Review

ANNALS OF MEDICINE & SURGERY

OPEN

Artificial intelligence in mental health: integrating opportunities and challenges of multimodal deep learning for mental disorder prevention and treatment

Mohammad Narimani, PhD, Mahdi Naeim, MA*

Background: Artificial intelligence (AI), through multimodal deep learning and predictive analytics, holds transformative potential in the prevention and treatment of mental disorders. This study explores the opportunities and challenges of these technologies.
Objective: To present a conceptual framework for the responsible application of AI in mental health care.
Methods: This integrative review analyzed selected sources from Google Scholar up to June 2025. Both qualitative and quantitative analyses were conducted to identify opportunities and challenges.
Results: Key opportunities include early detection, personalized treatment, and enhanced access to mental health services. Major challenges involve ethical concerns, algorithmic bias, and data quality issues.
Conclusion: AI can revolutionize mental health care, but it requires standardization and regulatory oversight. Future research should focus on addressing ethical dilemmas and improving data quality.

Keywords: artificial intelligence, multimodal deep learning, predictive analytics, mental health, mental disorders, ethical issues

Introduction

Mental disorders including depression, anxiety, and schizophrenia have significantly increased the global burden of disease, accounting for approximately 32% of years lived with disability[1]. With the rising demand for mental health services and a shortage of human resources, artificial intelligence (AI) technologies have emerged as innovative tools to enhance prevention and treatment[2]. Specifically, AI techniques such as multimodal deep learning (MDL) and predictive analytics enable the analysis of diverse data types including text, images, and audio creating opportunities for early detection, personalized care, and broader service access[3,4].

MDL utilizes neural networks such as Convolutional Neural Networks (CNNs) and transformers to process heterogeneous data sources (e.g. brain MRI scans, vocal patterns, and social media posts)[5]. These technologies have demonstrated high accuracy in detecting early symptoms of mental disorders like depression and anxiety[6]. For instance, natural language processing (NLP) algorithms can identify depression-related speech patterns in conversations or social media content[7]. Additionally, predictive analytics using longitudinal data such as medical histories and wearable sensor outputs can forecast relapse risk or treatment response[8].

One of the most promising applications of AI is the delivery of digital interventions through therapeutic chatbots like ChatGPT, which provides psychotherapeutic services in underserved regions[9]. Leveraging NLP and machine learning, these tools offer cognitive behavioral therapy (CBT)-based counseling and help bridge the mental health access gap[10]. Furthermore, AI can analyze big data from diverse sources, including social networks and wearable devices, to detect behavioral and psychological patterns[11]. These capabilities are particularly valuable in suicide prevention and managing severe conditions like schizophrenia[12].

However, the integration of AI into mental health care comes with significant challenges. Ethical concerns – such as data privacy and informed consent – remain unresolved[13]. Non-representative training datasets may lead to algorithmic bias, increasing the risk of misdiagnosis in diverse populations[14]. Moreover, the lack of data standardization and the low quality

HIGHLIGHTS

- Artificial intelligence (AI)-powered multimodal deep learning enables early detection and personalized treatment of mental disorders.
- This review proposes a conceptual framework for the ethical integration of AI into mental health care.
- Key barriers include algorithmic bias, low-quality data, and unresolved ethical dilemmas.
- AI tools like chatbots and predictive analytics show promise in expanding access to underserved populations.

Department of Psychology, University of Mohaghegh Ardabili, Ardabil, Iran

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

**Corresponding author. Address: Department of Psychology, University of Mohaghegh Ardabili, Ardabil 5613644447, Iran. Tel.: +989144514810. E-mail: mnaeim.psy@gmail.com (N. Mahdi).*

Annals of Medicine & Surgery (2025) 87:5757–5761

Received 21 April 2024; Accepted 4 July 2025

Published online 22 July 2025

http://dx.doi.org/10.1097/MS9.0000000000003624

of input data such as electronic health records (EHRs) can hinder AI model performance[15]. In addition, the opacity of complex AI models, often referred to as the "black box" problem, undermines the trust of clinicians and patients[16].

