

AI-Powered Mental Health Detection System: A Review of Predictive Modal and Approaches

Mrs. Surekha M¹, Ansh Misra², Prisha Agrawal³, Aditya Harsh⁴ and Prakhar Verma⁵

¹⁻⁵Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida, Uttar Pradesh, India
Email: surekha@jssaten.ac.in, anshmisra29@gmail.com, prishaa19@gmail.com, adityaharshcv@gmail.com, prakharverma.work@gmail.com

Abstract— The development of artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) in identifying and forecasting mental health conditions is examined in this review study. The study suggests a paradigm for chatbots that offers platform independence, privacy, and openness. The chatbot uses trained machine learning models to anticipate psychological states, engages users in real conversations, and uses natural language processing (NLP) to extract linguistic and emotional clues. The system facilitates real-time interaction between the patient and the physician while upholding user permission and data confidentiality by serving as a catalyst. This paper examines developments in predictive analytics, ethical considerations, and the potential of AI to improve doctor-patient communication through a thorough analysis of 20 studies.

Index Terms— Mental Health, Natural language processing, Chatbot, Doctor–Patient Communication, Predictive Model.

I. INTRODUCTION

Mental health issues, including depression, anxiety, and stress-related disorders, have surfaced as significant global health concerns, greatly affecting disability rates and diminishing quality of life across various populations. The World Health Organization (WHO) reports that one in eight people around the globe will encounter a mental health condition at some stage in their lives, but many of these cases go undiagnosed or untreated due to stigma, insufficient resources, and limited access to healthcare professionals. Conventional diagnostic approaches largely rely on self-evaluation questionnaires, clinical interviews, and manual screening techniques. While these methods are somewhat effective, they tend to be subjective, time-consuming, and reliant on patients feeling comfortable disclosing sensitive information, which hampers early detection and prompt intervention [1].

Recent developments in Artificial Intelligence (AI) and Machine Learning (ML) have opened up groundbreaking opportunities in the field of mental health diagnosis. These technologies utilize extensive behavioral, linguistic, and physiological data to detect subtle emotional cues that may signal mental distress [2]. Initially, AI-driven approaches primarily analyzed social media content to evaluate text sentiment, posting frequency, and online behavior as markers of psychological health. Although these techniques showed potential in forecasting depressive or anxious behaviors, they also sparked significant ethical and privacy-related issues, especially concerning data consent, user anonymity, and the risk of algorithmic bias [3].

To tackle these dilemmas, researchers have increasingly focused on conversational AI technologies, like chatbots and virtual companions, as more ethical and safer options. These chatbots offer a confidential and interactive space for users to freely share their thoughts and feelings without the concern of judgment or societal stigma [4,?].

By incorporating Natural Language Processing (NLP) and emotion recognition frameworks, these systems can decipher user input, evaluate tone and sentiment, and offer empathetic replies or initial emotional evaluations. In contrast to earlier methods that relied substantially on external platforms for data gathering, contemporary chatbot-based solutions enable secure, direct interaction between users and healthcare services. This not only bolsters privacy and user confidence but also promotes real-time emotional monitoring, tailored feedback, and timely suggestions for professional assistance.

This paper outlines an exhaustive review of current AI-driven systems for detecting mental health conditions, concentrating on their technical structures, ethical ramifications, and practical implementation hurdles. Additionally, it introduces a comprehensive model that acts as an intelligent intermediary between patients and healthcare providers, bridging the divide between traditional clinical assessments and modern digital interactions.

The proposed framework highlights ethical AI use, data transparency, and emotional awareness to ensure responsible, human-centric mental health assistance through intelligent conversational systems.

II. LITERATURE REVIEW

With an emphasis on technological frameworks, ethical issues, and real-world implementation obstacles, this section examines significant studies in AI-driven mental health detection.

