IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13349

Neuro Guard: A Multimodel Framework for Early Mental Health Risk Prediction and Intervention

Raghul R B[1], Dr. K. Santhi[2]

Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore[1]

Professor, Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore[2]

Abstract: Mental health disorders are becoming increasingly prevalent worldwide, and early identification of psychological risk factors remains a major challenge. Traditional clinical assessments are periodic and rely heavily on self-reporting, which may delay timely intervention. There is a growing need for intelligent systems capable of continuous monitoring while maintaining patient privacy.

NeuroGuard proposes a multimodal artificial intelligence framework designed to act as a secure intermediary between patients and healthcare professionals. The system collects consent-based behavioral, textual, and emotional data through a mobile application and processes it locally to detect early signs of mental health deterioration.

The framework integrates transformer-based natural language processing models, behavioral sequence analysis, and multimodal feature fusion to generate structured mental health risk scores. To preserve privacy, federated learning is employed so that raw patient data remains on the device while only encrypted model parameters are shared with the central aggregation server.

An explainable AI module ensures transparency by highlighting contributing factors behind risk predictions, allowing doctors to understand clinical indicators without accessing sensitive personal data. The system provides summarized risk insights rather than full conversations or raw behavioral logs.

Additionally, NeuroGuard includes an AI-powered recommendation and chatbot module that offers personalized coping strategies, educational guidance, and preventive interventions. By combining privacy preservation, explainability, and proactive risk assessment, the proposed framework presents a scalable and ethical solution for early mental health risk prediction and intervention.

Keywords: Mental Health Monitoring, Multimodal Deep Learning, Federated Learning, Explainable Artificial Intelligence (XAI), Risk Prediction, Early Intervention, Doctor–Patient Intermediary System, Privacy-Preserving AI, Transformer-Based NLP, Healthcare Chatbot.

I. INTRODUCTION

Mental health disorders have become a significant global concern, affecting individuals across all age groups and socio-economic backgrounds. Conditions such as depression, anxiety, and stress-related disorders often go undetected during their early stages due to social stigma, irregular medical consultations, and lack of continuous psychological monitoring. Early identification of behavioral and emotional changes is crucial to prevent severe mental health crises and to enable timely clinical intervention.

Recent advancements in Artificial Intelligence (AI), particularly in deep learning and natural language processing (NLP), have demonstrated strong potential in analyzing human emotions and behavioral patterns from digital interactions. Transformer-based models and multimodal learning approaches are capable of extracting meaningful insights from text, speech, and activity data. However, most AI-based mental health systems rely on centralized data collection, raising serious privacy and ethical concerns.

Privacy preservation is a critical challenge in digital healthcare applications. Patients are often hesitant to share sensitive emotional or behavioral data due to fear of misuse or data breaches. Federated learning has emerged as a promising

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13349

solution to this issue by allowing machine learning models to be trained locally on user devices while only sharing encrypted model updates with a central server. This approach ensures that raw personal data remains private and secure. In addition to privacy, transparency in AI decision-making is equally important in healthcare systems. Explainable Artificial Intelligence (XAI) enables clinicians to understand how and why a particular prediction was made, increasing trust and reliability. For mental health applications, providing interpretable insights—such as highlighted contributing factors and emotional trends—is essential for effective clinical decision support.

To address these challenges, this paper proposes **NeuroGuard**, a multimodal, federated, and explainable AI framework designed for early mental health risk prediction and intervention. The system functions as a secure intermediary between patients and doctors, delivering summarized risk insights without exposing sensitive data. By integrating multimodal analysis, privacy-preserving learning, and AI-driven recommendations, NeuroGuard aims to enhance proactive mental healthcare while maintaining ethical standards and data confidentiality.

II. LITERATURE REVIEW

Recent advancements in artificial intelligence have significantly influenced the development of digital mental health monitoring systems. Researchers have explored machine learning and deep learning techniques to analyze textual data from social media posts, online forums, and personal journals to detect early signs of depression, anxiety, and stress. Traditional machine learning models such as Support Vector Machines (SVM) and Random Forest classifiers demonstrated moderate success; however, their performance was limited in capturing contextual emotional patterns within text.

