International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 9 (2025) 123–135

Paper Type: Original Article

Artificial Intelligence for Detecting Mental Disorders: A Review

Osama A. Ghoneim [1], Saleh M. Mohamed [1,*], Mohamed Alaa Fahmy El-Ghamry [2], and Ahmed Sleem [1]

[1] Department of Computer Science, Faculty of Computers and Information, Tanta University, Tanta, 31527, Egypt; Emails: OsamaGhoneim@ics.tanta.edu.eg; saleh_mohamed@ics.tanta.edu.eg; Ahmed.selim@ics.tanta.edu.

[2] Department of Computer Science, Faculty of Computers and Information, Zagazig University, Zagazig, 44519, Egypt; ghamry11@gmail.com.

Received: 10 Sep 2025 **Revised:** 30 Oct 2025 **Accepted:** 28 Dec 2025 **Published:** 30 Dec 2025

Abstract

Mental disorders demonstrate a significant global health challenge, affecting millions of individuals and often leading to severe social and economic consequences. Traditional diagnostic methods, such as clinical interviews and self-report questionnaires, are limited by subjectivity and scalability issues. Recent advancements in artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), offer promising solutions for early detection and accurate classification of mental health conditions. This review explores the AI techniques in diagnosing disorders such as depression, anxiety, bipolar disorder, schizophrenia, PTSD, and ADHD. It summarizes state-of-the-art methodologies, highlights publicly available datasets, and discusses key challenges, including data scarcity, model interpretability, and bias. The paper concludes by outlining future research directions focused on multimodal models, explainable AI, privacy-preserving techniques, and clinical integration to enhance mental health care.

Keywords: Mental Disorders; Mental Health; Mental Illness; Machine Learning; Deep Learning

1 | Introduction

A mental disorder is a medical condition that surely affects a person's thoughts, feelings, thinking, and interactions with others. These problems have demonstrated the significant societal impacts of mental health disorders and the need for innovative mitigation and treatment methods. Miner et al. said that predictive medical models will change the medical health care industry as a whole [1]. Personalized self-reports, which include questionnaires, aim to identify certain psychological or social behaviors and are typically used to diagnose mental disorders [2]. Many people will be capable of recovering from disorders or mental diseases with the right care and therapy [3].

Mental disorders represent a growing global health concern, affecting millions of individuals across diverse populations. These conditions, ranging from anxiety and depression to schizophrenia and bipolar disorder, can significantly impair cognitive, emotional, and social functioning. Traditional diagnostic methods, such as clinical interviews and self-report questionnaires, while valuable, often suffer from subjectivity, limited scalability, and delayed detection. In recent years, the integration of Artificial Intelligence (AI) into mental health research has opened new avenues for early diagnosis, personalized treatment, and predictive modeling. AI technologies, particularly machine learning (ML) and deep learning (DL), have demonstrated remarkable

Corresponding Author: saleh_mohamed@ics.tanta.edu.eg.

capabilities in analyzing complex datasets, including brain imaging, social media activity, and physiological signals to uncover patterns indicative of mental health conditions. These approaches offer the potential to enhance diagnostic accuracy, reduce human bias, and support clinicians in making informed decisions.

This review explores the intersection of AI and mental health, focusing on how various AI methodologies are being applied to detect and classify mental disorders. Section 2 provides an overview of key disorder categories. In section 3, it summarizes recent related work advancements in AI-driven diagnostic models and tools. Section 4 highlights the publicly available datasets used for benchmarking mental health prediction. Section 5 discusses the challenges and limitations in mental disorders prediction. Finally, section 6 concludes the paper and Future direction to the growing field of computational psychiatry.

2 | Background

2.1 | Mental Disorders (MD)

Mental Disorder (MD), often referred to as mental illness or mental disease, describes a variety of psychological conditions that influence an individual's mood, behavior, and thinking. Simply put, when a person's usual emotional state and behavior are adversely altered, it is said that they are experiencing a mental illness [4]. Mental illnesses can have various causes and exhibit a wide range of symptoms.

The World Health Organization (WHO) has highlighted the importance of in-depth research on mental health conditions, given their increasing prevalence in modern society [5]. As modern technologies have progressed, new methods for diagnosing mental illness have emerged. Traditionally, assessments relied on questionnaires and interviews. Today, brain signals, social media activity, sensory information, and medical records are also utilized in evaluating mental health. Additionally, machine learning techniques are gaining popularity among researchers because they can analyze large datasets to identify patterns and offer more accurate predictions of an individual's mental health status [6]. In this section, we will begin by examining the classification of mental disorders. Next, we will consider recent studies on mental disorder detection using machine learning, focusing on the methodological aspects.

2.2 | Types of MD

WHO states that one in eight individuals experiences mental health problems. Although effective prevention and treatment options are available, most people remain unwilling to seek help. Additionally, many are unaware that they have a mental health condition. Mental health disorders can take various forms, including depression, anxiety disorders, bipolar disorder (BD), post-traumatic stress disorder (PTSD), suicidal tendency, and schizophrenia. Figure 1 illustrates the classification of mental disorder issues.

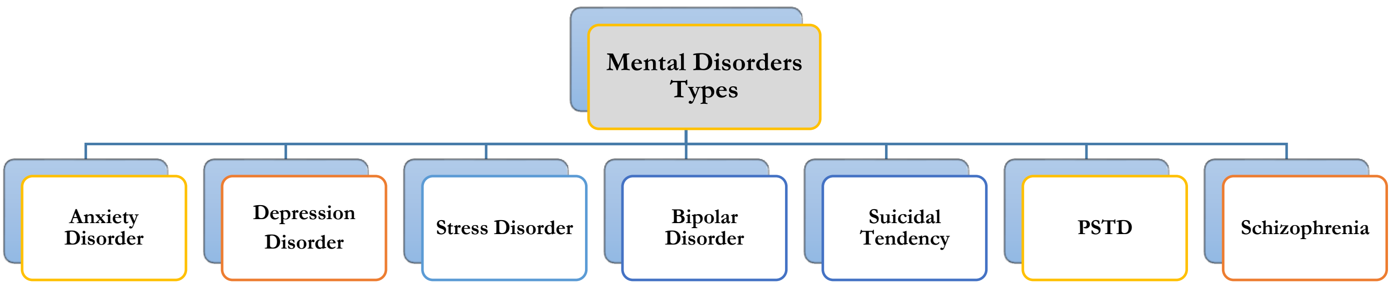

Figure 1. Types of mental disorders issues

2.2.1 | Anxiety Disorder (AD)

Anxiety, as outlined by the American Psychological Association (APA), includes pressure, worry, and physical expressions like heightened heart rate and muscle pressure. It begins with observed threats or stressors, real or not real [7]. It is represented among the most prevalent categories of mental health conditions within the general population, yet a substantial proportion of individuals experiencing these disorders remain unengaged with therapeutic interventions or clinical treatment modalities [8].