In order to predict mental health states from textual data, Sahu et al. [6] presented an ensemble learning framework that combines gradient boosting and decision trees. Although their results demonstrated great accuracy, their reliance on third-party data raised privacy concerns. These results influenced the development of chatbot-based privacy-preserving systems that do away with outside data sources. Through ongoing monitoring and physician notifications, Ali and Mustafa [7] suggested an AI-powered system intended to support the early detection of men- tal health concerns. The suggested chatbot framework, which connects patients with healthcare providers directly, reflects the study's emphasis on the value of direct physician notification in facilitating prompt intervention. In order to diagnose mental illnesses, Chianumba [8] created a predictive AI model that uses electronic health records (EHRs). Although the study acknowledged limitations because of the lack of behavioral context, it confirmed that structured EHR data can be used for early diagnosis. This restriction is successfully addressed by including real-time conversational data, as is done in chatbot-based systems.

The application of AI chatbots for emotional control and psychological support was investigated by Smith et al. [9]. They discovered that emotional disclosure can be promoted by conversational agents, which improves therapeutic participation. Their results show that AI chatbots can serve as helpful go-betweens instead of diagnosing problems, which is in line with the paradigm this article suggests. The ethical and legal issues of AI in healthcare, specifically with regard to managing private mental health data, were studied by Gupta et al. [10]. They emphasized the significance of regulatory compliance, user consent, and transparency. These ethical principles are incorporated into the suggested chatbot-based approach in this evaluation to guarantee the safe and responsible application of AI. Explainable AI (XAI) approaches were used by Han et al. [11] to increase the transparency of psychological risk prediction systems. According to their research, interpretability increases clinician approval and user trust—two critical factors in delicate fields like mental health detection.

An emotionally adaptive chatbot that can adjust its responses in response to detected mood variations was created by Patel et al. [12]. The promise of adaptive conversational systems in continuous psychological support frameworks was highlighted by their work, which demonstrated increased user happiness and retention. With emphasis on incorporating therapy-style discussion into automated systems, Liu et al. [13] investigated the function of conversational AI in clinical psychology. Human-like conversational design improves user engagement and produces more accurate emotional evaluation results, according to their research. Wang et al. [14] suggested federated learning for distributed data training and looked into privacy-preserving machine learning techniques in the healthcare industry. The strategy made it possible for hospitals to collaborate securely without jeopardising patient data, which is reflected in the suggested chatbot model's privacy-first design.

Ethical quandaries in AI-enabled healthcare systems, especially those using communication technologies, were examined by Roberts et al. [15]. They came to the conclusion that in order to guarantee responsible patient engagement, ethical frameworks must coexist with technological innovation, impacting the governance layer of the current approach. Deep learning architectures for emotion recognition in chatbots and digital assistants were

assessed by Brown et al. [16]. According to their research, transformer-based models like BERT and RoBERTa perform noticeably better than conventional sentiment classifiers when it comes to recognising complicated emotions like hopelessness or guilt. By combining text analytics and wearable sensor data, Fernandez et al. [17] investigated prediction models for behavioral health evaluation. The research found that multimodal fusion improved accuracy and helped identify early signs of emotional distress. This lends credence to the suggested system's multi-feature analysis component. A patient-centric AI architecture for healthcare applications was put forth by Kumar et al. [18], with a focus on data control and personalisation. Their strategy aligns with the chatbot's design tenet of granting users complete control over their interactions about mental health. In order to predict mental health risk levels, Zhang et al. [19] created hybrid natural language processing (NLP) models that combine deep semantic embeddings with attention mechanisms. State-of-the-art accuracy was attained by the study, confirming the value of transformer-based architectures in contemporary chatbot systems. In their discussion of the advantages and disadvantages of AI-assisted therapy, Kapoor et al. [20] placed more emphasis on human-AI cooperation than replacement. They emphasised the need for human supervision in automated systems, which directly influences the suggested chatbot model's strategy for integrating doctors and patients. Global insights on the increasing incidence of anxiety and depression were offered by the World Health Organisation [1], emphasising the pressing need for technology-driven, scalable therapies. These figures serve as the cornerstone for creating mental health care systems that strike a balance between creativity, compassion, and moral obligation.