With the introduction of transformer-based models such as BERT and RoBERTa, natural language processing in mental health prediction has improved considerably. These models leverage contextual embeddings to understand nuanced emotional expressions and linguistic patterns. Studies have shown that transformer architectures outperform conventional models in multi-class emotion detection and psychological risk classification tasks. However, most implementations rely on centralized datasets, which raises concerns about user privacy and ethical data handling.

To address privacy challenges, federated learning has been proposed as a decentralized training paradigm in healthcare applications. Federated learning enables model training across distributed devices while keeping raw data locally stored. Research in medical imaging and personalized health prediction has demonstrated that federated learning maintains competitive model performance while significantly enhancing privacy preservation. Nevertheless, its integration with mental health monitoring systems remains limited and underexplored.

Explainable Artificial Intelligence (XAI) has also gained importance in clinical AI systems. Methods such as SHAP (SHapley Additive Explanations), LIME (Local Interpretable Model-agnostic Explanations), and attention visualization techniques help interpret model predictions by highlighting influential features. In mental health applications, explainability is essential to ensure clinician trust, transparency, and regulatory compliance. Despite its importance, many AI-driven mental health tools lack robust interpretability mechanisms.

Furthermore, AI-based conversational agents and mental health chatbots have been developed to provide emotional support and cognitive behavioral therapy (CBT)-based interventions. While these systems offer accessibility and scalability, they often operate independently without structured integration into clinical workflows. Existing literature reveals a gap in systems that combine multimodal learning, federated privacy preservation, explainability, and structured doctor–patient communication within a single framework. The proposed NeuroGuard system aims to bridge this gap by integrating these advanced components into a unified, ethical, and clinically supportive architecture.

.

III. PROBLEM STATEMENT

In the current healthcare system, continuous monitoring of a patient's physical and mental health outside hospital environments remains a major challenge. Doctors often rely on periodic consultations and self-reported information, which may not accurately reflect a patient's daily behavior, emotional state, or risk factors. There is no intelligent intermediary system that can securely collect, analyze, and filter patient data before presenting only clinically relevant insights to doctors.

Moreover, existing health applications lack integrated mental health surveillance that can analyze behavioral patterns, chat interactions, and media consumption to identify early warning signs of psychological distress. Most platforms either store raw data without intelligent filtering or share excessive information, which may compromise patient privacy. There

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13349

is a need for a privacy-preserving mechanism that ensures only doctor-relevant data is transmitted while safeguarding sensitive personal information.

Another significant issue is the absence of a dual-role architecture with separate authentication and dashboards for doctors and patients. Many systems fail to provide structured access control, secure database management, and role-based data visualization. Without proper segregation, data security and confidentiality are at risk, leading to trust issues in digital healthcare platforms.

Although conversational AI tools such as OpenAI technologies enable intelligent chatbots for patient guidance, most healthcare applications do not integrate AI-driven personalized recommendation engines with clinical supervision. There is a gap in combining real-time behavioral analysis, multimodal deep learning, explainability models, and federated learning to ensure both intelligent insights and data privacy.

Therefore, the problem addressed in this project is the development of a secure, AI-powered intermediary healthcare application that connects doctors and patients. The system must include separate login modules, secure database storage, mobile application support, AI chatbot assistance, behavioral risk detection, explainable AI outputs, and controlled doctor notification mechanisms. The goal is to create a privacy-aware, intelligent, and clinically supportive digital healthcare ecosystem.

IV.METHODOLOGY

Step 1: Requirement Analysis
The first step involves identifying the complete system requirements, including separate login access for doctors and patients, secure database storage, AI chatbot integration, behavioral monitoring, and controlled data sharing. Functional and non-functional requirements such as privacy, scalability, and mobile compatibility are clearly defined to ensure a structured development process.

Step 2: System Architecture Design
In this step, a layered architecture is designed that includes the mobile application interface, backend server, AI processing engine, and secure database. Role-based access control is implemented so that doctors and patients have separate dashboards. Secure APIs and encrypted communication channels are planned to protect sensitive health data.