2.2.2 | Depression Disorder (DD)

Data derived from the 2019 Global Burden of Disease Study by the Institute for Health Metrics and Evaluation (IHME) demonstrates that major depressive disorders impact approximately 280 million people globally, representing a substantial proportion of the world's population [9]. It constitutes a prevalent and disabling mental health condition, disproportionately affecting women across age groups. Characterized by enduring affective and somatic symptoms, its etiology involves multifactorial interactions among psychosocial and biological determinants [10].

2.2.3 | Stress Disorder (SD)

Stress constitutes a significant etiological factor contributing to the manifestation of various mental health disorders. Stress is operationally defined as any environmental or internal stimulus that induces physical, emotional, or psychological disequilibrium within an individual. While stress itself does not qualify as a distinct psychiatric diagnosis, empirical evidence demonstrates a direct correlation between an individual's stress exposure levels and their overall psychological well-being and mental health status [11].

2.2.4 | Post-traumatic stress disorder (PTSD)

Experiencing fear during and following traumatic events is a normal human response. This fearful reaction stems from the body's inherent "fight-or-flight" mechanism, which serves as a protective system designed to help individuals either evade or confront threatening situations. Following traumatic experiences, people typically display various emotional and psychological responses, with the majority naturally recovering from these symptoms as time progresses. However, individuals whose symptoms persist over an extended period may receive a clinical diagnosis of post-traumatic stress disorder (PTSD)[12].

2.2.5 | Bipolar Disorder (BD)

Bipolar disorder represents a distinct mental health condition characterized by alternating episodes of mania and depression, with some instances presenting mixed episodes containing elements of both states. Manic episodes are distinguished by heightened irritability, elevated energy levels, and reduced sleep requirements. During manic phases, individuals frequently demonstrate impulsive and risky behaviors. Conversely, the depressive phases of bipolar disorder mirror the symptomatology observed in major depression. While research indicates that certain patients achieve functional recovery between episodes, a significant proportion continue to experience persistent symptoms that compromise their daily functioning and overall quality of life [13].

2.2.6 | Suicidal Tendency

A further mental health concern that is frequently associated with depressive disorders and bipolar disorder is the propensity for suicidal ideation [14]. Depressive symptomatology is frequently accompanied by profound feelings of worthlessness that predispose individuals to engage in deliberate self-harm, potentially culminating in suicidal behavior. Contemporary research in clinical mental health focuses on developing predictive models and screening instruments to identify at-risk populations through systematic evaluation of patient narratives, cognitive content analysis, and behavioral observations obtained during clinical assessments, thereby enabling proactive therapeutic intervention and suicide prevention measures.

2.2.7 | Schizophrenia

Schizophrenia is a serious mental health condition that changes how a person thinks, feels, and acts, as defined by the National Institute of Mental Health (NIMH). People with schizophrenia may seem disconnected from reality, which can be upsetting for them and for those close to them [15].

3 | Artificial Intelligence (AI) approaches

Artificial Intelligence (AI) means building systems that are skilled in executing duties and functions that normally need human-level intelligence [16]. These duties include analyzing data, identifying trends, coming to conclusions, and resolving challenging issues. These responsibilities include analyzing data, detecting general patterns, coming to conclusions, and resolving challenging issues [17], [18]. AI includes a range of approaches, each of which is appropriate for a particular set of problems and application areas. Machine learning (ML), deep learning (DL), natural language processing (NLP), and reinforcement learning (RL) are the main AI approaches that have been extensively applied in applications related to mental health [19], [20].

Machine Learning (ML) enables computers to gain insights from data; this makes it important in enhancing systems for global health and well-being. The primary goal of ML is to create systems that can learn without requiring detailed and specific programming [21]. There are three types of ML, which are supervised, unsupervised, and reinforcement learning. Supervised learning is like educating a little child, so the machine learns by examples based on a labeled dataset. Unsupervised learning does not require any predictions. So there is no need for the machine to be trained on a labeled dataset. It is used for grouping similar hidden items in the same group, which is known as clustering.

Ensemble methodologies constitute a prevalent machine learning paradigm wherein multiple constituent algorithms are systematically integrated to form a composite predictive model, thereby achieving enhanced forecasting accuracy and robustness. The implementation of sophisticated techniques, including stacking, bagging, and boosting, facilitates the orchestration of heterogeneous learning algorithms, enabling the exploitation of their complementary strengths to yield superior predictive performance relative to individual model implementations [22].

Deep Learning (DL), a branch of machine learning (ML) that focuses on analyzing enormous volumes of unstructured data, such as audio, image, video, and free text of individuals [20]. Natural Language Processing (NLP), which makes it possible for machines to understand and communicate with human language, is a helpful enhancement to these methods. Because it can investigate medical records, social media posts, and text messages, NLP is essential for global mental health care.

4 | Review of Related Work

This section provides a comprehensive review and discussion of the documents and information related to machine learning approaches employed by researchers for predicting or diagnosing mental health conditions. Furthermore, the performance of the applied machine learning algorithms will be critically evaluated and analyzed. Mental health issues will be categorized into distinct disorders, including schizophrenia, anxiety and depression, bipolar disorder, post-traumatic and stress disorder. Table 1 contains a list of different studies that use several data sources to detect mental health disorders in humans.