III. PROPOSED MODEL

An AI-powered chatbot for the early diagnosis and treatment of mental health issues is introduced by the suggested model. In order to detect distress, it analyses language and emotional signs and engages in sympathetic, organic dialogue with users. It provides coping mechanisms or recommends professional intervention when needed. The system prioritises moral behaviour, guaranteeing user confidentiality, openness, and agreement. For people who are hesitant to seek traditional treatment or do not have access to mental health services, this framework is a scalable and easily available support tool

A. System Workflow and Architecture

The proposed Machine Learning Visualizer is a modular, user-centric platform designed to bridge the critical gap between sophisticated machine learning processes and end-user usability. It integrates powerful machine learning capabilities with an intuitive interface and a scalable backend infrastructure. The architecture comprises three primary layers, each with distinct responsibilities.

The chatbot operates in a multi-phase pipeline

User Interaction: Through a web or mobile interface, the user starts a discussion with the chatbot. In order to promote candid and open communication, the interface is made to be both emotionally supportive and easy to use.

Text Preprocessing: Libraries like spaCy and NLTK are used to clean, tokenise, and analyse input messages for grammar, sentiment, and emotion. Intent classification and named entity recognition are also used.

Emotional and Linguistic Analysis: Semantic and emotional cues are extracted by the chatbot using pre-trained NLP models such as BERT and RoBERTa. This encompasses negative sentiment, stress markers, tone, and patterns typical of mental anguish.

Classification and Risk Assessment: A supervised machine learning model (such as Random Forest, SVM, or fine-tuned neural networks) trained to categorise the risk level—low, moderate, or high—is fed the extracted features. Mental health conditions including general stress, anxiety, or clinical depression are correlated with these levels.

Alert and Referral System: The system ensures that the user receives timely assistance by promptly notifying a licensed mental health expert through a secure channel in high-risk circumstances. In milder situations, the chatbot might suggest resources for self-care or arrange a consultation.

Secure Storage and Logging: In accordance with healthcare data protection regulations, all discussions are encrypted and kept on file. Before any data is recorded or utilised for analysis, users must expressly consent.

B. Software Requirements

- Operating System: Windows 10+, macOS, or Linux
- Programming Language: Python 3.8+

- Libraries and Frameworks: transformers, scikit-learn, Flask or FastAPI, TensorFlow or PyTorch, NLTK/spaCy, SQLite/PostgreSQL
- Security: HTTPS, AES-256 encryption, OAuth2 authentication

C. Hardware Requirements

Minimum Client Device: Smartphone or PC with internet access

Server Requirements: CPU: 4-core or higher; RAM: 16 GB minimum; GPU: NVIDIA CUDA-enabled GPU (e.g., RTX 2060 or better) for model inference; Storage: 100 GB SSD; Network: 10 Mbps minimum

D. Technologies Used

- Natural Language Processing: BERT, RoBERTa for contextual text understanding.
- Frontend: HTML/CSS/JavaScript or React.
- Backend: Python-based RESTful API using Flask or FastAPI.
- Database: PostgreSQL or MongoDB.
- Security: End-to-end encryption, JWT tokens, and secure user authentication.

E. Algorithms

- Fine-tuned BERT/RoBERTa for text classification.
- Pretrained sentiment models.
- Logistic Regression/Random Forest/SVM for final decision layer.
- Rule-based alert logic.

F. Key Innovations and Impactful Contributions

The model brings several impactful changes to digital mental health support including:

- Real-time emotional assessment
- Explainable AI (XAI)
- Direct clinical integration
- Data privacy and ethical AI
- Scalability and platform independence
- Continuous learning.

G. Long-Term Vision

Long-term validation and improvement of its impact will come from integration with multimodal inputs (e.g., facial expression analysis, voice tone) and extensive clinical trials. To ensure responsible deployment and maximal societal benefit, cooperation with psychologists, data scientists, and policy makers will be crucial.