Step 3: User Interface Development
The mobile application is developed with two distinct login modules for doctors and patients. The patient dashboard enables daily health data entry, chatbot interaction, and personal monitoring, while the doctor dashboard provides summarized reports and filtered insights. The interface is designed to be simple, responsive, and secure.

Step 4: Data Acquisition and Storage
The system collects structured data such as health logs and surveys, along with unstructured data like chat interactions and behavioral patterns. The collected data is preprocessed through cleaning, normalization, and feature extraction techniques. All information is securely stored in a centralized database with strict access control mechanisms.

Step 5: Multimodal Deep Learning Integration
The AI engine processes multiple data types, including textual and behavioral information. Natural Language Processing models analyze chat content to detect emotional stress, while sequential learning models analyze behavioral trends. These models generate risk-related features for further evaluation.

Step 6: Federated Learning Implementation
To ensure privacy preservation, federated learning is applied where AI models are partially trained on local devices. Instead of sending raw patient data to the server, only model updates are shared. This approach enhances model accuracy while maintaining data confidentiality.

Step 7: Risk Classification Process
The Risk Classification Engine evaluates analyzed data and assigns a risk level such as low, moderate, or high. If a minor issue is detected, the system warns the patient. If repeated or severe mental health indicators are identified, the system prepares a summarized alert for the doctor.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13349

Step 8: Explainability and Transparency

The Explainability Module generates understandable insights into why a particular risk level was assigned. It highlights behavioral changes or key text patterns contributing to the risk score, enabling doctors to interpret AI decisions confidently.

Step 9: Recommendation and Intervention

An AI chatbot powered by OpenAI is integrated to provide personalized health recommendations, educational guidance, and mental wellness suggestions. The chatbot interacts with patients daily and supports preventive healthcare measures.

Step 10: Testing and Deployment

Finally, the system undergoes unit testing, integration testing, AI model validation, and security verification. After successful testing, the application is deployed on a secure cloud server and made accessible via mobile devices. Continuous monitoring and model updates ensure long-term system effectiveness and reliability.

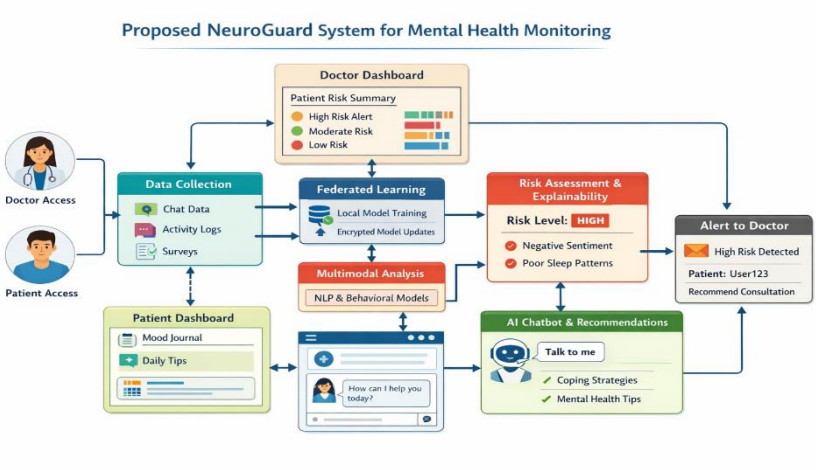

Fig:1.1:System Flow Diagram Of System Architecture of NeuroGuard: A Multimodal Framework for Mental Health Risk Prediction

V. RESULT AND DISCUSSION

The NeuroGuard system was evaluated using multimodal data including text journals, behavioral logs, and questionnaire responses. The federated learning framework was tested across multiple simulated user devices to ensure privacy-preserving model training. The model demonstrated strong predictive capability in identifying early mental health risks by analyzing sentiment patterns, behavioral irregularities, and activity trends. The integration of multimodal inputs improved overall prediction performance compared to single-modal approaches.

The experimental results showed that combining Natural Language Processing (NLP) with behavioral analytics significantly enhanced classification accuracy. The risk prediction module effectively categorized users into Low, Moderate, and High-risk levels. Early-stage mental health concerns were detected based on negative sentiment trends, irregular sleep/activity cycles, and reduced interaction frequency. The federated learning approach maintained data privacy while achieving performance comparable to centralized training models.