Chigagure and Sakala [23] evaluated multiple models on eight years of hospital data, applying techniques such as imputation, feature selection, and imbalance correction. Their findings show that ensemble methods, particularly XGBoost and AdaBoost, outperform traditional classifiers, achieving up to 93% accuracy and 97% recall.

The research conducted by Jetli and Jason [23] demonstrated that machine learning approaches can effectively predict and classify the problems of mental health. The study evaluated seven algorithms, including logistic regression, gradient boosting, neural networks, K-nearest neighbors, support vector machines, Extreme Gradient Boosting, and Deep Neural Networks, alongside a majority voting ensemble method. Gradient boosting achieved the highest individual performance at 88.80% accuracy, followed by neural networks at 88.00%, while the ensemble approach reached 85.60% accuracy. Contrary to expectations, the ensemble method underperformed compared to the best individual classifiers, though it outperformed simpler algorithms and demonstrated competitive sensitivity scores. The findings suggest that individual algorithms like gradient boosting can be more effective than ensemble approaches for this specific mental health prediction task.

A related study was conducted by Sau and Bhakta [25] in 2019 and works on anxiety and depression prediction among seafarers using socio-demographic, occupational, and health features. The study evaluated five machine learning classifiers: Catboost, logistic regression, naive Bayes, random forest, and support vector machine. The study involved 470 male seafarers from Haldia Dock Complex, India, with ground truth labels established using the Hospital Anxiety and Depression Scale (HAM-A and HAM-D), where 48.7% were classified as having anxiety and/or depression. Catboost emerged as the top-performing classifier with 82.6% accuracy and 84.1% precision during 10-fold cross-validation on the training set (n=414) and achieved superior performance on the test set (n=56) with 89.3% accuracy and 89.0% precision.

Another research by Srinivasa Gopalan et al. [26] works on diagnosing schizophrenia based on a novel proposed deep learning approach configured with three hidden layers. They applied their model on a functional magnetic resonance imaging (fMRI) dataset from the National Institute of Health to diagnose patients with schizophrenia automatically. Their findings demonstrate that the proposed deep learning approach achieves superior diagnostic accuracy compared to traditional methods. The authors conclude that deep learning techniques hold significant promise for improving schizophrenia diagnosis and may represent an important advancement in the field.

The authors in [27] showed that schizophrenia is a serious mental illness that takes time and careful evaluation to diagnose properly, yet identifying it early helps patients manage their symptoms more effectively. While deep learning methods have shown potential for diagnosing schizophrenia from brain scans, these research approaches often fail to work well in actual clinical practice. To bridge this gap, we developed a system using routine psychiatric assessments based on DSM-5 criteria—information that clinicians already collect during standard evaluations—making it practical for everyday use in mental health settings. Our model incorporates interpretable attention mechanisms that reveal which factors influenced each diagnosis, helping doctors understand and trust the system's recommendations while maintaining alignment with established diagnostic standards. Testing showed the model is reliable across different data conditions and achieved 98% accuracy, suggesting it could effectively support clinicians in real-world diagnostic decisions.

Another study [28] the use of network analysis and machine learning methods to diagnose schizophrenia using brain imaging data. The researchers analyzed brain connectivity patterns in 48 schizophrenia patients and 24 healthy controls by reconstructing brain networks from diffusion-weighted imaging and calculating various network properties. They found significant differences in global network properties, with schizophrenia patients showing reduced overall connectivity, efficiency, and clustering compared to healthy controls. Multiple machine learning algorithms, including support vector machine, random forest, and gradient boosting, were applied to classify patients and controls based on these network features, achieving accuracies up to 68.6% with overall connectivity identified as the most important distinguishing feature. The study demonstrates the potential for integrating brain network analysis with machine learning approaches to assist in schizophrenia diagnosis, though the researchers acknowledge limitations, including small sample size, and recommend further validation with larger populations.

Sau and Bhakta presented [28][29] a machine learning approach for schizophrenia detection using electroencephalogram (EEG) signals as the primary diagnostic tool. The researchers developed an automated decision support system that combines multiple signal processing techniques and feature extraction methods to analyze brain wave patterns in patients with schizophrenia compared to healthy controls. The study utilized advanced computational methods, including support vector machines and other machine learning algorithms, to classify EEG recordings and identify distinctive neural patterns associated with schizophrenia. The proposed hybrid system demonstrated promising results in accurately distinguishing between schizophrenic patients and healthy individuals, offering potential for early detection and diagnosis support in clinical settings. This work contributes to the growing field of computational psychiatry by providing an objective, technology-based approach to supplement traditional clinical diagnosis methods for schizophrenia.

Research by Niu et al. [30] introduces HCAG, a novel Hierarchical Context-Aware Graph attention model that mirrors the natural structure of depression assessment procedures and employs Graph Attention Networks to capture meaningful relationships between textual and audio information. Current approaches face significant limitations, as they depend on time-intensive topic modeling techniques and manual selection processes while failing to fully utilize the rich contextual information available in clinical interview data. The proposed model demonstrates exceptional performance on the DAIC-WOZ dataset, achieving an F1-score of 0.92, a mean absolute error of 2.94, and a root mean square error of 3.80, surpassing all existing state-of-the-art methods in the field. This advancement represents a significant step forward in developing reliable, automated tools that can assist healthcare professionals in identifying depression through comprehensive analysis of clinical interview content.

Xezonaki et al. [31] a machine learning approach for detecting depression from written transcripts of clinical interviews, recognizing that depression affects not only mood but also how people use language. The researchers developed a Hierarchical Attention Network that incorporates emotional language patterns from established affective word databases to improve classification accuracy. Their analysis revealed that people with depression tend to use emotionally charged language more frequently than those without depression. By integrating this affective information into their model, the system achieved strong performance on two depression datasets, reaching F1-scores of 71.6% and 70.3%, respectively. These results demonstrate that combining structural language analysis with emotional content can effectively support depression detection from clinical conversations.

