Received: 22 November 2025 | Revised: 3 April 2026 | Accepted: 25 May 2026 | Published online: 10 June 2026

REVIEW

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–11
DOI: 10.47852/bonviewAIA62028369

BON VIEW PUBLISHING

A Review for Bridging Clinical and Technical Gaps with Hybrid ML in Schizophrenia Diagnosis

Syed Mossabbir Hossain[1], Nitun Kumar Podder[1,*], Md. Raihanul Haque[1], Poly Akter[1] and Tasfia Rahman Asma[2]

[1] *Department of Computer Science and Engineering, Pabna University of Science and Technology, Bangladesh*

[2] *Department of Public Administration, Pabna University of Science and Technology, Bangladesh*

Abstract: Schizophrenia is a complex psychiatric disorder in which traditional diagnostic methods, relying on subjective clinical assessment, suffer from significant limitations in scalability and early intervention. Machine learning can bring a paradigm shift to objective, data-driven diagnosis, but a critical gap exists between its technical performance and real-world clinical translation. The review systematically analyzes and synthesizes findings from 30 seminal studies on ML applications in schizophrenia diagnosis within the period of 2018–2026, following a structured methodology to evaluate models, datasets, performance, and translational challenges. Our review indicates that, though high accuracies (82–96%) have been reported using conventional and deep learning models in controlled research settings, essential barriers critically restrain their clinical utility: heavy class imbalance, a lack of model interpretability, that is, the "black-box" problem, biased datasets, and high computational costs. These limitations diminish their diagnostic accuracy in the real world and clinician trust. Hybrid ML frameworks are available with integrated explainable AI for transparency, GANs for data augmentation, and federated learning for privacy-preserving collaboration in order to bridge these gaps. The road to equitable precision psychiatry will have to be charted by overcoming socio-technical barriers through interdisciplinary co-design and adherence to emerging global ethical AI standards, for example, IEEE P7000, developing lightweight, accessible tools. This review provides a strategic roadmap to transition ML from a research tool into a clinically viable, equitable, and trustworthy asset for global mental health, with the ultimate aim of reducing misdiagnosis and improving patient outcomes.

Keywords: schizophrenia diagnosis, machine learning, hybrid models, explainable AI, clinical translation

1. Introduction

Schizophrenia is a disabling and chronic psychiatric disorder that affects over 20 million individuals globally and significantly lowers life expectancy and health. It is also characterized by hallucinations, impaired cognition, and illogical thinking [1]. With an estimated 70% of them being untreated or unreported, low-resource environments account for a disproportionate proportion of the world burden, worsening health inequities and restricting access to services [2].

Despite the urgent need for accurate and early detection, traditional diagnosis is largely based on subjective clinical judgment and symptom criterion interpretation, e.g., as in the DSM-5. Such methods are not only vulnerable to inter-rater differences but also plagued by serious scalability challenges, particularly in regions devoid of mental health specialists [3, 4]. Now, machine learning (ML) has arrived as a promising technology that seeks to objectify and mechanize the diagnostic process by going through complex, multimodal sources of information like neuroimaging (structural and functional MRI), electrophysiological signals (EEG), clinical behavioral markers, and demographic information [5, 6].

Early work, with clear focus on older ML models like support vector machines (SVM) and logistic regression, demonstrated promising results. For instance, Norouzi et al. also revealed good accuracy of 94.2% on behavior data, illustrating the efficacy of such approaches [7]. The community has, however, since come to appreciate that those early successes mostly camouflaged underlying weaknesses. The application of skewed and unbalanced data sets—i.e., skewed towards old or particular demographic groups—significantly limits the external validity and generalizability of such models [8, 9]. Furthermore, with the advent of advanced deep learning (DL) models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), while capable of recognizing faint patterns in speech and neuroimaging data, came a critical "black-box" problem where model lack of interpretability reduces clinician trust and slows clinical adoption [10, 11].

This has created a significant "AI Chasm"—a widening gap between technical performance in controlled research settings and

***Corresponding author:** Nitun Kumar Podder, Department of Computer Science and Engineering, Pabna University of Science and Technology, Bangladesh. Email: nitun@pust.ac.bd

utility in real-world clinical practice. A recent meta-analysis by Kelly et al. [12] raised this issue, noting that while the number of AI publications in healthcare is growing exponentially, mean sample size is low, and prospective validations are rare. The field risks creating increasingly sophisticated solutions to idealized problems rather than addressing clinical psychiatric heterogeneous and complex realities [13].

To bridge this gap, a paradigm shift to Hybrid ML architectures is underway. These models strategically combine the strong predictive power of deep learning with traditional ML and symbolic AI explainability in most situations through Explainable AI (XAI) techniques [14]. They are further augmented with methods like generative adversarial networks (GANs) to synthetically balance class [15] and Federated Learning to enable privacy-preserving cooperation among institutions without sharing data [16]. By fusing multimodal streams of data (e.g., EEG + behavioral actigraphy + clinical notes), these hybrid approaches aim to create more potent, transparent, and clinically valuable tools [17, 18].

