



(RESEARCH ARTICLE)



AI4MentalWellness: Democratizing access to mood insights and professional mental health guidance

Durga Bhavani Sankabathula *, Tapaswini Chinde, Gousia Rahmani Shaik and Poojitha Dommeti

Department of Computer Science and Engineering, Aditya College of Engineering and Technology, Surampalem, Kakinada, Andhra Pradesh, India.

International Journal of Science and Research Archive, 2026, 18(03), 332–338

Publication history: Received on 27 January 2026; revised on 02 March 2026; accepted on 04 March 2026

Article DOI: <https://doi.org/10.30574/ijrsra.2026.18.3.0451>

Abstract

AI4MentalWellness is an AI based model that aims to level access to mood knowledge and professional mental health advice through a user-friendly and free-to-use application. The system helps users to monitor the emotional states on a regular basis, note the daily experiences and discover the pattern of stresses through smart analysis tools. The platform enhances self-awareness of mental health risks and early intervention by converting self-reported data into useful mood inferences. Another impact of the application is that it offers a safe platform upon which users are able to communicate with therapists in a bid to help the professionals to comprehend emotional patterns and provide customized support. The presented structure is centered on simplicity, privacy, and constant interaction to develop positive mental health habits. It will address the issues of stigma reduction, emotional self-monitoring, and enhance access to care among individuals who do not seek help early. This study shows that mood analysis using AI can assist in preventing mental health and enhancing emotional well-being among various groups in society. AI4MentalWellness is an AI based model that aims to level access to mood knowledge and professional mental health advice through a user-friendly and free-to-use application. The system helps users to monitor the emotional states on a regular basis, note the daily experiences and discover the pattern of stresses through smart analysis tools. The platform enhances self-awareness of mental health risks and early intervention by converting self-reported data into useful mood inferences. Another impact of the application is that it offers a safe platform upon which users are able to communicate with therapists in a bid to help the professionals to comprehend emotional patterns and provide customized support. The presented structure is centered on simplicity, privacy, and constant interaction to develop positive mental health habits. It will address the issues of stigma reduction, emotional self-monitoring, and enhance access to care among individuals who do not seek help early. This study shows that mood analysis using AI can assist in preventing mental health and enhancing emotional well-being among various groups in society.

Keywords: Multi-Modal Emotion Detection; Artificial Intelligence; Machine Learning; Mental Health; Emotional Well-Being; Mental Health Support; Professionals Help; Guidance and Support

1. Introduction

Mental health is very important in shaping the thoughts, feelings and behavior of individuals in their day-to-day lives. In the contemporary rapid world, stress, anxiety and emotional difficulties are common among a good number of people, yet they fail to seek help because of stigma, ignorance or inaccessibility to professional help. This means that the symptoms of mental health problems can go without detection and treatment.

As the sphere of Artificial Intelligence is evolving, digital technologies are produced to facilitate emotional well-being by continuously monitoring and providing smart analysis. In addition to promoting self-knowledge and early

* Corresponding author: S. Durga Bhavani

intervention, AI-based systems can be used to assist people to learn how to be emotionally balanced and help recognize their emotional patterns. Proposed system, which is AI4MentalWellness, is an AI-based framework that will offer a simple and safe platform through which mood tracking, emotional analysis, and connection with a therapist can be done. It allows users to track their daily feelings, obtain meaningful insights, and receive professional assistance in the cases.

The system will enhance emotional awareness, early help-seeking behaviour, and make mental healthcare more accessible and effective by combining AI-based analysis and mental health support.

2. Literature survey

The recent researches demonstrate the increasing role of AI and mobile apps in the field of mental health care, primarily in mood-tracking and emotional tracking. These web applications increase emotional intelligence and help people identify stress, anxiety, and depression symptoms at an early stage in their formation. Such interventions have been seen to raise emotional stability and reduce depressive symptoms as they help those who are not willing to go through conventional therapy to engage in constant self-reflection, consequently, upon the participation, they become emotionally stable.

Despite the mentioned benefits, the current applications could be referred to as passive data storage lacking sophisticated real-time analysis and long-term interaction with the users. Even though sensing on smartphones can be applied to obtain valuable behavioral pointers, the existing platforms can be described as a bridge between the data extraction and expert intervention. In addition, the privacy and security of data, as well as the trust of users, continue to be a significant challenge to the usage and effective applications of these digital health tools.