Prediction of depression is an essential goal among medical check-ups by applying machine learning. Cho et al. [32] a predictive model to identify depression risk using data from routine medical check-ups in South Korea, analyzing over 433,000 individuals with approximately 2.56% diagnosed with depression. The researchers employed a random forest machine learning algorithm with proper validation methods, achieving strong predictive performance with an area under the curve of 0.849 and specificity of 82.4%. The model revealed that benzodiazepine prescriptions were the strongest predictor of future depression onset, suggesting these medications may signal underlying mental health concerns. The system demonstrated potential for early intervention by identifying at-risk patients during standard health screenings, though the researchers noted that prospective validation studies are needed to confirm its real-world effectiveness. This approach could enable healthcare providers to proactively address depression risk by targeting preventive interventions to vulnerable individuals identified through routine medical visits.

Sharma et al. [33] explore using machine learning to improve depression diagnosis through biomarkers rather than traditional interview methods. The researchers applied an Extreme Gradient Boosting (XGBoost) algorithm to a Dutch dataset of 11,081 individuals, where only 570 self-reported depression, creating a challenging imbalanced classification problem. They addressed this imbalance using different resampling techniques oversampling, under sampling, and combined approaches—to balance the data before training the model. The study found that XGBoost models trained on oversampled and over-under sampled datasets achieved strong performance, with balanced accuracy, precision, recall, and F1 scores exceeding 0.90, identifying key biomarkers like thrombocytes, triglycerides, and neutrophil granulocytes as important predictors. The findings suggest that biomarker-based machine learning models could provide a supportive tool for diagnosing depression, especially when traditional interview methods yield uncertain results or when patients struggle to respond to standard questionnaires.

Ritcher et al. [34] presents a machine learning diagnostic support system designed to distinguish between anxiety and depression disorders in clinical patients using cognitive-behavioral assessments rather than self-reports. The researchers recruited 86 psychiatric patients with clinical anxiety, depression, or both, along with

25 healthy controls, and had them complete six cognitive tasks measuring biases in attention, memory, expectations, interpretation, and executive functions. Using a random forest machine learning algorithm with cross-validation, the system successfully differentiated clinical patients from healthy controls with 76.81% specificity and 69.66% sensitivity and distinguished anxiety from depression with 80.50% and 66.46% accuracy, respectively. The findings demonstrate that cognitive performance patterns can serve as an objective diagnostic tool to supplement traditional clinical interviews, potentially increasing diagnostic precision and enabling more personalized treatment approaches. This mechanism-based approach offers clinicians and patients a data-driven assessment that reduces reliance on subjective self-reporting and provides insight into the specific cognitive biases characterizing each patient's condition.

Li et al. [35] explore using machine learning combined with brain imaging to accurately identify individuals with bipolar disorder. The researchers studied 44 bipolar patients and 36 healthy individuals, using both structural brain scans (measuring gray matter volume) and functional scans (measuring brain activity patterns) to build a support vector machine classifier. They found that combining both types of imaging data produced much better results than using either type alone, achieving 87.5% accuracy in correctly identifying bipolar disorder patients. The study identified specific brain regions with abnormalities, including areas in the frontal lobe, temporal lobe, and limbic system, which are involved in emotion regulation and cognitive processing. This multimodal approach demonstrates that combining structural and functional brain imaging with machine learning could help clinicians diagnose bipolar disorder more accurately, potentially reducing the years of misdiagnosis that many patients currently experience.

Another study by Li et al. [36] addresses the challenge of using deep learning to diagnose mental illnesses by proposing a deep learning approach to automatically distinguish between first-episode psychosis, bipolar disorder, and healthy individuals. The researchers developed a convolutional neural network that analyzes structural brain imaging data, specifically gray matter volume images, from 89 first-episode psychosis patients, 40 bipolar disorder patients, and 83 healthy controls. They trained both a three-way classifier to distinguish all groups simultaneously and three separate binary classifiers to compare pairs of groups. The CNN-based method achieved better performance than traditional machine learning classifiers in both two-category and three-category classification tasks. The findings suggest that differences in gray matter volume are important brain features that can help distinguish between first-episode psychosis, bipolar disorder, and healthy individuals using automated deep learning techniques.

Abaei and Osman [37] developed a method to identify different BD states by analyzing emotional expressions captured in video recordings of patient interviews. We used a Convolutional Neural Network (CNN) to detect facial features and a Long-Short-Term Memory (LSTM) network to classify patients into three clinical states: remission, hypomania, and mania. Our approach achieved encouraging results on the Turkish Audio-Visual Bipolar Disorder Corpus, reaching an Unweighted Average Recall of 60.67% on the development dataset.

Schultebraucks et al. [38] investigated whether pre-deployment risk factors could predict post-traumatic stress disorder (PTSD) in Army personnel deployed to Afghanistan using machine learning approaches. Researchers collected comprehensive biological, psychological, and neurocognitive data from 473 soldiers before their deployment and followed them for 90-180 days after their return. The machine learning models achieved strong predictive accuracy, with support vector machines reaching 88% sensitivity and 79% specificity for predicting provisional PTSD diagnosis. The most important pre-deployment predictors included sleep quality, anxiety, depression, sustained attention, cognitive flexibility, and various blood-based biomarkers such as inflammatory markers and metabolites. These findings suggest that combining diverse pre-deployment measurements could help identify at-risk soldiers and inform targeted prevention strategies before deployment.

In research conducted by Campbell et al. [39], a decision tree algorithm was employed to predict unit-level mental health risk using data from the Combat and Operational Stress Control (COSC) survey. The study examined 2,290 U.S. Navy personnel, including both officers and enlisted sailors from intact battalions, ground-based aviation squadrons, and medical support units serving alongside the Marine Corps during deployments to Iraq and Afghanistan between 2007 and 2008. The decision tree model demonstrated strong predictive performance for identifying high-risk PTSD cases in the validation dataset, though it showed approximately 10% misclassification when applied to independent samples. This level of prediction error suggests that while the model achieved reasonable accuracy, there remains room for improvement in generalizing findings across different populations. The study highlights both the potential and limitations of machine learning approaches for mental health screening in military settings.

Detecting mental health problems that result from childhood sexual abuse remains a difficult task for both clinicians and researchers. Gokten and Uyulan [40] conducted a study to predict the development of post-traumatic stress disorder (PTSD) and depression in sexually abused children and adolescents. Following forensic examinations, psychiatrists evaluated each child using the DSM-V diagnostic criteria to assess their mental health status. The researchers then developed a predictive model using a Random Forest (RF) classifier to identify patterns in the data. This machine learning approach aimed to improve early identification of trauma-related mental disorders in young survivors of sexual abuse.

