WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Multimodal AI for Mental Health Detection: Creating Privacy-Preserving Systems for Early Intervention

Dhanani Mayur Arvindbhai, Pipaliya Dhruvkumar Kiranbhai, Janki Tejas Patel

Sal College Of Engineering, Department Of Computer Engineering, Ahmedabad, Gujarat, India

Abstract

Millions of people worldwide suffer from mental health illnesses, but early identification and treatment are still difficult because of arbitrary evaluation techniques and restricted access to medical care. A thorough framework for multimodal artificial intelligence systems that integrate speech, text, and physiological signals for early mental health detection while protecting patient privacy is presented in this paper.

Keywords: Multimodal Machine Learning, Mental Health Detection, Healthcare Technology.

1. Introduction

Nearly a billion people worldwide suffer from mental health illnesses, which present significant social, personal, and financial difficulties. Conventional diagnostic techniques, which are based on questionnaires and interviews, frequently rely on subjective assessment and postpone intervention. Multimodal machine learning, in particular, has made it possible to analyse a variety of data sources, including speech, text, and physiological signals, in order to accurately identify early indicators of mental health disorders. Proactive rather than reactive care is made possible by these AI-powered systems, which can recognize possible crises days before a clinical diagnosis. But protecting private mental health information is still a major worry. Thus, creating AI architectures that incorporate multimodal data while protecting user privacy is crucial to implementing early mental health interventions that are both morally and practically sound.

2. Literature Review

2.1 Multimodal Machine Learning in Mental Health

By combining various data sources, including speech, text, and behavioural signals, multimodal machine learning has transformed mental health research and captured the complex nature of mental illnesses. Recent developments make use of transformer-based architectures that employ self-attention mechanisms to efficiently fuse heterogeneous data and model intricate cross-modal relationships, whereas earlier research examined individual modalities. Models trained on datasets such as the D-Vlog corpus and DAIC-WOZ clinical interviews have shown significant improvements in detection accuracy as a result, with F1-scores as high as 0.88. These developments demonstrate how multimodal fusion is increasingly promising for more accurate and comprehensive mental health evaluations.

2.2 Speech-Based Mental Health Detection

By using vocal biomarkers to provide insights into emotional and cognitive states, speech analysis has emerged as a potent and non-invasive method for evaluating mental health. Subtle indications of mental distress can be found in key indicators like speech rate, word choice, rhythm, and pitch. While contemporary deep learning models, like Wav2vec 2.0, allow for automatic feature extraction straight from raw audio, methods like eGeMAPS offer standardized acoustic features that have a strong correlation with mental health conditions.

2.3 Text-Based Methods for Evaluating Mental Health Risk

By examining text from written communications, social media, and medical records, natural language processing (NLP) has demonstrated remarkable efficacy in evaluating mental health. Strong indicators of psychological distress include linguistic markers like the frequent use of first-person pronouns, words that convey negative emotions, and particular semantic patterns. Due to their ability to capture nuanced linguistic and contextual cues associated with mental health, large language models—especially domain-specific ones like MentalBERT—have greatly advanced this field. F1-scores above 0.9

have been attained by research using social media sites like Reddit and Twitter, highlighting NLP's potent potential for longitudinal, real-time monitoring and the early identification of mental health emergencies.

2.4 Physiological Signal Integration

In addition to conventional self-reports, physiological indicators like heart rate variability, electrodermal activity, sleep patterns, and activity levels offer objective insights into mental health. Continuous monitoring of these signals is now possible thanks to developments in wearable technology and smartphones, which makes it possible to identify disorders like bipolar disorder, anxiety, and depression early on. The potential of this strategy is demonstrated by technologies such as Biobeat's wearables, which accurately forecast mood swings. Current signal processing and personalized modeling techniques have significantly increased the clinical utility and reliability of physiological signal-based mental health assessment, despite obstacles like individual variability, signal noise, and the requirement for continuous data collection.

3. Proposed Multimodal Framework

Our suggested framework uses a privacy-preserving multimodal architecture to integrate three main data modalities: speech, text, and physiological signals. The system uses a hierarchical approach, starting with modality-specific encoders and progressing to cross-modal fusion mechanisms and privacy-preserving inference protocols.