To address these drawbacks, it is of dire necessity that integrated systems that can integrate multi-modal emotion analysis systems and establish secure connectivity of therapists. The AI4MentalWellness model offered in this proposal will cover those gaps and offer a complete-fledged, intelligent, and convenient platform. The focus on personalized advice and long-term and secure monitoring causes the framework to transform mental health assistance into a highly digitized and active experience.

3. Existing system

Nowadays, there is a multitude of digital mental health apps and mood-tracking devices that allow consumers to document their mood and engage in meditation, as well as a simple form of mental health guidance. These systems mostly concentrate on the self-assessment via questionnaires, journaling and other basic relaxation measures. They are targeted towards being more aware of the mental health and offer convenient access to the support resources via mobile devices.

Nevertheless, there are a number of limitations of the existing systems. There are a lot of apps that offer simple tracking options with no smart analysis and customized feedback. They tend to gather emotional information but fail to convert it into perceptive ideas and timely alerts. There are apps that do not have real-time tracking and intelligent suggestions and are not effectively integrated with professional therapists, making this aspect diminish their ability to offer prompt assistance.

The other significant issue is privacy and data security. Research has revealed that various health applications transmit their user data to third parties or do not have effective protection features that pose threats to sensitive personal information. There are also few apps on mental health that are clinically verified, and this is a cause of concern in terms of reliability and trust.

In that way, in spite of the fact that current systems assist in the simplest ways to track and monitor mood, to be more aware of it, they still do not offer intelligent emotional analysis, personal advice, good privacy level, and uninterrupted communication with a therapist, which shows the necessity of a sophisticated AI-based mental health system.

4. Methodology

The proposed system will take a step-by-step and secure method of identifying the dangers of early mental health by utilizing a multi-modal artificial intelligence model. It begins with the safe authentication of the users then the emotional data privacy and access is regulated. After a user is logged in, he/she can input emotions in a number of modalities, e.g.

text, facial pictures and voice recordings. This system is multi-modal and, therefore, is able to capture non-verbal emotional cues in addition to verbal thus increasing the level of reliability when compared to uni-modal systems.

All the collected inputs are done through preprocessing and feature extraction and finally emotion classification is done. Text data is processed using natural language processing techniques such as tokenization, elimination and normalization of stop words to extract semantic patterns. Emotion based facial expression is detected by processing the facial image and normalizing it and speech signal is filtered and the acoustic features of the speech i.e. variation in pitch and tone. The operations convert raw data to structured data that can be processed by deep learning models.

Each modality is analyzed by a separate deep learning model. Text emotion recognition is aided by language based model that is capable of the perception of the contextual nature, facial emotion recognition is aided by convolutional neural network, and speech emotion recognition is aided by the neural network models owing to vocal characteristics. The combination of the individual predictions is then performed at the decision level in order to produce the final state of the emotions and, therefore, to reduce the ambiguity and maximize the accuracy of the prediction.

The predicted emotional states are stored in user specific database which is repeatedly monitored to identify long term emotional trend. The system recommends professional consultation in case some negative patterns should be persistent, such as stress, anxiety, depression, etc. The user is allowed to access summarized emotional reports and this information can be used by authorized therapists to implement ethical control and improve the clinical fidelity of the artificial intelligence-based mental health monitoring system.