Mikolas and colleagues [41] developed an approach to identify patients with ADHD from a diverse group of individuals with various mental health conditions using anonymized hospital records. The researchers applied a support vector machine (SVM) classifier to analyze 30 different clinical features from patient data. They conducted three separate analyses: one using all features, another excluding demographic information like age and sex, and a third using only complete data without any missing values. The three classification models achieved accuracies of 66.1%, 65.1%, and 68.8%, respectively, showing that the model performed best when working with complete datasets. These results demonstrate that machine learning can help distinguish ADHD from other psychiatric conditions in real clinical settings, though the moderate accuracy levels suggest room for improvement.

Tachmazidis and colleagues [42] conducted a study to diagnose ADHD in adults who received clinical evaluations over an extended period. The researchers combined clinical assessment data with questionnaire responses and developed a unique hybrid approach that integrated both machine learning algorithms and knowledge-based reasoning systems. This innovative method achieved an impressive accuracy rate of 95% in identifying adult ADHD cases. The strong performance led to the implementation of this diagnostic tool in actual clinical settings for real-world testing. This work demonstrates that combining different computational approaches can significantly improve the accuracy of ADHD detection in adult populations.

Yin et al. [43] investigated whether neural flexibility could serve as a biological marker to distinguish children with ADHD from typically developing children using brain imaging data from multiple research sites. The study analyzed data from 236 participants at Peking University and 192 participants at New York University, both part of the larger ADHD-200 dataset. The researchers employed an extreme gradient boosting (XGBoost) algorithm to create two models: one to classify children as having ADHD or being typically developing, and another to predict the severity of ADHD symptoms. Their findings revealed that children with ADHD show different patterns of neural flexibility compared to typically developing children. These results suggest that measuring neural flexibility could be clinically useful for diagnosing ADHD, tracking treatment effectiveness, and assessing the severity of the condition.

Table 1. Summary of Recent Approaches in Multimodal Fake News Detection.

Ref.	Year	Author	Disorder	Used Methodology	Results
[23]	2025	Chigagure and sakala	Generic mental health	Traditional Models: Logistic Regression Support Vector Machine (SVM) K-Nearest Neighbors (K-NN) Ensemble Models: Bagging	Accuracy (%): LR = 88 KNN = 81.22 SVM = 88 XGBoost = 92 AdaBoost = 92

				Boosting (XGBoost, AdaBoost) Stacking	
[24]	2023	Jetli Chung and Jason Teo	Generic mental health	Logistic regression (LR) Gradient boosting (GB) Neural networks (NN) K-nearest neighbors (KNN) Support vector machine (SVM) Deep neural networks (DNN) Ensemble approach Voting classifier Extreme gradient boosting	Accuracy (%): LR = 79.63 GB = 81.22 NN = 78.57 KNN = 81.22 SVM = 80.69 DNN = 79.89 Voting classifier = 81.75 Extreme gradient boosting = 80.69
[25]	2019	Sau and Bhakta	anxiety and depression	Logistic Regression (LR) Naïve Bayes (NB) Random Forest (RF) Support Vector Machine (SVM) Catboost	LR = 77.8 NB = 75.8 RF = 81.2 SVM = 76.1 Catboost = 82.6
[26]	2019	Srinivasagopalan et al.	Schizophrenia	LR SVM RF NN	Accuracy (%): LR: 82.77 SVM: 82.68 RF: 83.33 NN: 94.44
[27]	2022	Organisciak et al.	Schizophrenia	Interpretable framework based on deep network and Self-Attention	Accuracy (%): 98
[28]	2020	Jo et al.	Schizophrenia	SVM Multinomial naïve Bayes RF XGBoost	Accuracy (%): Multinomial NB: 66.9 RF: 68.6 SVM: 58.2 XGBoost: 66.3
[44]	2021	Niu et al.	Depression	Context-Aware Graph Attention Model	F1-Score: 92 % RMSE: 3.80
[45]	2022	Yoon et al.	Depression	Multimodal cross-attention mechanism	Precision: 65.40 Recall: 65.57 F1-Score: 63.50
[31]	2020	Xezonaki et al.	Depression	Hierarchical attention networks	F1-Score (%): GPC dataset: 71.6 DAIC-WOZ: 68.6
[46]	2020	Cho et al.	Depression	RF	Accuracy: 78%
[33]	2020	Sharma et al.	Depression	XGBoost	Accuracy: 97%
[34]	2021	Richter et al.	Anxiety Depression	RF	Anxiety vs. Depression Specificity: 80.50%, Recall: 66.46%
[35]	2020	Li et al.	Bipolar disorder	SVM	Accuracy: 87.5%
[36]	2021	Li et al.	Bipolar disorder	CNN	Accuracy: 99.15%
[37]	2020	Abaei and Osman	Bipolar Disorder	CNN-LSTM	Accuracy: 63.32%
[38]	2021	Schultebraucks et al.	PTSD	RF SVM	Accuracy: RF: 78% SVM: 88%
[39]	2019	Campbell et al.	PTSD	Decision tree	Sensitivity: 42.5% Specificity: 88%
[40]	2021	Gokten and Uyulan	PTSD/Depression	RF	Accuracy: Depression: 88.0% PTSD: 76.0%
[41]	2022	Mikolas et al.	ADHD	SVM	Accuracy: 65.1%

[42]	2021	Tachmazidis et al.	ADHD	ML + knowledge-based hybrid model	Accuracy: 95%
[47]	2022	Yin et al.	ADHD	XGBoost for classification of ADHD from TDC and Regression for ADHD severity	Accuracy for classification: 77% R^2 for regression: 0.2794

5 | Open access dataset

We have included a number of datasets that are utilized in various works on mental disorders. Table 2 contains a list of these publicly open-access datasets.