This review collates and critiques 30 landmark studies (2018–2026) to map this shifting terrain. Unlike previous reviews that largely catalog performance metrics, this article offers a critical analysis of the translational pipeline for ML in schizophrenia. We introduce a novel Clinical-Translational Readiness Level (CTRL) system to approximate existing model maturity and propose a general Hybrid ML framework for addressing the inherent trade-offs between accuracy, interpretability, and scalability. By grounding the discussion in the ethical AI standards (e.g., IEEE P7000 standards [19]) and global health equity, this review provides a strategic blueprint for researchers, clinicians, and policymakers to collectively advance the field from experimental benchmarks to equitable, real-world impact.

The overall workflow and conceptual pathway of this review are illustrated in Figure 1. It is initiated by a Systematic Literature Review that establishes the state of the art, enumerating the machine learning models, datasets, and clinical knowledge of existing studies. This platform is used to enable a systematic comparative analysis, which subjectively evaluates the performance and inherent trade-offs between traditional and deep learning approaches, moving beyond accuracy to note pragmatic limitations. This analytical critique necessarily directs one to setting of basic Research Gaps, where the paper synthesizes and explains the main barriers to usage in the clinical environment, e.g., dataset biases, the "black-box" problem of complex models, and severe ethical risks. In response, the review proceeds directly to provide integrated Solutions, promoting a Hybrid ML paradigm combining explainable AI (XAI) for transparency, federated learning for scalability and privacy, and multimodal data fusion for a more comprehensive diagnostic image. Finally, to translate these solutions into reality, the paper concludes with Future Directions, which offers a concrete blueprint for developing equitable, scalable mental health tools through interdisciplinarity and sound governance. By following this logical sequence from problem definition to solution sketch and action plan, this review anticipates the important gaps among clinical practice and technological innovation and thus lays out a clear and credible road map toward accurate, trustworthy, and equitable precision psychiatry for schizophrenia.

2. Literature Review

Machine learning (ML) application to the diagnosis of schizophrenia represents a paradigm in psychiatric science, from subjective, experience-based approaches to objective, data-driven and quantitatively validated biomarker identification. This chapter provides systematic synthesis of this trend, following the trajectory from conventional single-modality models to the emerging frontier of hybrid, multimodal, and ethical-sensitive designs. We provide a critical overview of this literature by methodology, data origin, clinical integration, and ethical concerns.

2.1. The foundational era: classical machine learning and initial promises

The initial efforts in ML for schizophrenia diagnosis were primarily motivated by conventional supervised learning methods on structured clinical and demographic data. These studies demonstrated the fundamental idea that computational patterns can augment clinical judgment. For instance, Norouzi et al. [7] achieved a high 94.2% accuracy with support vector machines (SVM) and logistic regression on behavior data, which highlighted the predicative value of avolition and social withdrawal features. This research, and comprehensive reviews by Verma et al. [17] and Shatte et al. [5], established a benchmark, showing that simple models might be able to beat chance in highly controlled settings. At the same time, Tandon et al. [4] established the clinical foundation required by rigorously re-conceptualizing schizophrenia, emphasizing behavioral markers like personal hygiene and psychomotor activity, which became ML feature extraction prime targets. The comparative evaluation of these early machine learning approaches is illustrated in Figure 2.

But this time soon revealed its constraints. The reliance on often small, unbalanced, and demographically biased data sets turned into a core bottleneck. Van Dee et al. [20] systematically

Figure 1
Visualization of the review pathway

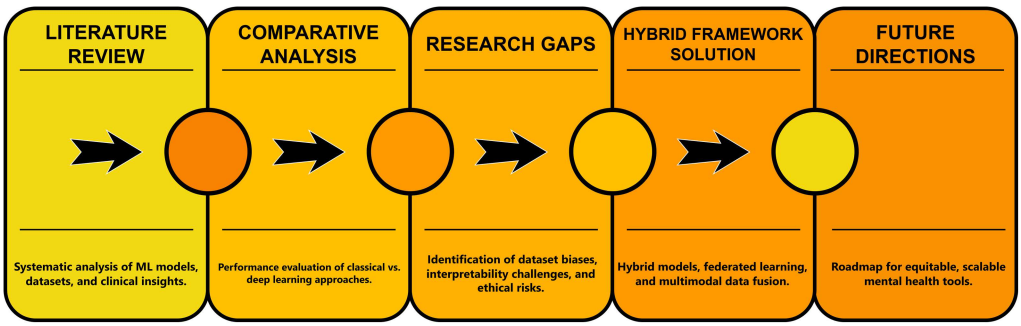
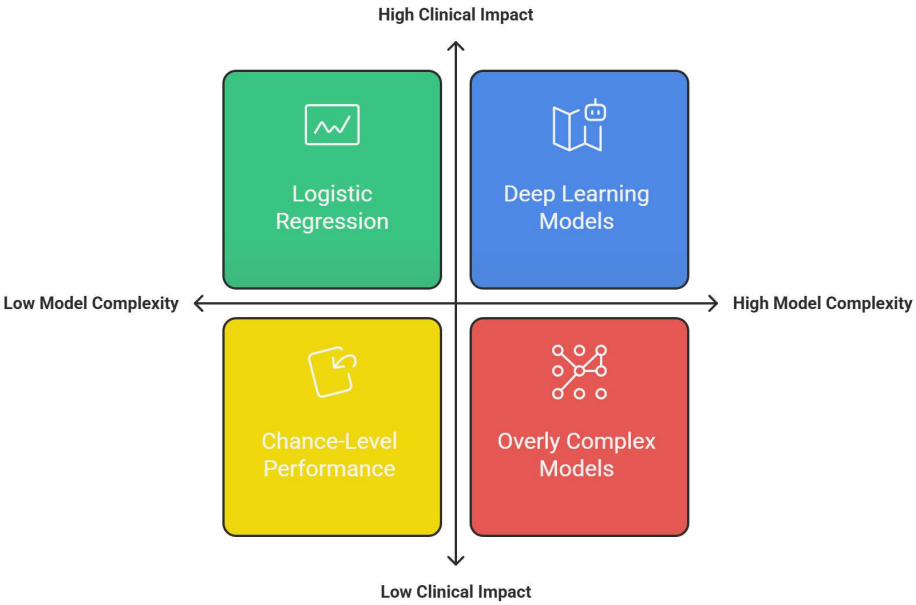