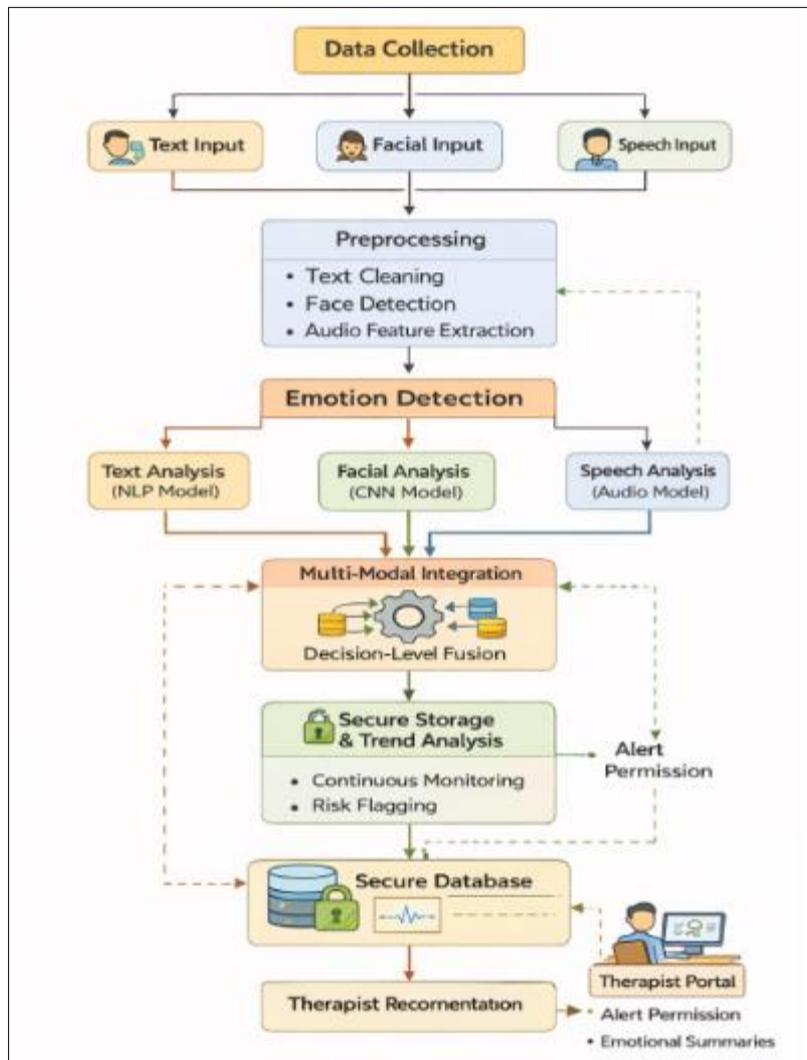


Figure 1 Flowchart of the Proposed Multi-Modal Mental Health Monitoring System

5. Experiments and Results

5.1. Experimental Setup

The proposed multi-modal system of mental health monitoring was developed with the help of Python and deep learning systems. The publicly available datasets on text sentiment analysis, facial expression recognition, and speech emotion detection were used to test the system. Emotion classification was done using deep learning models that were pre trained in order to get reliable performance.

The assessment was done by undergoing tests in each modality, and later testing the multi- modal framework. Validation of real-time prediction capability, emotional trend storage and therapist alert feature were tested using sample user inputs. The system was implemented in a typical computing environment which can be used in academic research and prototype validation.

5.2. Dataset Description

In the case of text emotion detection, sentiment and emotion datasets which were publicly available were used to test the performance of linguistic classification. Labelled facial expression datasets of categories happy, sad, angry and neutral were used to test facial emotion recognition. Speech emotion recognition was assessed on conventional emotional audio collection of different tone, pitch and intensity.

In order to develop the models, the datasets were separated into training and testing sets. To test at the system level, pre-trained models were incorporated into the application to do real-time inference.

5.3. Performance Evaluation Metrics

In the current study, the major evaluation measure was the classification accuracy. Other than accuracy, prediction consistency and integration reliability were also observed in the process of multi-modal fusion. Individual emotion detection modules were also compared to the integrated system to determine whether the system could improve performance on the level of decision-level fusion.

5.4. Results of Individual Modules

The performance of individual emotion detection modules is summarized in Table 1.

Table 1 Performance Accuracy of Individual Emotion Detection Modules

Emotion Detection Module	Accuracy (%)
Text Emotion Detection	93.5
Facial Emotion Recognition	94.1
Speech Emotion Detection	92.8

Modalities showed good individual performance. Nevertheless, there were slight differences that were found when there are differences in the clarity of the inputs, or the environmental conditions like light and background noise.

5.5. Integrated Multi-Modal System Performance

The system as a whole with the decision-level fusion of predictions produced a better classification accuracy of 96.2%. The combination of verbal and non-verbal affective cues diminished the ambiguity and enhanced the degree of prediction.

Multi-modal system was more stable and robust than single modality approaches, especially when one input modality gave uncertain predictions.

