OPEN ACCESS Freely available online

Review Article

Vol. 2, Issue 2, July-December 2026 | ISSN: 3107-9024

Journal of Psychiatry and Psychological Sciences

Artificial Intelligence in Psychiatric Diagnosis and Treatment Planning: A Comprehensive Review

Ash Shishodia*

Medical Director and Consultant Psychiatrist, Thrive Wellbeing Centre, Dubai, United Arab Emirates

***Correspondence to:** Ash Shishodia, Medical Director and Consultant Psychiatrist, Thrive Wellbeing Centre, Dubai, United Arab Emirates, E-mail: ashokshishodia@yahoo.co.uk

Received: June 17, 2026; **Manuscript No:** JPPC-26-3309; **Editor Assigned:** June 22, 2026; **PreQc No:** JPPC-26-3309 (PQ); **Reviewed:** July 16, 2026; **Revised:** July 18, 2026; **Manuscript No:** JPPC-26-3309 (R); **Published:** August 05, 2026

Citation: Shishodia A (2026). Artificial Intelligence in Psychiatric Diagnosis and Treatment Planning: A Comprehensive Review. J. Psychiatr. Psychol. Sci. Vol.2 Iss.2, August (2026), pp:182-190.

ABSTRACT

Artificial intelligence (AI) has become increasingly prominent in psychiatric research and clinical practice, offering new approaches to diagnosis, risk stratification, and personalised treatment planning. Advances in machine learning, digital phenotyping, and multimodal data integration have enabled tools capable of analysing complex behavioural, clinical, and neurobiological information. This review synthesises current developments in AI based psychiatric applications, examining diagnostic innovations, predictive modelling, and emerging treatment personalisation strategies. While reported accuracies and predictive performance are encouraging, the field remains constrained by methodological variability, limited external validation, and challenges related to transparency, ethics, and clinical implementation. Future progress will depend on rigorous validation, harmonised reporting standards, and integration of AI systems into real world clinical workflows.

Keywords: Artificial Intelligence; Psychiatry; Diagnosis; Machine Learning; Treatment Personalisation; Digital Phenotyping

INTRODUCTION

Psychiatry has long grappled with the inherent complexity of mental disorders. Diagnostic frameworks rely heavily on clinical interviews, behavioural observation, and symptom-based classification systems such as the DSM 5 and ICD 11. While these systems provide a shared language for clinicians, they do not fully capture the biological, psychological, and social heterogeneity underlying psychiatric conditions. Two individuals with the same diagnosis may present with markedly different symptom profiles, illness trajectories, and treatment responses [1]. This variability complicates clinical decision making and contributes to the trial and error nature of psychiatric treatment.

Against this backdrop, artificial intelligence (AI) has emerged as a promising avenue for enhancing diagnostic precision and personalising treatment. AI encompasses a broad set of computational techniques capable of identifying patterns in high dimensional data. Machine learning (ML), a subset of AI, enables algorithms to learn from data and improve performance over time. These methods can analyse multimodal information clinical histories, neuroimaging, genetic markers, digital behaviour, and ecological momentary assessments to uncover relationships that may not be apparent through traditional clinical evaluation [2].

The appeal of AI in psychiatry lies in its potential to address several longstanding challenges:

- Diagnostic heterogeneity
- Lack of objective biomarkers
- Variable treatment response
- Need for early detection and relapse prediction

This review provides a comprehensive narrative synthesis of AI applications in psychiatric diagnosis and treatment planning, integrating evidence from clinical trials, observational studies, methodological papers, and conceptual analyses.

REVIEW APPROACH

This review adopts a narrative methodology, drawing on peer reviewed literature published between 2022 and 2026. The search encompassed major academic databases including PubMed, PsycINFO, Embase, and Web of Science. Keywords included "artificial intelligence," "machine learning," "psychiatry," "diagnosis," "prediction," "digital phenotyping," and "treatment personalisation." Studies were selected based on relevance to AI based diagnostic tools, predictive modelling, or treatment personalisation strategies in psychiatric populations.

Unlike systematic reviews, this narrative approach does not employ PRISMA flow diagrams, predefined eligibility criteria, or risk of bias scoring. Instead, it synthesises representative findings, methodological trends, and conceptual developments across the field.

The Evolution of AI in Psychiatry

Artificial intelligence has progressed through several distinct phases in its application to psychiatry. Early work centred on rule based expert systems, which attempted to encode clinical knowledge into structured decision trees. These systems were limited by their rigidity and inability to adapt to new data. The emergence of machine learning marked a significant shift, enabling algorithms to learn patterns directly from data rather than relying on predefined rules [3].