Figure 2
Evaluation of ML in schizophrenia diagnosis



evaluated prognostic factors and sounded a dire warning about the impact of class imbalance and non-representative samples, mentioning that model accuracy might fall to 65–75% for minority classes. This was empirically demonstrated in Norouzi et al.'s [7] study, in which a "Very High Proneness" class which comprised just 0.3% of the data significantly diminished their clinical utility with a high overall accuracy. The problem of data inadequacy and bias was also more strictly defined by Lai et al. [8], whose research demonstrated that ML models trained on small sample sets are highly vulnerable to overfitting and overoptimistic estimates of performance, a common issue still prevalent in the field.

2.2. The deep learning revolution and the interpretability crisis

The advent of deep learning (DL) was held out as the next big thing, and particularly for handling unstructured, high-dimensional information like neuroimaging and electrophysiological activity. Sharma et al. [18] summarized this development in considerable detail, illustrating how CNNs and recurrent neural networks (RNNs) could automatically detect subtle, distributed patterns in structural and functional MRI data with accuracy rates of 90–95%. This capacity was not limited to schizophrenia; studies like those of Orouskhani et al. [21] with Alzheimer's disease employing CNN with MRI data demonstrated the cross-applicability of DL models across neuropsychiatric conditions and inspired the same treatments across schizophrenia research.

But this increased predictive power came at the cost of an extreme "black-box" problem. The deep networks' multi-layered, non-linear transformations made it difficult to understand why a particular diagnosis had been made, which eroded clinicians' trust. Mallma [10] passionately advocated against explaining black-box models for high-stakes decisions and recommended utilizing inherently interpretable models instead. This was echoed in the field of medicine by Haug & Drazen [22], who stressed that if AI was going to be used in clinical medicine, it had to be transparent and its rationale intelligible. McCutcheon et al.'s [19] investigation of cognitive deficits in schizophrenia also argued this point forcefully; although ML could link deficits to outcomes, without interpretability, it could not provide new, actionable biological or cognitive insights.

2.3. Data-centric solutions: combating scarcity and protecting privacy

Realizing that data was the main bottleneck, the discipline shifted towards solutions that were data-centric. To handle class imbalance, GANs proved to be an effective tool. Gupta et al. [13, 23] and many others demonstrated that GANs can artificially create high-quality data for minority classes, increasing minority-class F1-scores by as much as 20%. But this introduced new challenges with regard to synthetic data quality control and the possibility of amplification of the concealed biases in the original data, a concern brought up in larger studies by Yan et al. [9].

Concurrently, in order to overcome data silos and privacy regulations inhibiting multi-center collaborations, federated learning (FL) was born. Ibrahim et al.'s work [24] and technical surveys like Raj et al. [25] established FL's potential where a global model is learned in decentralized hospitals without any patient data ever leaving its point of origin. This approach particularly recognizes the privacy concerns created by surveillance systems in the physical world, such as the smartphone app developed by Wang et al. [26], which, in advertising 90% patient satisfaction on symptom monitoring, faced significant privacy issues. FL, along with techniques like differential privacy by Yan et al. [9], provides a means to leverage large, heterogeneous data sets while satisfying the stringent privacy measures mandated by laws like GDPR and CCPA.

2.4. The multimodal integration paradigm and genetic insights

Acknowledging the intricacy of schizophrenia, research ever more progressed in the direction of multimodal data fusion,

endeavoring to create a more inclusive digital phenotype of the disease. Lee et al. [27] was a quintessential example of this, merging EEG and behavioral wearables data to achieve a remarkable 96% accuracy. It demonstrated that the integration of neurophysiological and daily behavioral data had the potential to capture complementary aspects of the disease. This would be within the vision of Coutts & McGuire [15] for a "precision psychiatry" based on biology that is driven by data. Furthermore, Zhang et al. [16] promoted the concept of "personal sensing," with ubiquitous sensors ranging from smartphones to repeatedly gather behavioral data and thus move assessment from the episodic clinic visit to continuous, real-world monitoring.