While previous studies have explored individual applications of AI in mental health, they often lack a comprehensive and interdisciplinary perspective that integrates technological, clinical, and ethical dimensions. This article contributes uniquely to the existing literature by: (1) synthesizing two critical domains MDL and Predictive Analytics in the context of mental health prevention and treatment; (2) proposing a conceptual framework for the responsible integration of AI technologies into mental health systems; and (3) presenting a structured, multi-dimensional analysis using comparative tables, data flow diagrams, and a SWOT framework tools rarely combined in prior research.

The educational value of this study lies in three core areas: first, it offers an in-depth understanding of AI's potential to enhance early diagnosis, personalize treatment, and expand access to care; second, it systematically identifies ethical and technical challenges such as algorithmic bias, data quality limitations, and the opacity of AI models; and third, it provides a practical and theoretical foundation for policymakers, clinicians, and researchers seeking to implement AI-driven tools in an evidence-based, ethically sound, and contextually sensitive manner. By bridging the gap between data science and clinical psychology, this work advances a more holistic vision for the future of AI in mental health.

Method

This study employed an integrative review approach to examine the opportunities and challenges of applying AI in the prevention and treatment of mental disorders. The integrative review, by enabling the combination of qualitative and quantitative data, is considered an appropriate approach for analyzing complex and multilayered issues such as the role of advanced technologies in mental health[17]. Unlike systematic reviews, this method offers greater flexibility in the selection of sources and allows for a broader examination of the literature.

Data collection

Data were collected from reputable scientific databases, including Google Scholar, PubMed, and Scopus, up to June 2025. Inclusion criteria consisted of English-language articles related to the application of AI in the field of mental health, with a focus on emerging technologies such as MDL including CNNs and Transformer models as well as predictive analytics such as random forests and support vector machines. The keywords used were: "Artificial Intelligence," "Mental Health," "MDL," "Predictive Analytics," and "Ethical Challenges." Articles that lacked relevance to mental health or did not focus on advanced technologies were excluded from the review.

Data analysis process

Data were analyzed at two levels:

1. Qualitative analysis: Core themes were extracted through qualitative content analysis. This process included initial coding, identification of patterns, and categorization of key concepts within the study's thematic framework.
2. Quantitative analysis: From the selected studies, quantitative data related to the performance of AI models (such as accuracy, sensitivity, and specificity) were extracted and analyzed to evaluate the effectiveness of these technologies in diagnosis and treatment.

Data presentation

To structure and present the results in an organized manner, tools such as comparative tables, data flow diagrams, and SWOT analysis were designed and prepared for the reporting stage. These tools were selected to facilitate a clearer explanation of the various dimensions of the findings.

Validation

To enhance the credibility and consistency of the analysis, sources were selected from articles published in reputable ISI-indexed journals. Furthermore, the data analysis process involved repeated theme reviews and cross-validation among research team members.

Among the limitations of this study are the lack of adherence to systematic review protocols such as PRISMA and the restriction of the source pool to selected articles, which may have resulted in the omission of some relevant studies.

Let me know if you'd like help designing the data flow diagrams, structuring your SWOT analysis, or finding real PRISMA-compliant alternatives for future reviews.

Results

Key opportunities

AI presents several significant opportunities for the prevention and treatment of mental disorders. Table 1 summarizes these opportunities.

Early detection

MDL models such as CNNs and transformer architectures analyze diverse datasets (e.g. MRI images, vocal patterns, and social media text) to identify early signs of depression, anxiety, or post-traumatic stress disorder[5]. For instance, one study found that NLP models analyzing Twitter posts achieved 89% accuracy in detecting depression[7].