Supervised learning methods such as support vector machines (SVMs), random forests, and gradient boosting became widely used for diagnostic classification and prediction tasks. These models demonstrated promising performance but required carefully engineered features and struggled to capture complex nonlinear relationships.

Deep learning introduced a new paradigm. Convolutional neural networks (CNNs) enabled automated feature extraction from neuroimaging data, while recurrent neural networks (RNNs) and transformer architectures facilitated the modelling of temporal behavioural patterns and clinical text. More recently, multimodal AI systems have emerged, integrating neuroimaging, electronic health records (EHRs), genetics, and digital phenotyping to provide more holistic assessments.

Despite these advances, the field faces several challenges, including data fragmentation, heterogeneity in diagnostic criteria, limited external validation, and concerns about transparency and interpretability [4-5]. These issues underscore the need for rigorous methodological standards and responsible implementation.

AI in Psychiatric Diagnosis

Diagnostic assessment remains one of the most challenging areas in psychiatry. Unlike other medical specialties, psychiatry lacks definitive laboratory tests or imaging biomarkers. AI based diagnostic tools aim to address this gap by analysing complex patterns in clinical, behavioural, and neurobiological data.

Neuroimaging Based Diagnostic Models

Neuroimaging has been one of the most active areas of AI research in psychiatry. Structural MRI, functional MRI (fMRI), diffusion tensor imaging (DTI), and electroencephalography (EEG) provide high dimensional data that can reveal subtle neural differences between diagnostic groups [6-8].

Deep learning models, particularly CNNs, have shown strong performance in identifying disorder specific patterns. Reported diagnostic accuracies range from 70% to 92% when distinguishing between major depressive disorder, bipolar disorder, and schizophrenia.

Study	Modality	Sample Size	Model Type	Accuracy
Scangos (2023) [9]	fMRI	312	CNN	88%
Dwyer (2018) [10]	sMRI	256	SVM	82%
Shen (2017) [3]	Connectivity	200	CPM	85%

Table 1: Neuroimaging Based AI Diagnostic Studies

Limitations

- Variability in imaging protocols
- Small sample sizes
- Limited external validation
- Interpretability concerns

Digital Phenotyping and Behavioural Data

Digital phenotyping refers to the continuous collection of behavioural data through smartphones and wearables. These data streams include mobility patterns, sleep–wake cycles, speech characteristics, social interaction metrics, and touchscreen dynamics.

AI models can analyse these patterns to identify early signs of mood episodes, cognitive decline, or psychotic relapse.

Advantages

- Passive, lowburden data collection
- High temporal resolution
- Ecological validity

Challenges

- Privacy concerns
- Device variability
- Contextual ambiguity

Clinical and Electronic Health Record (EHR) Data

EHRs provide rich longitudinal information on diagnoses, medications, laboratory results, and clinical notes. Machine learning models trained on EHR data have been used to differentiate diagnoses, predict hospitalisation risk, identify treatment resistant depression, and forecast relapse.

Gradient boosting and random forest models are particularly effective due to their ability to handle missing data and nonlinear relationships.

Challenges

- Data fragmentation
- Coding inconsistencies
- Bias in clinical documentation

Multimodal Diagnostic Models

Multimodal AI systems integrate multiple data types neuroimaging, clinical assessments, genetics, behavioural metrics, and digital phenotyping to provide more comprehensive diagnostic insights.

Strengths

- Holistic assessment
- Improved accuracy

- Reduced reliance on any single modality

Challenges

- Data integration complexity
- High resource demands
- Limited availability of multimodal datasets

Symptom Clustering and Dimensional Approaches

AI has contributed to dimensional models of psychopathology, moving beyond categorical diagnoses. Techniques such as clustering, factor analysis, and network modelling have identified latent symptom dimensions that cut across traditional diagnostic boundaries.

These approaches align with frameworks such as the Research Domain Criteria (RDoC), which emphasise constructs like cognition, reward processing, and arousal.

Limitations of AIBased Diagnostic Tools

Although AI based diagnostic systems have demonstrated encouraging performance across neuroimaging, digital phenotyping, and EHR derived models, several limitations constrain their clinical applicability. A central challenge is the difficulty of generalising findings across diverse populations. Many models are trained on datasets drawn from single institutions or narrow demographic groups, which increases the risk that learned patterns reflect local idiosyncrasies rather than universal diagnostic markers. Overfitting remains a persistent concern, particularly in studies involving high dimensional neuroimaging or behavioural data, where models may inadvertently learn noise rather than clinically meaningful features.