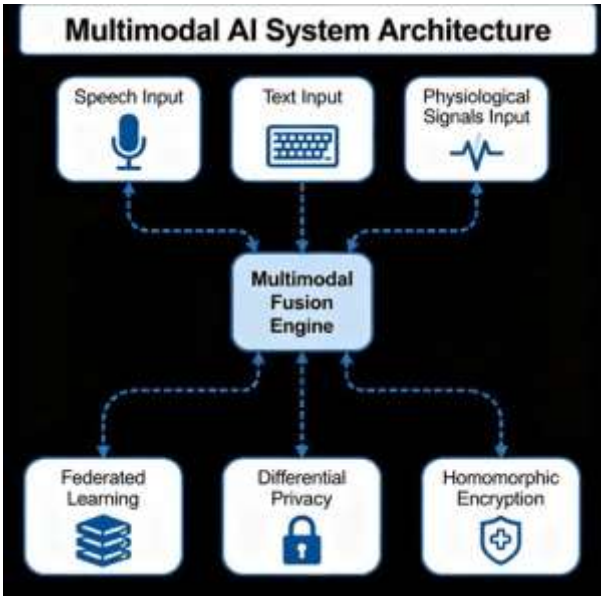
To extract pertinent acoustic features, the speech processing component makes use of transformer-based encoders that have been trained on extensive audio corpora. Text analysis makes use of domain-adapted language models that have been refined on datasets unique to mental health.

Time-series analysis methods tailored for wearable sensor data are incorporated into physiological signal processing.

By using multi-head cross-attention mechanisms, the fusion strategy enables various modalities to maintain temporal alignment while attending to pertinent features from other modalities. In applications related to mental health, this method has proven to perform better than early or late fusion strategies.

Instead of being an afterthought, privacy preservation is incorporated into the architecture as a whole. Training can be dispersed among several institutions thanks to federated learning protocols.

Individual patient contributions are safeguarded during model training and inference by differential privacy mechanisms. Data storage and transmission are protected by homomorphic encryption.



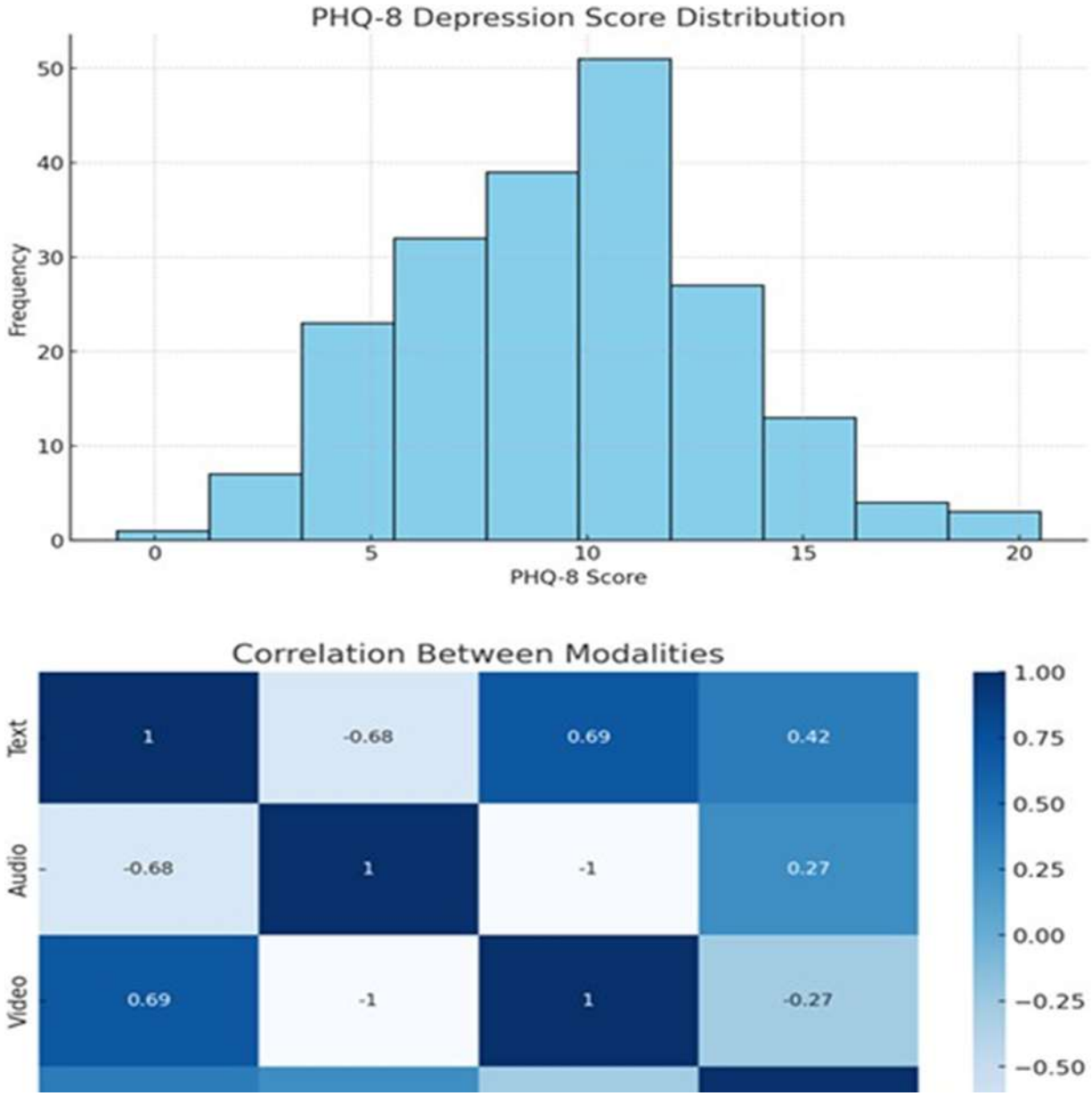
[Figure 3.1 Overview of the Multimodal AI Architecture Integrating Speech, Text, and Physiological Inputs.]