Table 1

Key opportunities of AI in mental health

Domain	Application	Example
Early detection	Multimodal data analysis	Detecting depression using CNN and NLP[4]
Personalized treatment	Predicting treatment response	Using random forest for depression prognosis[7]
Improved accessibility	Therapeutic chatbots	ChatGPT for managing anxiety symptoms[8]

Personalized treatment

Predictive analytics using longitudinal data (e.g. medical records and wearable sensor outputs) can anticipate patients' responses to pharmacological or CBT[8]. This approach enables clinicians to tailor treatments based on individual patient profiles.

Improved accessibility

AI-driven chatbots provide mental health interventions in underserved regions[9]. These tools are available 24/7 and help reduce treatment costs[10].

Key challenges

The key challenges of integrating AI into mental health care are summarized in Table 2.

Ethical concerns

Issues related to data privacy and patients' informed consent remain unresolved[13]. For example, using social media data without explicit consent raises serious legal and ethical questions.

Algorithmic bias

Non-representative training datasets can lead to misdiagnoses, particularly among minority populations[14]. A study found that AI models demonstrated reduced diagnostic accuracy for depression in certain ethnic groups[18].

Data quality

Lack of standardization in input data such as EHRs can limit AI performance[15]. Incomplete or inconsistent datasets may lead to diagnostic errors.

Figure 1 illustrates the data processing workflow in multimodal AI systems. Input data includes text (e.g. social media posts), images (e.g. brain MRIs), and audio (e.g. speech patterns), which are analyzed by deep learning models such as CNNs and transformers. The outcome comprises diagnostic outputs (e.g. depression detection) or therapeutic recommendations (e.g. personalized intervention suggestions).

Table 2

Key challenges of AI in mental health

Challenge	Description	Example
Ethical concerns	Data privacy	Use of data without informed consent[12]
Algorithmic bias	Misdiagnosis in minorities	Underrepresentation in training data[13]
Data quality	Lack of standardization	Incomplete or inconsistent EHR data[14]

SWOT analysis

The diagram below presents a SWOT analysis of the opportunities and challenges related to the application of AI in mental health.

Figure 2 presents a SWOT analysis: Strengths, Weaknesses, Opportunities, and Threats based on the findings of the study. The identified strengths include high diagnostic accuracy[7] and improved accessibility[9]; weaknesses consist of algorithmic biases[14] and data quality issues[14]; opportunities involve the development of explainable AI (XAI)[17] and data standardization[19]; and threats include ethical concerns[13] and model opacity[16]. These factors are prioritized based on their relative importance and supported by both qualitative and quantitative evidence from the text.

Discussion

The findings indicate that AI, particularly MDL, can significantly enhance diagnostic accuracy and treatment effectiveness in mental health[5]. For instance, NLP models can identify symptoms of depression with high precision by analyzing speech patterns[7]. Such capabilities are particularly valuable in regions with limited access to psychologists[9]. However, ethical and technical challenges hinder the widespread adoption of AI technologies.

Data privacy is one of the primary concerns. Using sensitive data such as social media posts without explicit consent can lead to patient rights violations[13]. International regulations, such as the GDPR, emphasize the need for informed consent, yet implementation in complex AI systems remains challenging[20]. Additionally, the opacity of AI models (i.e. the "black box"

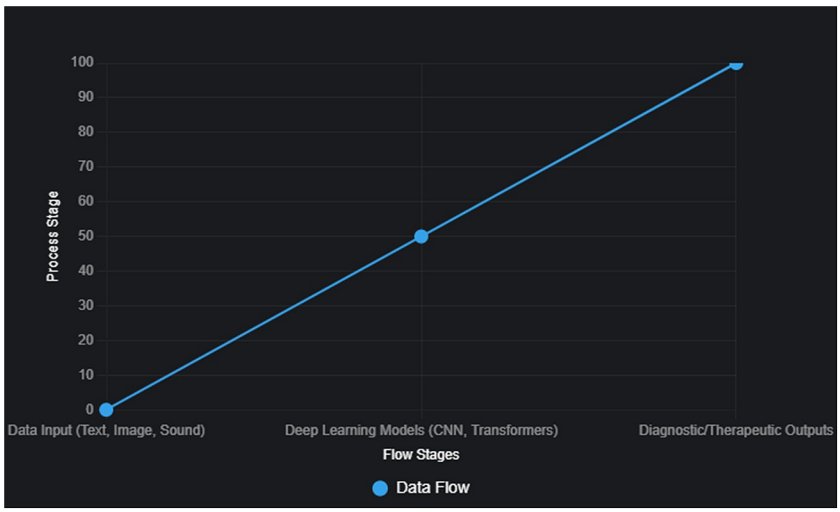

Figure 1. Data flow in multimodal artificial intelligence systems.