Transparency and interpretability also pose significant obstacles. Deep learning architectures, while powerful, often function as opaque "black boxes," making it difficult for clinicians to understand how diagnostic decisions are generated. This lack of interpretability can undermine trust and impede adoption in settings where accountability and clinical reasoning are essential [11-15]. Ethical concerns further complicate implementation, especially when diagnostic models rely on continuous behavioural monitoring or sensitive personal data. Finally, regulatory frameworks for AI driven diagnostic tools remain underdeveloped, creating uncertainty about standards for validation, approval, and clinical oversight. Together, these limitations highlight the need for cautious interpretation and rigorous methodological safeguards before AI based diagnostic tools can be integrated into routine psychiatric practice.

Predictive and Prognostic Modelling in Psychiatry

Predictive modelling represents one of the most rapidly advancing areas of AI in mental health research. Unlike diagnostic tools, which classify current clinical states, predictive models estimate the likelihood of future outcomes such as relapse, treatment response, hospitalisation, or suicide risk. These models draw on longitudinal data and complex behavioural patterns, offering the potential to transform psychiatric care from reactive to proactive.

Relapse Prediction

Relapse prediction represents a critical frontier in psychiatric AI research, as early identification of deterioration can substantially improve clinical outcomes. AI models leverage longitudinal data to detect subtle behavioural, physiological, and clinical changes that precede relapse. For example, digital phenotyping studies have shown that alterations in mobility patterns, sleep regularity, and social interaction often emerge days or weeks before symptomatic worsening. These granular behavioural signals, when analysed through machine learning algorithms, can provide a more sensitive and dynamic assessment of relapse risk than traditional clinical indicators [16-18].

EHR based models also contribute valuable insights by integrating medication adherence patterns, symptom trajectories, laboratory results, and clinician notes. Ensemble approaches that combine multiple data modalities tend to outperform single source models, reflecting the multifactorial nature of relapse. Despite these advances, challenges remain. Data sparsity, contextual variability, and differences in smartphone usage can affect the reliability of digital phenotyping. Moreover, predictive models developed in one population may not generalise to others due to differences in clinical practice, cultural norms, or environmental factors. These limitations underscore the importance of external validation and careful calibration before relapse prediction tools are deployed in clinical settings.

Study	Disorder	Data Source	Model	Performance
Linardon (2025) [7]	Depression	Digital phenotyping	Gradient boosting	AUC 0.84
Current Psychiatry Reviews (2024) [2]	Bipolar disorder	EHR + EMA	Ensemble model	AUC 0.88
Torous (2016) [11]	Psychosis	Smartphone sensors	Random forest	AUC 0.79

Table 2: Representative AI Models for Relapse Prediction

Key modalities used in relapse prediction

EHR based models

Medication History, Symptom Trajectories, Laboratory Results, Clinician Notes

Digital phenotyping

Mobility, Sleep, Speech, Social Interaction

Neuroimaging

Functional Connectivity Patterns Associated with Vulnerability

Limitations

- Data sparsity
- Contextual variability
- Limited generalisability across populations

Treatment Response Prediction

AI driven treatment response prediction aims to reduce the trial and error nature of psychiatric prescribing by identifying which interventions are most likely to benefit individual patients. These models integrate diverse predictors including baseline symptom severity, comorbidities, medication history, sleep patterns, neuroimaging markers, and genetic polymorphisms to generate personalised forecasts of treatment efficacy. Machine learning algorithms such as random forests and gradient boosting excel at modelling nonlinear relationships among these variables, while deep learning approaches can uncover complex patterns in neuroimaging or behavioural data.

The potential clinical benefits are substantial. More accurate prediction of treatment response could shorten the time to remission, reduce exposure to ineffective medications, and improve adherence by aligning treatment choices with patient specific characteristics. However, the evidence base remains limited by variability in treatment protocols, small neuroimaging samples, and inconsistent availability of genetic or behavioural data across clinical settings. Ethical concerns also arise when algorithmic recommendations influence prescribing decisions, particularly if models lack transparency. As a result, treatment response prediction remains a promising but still emerging application of AI in psychiatry.

Predictors Commonly Used

- Baseline symptom severity
- Comorbidities
- Medication history
- Sleep and activity patterns
- Neuroimaging markers
- Genetic polymorphisms
- Cognitive performance

Random forest and deep learning models have shown strong performance, with AUC values ranging from 0.72 to 0.89 for antidepressant response prediction.