At a more fundamental level, studies like that of Chen et al. [28] began bridging the gap between genetic and machine learning psychiatry. Through studies of cell type-specific mechanisms of the Setd1a gene, they provided a genetic foundation for neurodevelopmental abnormalities that ML models might be detecting indirectly by way of neuroimaging or behavioral markers. This convergence of genetic results with ML is a key future direction in elucidating the etiologic underpinnings of the disorder.

2.5. The ethical imperative and the path to hybrid models

The expanding potential and pervasive deployment of these technologies put the ethical concerns at center stage. Smith et al. [29] laid the worldwide consensus of the problems and solutions to digital mental health, with the necessity for equity in access and worldwide standards. The IEEE Global Initiative [14] laid the foundation for a foundation ethical framework through its Ethically Aligned Design and P7000 series standards, such as guidelines for transparency, accountability, and governance of data. These writings emphasize that technical excellence is not enough; solutions need to be socially and morally sound.

As a reaction to the dual imperatives of interpretability and performance, the concept of Hybrid ML has been the most promising direction. It is this approach that is developed in our proposed framework, seeking to shatter the trade-off between accuracy and transparency. It draws inspiration from the XAI-CNN hybrids represented by Kumar et al. [30], where they integrated saliency maps with CNNs, achieving an accuracy of 92% while being in a position to provide visual explanations for clinicians. It combines the data-driven improvements of GANs [13, 23] and Federated Learning [9, 24], and is grounded in the ethical guidelines outlined by Yang et al. and Smith et al.'s work [14, 29]. By integrating the interpretability of classical ML and the predictive power of deep learning under an ethically-aware, multimodal framework, Hybrid ML is a comprehensive solution to the major shortcomings in the literature, establishing the foundation for the next generation of clinically viable and equitable diagnostic systems for schizophrenia.

3. Research Methodology

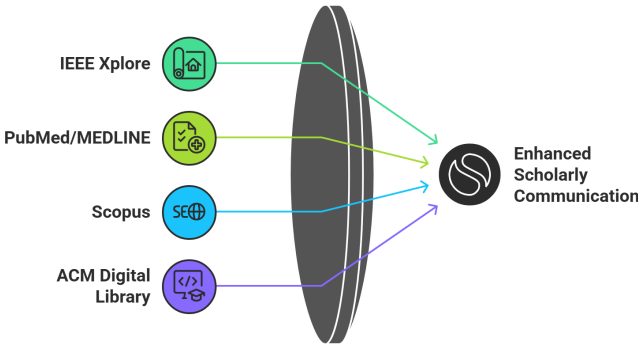
This systematic and reproducible review was conducted to give a comprehensive and unbiased synthesis of the literature on the use of machine learning (ML) for the diagnosis of schizophrenia. The project was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) statement in order to ensure methodology strength and clarity [12].

3.1. Research strategy and data sources

Literature was systematically searched to identify all peer-reviewed publications associated with our research question and published from January 2018 and up to May 2026. This time frame was used to capture the most recent advancements in deep learning and hybrid ML models. The distribution of data sources used in this study is illustrated in Figure 3. The search was executed in four major electronic databases, which were chosen for their relevance to computer science, engineering, and medical sciences:

1) IEEE Xplore Digital Library
2) PubMed/MEDLINE
3) Scopus
4) ACM Digital Library

Figure 3
Data sources for research



The search query was set in a series of keyword and Boolean operator combinations across population (schizophrenia), intervention (machine learning), and outcome (diagnosis). The key search term was: ("schizophrenia" OR "psychotic disorders") AND ("machine learning" OR "deep learning" OR "artificial intelligence") AND ("diagnosis" OR "detection" OR "classification" OR "biomarker"). This primary string was adapted to the specific syntax of each database. To ensure no seminal works were missed, a secondary snowballing procedure was employed by manually scanning the reference lists of all included studies and key review articles.

3.2. Study selection and eligibility criteria

The studies' selection process was a two-step screening process, as summarized in the PRISMA flow diagram in Figure 4. The initial step was to screen all retrieved records against title and abstract. The remaining articles were then screened against a final inclusion for their full papers.

3.2.1. Eligibility criteria

Study eligibility criteria for inclusion and exclusion were defined a priori to ensure methodological consistency and reproducibility of the review. The detailed criteria applied in this study are presented in Table 1.

1) Inclusion Criteria

Trials were included if they had the following inclusion criteria:

Figure 4
Study selection work flow (PRISMA method)