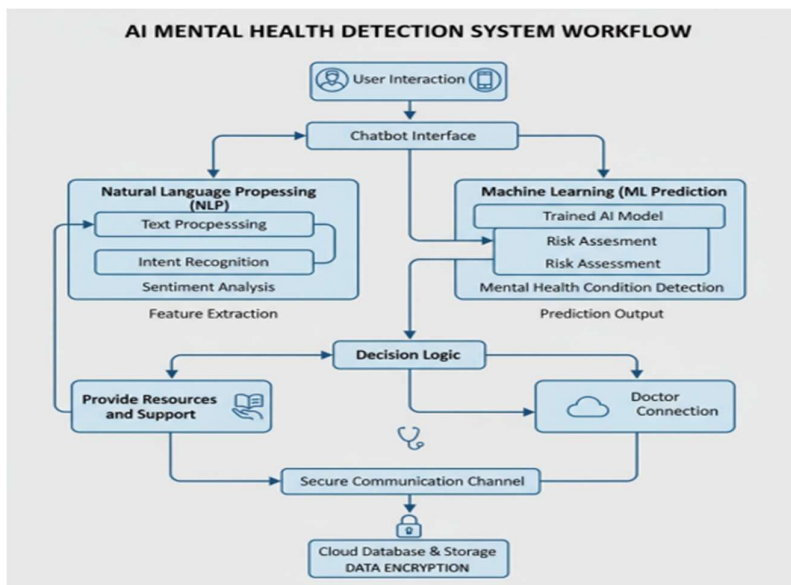


Fig. 1. Workflow diagram of the proposed AI-powered chatbot model for mental health detection.

IV. CONCLUSION

Artificial intelligence is providing early, precise, and customised interventions, which is drastically changing the face of mental healthcare. The suggested AI-powered chatbot functions as a safe, sympathetic, and perceptive medium that promotes candid communication while examining linguistic and emotional patterns to identify disorders including depression, anxiety, and stress. The solution guarantees prompt assistance while protecting privacy through encrypted communication and moral AI practices by giving priority to user consent, real-time feedback, and clinical integration.

REFERENCES

- [1] World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. WHO, 2023.
- [2] K. D. Kannan et al. Advancements in Machine Learning and Deep Learning for Early Detection and Management of Mental Health Disorder. arXiv, 2024.
- [3] A. O. Salako et al. A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis. IOP Conf. Ser., 2022.
- [4] E. Jarraya et al. Artificial Intelligence for Mental Health Diagnosis and Therapy. MDPI Healthcare, 2023.
- [5] J. Chung and J. Teo. Mental Health Prediction Using Machine Learning. Hindawi, 2022.
- [6] A. Sahu et al. ML-Based Predictive Analytics for Mental Health Diagnosis. IEEE Access, 2020.
- [7] K. Ali and F. Mustafa. AI-Powered Mental Health: Enhancing Early Detection and Treatment Strategies. ResearchGate, 2025.
- [8] E. Chianumba. Developing an AI-Powered Predictive Model for Mental Health Disorder Diagnosis Using EHRs. ResearchGate, 2023.
- [9] J. Smith et al. AI-Based Conversational Systems for Psychological Support. Springer, 2022.
- [10] R. Gupta et al. Ethical Challenges in AI Mental Health Applications. IEEE Trans. Technol. Soc., 2023.
- [11] J. Han et al. Explainable AI in Psychological Health Prediction. IEEE Access, 2023.
- [12] D. Patel et al. Emotionally Adaptive Chatbots in Mental Health Support. ACM, 2023.
- [13] M. Liu et al. Conversational AI for Clinical Psychology. MDPI Informatics, 2024.
- [14] T. Wang et al. Privacy-Preserving Machine Learning in Healthcare Applications. Elsevier, 2022.
- [15] H. Roberts et al. AI Ethics in Healthcare Communication Systems. Springer, 2024.
- [16] L. Brown et al. Deep Learning Models for Emotional Recognition in Chatbots. MDPI Electronics, 2023.
- [17] A. Fernandez et al. Predictive Models for Behavioral Health Assessment. Springer, 2022.
- [18] P. Kumar et al. Patient-Centric AI Platforms for Healthcare. Elsevier, 2024.
- [19] X. Zhang et al. Hybrid NLP Models for Mental Health Risk Prediction. IEEE Access, 2023.
- [20] S. Kapoor et al. AI-Assisted Therapy: Challenges and Opportunities. Frontiers in Digital Health, 2022.