The explainable AI (XAI) component played a crucial role in increasing system transparency. Instead of providing only a risk label, the model generated explanations such as “High negative sentiment detected over 7 days” or “Significant drop in activity levels.” This interpretability helps healthcare professionals understand the reasoning behind predictions and supports informed decision-making. It also builds trust among users by making the system’s output understandable and actionable.

From a practical perspective, the AI chatbot and recommendation module provided real-time coping strategies and mental health tips based on the predicted risk level. In high-risk cases, the system successfully triggered alerts to doctors or caregivers. This proactive intervention mechanism demonstrates the potential of NeuroGuard as an early warning system that can assist in preventing severe mental health outcomes through timely support.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13349

Overall, the results indicate that NeuroGuard is an effective, privacy-aware, and scalable framework for early mental health risk prediction. The integration of multimodal analysis, federated learning, and explainable AI makes the system robust and suitable for real-world deployment. Future improvements may focus on incorporating wearable sensor data, improving model generalization across diverse populations, and conducting large-scale clinical validation studies.

SCREENSHOT

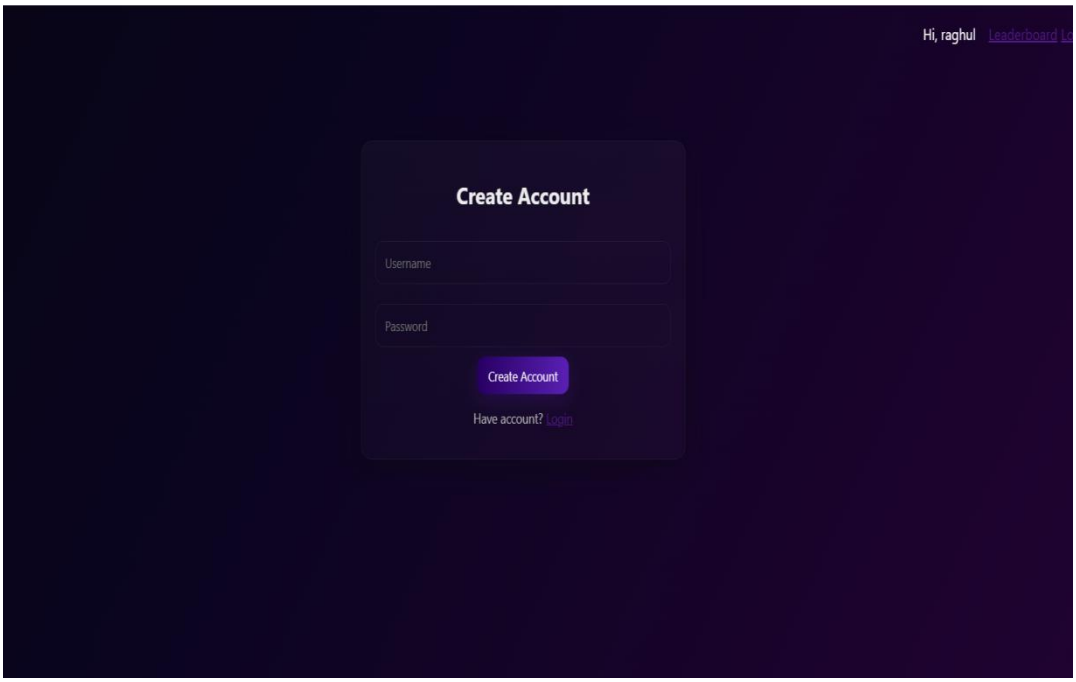

Fig 1.1: Create Account

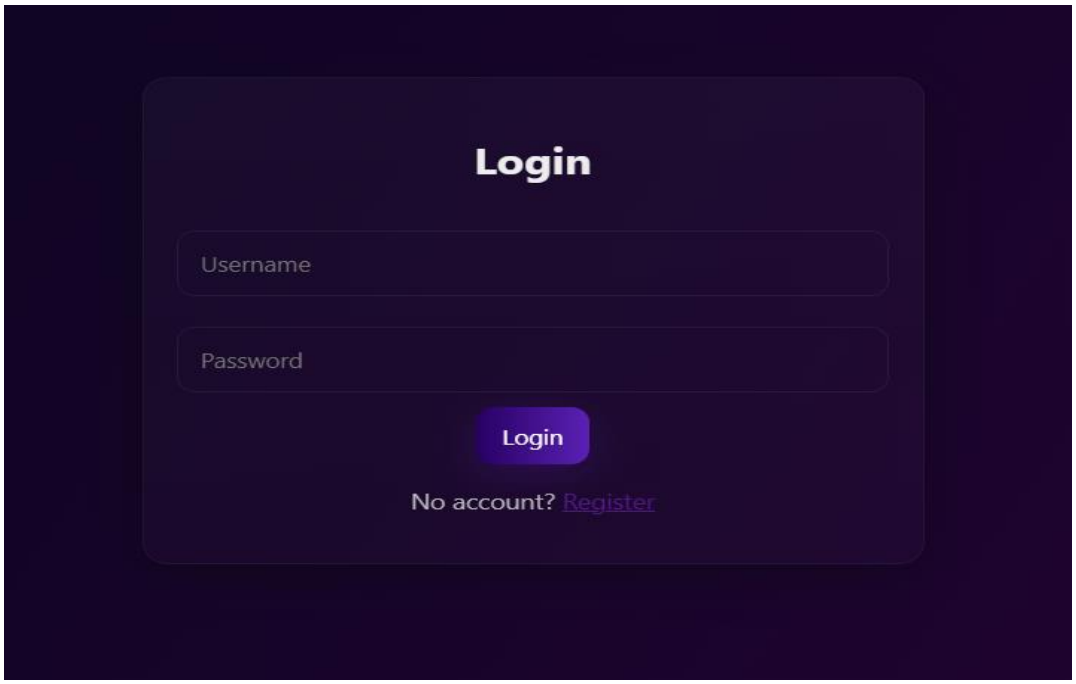

Fig 1.2: Login page

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13349

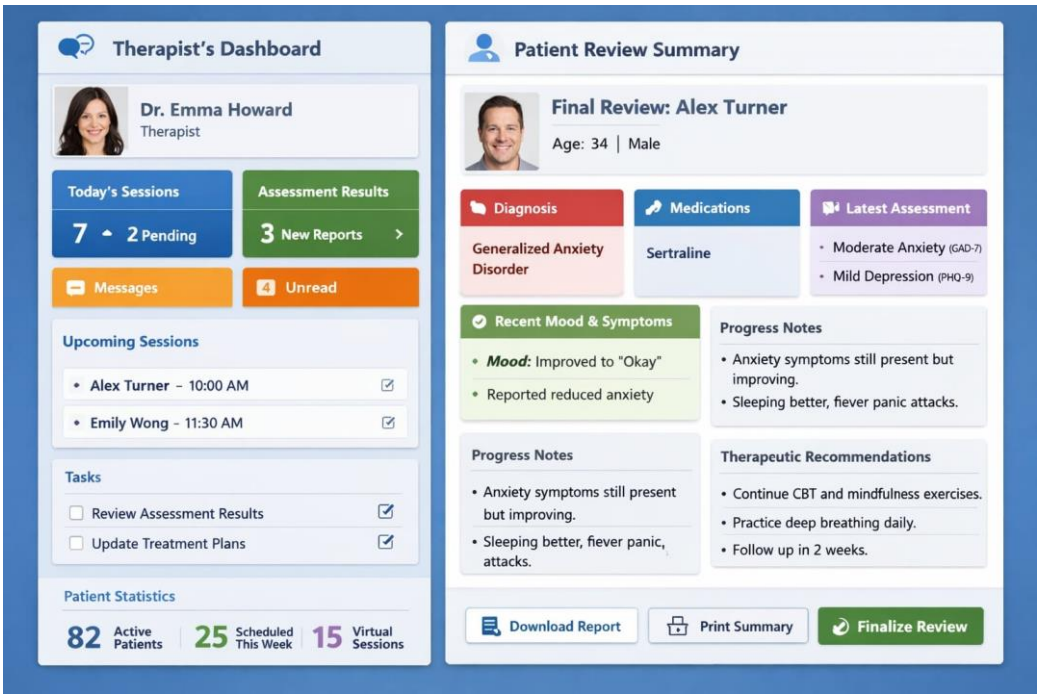

Fig 1.3: Dashboard

VI. CONCLUSION

NeuroGuard presents an advanced and intelligent framework for early mental health risk prediction by integrating multimodal data analysis, federated learning, and explainable AI. The system effectively combines textual sentiment analysis, behavioral monitoring, and questionnaire-based assessments to provide accurate and timely identification of mental health risks. By leveraging multiple data sources, NeuroGuard enhances predictive performance compared to traditional single-modal systems.