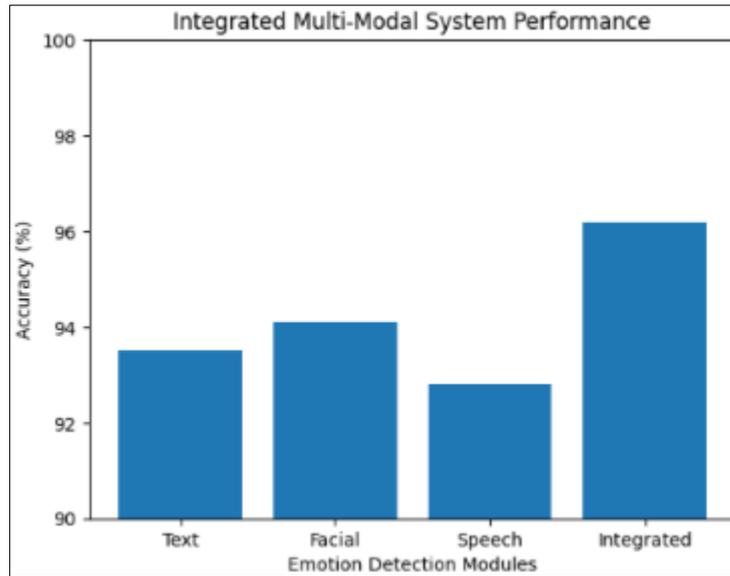


Figure 2 Accuracy of Proposed Multi-Modal Emotion Detection System

6. Discussion

The experimental test proves the fact that the combination of the text, facial and speech emotion recognition bolsters the overall system reliability. Decision-level fusion strategy boosts the accuracy of prediction by combining emotional predictions of various sources which are complementary. Further on, the system was also able to execute the secure storage of emotional data as well as perpetual trend monitoring.

The findings suggest that the presented framework can be used in real-time mental health monitoring software based on digital tools, especially in academic and prototype implementations.

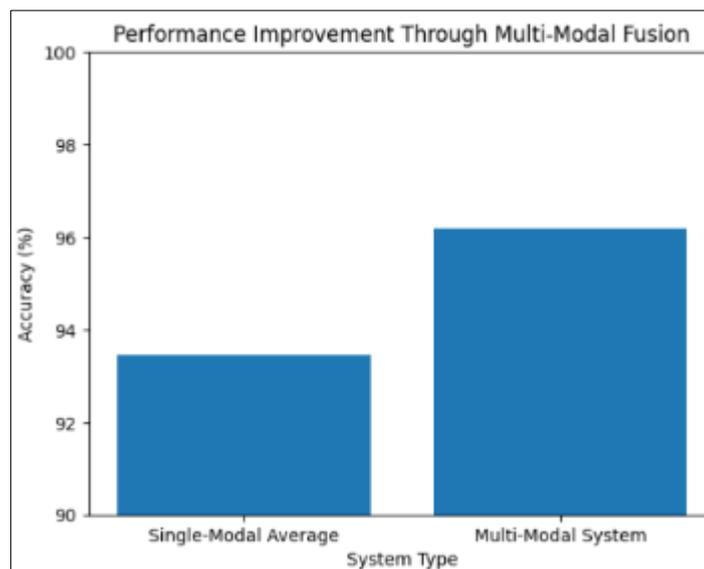


Figure 3 Accuracy of Proposed Single-Modal Vs Multi-Modal Emotion Detection System

6.1. Comparison with Existing Frameworks

Some of the current mental health monitoring models are either single-modal or low-input based like text-based mood analysis or questionnaire-based. Although these systems have simple emotional feedbacks, they do not always pick non-verbal behavior and do not have 24-hours monitoring abilities. The comparative analysis table is given in table 2 where the proposed system is compared with current frameworks.

Table 2 Comparison with Existing Mental Health Monitoring Frameworks

Feature / Parameter	Existing Frameworks	Proposed System
Input Modalities	Single or dual modality (mostly text or surveys)	Multi-modal (text, facial, speech)
Emotion Detection Approach	Rule-based or single-model ML	Deep learning-based models
Non-Verbal Emotion Analysis	Not supported or limited	Fully supported (facial & speech)
Multi-Modal Fusion	Not available	Decision-level fusion
Continuous Trend Monitoring	Not supported	Supported
Real-Time Prediction	Limited	Supported
Therapist Interaction	Not integrated	Integrated
Application Scope	Short-term analysis	Long-term mental health monitoring

Future Scope

Although the proposed multi-modal mental health monitoring system has already been more accurate and reliable, there are ways that some of these areas can be improved through future researchers. The other significant extension would be the inclusion of hi-tech transformer-based language models that would be used to increase the contextual understanding of text emotion recognition. That would enhance its capacity to discern the complicated emotive expression and sarcasm.