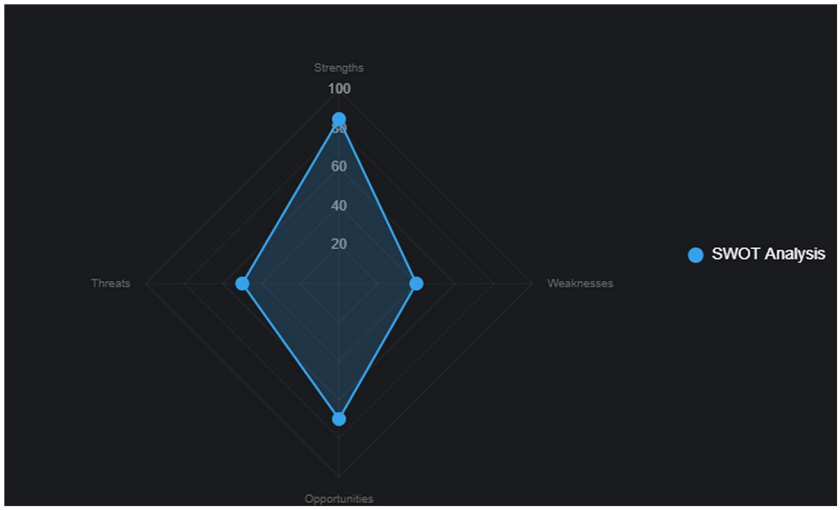

Figure 2. SWOT analysis of AI applications in mental health.

problem) undermines trust among clinicians and patients[16]. XAI may offer a solution, though it is still in the early stages of development[17].

Non-representative training datasets can lead to misdiagnoses, especially among minority populations[14]. One study found that AI models showed up to 20% lower accuracy in diagnosing depression among certain ethnic groups[18]. This issue calls for the inclusion of more diverse datasets and the development of bias-resilient algorithms.

A major obstacle is the lack of standardization in input data such as EHRs[15]. Incomplete or inconsistent data can lead to diagnostic errors. Proposed solutions include the creation of global standards for mental health data and the use of data-cleaning techniques[19].

Comparison with previous studies reveals that recent advances in MDL have enhanced diagnostic accuracy[5], yet ethical challenges remain unresolved[13]. This study proposes that interdisciplinary collaboration among data scientists, psychologists, and policymakers is essential to develop standard protocols and ensure ethical oversight in AI applications. Furthermore, investments in XAI could help increase transparency and trust[17].

Conclusion

Advanced AI, particularly MDL and predictive analytics, offers promising opportunities for the prevention and treatment of mental health disorders. Early diagnosis, personalized treatment, and improved access to care through AI-driven therapeutic chatbots illustrate the transformative potential of AI in mental health. However, challenges such as ethical concerns, algorithmic bias, and data quality limitations continue to hinder the widespread adoption of these technologies. This study proposes a conceptual framework for the responsible integration of AI into mental health systems, emphasizing data standardization, the development of XAI, and ongoing ethical oversight. Future research should focus on real-world evaluations of AI models, the development of standardized protocols for mental health data, and the resolution of ethical dilemmas. Given the rising demand for mental health services, AI has the potential to play a pivotal role in reducing the global burden of psychiatric disorders provided that existing challenges are effectively addressed.

Ethical approval

Ethics approval was not required for this review.

Consent

Informed consent was not required for this review.

Sources of funding

The current study received no funding.