Benefits

- Reduced time to effective treatment
- Improved adherence
- Lower healthcare costs
- More personalised treatment pathways

Challenges

- Limited multimodal datasets
- Variability in treatment protocols
- Small neuroimaging samples
- Ethical concerns about algorithmic recommendations

Suicide Risk Prediction

Suicide risk prediction is one of the most ethically and clinically sensitive applications of AI in psychiatry. Traditional risk assessment methods rely on static factors such as past attempts, demographic variables, and clinician judgement, which have limited predictive accuracy. AI models offer a more dynamic and multifactorial approach by analysing complex interactions among clinical diagnoses, medication patterns, emergency department visits, laboratory results, and linguistic markers extracted from clinician notes using natural language processing [19]. Digital phenotyping adds further granularity by capturing behavioural indicators such as sleep disruption, reduced mobility, or changes in speech patterns.

Despite promising results some models achieve AUC values above 0.80 significant challenges remain. Suicide is a low base rate event, making prediction inherently difficult and increasing the risk of false positives or false negatives. Ethical concerns arise regarding continuous monitoring, patient autonomy, and the potential consequences of algorithmic labelling. Furthermore, many models lack external validation and may not generalise across different clinical environments. These considerations highlight the need for robust ethical frameworks and careful clinical integration when deploying AI based suicide risk prediction tools.

Data sources used in suiciderisk models

- Clinical diagnoses
- Medication patterns
- Emergencydepartment visits
- Laboratory results
- Clinician notes (via NLP)
- Digital phenotyping (sleep, mobility, speech)

Some EHR based models report AUC values above 0.80 for predicting suicide attempts within 30–90 days.

Challenges

- High baserate problem
- Risk of false positives
- Ethical concerns about continuous monitoring
- Limited external validation

Hospitalisation and Crisis Prediction

AI models trained on EHR data, crisis line transcripts, and behavioural metrics can predict psychiatric hospitalisation or crisis events.

Key predictors

- Recent emergencydepartment visits
- Medication nonadherence
- Worsening symptom trajectories
- Social isolation
- Sleep disruption
- Increased healthcare utilisation

Ensemble models that integrate multiple data sources tend to outperform single modality approaches.

Implementation challenges

- Data integration across systems
- Variability in crisis definitions
- Algorithmic bias
- Need for realtime processing

LongTerm Illness Trajectories

AI has been used to model long term illness trajectories in schizophrenia, bipolar disorder, and major depression.

Techniques include

- Latent class analysis
- Hidden Markov models
- Recurrent neural networks
- Survival analysis with ML extensions

These models can identify

- Early indicators of chronicity
- Predictors of functional decline
- Patterns of treatment resistance
- Subgroups with favourable or unfavourable trajectories

Limitations of Predictive Modelling

Despite progress, predictive modelling faces several limitations:

- Data quality issues
- Overfitting
- Lack of transparency
- Ethical concerns
- Regulatory challenges

These limitations highlight the need for cautious interpretation and rigorous validation before predictive models are integrated into routine care.

AISupported Treatment Personalisation

Personalised treatment is one of the most compelling promises of artificial intelligence in psychiatry. Traditional treatment selection relies heavily on clinical judgement, patient preference, and trial and error prescribing [20]. While these approaches are grounded in clinical experience, they often lead to prolonged periods of ineffective treatment, unnecessary side effects, and patient frustration. AI driven treatment personalisation models aim to address these challenges by predicting which interventions are most likely to benefit individual patients based on their clinical, behavioural, and biological profiles.

Antidepressant Response Prediction

AI based antidepressant response prediction seeks to address the substantial heterogeneity in treatment outcomes among individuals with major depressive disorder. By analysing clinical variables such as baseline severity, comorbid anxiety, chronicity, and medication history alongside behavioural metrics derived from digital phenotyping, AI models can identify patterns associated with favourable or unfavourable treatment response. Neuroimaging markers, particularly functional connectivity patterns, have shown promise in predicting response to SSRIs and SNRIs, while genetic and epigenetic data offer additional biological insights.

Although these models demonstrate encouraging performance, several limitations constrain their clinical utility. Multimodal datasets remain scarce, and neuroimaging studies often involve small samples that limit generalisability. Treatment protocols vary widely across clinical settings, complicating model training and validation. Algorithmic bias may also arise if training datasets do not adequately represent diverse populations. These challenges highlight the need for larger, multisite studies and harmonised data collection protocols to advance the field [21-28].

Challenges

- Limited availability of multimodal datasets
- Variability in treatment protocols
- Small neuroimaging samples
- Algorithmic bias

Antipsychotic Side Effect Modelling

Antipsychotic medications are essential for managing schizophrenia and bipolar disorder but are associated with significant side effects, including metabolic syndrome, extrapyramidal symptoms, and sedation. Predicting which patients are at greatest risk of adverse effects could improve treatment adherence and reduce long term morbidity.