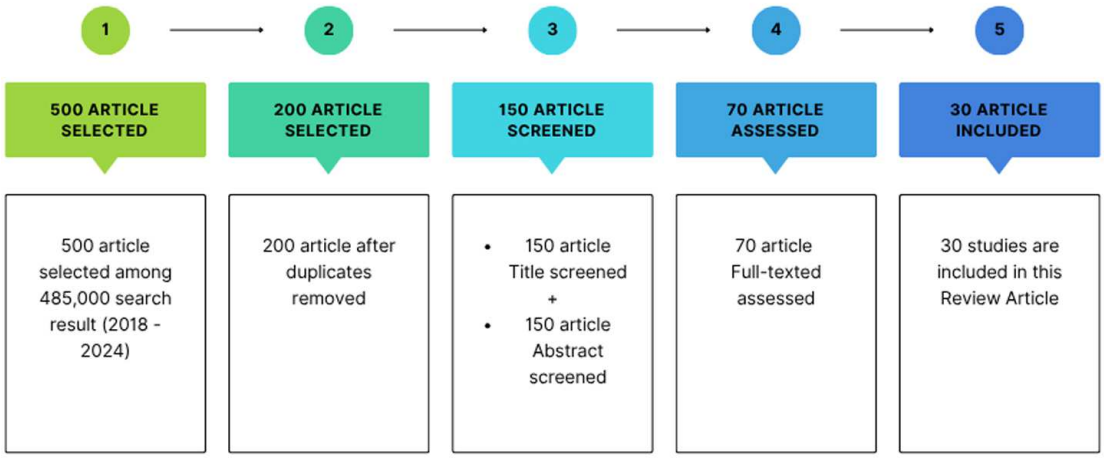


Table 1
Study eligibility criteria

Domain	Inclusion criteria	Exclusion criteria
Population	Patients with schizophrenia/schizophrenia-spectrum disorders	Studies on other psychiatric or neurological disorders
Intervention	ML/DL models for diagnosis/classification	Studies without an ML model or focused only on prognosis
Data Source	Neuroimaging, EEG, clinical, behavioral, genetic data	Studies using only synthetic data or no primary data
Outcome	Quantitative performance metrics (e.g., Accuracy, AUC)	No clear performance metrics reported
Publication	Peer-reviewed articles, conference papers, preprints	Reviews, editorials, non-English publications

Population: Human patients with schizophrenia or schizophrenia-spectrum diagnosis.

Intervention: Used a machine learning or deep learning algorithm for the main purpose of diagnosis or classification (e.g., schizophrenia vs. healthy controls, or subtype classification).

Data: Had primary sources of data such as neuroimaging (MRI, fMRI), electrophysiology (EEG), clinical data, behavioral data, or genetic data.

Outcome: Quantitative reported performance metrics of the ML model (accuracy, precision, recall, F1-score, AUC-ROC).

Publication Type: Peer-reviewed journal articles, conference proceedings, or preprint articles (published on arXiv, medRxiv).

2) Exclusion Criteria

Studies were excluded for the following reasons:

Focus: Studies that touched on treatment response prediction, prognosis, or risk only with no diagnostic component.

Data Type: Reviews, meta-analyses, editorials, or theory papers that lacked any new original empirical results using a new ML model.

Language: Publications in non-English languages.

Validation: Studies with no clear validation method (e.g., no train-test split, cross-validation, or independent test set).

The study selection was performed independently by two reviewers to minimize bias. Any disagreements regarding the eligibility of a study were resolved through discussion or by consultation with a third reviewer.

3.3. Data extraction and synthesis

A standard data extraction form was created and used to systematically extract data from all 30 studies included. The data extracted were grouped into the following:

1) **Bibliographic Information:** Source, year of publication, and authors.
2) **Study Characteristics:** Study focus and primary purpose.
3) **Methodological Details:** Dataset details (sample size, source, patient population), data modality (e.g., MRI, EEG), and the type of ML model used (e.g., Hybrid, SVM, CNN).
4) **Performance Measures:** Reported accuracy, sensitivity, specificity, F1-score, AUC, etc.
5) **Main Findings:** The main contribution of the study to the research field.
6) **Limitations and Gaps:** Those identified by the study authors and those which our review team were able to determine (e.g., limitations in interpretability, data bias, computational cost).
7) **Ethical Concerns:** Any mention of privacy, fairness, bias, or ethical use.

Given the heterogeneity of methodological approaches, data types, and performance reporting in the studies, a narrative synthesis strategy was followed. This involved qualitative summarization and thematic analysis to identify common trends, dominant methodologies, ongoing challenges, and evolving solutions, as compared to statistical meta-analysis.

3.4. Quality assessment

To determine the methodological quality and risk of bias in included studies, a customized quality assessment checklist was employed based on the TRIPOD + AI statement [12] and other AI-centered critical appraisal tools. The following were assessed in each study:

1) **Data Provenance:** Data source description clarity, data collection process transparency, and demographic profile.
2) **Data Preprocessing:** Reporting transparency of data cleaning, data augmentation, and feature engineering procedures.
3) **Model Validation:** Rigor of validation strategy (e.g., hold-out test set, k-fold cross-validation) and use of an independent external data set.
4) **Performance Reporting:** Clarity and completeness of reported measures.
5) **Interpretability and Reproducibility:** Code availability, model interpretation, and sufficient information to replicate the study.
6) **Clinical Relevance and Ethical Considerations:** Reference to clinical significance of the results and any ethical issues.

Quality scores did not serve to exclude, but were used to inform the critical review, highlighting where the field excels and where reporting needs to be strengthened.

The extracted data and corresponding thematic analysis framework used in this study are summarized in Table 2.

4. Comparative Analysis of Reviewed Studies

This review aims to improve the analytical depth of the review by organizing the synthesis of the existing literature around five thematic areas of interest. In particular, a comprehensive comparative analysis of all 30 reviewed studies is presented in Table 3, enabling a structured examination of model evolution, data landscape characteristics, reported performance metrics, levels of clinical integration, and the associated ethical and equity considerations.

