI agents like chatbots and virtual friends as a means to bridge the gap in mental health care access across the world. The authors point out the importance of using emotionally intelligent and culture-adaptive AI models to improve user engagement and trust. The authors emphasize the importance of emotionally intelligent and culture-adaptive AI models to improve user engagement and trust. The authors conclude that the use of conversational agents can help bridge the treatment gap by providing continuous support to users.
In [5], the authors, Danieli et al., studied the integration of a Conversational Agent (CA) in a mobile psychotherapy platform for patients with stress-related and compulsive disorders. The CA was a supportive complement to traditional therapy, engaging patients between sessions, supporting coping strategies, and providing emotional support. The authors found increased patient engagement, therapy compliance, and overall satisfaction among patients and therapists.

B. Research Gap

Most mental health models have been developed for either physiological data, text analysis, or conversation support. However, very few models have been developed that consider multi-modal data (facial expressions, text, and behavior) for overall emotional analysis. Most models have not been designed for realtime personalization, which means that they do not continuously adapt to changes in user mood. Some models still have not adequately addressed issues related to ethical considerations, such as data privacy, bias, and transparency.
Moreover, AI models have not been designed to provide highly empathetic and culturally competent interactions, which are provided by human support.

III. PROPOSED METHODOLOGY

The proposed methodology describes the development of a multi-modal system that intends to monitor and analyze the emotional states of the users in real-time. The proposed system integrates Computer Vision, Natural Language Processing (NLP), and Machine Learning models to analyze facial expressions and text inputs. The emotional analysis of facial expressions is performed using a convolutional neural network, and text-based emotional and sentiment analysis is performed using a fine-tuned transformer model. Based on the analyzed emotional state, the system provides empathetic responses and music recommendations according to the user's mood. Each module is individually developed and trained for accuracy and reliability before being integrated into a single system.
The flowchart below explains the processing of the AI Agent based on the user input from facial and text emotion analysis

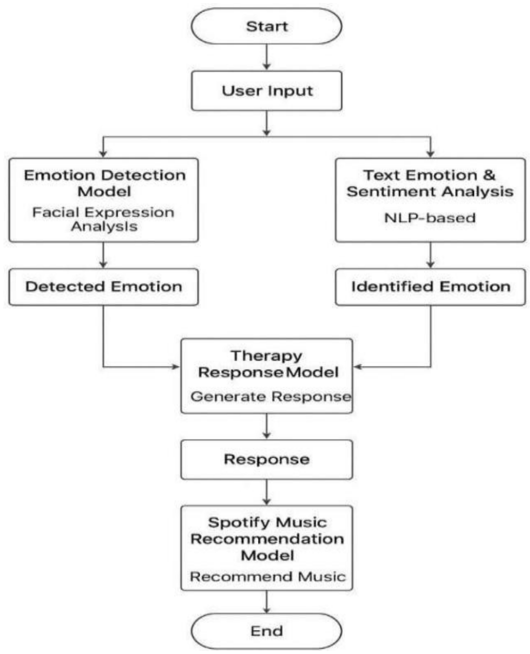

Fig. 1 System Methodology

The proposed system is designed as a multi-modal intelligent framework that combines Computer Vision, Natural Language Processing (NLP), and Machine Learning to monitor, interpret, and respond to the users' emotional states in real time. The methodology includes five major stages: Data Collection, Data Preprocessing, Model Development, Model Evaluation, and System Implementation.

A. Data Collection

The type of data needed for the proposed system is collected from various publicly available and domainspecific datasets to enable the various modules of the system architecture. For facial emotion recognition, the FER2013 dataset is used to train the convolutional neural network model. The FER2013 dataset is a collection of grayscale facial images that are labeled into seven basic emotions: happiness, sadness, anger, fear, surprise, disgust, and neutral, which helps the system to correctly identify the visual emotional expression. For textbased emotion and sentiment analysis, the GoEmotions dataset is used in conjunction with selected conversational text samples for multi-class emotion classification, and the Stanford Sentiment Treebank is used to identify the sentiment polarity as positive or negative.