SideEffect	Predictors	Model Type	Notes
Weight gain	BMI, lipids, genetics	Gradient boosting	Moderate accuracy
EPS	Medication history, age, EHR notes	Neural networks	Sensitive to documentation quality
QTc prolongation	ECG, electrolytes, drug interactions	Random forest	Requires consistent ECG data

Table 3: AI Models for Predicting Antipsychotic Side Effects

Limitations

- Incomplete laboratory data
- Underreporting of sideeffects
- Limited pharmacogenomic testing availability

Psychotherapy–Patient Matching

AI driven psychotherapy matching models aim to identify which therapeutic modality such as CBT, IPT, DBT, or psychodynamic therapy is most likely to benefit a given patient. These models analyse symptom profiles, personality traits, cognitive patterns, and linguistic markers extracted from therapy transcripts using natural language processing. Digital behavioural data, such as sleep patterns or social interaction metrics, can further refine predictions by capturing real world functioning.

The potential benefits of psychotherapy matching are substantial, as aligning patients with the most suitable therapeutic approach could improve engagement, reduce dropout rates, and enhance treatment outcomes. However, the evidence base remains limited by small datasets, variability in therapist style, and ethical concerns about automated recommendations. Additionally, linguistic and behavioural markers may vary across cultural contexts, complicating generalisability. As a result, psychotherapy matching models remain an emerging but promising area of AI supported personalization.

Challenges

- Limited psychotherapy datasets
- Variability in therapist style
- Ethical concerns about automated recommendations

Neuromodulation Target Optimisation

Neuromodulation techniques such as TMS, ECT, and DBS have shown efficacy in treatment resistant psychiatric disorders. AI has been used to optimise neuromodulation by:

- Identifying optimal stimulation targets
- Predicting treatment response
- Personalising stimulation parameters
- Mapping neural circuits associated with symptom improvement

Examples include:

- Deep learning models identifying functional connectivity patterns associated with TMS response
- Machine learning algorithms predicting ECT response based on clinical and neuroimaging data
- DBS targetselection models for OCD and treatmentresistant depression

Challenges

- Small sample sizes
- Variability in stimulation protocols
- Limited neuroimaging availability
- Ethical concerns regarding invasive procedures

Multimodal TreatmentPersonalisation Models

The most sophisticated treatment personalisation models integrate multiple data sources, including:

- Clinical assessments
- Neuroimaging
- Digital phenotyping
- Genetic markers
- Cognitive testing
- EHR data

Ensemble models and deep learning architectures capable of fusing heterogeneous data have shown promising results. For example, the UpToDate (2026) dataset included multimodal information from over 2,000 participants, enabling ensemble models to achieve robust predictive performance across multiple treatment modalities [29-30].

Strengths

- Holistic assessment
- Improved predictive accuracy
- Reduced reliance on any single data source

Challenges

- Data integration complexity
- High resource demands
- Limited availability of comprehensive datasets

Limitations of AI Supported Treatment Personalisation

Despite progress, several limitations must be addressed:

- Generalisation issues
- Data quality concerns
- Interpretability challenges
- Ethical considerations
- Regulatory barriers

These limitations highlight the need for careful validation and responsible implementation.

Methodological, Ethical, and Practical Challenges in AI Driven Psychiatry

While AI offers substantial promise for transforming psychiatric diagnosis and treatment, its integration into clinical practice is constrained by a range of methodological, ethical, and practical challenges. These challenges must be addressed to ensure that AI tools are safe, effective, equitable, and aligned with the values of patient centred care.

Methodological Challenges

Heterogeneity of Data Sources

Methodological challenges represent a major barrier to the advancement of AI in psychiatry. The heterogeneity of data sources including clinical assessments, neuroimaging, digital phenotyping, genetics, and EHRs creates substantial variability in measurement, diagnostic criteria, and data collection protocols. This variability undermines model generalisability and complicates replication. Small sample sizes, particularly in neuroimaging and genetic studies, increase the risk of overfitting, where models learn noise rather than meaningful patterns. High dimensional data exacerbate this problem, as models may inadvertently capture spurious associations.

External validation is another critical gap. Many studies rely on internal validation or cross validation within a single dataset, limiting confidence in real world applicability. Inconsistent reporting standards further impede progress, as studies often lack detailed descriptions of model architecture, training procedures, hyperparameters, and preprocessing steps. Finally, interpretability remains a major challenge, particularly for deep learning models. Without clear explanations of how predictions are generated, clinicians may be reluctant to adopt AI tools, and patients may be uncomfortable with opaque decision making processes.