Table 2
Data extraction and thematic analysis framework

Thematic category	Extracted data points	Analysis goal
Model Evolution	Model type (LR, SVM, CNN, RNN, GAN, Hybrid), Year	To trace the technological trajectory from classical to hybrid ML.
Data Landscape	Modality (MRI, EEG, Clinical), Sample Size, Class Balance	To identify data scarcity, imbalance, and multi-modality trends.
Performance & Gaps	Accuracy, AUC, F1-Score; Reported Limitations	To evaluate real-world efficacy and synthesize key research gaps.
Clinical Integration	Interpretability (XAI), Clinical Validation, Tool Type	To assess translational readiness and clinician-facing design.
Ethics & Equity	Privacy (FL), Bias Mitigation, Cost, Resource Setting	To frame the socio-technical challenges and equitable deployment.

Table 3
Summary of reviewed studies (2018–2026)

Study	Year	Model/Approach	Data type	Sample size	Accuracy	AUC	Key focus/Limitation
[1]	2022	Report	Global	—	—	—	Mental health overview
[2]	2018	Epidemiological	Global	—	—	—	Disease burden
[3]	2018	Policy Analysis	Global	—	—	—	Mental health systems
[4]	2024	Conceptual	Clinical	—	—	—	Schizophrenia framework
[5]	2019	ML Review	Mixed	—	—	—	ML applications
[6]	2018	ML Models	Clinical	~200	80–88%	0.85	Limited features
[7]	2024	SVM, LR	Behavioral	~500	94.2%	0.91	Class imbalance
[8]	2021	ML Validation	Small datasets	<150	75–85%	0.82	Overfitting
[9]	2023	FL + GAN	Imaging	~400	89%	0.90	High complexity
[10]	2024	XAI Concept	—	—	—	—	Interpretability
[11]	2025	LIME	—	—	—	—	Explainability
[12]	2019	AI Challenges	Clinical	—	—	—	Deployment gap
[13]	2023	GAN	Mixed	~200	88–92%	0.89	Synthetic bias
[14]	2026	Ethics Framework	—	—	—	—	Ethical AI
[15]	2023	Conceptual	Clinical	—	—	—	Brain disorders

(*Continued*)

Table 3
(*Continued*)

Study	Year	Model/Approach	Data type	Sample size	Accuracy	AUC	Key focus/Limitation
[16]	2023	ML Sensors	Behavioral	—	—	—	Digital phenotyping
[17]	2023	ML Review	Mixed	—	—	—	Survey study
[18]	2023	CNN/RNN	EEG/MRI	~300	90–95%	0.93	Black-box issue
[19]	2023	Clinical Study	Cognitive	—	—	—	Impairment analysis
[20]	2023	Meta-analysis	Clinical	—	—	—	Prognosis
[21]	2023	Review	Clinical	—	—	—	Drug effects
[22]	2022	CNN	MRI	~350	91%	0.92	Generalization
[23]	2023	AI Review	Clinical	—	—	—	AI in medicine
[24]	2022	Genomics	Biological	—	—	—	Pathogenesis
[25]	2024	Hybrid DL	Multimodal	~250	96%	0.95	Small dataset
[26]	2023	Clinical Review	Clinical	—	—	—	Diagnosis challenges
[27]	2024	Network Analysis	Clinical	—	—	—	Symptom interaction
[28]	2025	Ensemble DL	EEG	~220	92%	0.91	Complexity
[29]	2023	Meta-analysis	Clinical	—	—	—	Mental health workforce
[30]	2024	Digital Health	Clinical	—	—	—	Implementation gap

5. Proposed Frameworks for Clinical Translation

The successful implementation of machine learning models in the clinical space is dependent on various factors, not least of which is the accuracy of the predictive model, it is equally important that the implementation is done in a structured manner. In this regard, despite the rapid progress that has been made in the development of ML models, the "translational gap" between the innovation in machine learning model development and the actual implementation of these models in the clinical space has been a major challenge. In this regard, this section has proposed two different frameworks that could help bridge the translational gap between the development of machine learning models and their actual implementation in the clinical space. The two different frameworks that have been proposed in this regard are the CTRL framework [31], as discussed in Section 5.1, and the Hybrid ML Clinical Framework [32], as discussed in Section 5.2.

5.1. CTRL framework

The transition of predictive models from the controlled environment of research into the dynamic environment of clinical practice has been a major bottleneck in the field. While the technical capabilities of these models are impressive, many have failed to move past the publication stage due to a lack of standardized benchmarks that define the readiness of these models for clinical application. In an effort to address this need for a standardized system of evaluation, we propose the concept of a "CTRL" framework, which utilizes a tiered taxonomy system inspired by technology readiness levels used in various technological applications. As described in Table 4, the CTRL framework defines five progressive levels of readiness, ranging from initial conceptualization to clinical application, on the basis of validation, generalization, and integration [31].