B. Data Preprocessing

Preprocessing of the data ensures that the data is clean, normalized, and correctly formatted to ensure that the training and testing processes run smoothly. For the facial emotion recognition model, the images are standardized to 48x48 pixels and then converted to grayscale format. This ensures that the pixel values fall within the range of 0 to 1, making the computations simpler. To improve the ability of the model to generalize and prevent overfitting, the images are augmented to include rotation, flip, and zoom effects. There are seven distinct emotions in the dataset. The data is split to include 80% for training and 20% for testing. The emotion labels are encoded using one-hot encoding for the seven classes of emotions, and the dataset is divided into 80% training and 20% testing sets.

In text-based emotion and sentiment analysis, preprocessing involves the removal of URLs, special characters, hashtags, and unnecessary spaces, followed by the transformation of text to lowercase for uniformity. The preprocessed text is tokenized using the DistilBERT tokenizer, and padding and truncation are carried out to ensure a fixed input size for the model. The emotion and sentiment labels are encoded numerically, and the dataset is divided into training, validation, and testing sets in an 80:10:10 split. Class weighting or oversampling techniques are employed to account for the potential issue of class imbalance.

C. Model Development

The system consists of a range of independently developed modules to achieve optimal performance. The facial emotion recognition module is developed using a Convolutional Neural Network (CNN) model architecture that comprises convolutional, batch normalization, max-pooling, dropout, and fully connected layers with the Adam optimizer and categorical cross-entropy loss functions to predict the seven facial emotions from the webcam. The text emotion analysis module uses a DistilBERT model fine-tuned for multi-class emotion classification with a softmax output layer to output probability distributions over the emotional classes. Moreover, the sentiment analysis task is carried out using a DistilBERT model that is fine-tuned on the Stanford Sentiment Treebank for binary classification of positive and negative sentiments. A conversational model built using the Transformer architecture, inspired by GPT models, is trained on counseling dialogue datasets to generate empathetic and contextually aware responses for therapy. Moreover, a hybrid music recommendation system built using content-based and collaborative filtering techniques is designed to recommend mood-synced songs based on the identified emotional states and their corresponding audio features.

D. Model Evaluation

The performance of the individual models is measured using the standard classification performance metrics like accuracy, precision, recall, and F1-score. The performance on the validation set is tracked during the training process to ensure proper generalization.

E. System Implementation

The entire system is developed as a web application that integrates all the modules into a complete system. The front-end is developed using React.js, HTML, CSS, and Material-UI to facilitate live chatbot dialogue, facial emotion recognition using the webcam, mood tracking dashboards, relaxation methods, and music suggestion. The back-end is developed using Flask or FastAPI with Python to facilitate real-time data processing, interaction between models, and API integration, including the Spotify Web API for music suggestions. MongoDB or Firebase is utilized for secure storage of user interactions and mood data. The modules are capable of communicating with each other using RESTful APIs to facilitate smooth data transfer and real-time empathetic response and music suggestion generation.