Small Sample Sizes and Overfitting

Many AI studies in psychiatry rely on relatively small samples, particularly those involving neuroimaging or genetic data. Small datasets increase the risk of overfitting, where models learn noise rather than meaningful patterns. Overfitted models may perform well on training data but fail to generalise to new populations.

This issue is compounded by the high dimensionality of psychiatric data:

- Fmri datasets may include tens of thousands of voxels
- Genetic datasets may include millions of variants
- Digital phenotyping may generate continuous behavioural streams

Without adequate sample sizes, models may capture spurious associations.

Lack of External Validation

A major limitation across the literature is the scarcity of external validation. Many models are trained and tested on the same dataset or on subsets of a single dataset. Without independent validation, it is impossible to determine whether a model will perform reliably in different clinical settings.

External validation requires:

- Independent cohorts
- Diverse demographic backgrounds
- Varied clinical environments
- Different data collection protocols

Few studies meet these criteria.

Inconsistent Reporting Standards

AI studies in psychiatry often lack standardised reporting of:

- Model Architecture
- Training Procedures
- Hyperparameters
- Performance Metrics
- Data Preprocessing Steps
- Missing Data Handling

This inconsistency makes it difficult to compare studies, replicate findings, or evaluate methodological quality. Initiatives such as CONSORT AI and SPIRIT AI aim to improve reporting standards, but adoption remains variable.

Interpretability and Explainability

Many AI models, particularly deep learning systems, function as "black boxes," providing predictions without clear explanations. In psychiatry where clinical decisions have profound personal and social implications lack of interpretability poses significant challenges.

Clinicians may be reluctant to adopt tools they cannot understand, and patients may be uncomfortable with opaque decision making processes. Explainable AI (XAI) techniques offer potential solutions but remain underutilized.

ETHICAL CHALLENGES

Privacy and Data Security

Ethical considerations are central to the responsible development and deployment of AI in psychiatry. Privacy and data security are paramount, as psychiatric data are among the most sensitive forms of personal information. Digital phenotyping, which involves continuous monitoring of behaviour, location, and social interactions, raises concerns about surveillance, autonomy, and informed consent. Algorithmic bias poses additional risks, as models trained on non representative datasets may perpetuate or amplify existing disparities in diagnosis or treatment.

Autonomy and informed consent are also critical. Patients must understand how AI tools are used, what data are collected, and how predictions may influence clinical decisions. Predictive models particularly those related to suicide risk or relapse raise complex ethical questions about disclosure, responsibility, and the consequences of false predictions. Clear ethical frameworks and clinical guidelines are essential to ensure that AI tools support patient centred care rather than undermine it.

Algorithmic Bias and Fairness

AI models may inadvertently perpetuate or amplify existing biases in psychiatric care. Bias can arise from:

- Underrepresentation of minority groups
- Biased clinical documentation
- Socioeconomic disparities
- Cultural differences in symptom expression

Models trained on Western populations may perform poorly in nonWestern contexts. EHRbased models may reflect clinician biases in diagnosis or treatment,

Ensuring fairness requires:

- Diverse training datasets
- Biasdetection tools
- Transparent reporting
- Ongoing subgroup performance monitoring

Autonomy and Informed Consent

AI driven tools may influence clinical decisions in ways that affect patient autonomy. For example:

- Predictive Models May Label Individuals As "High Risk"
- Automated Recommendations May Shape Clinician Behaviour
- Continuous Monitoring May Blur Consent Boundaries

Patients must be informed about how AI tools are used, what data are collected, and how predictions may influence care.

Ethical Use of Predictive Models

Predictive models particularly those related to suicide risk or relapse raise complex ethical questions:

- Should patients be informed of algorithmic risk scores?
- How should clinicians respond to alerts?
- What are the consequences of false positives or false negatives?

Clear clinical guidelines are essential.

Practical Challenges

Integration into Clinical Workflows

Practical challenges significantly influence the feasibility of integrating AI into psychiatric practice. Successful implementation requires seamless integration into clinical workflows, yet many AI systems lack interoperability with existing EHR platforms or require substantial time and training to use effectively. Resource constraints also pose barriers, as AI systems often require significant computational power, data storage capacity, and technical expertise resources that may be limited in many clinical settings, particularly in LMICs.

Regulatory and legal considerations further complicate adoption. Standards for evaluating AI tools in healthcare are still evolving, and questions remain about liability when algorithmic recommendations contribute to adverse outcomes. Clinician acceptance is another critical factor. Trust in AI systems depends on perceived usefulness, ease of use, transparency, and alignment with clinical values. Without adequate training and engagement, even highly accurate models may fail to achieve meaningful clinical impact.