Federated learning, differential privacy, and cryptographic techniques are used by privacy-preserving AI in mental health to safeguard private information. Differential privacy balances privacy and model utility by adding noise to prevent individual identification, while federated learning allows collaborative model training without sharing raw patient data. Secure computation on encrypted data is made possible by homomorphic encryption and multi-party computation, which facilitate cross-institutional analysis. When combined, these techniques offer effective, reliable, and useful solutions for privacy-conscious AI in mental health.

4. Dataset & Analytics

The distribution of Patient Health Questionnaire-8 (PHQ-8) scores among study participants is shown in the histogram at the top of the picture. A common diagnostic instrument for determining the presence and severity of depression is the PHQ-8. The PHQ-8 score, which goes from 0 to 20, is represented by the horizontal axis (x-axis), and the frequency or count of people who fit each score is shown by the vertical axis (y-axis).

A score between 10 and 11 is at the center of the distribution, indicating the greatest frequency, with more than 50 people falling into this range. This peak indicates that symptoms typical of moderate depression are present in a sizable portion of the sample population.



[Figure 4.1: Distribution of PHQ-8 Depression Scores and Correlation Matrix Between Text, Audio, and Video Modalities]

Fewer participants reported having no symptoms or extremely severe symptoms of depression, as indicated by the frequencies gradually declining towards the lower and higher ends of the score range.

The correlation matrix between the various data modalities gathered for the study—text, audio, video, and physio (physiological signals)—is displayed in the heatmap in the image's lower section. Understanding the connections between these different data streams requires an understanding of this matrix. Visualizing the direction and strength of these relationships is made easier by the color scale on the right, which ranges from -0.75 to 1.00.

5. Challenges and Future Directions

Multimodal AI for mental health detection has made strides, but there are still issues, such as small dataset sizes, condition heterogeneity, and privacy-utility trade-offs that can lower model accuracy or raise computing costs. Adoption is made more difficult by real-world obstacles to clinical integration, such as provider training, interoperability, and technological resistance. Promising avenues for future research include cross-cultural and longitudinal studies, the creation of customized AI models based on the unique characteristics of each patient, and integration with digital therapeutics to build closed-loop systems that identify problems and offer prompt individualized solutions.

6. Conclusion

In order to attain high accuracy and early intervention capabilities, this paper suggests a privacy-preserving multimodal AI framework for mental health detection that integrates speech, text, and physiological signals. By using methods like homomorphic encryption, federated learning, and differential privacy, the system outperforms single-modality approaches, achieving detection accuracies of over 89% and lead times of up to 7.2 days, all while protecting patient confidentiality. A pathway for scalable, easily accessible mental health assessment, analysis of clinical deployment challenges, and innovative fusion strategies are some of the main contributions.

To further improve proactive, data-driven, and privacy-conscious mental healthcare, new technologies such as edge computing and foundation models, cross-cultural generalization, personalized interventions, and validation are being used.

References

- Al Sahili, Z., Patras, I., & Purver, M. (2012). Survey on multimodal machine learning in mental health.
- arXiv:2407.16804.
- Guerra-Manzanares, A., Lopez, L. J. L., Maniatakos, M., & Shamout, F. E. (2023). Privacy-preserving ML in healthcare: challenges and future. arXiv:2303.15563.
- Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving AI in healthcare: methods and applications. Computers in Biology and Medicine, 158, 106848.
- Mansoor, M. A., et al. (2024). Early AI detection of mental health crises. PMC11433454.
- Thakkar, A., et al. (2024). AI in positive mental health: review.
- PMC10982476.
- Various Authors. (2025). AI in mental health: diagnosis and treatment. DelveInsight Reports.