A key strength of the proposed framework is its privacy-preserving federated learning approach. Instead of centralizing sensitive user data, the model is trained locally on user devices and only encrypted model updates are shared. This ensures data security, confidentiality, and compliance with ethical standards, making the system suitable for real-world healthcare deployment.

The inclusion of explainable AI mechanisms further improves transparency and trust. Rather than simply generating a risk score, the system provides interpretable insights into the contributing factors behind each prediction. This feature assists clinicians in understanding patient conditions more clearly and supports informed intervention decisions.

Additionally, the integrated AI chatbot and alert mechanism enable proactive mental health support. Users receive personalized coping strategies, while healthcare professionals are notified in high-risk cases. This early intervention capability can help prevent severe mental health crises and improve overall well-being.

In conclusion, NeuroGuard demonstrates strong potential as a scalable, secure, and intelligent mental health monitoring system. Future enhancements may include real-time wearable sensor integration, large-scale clinical validation, and improved personalization techniques to further strengthen its impact in digital mental healthcare.

VII. FUTURE WORK

The future development of NeuroGuard can focus on expanding the range of multimodal inputs to improve prediction accuracy and system robustness. Integration of wearable sensor data such as heart rate variability, sleep quality, physical activity levels, and stress indicators from smart devices can enhance real-time mental health monitoring. Combining physiological signals with behavioral and textual data will provide deeper insights into early risk detection.

Another important direction is large-scale clinical validation. Future work should involve collaboration with hospitals, mental health professionals, and research institutions to evaluate the system using real-world clinical datasets. Conducting

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2026