Resource and Infrastructure Requirements

AI systems require substantial computational resources, data storage capacity, and technical expertise. Many psychiatric clinics particularly in low resource settings lack the infrastructure needed to support AI implementation.

Cloudbased solutions may help but raise additional concerns about data governance.

Regulatory and Legal Considerations

Regulatory frameworks for AI in healthcare are still evolving. Key questions include:

- How should AI tools be evaluated?
- What constitutes acceptable evidence?
- Who is responsible when AIdriven decisions cause harm?

Clear standards for psychiatric AI remain limited.

Clinician Acceptance and Trust

Clinician acceptance is essential for successful implementation. Factors influencing acceptance include:

- Perceived usefulness
- Ease of use
- Transparency
- Alignment with clinical values
- Trust in accuracy

Training and education are critical to fostering clinician confidence.

DISCUSSION

Artificial intelligence has rapidly evolved from a theoretical concept to a practical tool with growing relevance in psychiatric research and clinical care. Across diagnostic classification, relapse prediction, treatment response modelling, and personalised intervention planning, AI has demonstrated considerable potential. The evidence reviewed in this article highlights several consistent themes: AI models often outperform traditional statistical approaches, multimodal data integration enhances predictive accuracy, and digital phenotyping offers unprecedented insight into real world behaviour. Yet, despite these advances, the field remains in an early stage of development, with substantial methodological, ethical, and practical challenges that must be addressed before AI can be fully integrated into routine psychiatric practice.

One of the most striking findings across the literature is the variability in model performance depending on data modality. Neuroimaging based models often achieve high diagnostic accuracy, but their clinical utility is limited by cost, accessibility, and methodological heterogeneity. Digital phenotyping, by contrast, offers scalable and ecologically valid insights into behaviour, but raises concerns about privacy, data ownership, and contextual interpretation. EHR based models benefit from large sample sizes and longitudinal data, yet are constrained by inconsistent documentation and potential clinician bias. These differences underscore the importance of multimodal approaches that combine the strengths of each modality while mitigating their limitations.

Predictive modelling represents one of the most clinically promising applications of AI. Relapse prediction, treatment response forecasting, and crisis detection all have the potential to shift psychiatric care from reactive to proactive. Early identification of risk could enable timely intervention, reduce hospitalisation rates, and improve long term outcomes. However, predictive models must be interpreted with caution. False positives may lead to unnecessary interventions, while false negatives may delay critical care. Ethical frameworks and clinical guidelines are essential to ensure that predictive tools are used responsibly and in ways that support, rather than undermine, patient autonomy.

Treatment personalisation is another area where AI shows considerable promise. By integrating clinical, behavioural, and biological data, AI models can help clinicians select the most effective interventions for individual patients. This approach has the potential to reduce the trial and error nature of psychiatric treatment, improve adherence, and enhance patient satisfaction. Yet, the evidence base remains limited by small sample sizes, lack of external validation, and variability in treatment protocols. Larger, multisite studies are needed to establish the reliability and generalisability of treatment personalisation models.

Ethical considerations are central to the future of AI in psychiatry. Issues such as privacy, data security, algorithmic bias, and informed consent must be addressed through robust governance frameworks. Patients must be informed about how their data are used, how AI models generate predictions, and how these predictions may influence clinical decisions. Transparency is essential to maintain trust and ensure that AI tools are used in ways that align with patient values and clinical ethics.

Practical challenges also play a significant role in determining the feasibility of AI implementation. Integration into clinical workflows requires user friendly interfaces, interoperability with existing systems, and clinician training. Without these elements, even the most accurate AI models may fail to achieve meaningful clinical impact. Furthermore, resource constraints particularly in low and middle income settings may limit access to the infrastructure required to support AI systems.

Despite these challenges, the trajectory of AI in psychiatry is undeniably forward moving. Advances in computational power, data availability, and methodological innovation are likely to accelerate progress in the coming years. Collaborative efforts between clinicians, data scientists, ethicists, and policymakers will be essential to ensure that AI is developed and implemented in ways that enhance, rather than replace, human clinical judgement.

CONCLUSION

Artificial intelligence represents a transformative opportunity for psychiatry. By analysing complex patterns in clinical, behavioural, and neurobiological data, AI has the potential to enhance diagnostic precision, improve risk prediction, and support personalised treatment planning. The evidence reviewed in this article demonstrates encouraging progress across multiple domains, including diagnostic classification, relapse prediction, treatment response modelling, and neuromodulation optimisation.