Ref.	Dataset	Disorder	Domain of data
[24]	OSMI Mental Health in Tech Survey	Generic mental health	Survey data
[48]	DAIC	depression	Interviews data (audio & video & questionnaire)
[49]	Depression Dataset	Depression	Sensor & device data
[50]	Depression, Anxiety, and Stress dataset	Depression, Anxiety, Stress	Survey data
[51]	Reddit suicidal dataset	Suicidal Tendency	Text data
[52]	Human Stress dataset	stress	
[53]	Depression audio dataset	Depression	Audio data
[54]	Nurse Stress dataset wearable sensors	stress	Sensor & device data
[55]	MEMO dataset	Depression	Multi-modal
[56]	Distress Analysis Interview Corpus (DAIC)	Anxiety, Depression, PTSD	Audio/Video
[57]	Turkish Audio-visual Bipolar Disorder Corpus	Bipolar Disorder	Audio/Video
[58]	ADHD-200	ADHD	Images

6 | Challenges and Limitations

Despite the promising advancements in applying Artificial Intelligence (AI) to mental health diagnostics, several challenges and limitations persist that hinder widespread adoption and clinical integration. This review faces several constraints. First, the availability and quality of mental health datasets remain limited, often suffering from imbalance and lack of diversity. Second, many AI models, particularly deep learning architectures, operate as "black boxes," reducing interpretability and clinician trust. Third, bias in training data can lead to inequitable predictions across demographic groups. Finally, the absence of standardized benchmarks and the difficulty of integrating AI tools into clinical workflows hinder real-world adoption.

7 | Conclusion and future work

This review highlights the transformative potential of Artificial Intelligence (AI) in detecting and managing mental disorders. AI techniques, particularly machine learning and deep learning, have demonstrated strong capabilities in analyzing complex, multimodal data to improve diagnostic accuracy, reduce subjectivity, and support personalized care. However, challenges such as limited and imbalanced datasets, lack of interpretability, bias in predictions, and integration barriers into clinical workflows remain significant obstacles to real-world adoption.

Future research should focus on developing multimodal and context-aware models that integrate diverse data sources, improving interpretability through explainable AI techniques, and implementing fairness-aware algorithms to mitigate bias. Privacy-preserving methods such as federated learning should be explored to enable secure data sharing. Additionally, establishing standardized benchmarks and conducting large-scale clinical validations will be critical for ensuring reliability. Integration with telehealth platforms and wearable devices can further support continuous monitoring and early intervention.

Acknowledgments

The authors are grateful to the editorial and reviewers, as well as the correspondent author, who helped in the form of advice, assessment, and checking during the study period.

Funding

This research has no funding source.

Conflicts of Interest

The authors declare that there is no conflict of interest in research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] "Practical Predictive Analytics and Decisioning Systems for Medicine ... - Gary D. Miner, Linda A. Miner, Mitchell Goldstein, Robert Nisbet, Nephi Walton, Pat Bolding, Joseph Hilbe, Thomas Hill - Google Books." Accessed: Sep. 01, 2025. [Online]. Available: https://books.google.com.eg/books?hl=en&lr=&id=qHgRBAAAQBAJ&oi=fnd&pg=PP1&ots=5VzZ21-KXZ&sig=XFFppaEw8J9pyRg8Vwkn9TdXJsw&redir_esc=y#v=onepage&q&f=false

[2] M. HAMILTON, "Development of a Rating Scale for Primary Depressive Illness," British Journal of Social and Clinical Psychology, vol. 6, no. 4, pp. 278–296, Dec. 1967, doi: 10.1111/J.2044-8260.1967.TB00530.X;WGROUP:STRING:PUBLICATION.

[3] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science (1979), vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/SCIENCE.AAA8415;REQUESTEDJOURNAL:JOURNAL:SCIENCE;CTYPE:STRING:JOURNAL.

[4] T. R. Insel and P. S. Wang, "Rethinking mental illness," JAMA, vol. 303, no. 19, pp. 1970–1971, 2010.

[5] W. H. Organization. R. O. for the E. Mediterranean, "Summary report on the regional consultation on the development of the global mental health action plan, Cairo, Egypt, 2–4 July 2012," 2012. [Online]. Available: https://iris.who.int/handle/10665/116125

[6] D. Bzdok and A. Meyer-Lindenberg, "Machine learning for precision psychiatry: opportunities and challenges," Biol Psychiatry Cogn Neurosci Neuroimaging, vol. 3, no. 3, pp. 223–230, 2018.

[7] "Anxiety." Accessed: Aug. 05, 2025. [Online]. Available: https://www.apa.org/topics/anxiety

[8] M. Olfson, R. C. Kessler, P. A. Berglund, and E. Lin, "Psychiatric disorder onset and first treatment contact in the United States and Ontario," American Journal of Psychiatry, vol. 155, no. 10, pp. 1415–1422, 1998, doi: 10.1176/AJP.155.10.1415/ASSET/IMAGES/AF17F1.JPEG.

[9] G. W. Comstock and K. J. Helsing, "Symptoms of depression in two communities," Psychol Med, vol. 6, no. 4, pp. 551–563, 1977, doi: 10.1017/S0033291700018171.

[10] "Depressive disorder (depression)." Accessed: Aug. 05, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[11] "Life stress and mental health: II. The midtown Manhattan study." Accessed: Aug. 05, 2025. [Online]. Available: https://psycnet.apa.org/record/1965-02390-000

[12] "Post-Traumatic Stress Disorder - National Institute of Mental Health (NIMH)." Accessed: Aug. 12, 2025. [Online]. Available: https://www.nimh.nih.gov/health/publications/post-traumatic-stress-disorder-ptsd

[13] J. Angst and R. Sellaro, "Historical perspectives and natural history of bipolar disorder," Biol Psychiatry, vol. 48, no. 6, pp. 445–457, Sep. 2000, doi: 10.1016/S0006-3223(00)00909-4.

[14] L. Brådvik, "Suicide Risk and Mental Disorders," International Journal of Environmental Research and Public Health 2018, Vol. 15, Page 2028, vol. 15, no. 9, p. 2028, Sep. 2018, doi: 10.3390/IJERPH15092028.