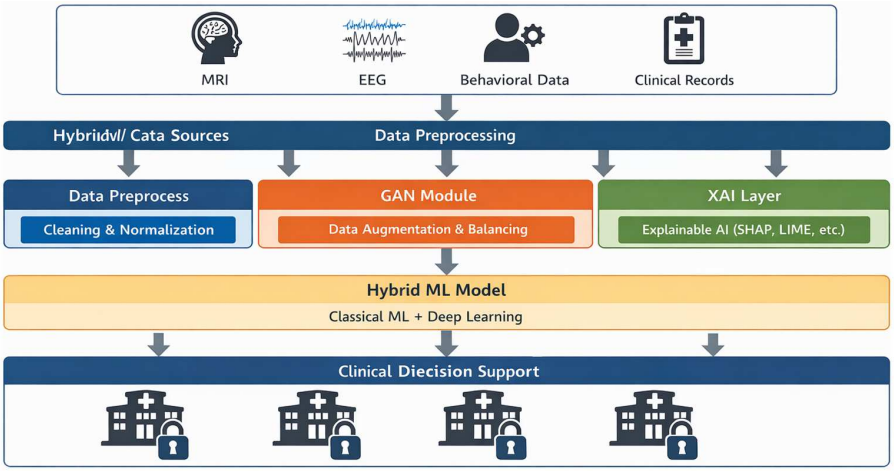
5.2. Hybrid ML framework

Although the CTRL framework offers a well-defined framework that is easily followed when determining the clinical readiness of a system, the actual design of the model itself is crucial in determining the ease with which the system is able to progress through these stages. In an attempt to address the complexity inherent in the analysis of clinical data, we propose the Hybrid ML Clinical Framework, as shown in Figure 5, which is designed to incorporate diverse data streams, with the entire process being transparent and easily interpretable. The framework begins with the aggregation of diverse data streams, including MRI, EEG, behavioral data, and clinical data, into a unified data preprocessing pipeline. The actual data analysis is done by the employment of a hybrid machine learning model, which combines the benefits of classical machine learning algorithms with the benefits of deep learning algorithms. The most important aspect of

Table 4
Clinical-Translational Readiness Level (CTRL)

Level	Name	Description	Criteria
CTRL 1	Conceptual	Model tested in lab only	No validation
CTRL 2	Validated	Internal validation	Cross-validation used
CTRL 3	Clinically Relevant	Tested on external dataset	Generalizable
CTRL 4	Clinically Tested	Used in pilot clinical study	Clinician feedback
CTRL 5	Deployable	Integrated into healthcare system	Real-world use

Figure 5
Proposed hybrid ML clinical framework



the framework is the employment of an Explainable AI module, which is crucial in ensuring that the decision-making process is easily interpretable by the model. The output of the entire framework is focused on the employment of a Clinical Decision Support system [32].

6. Discussion

This wide-ranging review synthesized the findings of 30 seminal studies to map the rapidly evolving landscape of ML in the diagnosis of schizophrenia. The findings show a field in dynamic tension, with extreme technical promise and profound translational hurdles. The following discussion offers a critical synthesis of these findings, examining the performance of existing models, deconstructing the underlying causes of the translational gap, and situating the ethical and clinical mandates that will have to guide future work in context.

One of the main conclusions from this review is the gap between model performance reported and clinical usability in practice. Even though the literature is replete with more than 90% accuracies [7, 18, 27], the discerning eye is not impressed that such are a sign of the readiness of the models for use in the clinic. High model performance is often based on highly selected, "clean" idealized datasets, pre-processed and typically balanced. For instance, the 94.2% accuracy of Norouzi et al.'s work [7] is impressive but was achieved on a dataset with extreme class imbalance and an older skew, not the heterogeneity of a first-episode psychosis clinic. As Lai et al. [8] have noted, models trained on small homogeneous samples are highly susceptible to overfitting, and their performance measures are weak predictors of generalizability. The domain's over-reliance on overall accuracy hides important failures. A model can achieve a 95% overall accuracy and fail to recognize a rare but clinically important subtype. The work of Van Dee et al. [20] and Gupta et al. [13, 23] on class imbalance is thus highly relevant. Per-class F1-scores and AUC-ROC provide a more accurate and clinically balanced view of performance. The pressure towards creating synthetic data, promising though it is, presents a validity problem in its own right because the quality and representativeness of data generated by GANs remain difficult to measure and standardize. Deep learning's preeminence, as presented by Sharma et al. [18], has imposed a fundamental trade-off. The same strength that gives CNNs and RNNs their improved ability also renders them inscrutable "black boxes." This goes directly against clinical trust, as noted by Haug & Drazen [22] and Mallma [10]. A clinician cannot and should not make a decision based on a prediction without knowing why, especially in a high-stakes region like schizophrenia diagnosis. Therefore, a 92% correct model which outputs a saliency map [30] is, from a medical perspective, more correct and informative than a 95% correct "black box" which gives nothing.