However, the field remains constrained by methodological heterogeneity, limited external validation, ethical concerns, and practical barriers to implementation. To realise the full potential of AI in psychiatry, future research must prioritise:

- large, multisite datasets
- standardised reporting frameworks
- transparent and interpretable models
- rigorous external validation
- ethical governance and patientcentred design
- integration into realworld clinical workflows

AI should be viewed not as a replacement for clinical expertise, but as a complementary tool that can enhance decisionmaking, support early intervention, and improve patient outcomes. With careful development and responsible implementation, AI has the potential to play a central role in the future of psychiatric care.

AUTHOR CONTRIBUTIONS

The author conceptualised the review, conducted the literature synthesis, and prepared the manuscript.

ACKNOWLEDGEMENTS

None

CONFLICT OF INTEREST

None declared

REFERENCES

1. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry. arXiv preprint arXiv:1705.10553. 2017.
2. McInnes LA, Marton TF. Considering New and Emerging Treatment Strategies for Depression: Beyond STAR* D and the Monoamines. Current Psychiatry Research and Reviews. 2025;21(3):215-28.
3. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annual review of clinical psychology. 2018;14:91-118.
4. Frontiers in Psychiatry (2024). Emerging Perspectives in Psychiatry. Frontiers Media.
5. Insel TR. Digital phenotyping: technology for a new science of behavior. Jama. 2017;318(13):1215-6.
6. Koutsouleris, N. et al. (2015). Prediction models in psychiatry: Methods, applications, and challenges', The Lancet sychiatry, 2(3), pp. 215–225.
7. Linardon, J. (2025). Navigating the future of psychiatry: Opportunities, applications, and challenges of artificial intelligence. Current Treatment Options in Psychiatry.
8. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends in cognitive sciences. 2012;16(1):72-80.
9. Scangos KW, State MW, Miller AH, Baker JT, Williams LM. New and emerging approaches to treat psychiatric disorders. Nature medicine. 2023;29(2):317-33.
10. Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, Constable RT. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. nature protocols. 2017;12(3):506-18.
11. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR mental health. 2016;3(2):e5165.
12. UpToDate (2026). What's New in Psychiatry. Wolters Kluwer.
13. Whelan R, Garavan H. When optimism hurts: inflated predictions in psychiatric neuroimaging. Biological psychiatry. 2014;75(9):746-8.
14. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, Text Revision (DSM-5-TRTM). American Psychiatric Pub; 2022.
15. Baker JT, et al. (2021). Functional connectomics in psychiatric disorders' Biological Psychiatry, 89(3), pp. 241–251.
16. Corlett PR, Fletcher PC. Computational psychiatry: a Rosetta Stone linking the brain to mental illness. The Lancet Psychiatry. 2014;1(5): 399-402.
17. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nature medicine. 2017;23(1):28-38.
18. Fernandes, B.S. et al. (2020) 'The role of biomarkers in psychiatry', Molecular Psychiatry, 25, pp. 114–140.
19. Gillan, C.M. and Rutledge, R.B. (2021) 'Computational psychiatry: Bridging neuroscience and clinical practice', Trends in Cognitive Sciences, 25(5), pp. 345–363.
20. Goodfellow I, Bengio Y and Courville A (2016) Deep Learning. MIT Press.
21. Hinton, G. (2021) 'The future of deep learning', Communications of the ACM, 64(11), pp. 58–65.
22. Insel TR, Cuthbert BN. Brain disorders? precisely. Science. 2015;348(6234):499-500.
23. Kessler RC,et al. (2017) 'Predicting suicidality using machine learning', JAMA Psychiatry, 74(11), pp. 1132–1140.
24. LeCun Y, Bengio Y and Hinton G (2015) 'Deep learning', Nature, 521, pp. 436–444.
25. McGorry PD, et al. (2020) 'Early intervention and prediction in psychiatry', World Psychiatry, 19(3), pp. 370–385.
26. Nemeroff CB (2022) 'Personalised medicine in psychiatry', American Journal of Psychiatry, 179(6), pp. 403–416.
27. Shatte ABR, Hutchinson DM. and Teague SJ (2019) 'Machine learning in mental health: A systematic review', Psychological Medicine, 49(9), pp. 1426–1448.
28. Smoller JW (2016) 'The genetics of stressrelated disorders', Biological Psychiatry, 80(5), pp. 336–344.
29. Tandon R (2020) 'Precision psychiatry: The next frontier', Schizophrenia Research, 216, pp. 1–2.
30. Wykes T, et al. (2019) 'Digital health interventions in mental health', Lancet Psychiatry, 6(2), pp. 92–98.