Author contributions

All authors have been involved in the writing of the article.

Conflicts of interest disclosure

No conflicts of interest of all authors in this study.

Research registration unique identifying number (UIN)

This is a review article. Therefore, it is not an interventional study and does not require research registration or a unique identifying number (UIN). According to institutional regulations, no special permit is needed for this type of article.

Guarantor

All authors.

Provenance and peer review

Not commissioned, externally peer reviewed.

Data availability statement

Data are available from authors on request. The data that support the findings of this study are available on request from the corresponding author.

Acknowledgements

None.

References

[1] Kieling C, Buchweitz C, Caye A, *et al*. Worldwide prevalence and disability from mental disorders across childhood and adolescence: evidence from the global burden of disease study. JAMA Psychiatry 2024;81:347–56.
[2] Alhuwaydi AM. Exploring the role of artificial intelligence in mental healthcare: current trends and future directions–a narrative review for a comprehensive insight. Risk Manag Healthc Policy.2024;17:1339–48.
[3] Razavi M, Ziyadidegan S, Mahmoudzadeh A, *et al*. Machine learning, deep learning, and data preprocessing techniques for detecting, predicting, and monitoring stress and stress-related mental disorders: scoping review. JMIR Ment Health 2024;11:e53714.
[4] Agha RA, Mathew G, Rashid R, *et al*. Transparency in the reporting of Artificial INtelligence – the TITAN guideline. Prem J Sci 2025;10:100082.
[5] Esteva A, Robicquet A, Ramsundar B, *et al*. A guide to deep learning in healthcare. Nat Med 2019;25:24–29.
[6] Martínez G, Trujillo A, Núñez J, *et al*. Enhancing clinical psychology practice through data-driven machine learning monitoring systems.
[7] Rai S, Stade EC, Giorgi S, *et al*. Key language markers of depression on social media depend on race. Proc National Acad Sci 2024;121:e2319837121.
[8] Chekroud AM, Zotti RJ, Shehzad Z, *et al*. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 2016;3:243–50.
[9] Chandra P, Joshi G, and Bhagwat R. Chatgpt's evolution in reshaping cognitive behavioral therapy. IEEE Engineering Informatics Society; 2023: 1–9.
[10] Manole A, Cârciumaru R, Brînzaş R, *et al*. An exploratory investigation of chatbot applications in anxiety management: a focus on personalized interventions. Information 2024;16:11.
[11] Drira M, Ben Hassine S, Zhang M, *et al*. Machine learning methods in student mental health research: an ethics-centered systematic literature review. Appl Sci 2024;14:2076–3417.
[12] Pigoni A, Delvecchio G, Turtulici N, *et al*. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. Transl Psychiatry 2024;14:140.
[13] Farmer RL, Lockwood AB, Goforth A, *et al*. Artificial intelligence in practice: opportunities, challenges, and ethical considerations. Prof Psychol Res Pract 2025;56:19–25.
[14] Coots M, Linn KA, Goel S, *et al*. Racial bias in clinical and population health algorithms: a critical review of current debates. Annu Rev Public Health. 2025;46:507–523.
[15] Sardar TH, Khatun A, Sengupta S, Alam Y, Ara T. Machine learning in the healthcare sector and the biomedical big data: techniques, applications, and challenges. Big Data Comput 2024;17:336–52.
[16] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56.
[17] Sadeghi Z, Alizadehsani R, Cifci MA, *et al*. A review of explainable Artificial Intelligence in healthcare. Comput Electr Eng 2024;118:109370.
[18] Sufian MA, Alsadder L, Hamzi W, *et al*. Mitigating algorithmic bias in AI-driven cardiovascular imaging for fairer diagnostics. Diagnostics 2024;14:2675.
[19] Ciobanu-Caraus O, Aicher A, Kernbach JM, *et al*. A critical moment in machine learning in medicine: on reproducible and interpretable learning. Acta Neurochir (Wien) 2024;166:14.
[20] Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2019;25:37–43.