[15] "Schizophrenia - National Institute of Mental Health (NIMH)." Accessed: Aug. 12, 2025. [Online]. Available: https://www.nimh.nih.gov/health/topics/schizophrenia

[16] J. E. (Hans) Korteling, G. C. van de Boer-Visschedijk, R. A. M. Blankendaal, R. C. Boonekamp, and A. R. Eikelboom, "Human- versus Artificial Intelligence," Front Artif Intell, vol. 4, p. 622364, Mar. 2021, doi: 10.3389/FRAI.2021.622364/BIBTEX.

[17] P. Mikalef, N. Islam, V. Parida, H. Singh, and N. Altwaijry, "Artificial intelligence (AI) competencies for organizational performance: A B2B marketing capabilities perspective," J Bus Res, vol. 164, p. 113998, Sep. 2023, doi: 10.1016/J.JBUSRES.2023.113998.

[18] "Opportunities of artificial intelligence for supporting complex problem-solving: findings from a scoping review - MADOC." Accessed: Sep. 02, 2025. [Online]. Available: https://madoc.bib.uni-mannheim.de/64680/

[19] L. Sawalha, T. A.-S. L. vs. D. L. A. Practical, and undefined 2024, "Shallow learning versus deep learning in natural language processing applications," SpringerL Sawalha, TC AkinciShallow Learning vs. Deep Learning: A Practical Guide for Machine Learning, 2024•Springer, pp. 179–206, 2024, doi: 10.1007/978-3-031-69499-8_8.

[20] N. Rane, S. Mallick, Ö. Kaya, and J. Rane, Applied Machine Learning and Deep Learning: Architectures and Techniques. 2024. Accessed: Oct. 02, 2025. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=kuYyEQAAQBAJ&oi=fnd&pg=PP1&dq=Rane,+N.L.,+et+al.+(2024).+Applied+Machine+Learning+and+Deep+Learning:+Architectures+and+Techniques:+Deep+Science+Publishing.&ots=scFtfwwx-1&sig=ioJeYPqSX90BFTtCr_MMPMYDMU4

[21] S. Mustary, M. Abul Kashem, M. N. I. Khan, F. A. Jewel, M. M. Islam, and S. Islam, "LEACH Based WSN Classification Using Supervised Machine Learning Algorithm," 2021 International Conference on Computer Communication and Informatics, ICCCI 2021, vol. 2021-January, Jan. 2021, doi: 10.1109/ICCCI50826.2021.9457001.

[22] I. Mienye, Y. S.-I. Access, and undefined 2022, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," ieeexplore.ieee.orgID Mienye, Y SunIeee Access, 2022•ieeexplore.ieee.org, Accessed: Sep. 25, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9893798/

[23] H. Chigagure and L. C. Sakala, "A machine learning approach for early prediction of mental health crises," Computer Science and Information Technologies, vol. 6, no. 3, pp. 335–345, Nov. 2025, doi: 10.11591/CSIT.V6I3.P335-345.

[24] J. Chung and J. Teo, "Single classifier vs. ensemble machine learning approaches for mental health prediction," Brain Inform, vol. 10, no. 1, p. 1, 2023.

[25] A. Sau and I. Bhakta, "Screening of anxiety and depression among seafarers using machine learning technology," Inform Med Unlocked, vol. 16, p. 100228, Jan. 2019, doi: 10.1016/J.IMU.2019.100228.

[26] S. Srinivasagopalan, J. Barry, V. Gurupur, and S. Thankachan, "A deep learning approach for diagnosing schizophrenic patients," Journal of Experimental & Theoretical Artificial Intelligence, vol. 31, no. 6, pp. 803–816, Nov. 2019, doi: 10.1080/0952813X.2018.1563636.

[27] D. Organisciak, H. P. H. Shum, E. Nwoye, and W. L. Woo, "RobIn: A robust interpretable deep network for schizophrenia diagnosis," Expert Syst Appl, vol. 201, p. 117158, Sep. 2022, doi: 10.1016/J.ESWA.2022.117158.

[28] Y. T. Jo, S. W. Joo, S. H. Shon, H. Kim, Y. Kim, and J. Lee, "Diagnosing schizophrenia with network analysis and a machine learning method," Int J Methods Psychiatr Res, vol. 29, no. 1, p. e1818, Mar. 2020, doi: 10.1002/MPR.1818.

[29] A. Sau, I. B.-H. T. Letters, and undefined 2017, "Predicting anxiety and depression in elderly patients using machine learning technology," Wiley Online LibraryA Sau, I BhaktaHealthcare Technology Letters, 2017•Wiley Online Library, vol. 4, no. 6, pp. 238–243, 2017, doi: 10.1049/HTL.2016.0096.

[30] M. Niu, K. Chen, Q. Chen, and L. Yang, "HCAG: A Hierarchical Context-Aware Graph Attention Model for Depression Detection," in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Jun. 2021, pp. 4235–4239. doi: 10.1109/ICASSP39728.2021.9413486.

[31] D. Xezonaki, G. Paraskevopoulos, A. Potamianos, and S. Narayanan, "Affective Conditioning on Hierarchical Attention Networks Applied to Depression Detection from Transcribed Clinical Interviews," Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, vol. 2020-October, pp. 4556–4560, 2020, doi: 10.21437/INTERSPEECH.2020-2819.

[32] S. E. Cho, Z. W. Geem, and K. S. Na, "Prediction of depression among medical check-ups of 433,190 patients: A nationwide population-based study," Psychiatry Res, vol. 293, p. 113474, Nov. 2020, doi: 10.1016/J.PSYCHRES.2020.113474.

[33] A. Sharma and W. J. M. I. Verbeke, "Improving Diagnosis of Depression With XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081)," Front Big Data, vol. 3, p. 523466, Apr. 2020, doi: 10.3389/FDATA.2020.00015/BIBTEX.

[34] T. Richter, B. Fishbain, E. Fruchter, G. Richter-Levin, and H. Okon-Singer, "Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders," J Psychiatr Res, vol. 141, pp. 199–205, Sep. 2021, doi: 10.1016/J.JPSYCHIRES.2021.06.044.

[35] H. Li et al., "Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques," BMC Psychiatry, vol. 20, no. 1, pp. 1–12, Oct. 2020, doi: 10.1186/S12888-020-02886-5/FIGURES/4.