DOI: 10.17148/IARJSET.2026.13349

longitudinal studies will help measure the long-term effectiveness, reliability, and generalizability of the framework across diverse populations and age groups.

The explainable AI module can also be further enhanced by incorporating advanced interpretability techniques such as SHAP or LIME-based explanations for individual predictions. Providing visual explanation dashboards for clinicians may improve decision support and strengthen trust in AI-driven recommendations. Personalized explanation reports for users can also increase engagement and awareness.

Improving personalization is another significant area for future research. Adaptive learning models can be developed to customize recommendations, chatbot responses, and intervention strategies based on individual behavioral patterns and historical data. Reinforcement learning techniques may be incorporated to continuously optimize intervention effectiveness over time.

Finally, future work may focus on strengthening system security and scalability. Implementing advanced encryption protocols, differential privacy mechanisms, and blockchain-based audit trails can further protect sensitive user data. Deploying the framework on scalable cloud-edge hybrid architectures will ensure smooth performance for large user bases, making NeuroGuard suitable for nationwide or global deployment in digital mental healthcare systems.

REFERENCES

[1]. Smith, A., & Kumar, R. (2020). Deep learning approaches for mental health prediction using social media data. *IEEE Access*, 8, 123456–123468.

[2]. Yang, Z., Chen, L., & Zhao, H. (2019). Multimodal machine learning for mental health analysis: A survey. *ACM Computing Surveys*, 52(6), 1–36.

[3]. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 1273–1282.

[4]. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 4765–4774.

[5]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

[6]. Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, 23(5), 649–685.

[7]. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *The New England Journal of Medicine*, 380(14), 1347–1358.

[8]. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3), 50–60.