The future can be improved with the introduction of real-time implementation of the mobile applications to make them more accessible and widespread. The emotional analysis can be further strengthened with the help of the physiological data such as the heart rate variability or the stress indicators, which can be measured with the help of wearable sensors. Such multi sensor integration would be more usable and robust.

The introduction of adaptive learning systems would be the other potential enhancement with the system being personalised to the trends of each user in the long run by personalising emotional prediction models. This would allow tracking of long-term monitoring and early detection of the risks. Moreover, the privacy could be increased by the federated learning techniques since the models might be trained on the centralized servers without the transfer of user-sensitive information.

Finally, the system would also need to be subjected to a massive clinical validation which involves mental health professionals to establish the system reliability within the healthcare settings. Further improvement of the data assortment and the cross-cultural emotional research would facilitate the generalization competence and the preparedness to implement it practically.

7. Conclusion

In this paper, a secure and multi-modal AI framework was proposed to facilitate the detection of mental health risk in the early stages. By using text, facial, and speech-based emotion detection, the proposed system was able to effectively detect both verbal and non-verbal emotional characteristics of the user, which addresses the limitations of the traditional single-modal mental health detection methods. Deep learning algorithms were used to improve the overall classification accuracy of the proposed system and to avoid ambiguity in the prediction of the user's emotions.

Moreover, the proposed system was experimentally evaluated to prove that the overall system performance was better than the performance of the individual emotion detection modules used in the proposed system. Moreover, the proposed system was also capable of monitoring the overall emotional trends of the user, which helps to identify the prolonged negative emotional state of the user.

Based on the overall performance of the proposed system, it was found that the proposed system has strong potential to be used in the digital mental health monitoring of users in the academic and prototype levels. By enhancing the proposed system, the overall framework can be extended to be used in the development of mental health support systems.

Compliance with ethical standards

Acknowledgments

The authors would like to express their gratitude to the Department of Computer Science and Engineering for their academic support and there was no external funding was received for this research

Disclosure of conflict of interest

The authors declare that there is no conflict of interest.

Statement of ethical approval

This research work used publicly available data sets, and there were no direct human subjects involved.

Statement of informed consent

Informed consent was not required as no identifiable person data was used in this study.

References

- [1] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, “Delivering cognitive behavior therapy using a fully automated conversational agent,” *JMIR Mental Health*, vol. 4, no. 2, pp. 1–10, 2017, doi: 10.2196/mental.7785.
- [2] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, “Machine learning in mental health: A scoping review of methods and applications,” *Psychological Medicine*, vol. 50, no. 9, pp. 1426–1448, 2020, doi: 10.1017/S0033291719000151.
- [3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [4] R. A. Calvo, D. N. Milne, M. S. Hussain, and H. Christensen, “Natural language processing in mental health applications using non-clinical texts,” *Natural Language Engineering*, vol. 23, no. 5, pp. 649–685, 2017, doi: 10.1017/S1351324916000383.
- [5] S. Koelstra et al., “DEAP: A database for emotion analysis using physiological signals,” *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012, doi: 10.1109/T-AFFC.2011.15.
- [6] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [7] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.
- [8] P. Ekman, “An argument for basic emotions,” *Cognition and Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.
- [9] C. Busso et al., “IEMOCAP: Interactive emotional dyadic motion capture database,” *Language Resources and Evaluation*, vol. 42, no. 4, pp. 335–359, 2008.
- [10] S. R. Livingstone and F. A. Russo, “The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS),” *PLOS ONE*, vol. 13, no. 5, pp. 1–35, 2018, doi: 10.1371/journal.pone.0196391.
- [11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
- [12] T. Baltrušaitis, C. Ahuja, and L. P. Morency, “Multimodal machine learning: A survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019, doi: 10.1109/TPAMI.2018.2798607.
- [13] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, “Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge,” *Speech Communication*, vol. 53, no. 9–10, pp. 1062–1087, 2011.
- [14] S. Li, W. Deng, and J. Du, “Reliable crowdsourcing and deep locality-preserving learning for facial expression recognition in the wild,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2852–2861.
- [15] H. Lian et al., “A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face,” *Entropy*, vol. 25, no. 10, p. 1440, 2023.
- [16] A. Xiang et al., “Transformer-Based Multimodal Fusion Network for Emotion Recognition,” *IEEE Access*, vol. 12, pp. 23456–23470, 2024.