IV. RESULTS AND DISCUSSION

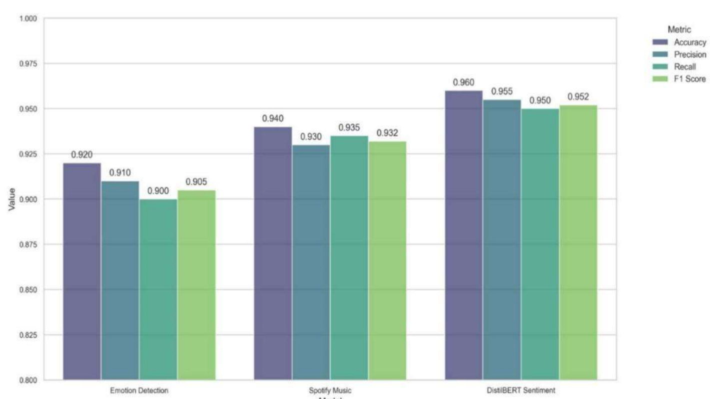

Fig 2. Model Performance Metrics

We tested accuracy, consistency and real time performance in each of the modules to ensure that the system is legitimate. To confirm the outputs of the facial emotion detection and sentiment analysis, we employed standard measures, and the therapy-response component was assessed in terms of relevance, consistency, and empathy, and we ensured that that the music- recommendation section of the app does reflect the mood of the user.
The chart provides a comparison of the three key models, which include Emotion Detection, Spotify Music Recommendation, and DistilBERT Sentiment Analysis. We considered Accuracy, Precision, Recall, and F1 Score. DistilBERT is on the top of the rating in all measures due to the possibility to select context and feeling.
The recommendation algorithm of Spotify is also performing well since it matches up to emotions. Emotion Detection model is a reliable one, though somewhat less powerful, and the primary reason is that real-life lighting, as well as angles of the face, may confuse it. All in all, the scores indicate that text sentiment analysis is the most precise, which validates the system.

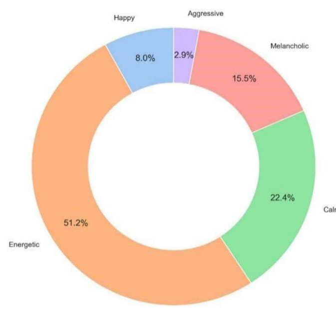

Fig 3. Class Distribution of Spotify Music Recommendations

The figure indicates the genre distribution of the music that we suggest Energetic, Calm, Melancholic, Happy and Aggressive. It informs us that the system can match affective reactions with appropriate songs, bearing in mind the valence and the energy. The combination of slow and fast songs accompanies the mood- elevating or relaxing process. Due to the appropriate balance of genres, the system is not confined to one emotional bias; it can suit any mood of the user.

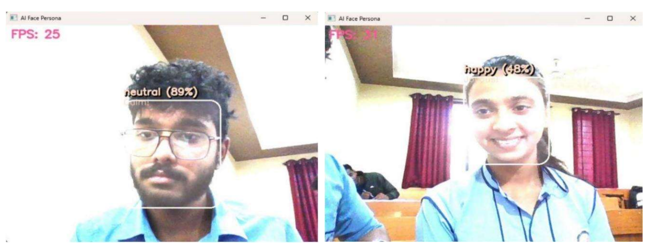

Fig 4. Real-Time Facial Emotion Detection Results

The figure shows real-time emotion detection outputs captured through webcam input. The system successfully identifies facial expressions such as Neutral and Happy, along with confidence percentages, while maintaining stable frame-per-second (FPS) performance.

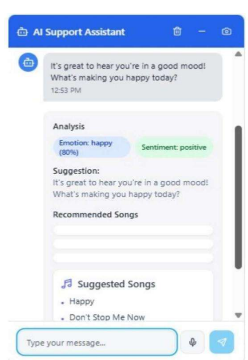

Fig 5. Chatbot Interface with Emotion and Sentiment Analysis

The image depicts the real time face emotion detection on plugging the web camera. The system is capable of identifying happy and other simple faces with a constant FPS and is accompanied with confidence scores. The findings support the claim that CNN model is resistant to various angles of faces and minor light variations. This implies that the emotion recognition will be able to monitor moods in the real world.
The scores on the test demonstrate the high emotion-detecting and user-supporting capacity of the system. It reads statements, calculates written text, and a real-time stream of code, therefore, conversations are lagless. The system follows moods, has interactive conversations, suggests the music that members like, making it look like friendly personal robot, the one that keeps the members happy at all times, as long as they use it.