[36] Z. Li et al., "Deep learning based automatic diagnosis of first-episode psychosis, bipolar disorder and healthy controls," Computerized Medical Imaging and Graphics, vol. 89, p. 101882, Apr. 2021, doi: 10.1016/J.COMPMEDIMAG.2021.101882.
[37] N. Abaei and H. Al Osman, "A hybrid model for bipolar disorder classification from visual information," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2020-May, pp. 4102–4106, May 2020, doi: 10.1109/ICASSP40776.2020.9054648.
[38] K. Schultebraucks et al., "Pre-deployment risk factors for PTSD in active-duty personnel deployed to Afghanistan: a machine-learning approach for analyzing multivariate predictors," Mol Psychiatry, vol. 26, no. 9, pp. 5011–5022, Sep. 2021, doi: 10.1038/S41380-020-0789-2;TECHMETA.
[39] J. S. Campbell, M. L. Wallace, A. Germain, and R. L. Koffman, "A predictive analytic approach to planning combat stress control operations.," Int J Stress Manag, vol. 26, no. 2, pp. 120–131, May 2019, doi: 10.1037/str0000092.
[40] E. S. Gokten and C. Uyulan, "Prediction of the development of depression and post-traumatic stress disorder in sexually abused children using a random forest classifier," J Affect Disord, vol. 279, pp. 256–265, Jan. 2021, doi: 10.1016/J.JAD.2020.10.006.
[41] P. Mikolas et al., "Training a machine learning classifier to identify ADHD based on real-world clinical data from medical records," Sci Rep, vol. 12, no. 1, p. 12934, Jul. 2022, doi: 10.1038/S41598-022-17126-X;SUBJMETA.
[42] I. Tachmazidis, T. Chen, M. Adamou, and G. Antoniou, "A hybrid AI approach for supporting clinical diagnosis of attention deficit hyperactivity disorder (ADHD) in adults," Health Inf Sci Syst, vol. 9, no. 1, pp. 1–8, Dec. 2021, doi: 10.1007/S13755-020-00123-7/TABLES/8.
[43] W. Yin et al., "Altered neural flexibility in children with attention-deficit/hyperactivity disorder," Mol Psychiatry, vol. 27, no. 11, pp. 4673–4679, Nov. 2022, doi: 10.1038/S41380-022-01706-4;TECHMETA.
[44] M. Niu, K. Chen, Q. Chen, and L. Yang, "HCAG: A Hierarchical Context-Aware Graph Attention Model for Depression Detection," in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Jun. 2021, pp. 4235–4239. doi: 10.1109/ICASSP39728.2021.9413486.
[45] J. Yoon, C. Kang, S. Kim, and J. Han, "D-vlog: Multimodal Vlog Dataset for Depression Detection," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 11, pp. 12226–12234, Jun. 2022, doi: 10.1609/AAAI.V36I11.21483.
[46] S. E. Cho, Z. W. Geem, and K. S. Na, "Prediction of depression among medical check-ups of 433,190 patients: A nationwide population-based study," Psychiatry Res, vol. 293, p. 113474, Nov. 2020, doi: 10.1016/J.PSYCHRES.2020.113474.
[47] W. Yin et al., "Altered neural flexibility in children with attention-deficit/hyperactivity disorder," Mol Psychiatry, vol. 27, no. 11, pp. 4673–4679, Nov. 2022, doi: 10.1038/S41380-022-01706-4;TECHMETA.
[48] "DAIC-WOZ Database." Accessed: Sep. 13, 2025. [Online]. Available: https://dcapswoz.ict.usc.edu/
[49] "The Depression Dataset." Accessed: Sep. 13, 2025. [Online]. Available: https://www.kaggle.com/datasets/arashnic/the-depression-dataset
[50] "Depression Anxiety Stress Scales Responses." Accessed: Sep. 13, 2025. [Online]. Available: https://www.kaggle.com/datasets/lucasgreenwell/depression-anxiety-stress-scales-responses
[51] "Suicidality on Reddit." Accessed: Sep. 13, 2025. [Online]. Available: https://www.kaggle.com/datasets/thedevastator/c-ssrs-labeled-suicidality-in-500-anonymized-red
[52] "Human Stress Prediction." Accessed: Sep. 13, 2025. [Online]. Available: https://www.kaggle.com/datasets/kreeshrajani/human-stress-prediction
[53] "Depression_audio_dataset." Accessed: Sep. 13, 2025. [Online]. Available: https://www.kaggle.com/datasets/sanskjain/depression-audio-dataset/data?select=473_AUDIO.wav
[54] "Nurse Stress Prediction Wearable Sensors." Accessed: Sep. 13, 2025. [Online]. Available: https://www.kaggle.com/datasets/priyankraval/nurse-stress-prediction-wearable-sensors
[55] A. Srivastava, T. Suresh, S. P. Lord, M. S. Akhtar, and T. Chakraborty, "Counseling Summarization Using Mental Health Knowledge Guided Utterance Filtering," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 3920–3930, Aug. 2022, doi: 10.1145/3534678.3539187.
[56] J. Gratch, R. Artstein, G. Lucas, G. Stratou, S. S.- Lrec, and undefined 2014, "The distress analysis interview corpus of human and computer interviews.," academia.eduJ Gratch, R Artstein, GM Lucas, G Stratou, S Scherer, A Nazarian, R Wood, J BobergLrec, 2014•academia.edu, Accessed: Sep. 19, 2025. [Online]. Available: https://www.academia.edu/download/98764174/508_Paper.pdf
[57] E. Çiftçi, H. Kaya, … H. G.-2018 F. A., and undefined 2018, "The turkish audio-visual bipolar disorder corpus," ieeexplore.ieee.orgE Çiftçi, H Kaya, H Güleç, AA Salah2018 First Asian Conference on Affective Computing and Intelligent, 2018•ieeexplore.ieee.org, Accessed: Sep. 19, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8470362/
[58] D. P. Dickstein et al., "Fronto-temporal spontaneous resting state functional connectivity in pediatric bipolar disorder," Biol Psychiatry, vol. 68, no. 9, pp. 839–846, Nov. 2010, doi: 10.1016/j.biopsych.2010.06.029.