V. Conclusion

Using the power of instantaneous interaction to meet the requirements of the user, this AI-based solution offers a comprehensive and informative solution to improve the level of mental health awareness. This system can provide the responsive answers by integrating the methods of computer vision, sentiment analysis through natural language processing and contextual understanding through machine learning to give the empathetic responses that are precisely defined to user requirements, similar to actual cognitive skills. The way to improve the accuracy is to combine textual and visual sentiment detection algorithms to allow the systems to more effectively and empathetically measure the emotional states of the users.
The improvement of the worth of such system implies one auxiliary element: the therapeutic response module provides reassuring advice, inspirational quotations, and emotional support. Furthermore, the musical suggestion module offers a supplementary healing technique, which consists of the judiciously chosen aural content by means of which emotional stability is reached, as well as the anxieties are reduced.

References

[1] J. De Freitas, A. K. Uğuralp, Z. Oğuz‐Uğuralp, and S. Puntoni, “Chatbots and mental health: Insights into the safety of generative AI,” J. Consumer Psychol., vol. 34, no. 3, pp. 481–491, 2024.
[2] R. Doshi et al., “TheraGen: A LLaMA-based conversational AI for personalized mental health support,” in Proc. NeurIPS Workshop on AI for Wellbeing, 2024.
[3] L. Fernandez et al., “AI-enhanced VR therapy using real-time emotion recognition,” Virtual Health Technologies Journal, 2024.
[4] R. Gupta et al., “Transformer-based multimodal models for mental health diagnostics,” in Proc. Int. Conf. Computational Neuroscience, 2024.
[5] G. Espejo, W. Reiner, and M. Wenzinger, “Exploring the role of artificial intelligence in mental healthcare: Progress, pitfalls, and promises,” Cureus, vol. 15, no. 9, Art. no. e44748, Sep. 2023.
[6] M. Irfan et al., “Personalized emotional health monitoring using ML and NLP on user-generated text,” J. Computational Mental Health Analytics, 2023.
[7] M. Bunge et al., “The Thera-Turing test: Evaluating therapeutic capabilities of mental health AI agents,” in Proc. Conf. Artificial Intelligence in Psychology, 2023.
[8] S. Martinez et al., “Federated learning for privacy-preserving mental health monitoring,” J. Privacy-Preserving AI, 2023.
[9] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, “Natural language processing applied to mental illness detection: A narrative review,” npj Digit. Med., vol. 5, no. 1, Apr. 2022.
[10] E. E. Lee, J. Torous, M. De Choudhury, C. A. Depp, S. A. Graham, H. C. Kim, M. P. Paulus, J. H. Krystal, and D. V. Jeste, “Artificial intelligence for mental health care: Clinical applications, barriers, facilitators, and artificial wisdom,” Biol. Psychiatry Cogn. Neurosci. Neuroimaging, vol. 6, no. 9, pp. 856–864, Feb. 2021.
[11] K. Denecke, S. Vaaheesan, and A. Arulnathan, “A mental health chatbot for regulating emotions (SERMO) – Concept and usability test,” IEEE Trans. Emerg. Topics Comput., vol. 9, no. 3, pp. 1170–1182, Jul. 2021.
[12] A. Basil, “ELIZA: The chatbot who revolutionised human-machine interaction – An introduction,” Nerd For Tech, 2021.
[13] S. Graham, C. Depp, E. E. Lee, C. Nebeker, X. Tu, H.-C. Kim, and D. V. Jeste, “Artificial intelligence for mental health and mental illnesses: An overview,” Curr. Psychiatry Rep., vol. 21, no. 11, Nov. 2019.
[14] A. N. Vaidyam, H. Wisniewski, J. D. Halamka, M. S. Kashavan, and J. B. Torous, “Chatbots and conversational agents in mental health: A review of the psychiatric landscape,” Can. J. Psychiatry, vol. 64, no. 7, pp. 456–464, Mar. 2019.