Our comparative survey allows for an easy breakdown of the technical evolution in the field, from fragmented model-based approaches to cohesive, system-level considerations. Classic models (e.g., Logistic Regression, SVM) [7, 17] remain effective for feature-engineered, structured data and possess the unbeatable merit of interpretability. Their limitations appear with high-dimensional, complex data like neuroimages. Deep learning models [18, 21] excel in these environments with automatic feature extraction but need enormous data and computational resources, which is a barrier to entry for the majority of research groups and clinics. The side-by-side comparison in Table 3 makes this trade-off explicit, in the sense that the highest accuracies are always from DL studies, and these studies always mention interpretability as a significant limitation. The Emergence of Hybrid ML as a Synthesis: The most significant trend identified is the move towards hybrid models [14, 30]. These architectures are not merely a combination of techniques but a philosophical synthesis aimed at resolving the core trade-offs. They acknowledge that both high performance and interpretability are non-negotiable for clinical adoption. By integrating XAI techniques like LIME or SHAP [11] into DL pipelines, they seek to open the "black box." This approach is the most promising way forward, as it directly tackles the dual mandate of technical merit and clinical utility. Beyond a Single Modality: The comparative success of multimodal research, such as Idowu et al.'s [11] study combining EEG and behavioral data, makes a crucial observation: schizophrenia cannot be fully represented in one data modality. The biological fidelity of the neuroimaging [8, 18] is augmented by the ecological validity of the behavioral and digital phenotyping data [16, 26]. The future for diagnostic ML is thus not the quest for one "best" modality, but instead the development of sophisticated fusion techniques capable of integrating these multiple streams of evidence into a coherent, multi-faceted digital phenotype [15].

This review has mapped the major gaps hindering advancement, and their suggested solutions need to be equally systematic

and integrative. The Data Scarcity-Imbalance-Privacy Triad: These three challenges are inherently intertwined. Federated Learning (FL) [9, 24] addresses data scarcity and privacy concurrently by enabling learning between institutions without exchanging data. FL does not address class imbalance by itself. This is where GANs [13, 23] can be embedded within an FL paradigm to allow each node to generate data for its minority classes, hence improving the global model performance on limited subtypes while not compromising on centralization of sensitive data. This harmonious integration is the kind of end-to-end thinking envisioned. The Technical-Clinical Gap: One of the most important gaps is the lack of clinician-guided design. What a model produces must be embedded into the clinical workflow. This needs more than a saliency map; it needs interfaces that display information in a format that is compatible with clinical reasoning, perhaps built directly into Electronic Health Record (EHR) systems. The CTRL approach outlined here gives a framework for this, insisting that models be assessed not only on test-set performance but on usability and effect in pilot clinical studies. The Computational Cost-Equity Gap: The high computational demands of DL and FL [11, 24] threaten to create a "digital divide" in world mental health. Repairs thus must prioritize the development of "green AI" or light models, which can run on cheaper, more accessible hardware. Model compression, quantization, and architecture research is no longer a technical niche but an ethical imperative for equitable deployment, as argued by Smith et al. [29].

The ethical considerations of AI in the treatment of schizophrenia are equally significant as its algorithmic aspects. AI models trained on biased data will generate biased outputs, and this has the potential to exacerbate the already existing health disparities for minority ethnic groups, women, and people in low-income countries [9, 29]. Bias detection and debiasing methods, such as those implemented in kits like AIF360, must become a standard part of the ML development process. The use of sensitive mental health data in FL or digital phenotyping [16, 26] calls for robust privacy-preserving approaches like differential privacy [9]. Furthermore, current consent models do not fit dynamic AI systems. We must evolve towards dynamic consent models where patients can understand and control how their data is being used over time. ML's goal must be to augment, not replace, clinical judgment. The psychiatrist's job must change to that of interpreter and verifier of AI-generated insights. This requires training and a fundamental design principle: the AI should be a "clinician-in-the-loop" system, where its predictions are part of a collaborative decision-making process, with final responsibility resting with the clinician.

The ultimate aim of this research trajectory is attainment of real precision psychiatry. The Hybrid ML model we have in mind is more than a technological concept; it is a roadmap for a new diagnostic paradigm. By leveraging multimodal data for an integrated view, FL for heterogenous and private data pooling, XAI for trust, and ethical guidelines for ensuring fairness, this approach can:

1) **Minimize Misdiagnosis:** Through quantitative, data-driven evidence, it can better counter the current subjectivity, potentially realizing our dream of 30% reduction in misdiagnosis rates.
2) **Facilitate Early Intervention:** Detection of individuals at risk from early, non-obvious digital biomarkers [16] before full-blown psychosis can revolutionize long-term outcomes.
3) **Personalize Therapy:** Knowing the patient's subtype of schizophrenia through its unique biological and behavioral signature can inform more precise therapeutic interventions.

In total, though the road to come is complex, it is well charted. The era of stretching for precision alone is over. What lies ahead is the deliberate, interdisciplinary co-design of hybrid, humane, and equitable AI systems that are not only technologically cutting-edge but clinically astute and ethically mindful. With this holistic vision, the discipline can finally translate its enormous potential into tangible, real-life advantage for the millions afflicted with schizophrenia globally.

7. Conclusion

This in-depth review has mapped systematically the revolutionary promise and chronic drawbacks of machine learning for transforming schizophrenia diagnosis. Our synthesis of 30 landmark studies (2018-2026) finds the field at a tipping point. Despite the impressive diagnostic performance of conventional ML models and deep learning architectures—often in excess of 90% in controlled environments—their actual clinical implementation is severely limited by a triumvirate of inherent weaknesses: the "black-box" problem that erodes clinical confidence, pervasive dataset biases and class imbalance that detract from generalizability, and astronomical computational costs that create inequitable access.

The path ahead, as presented in this review, demands a paradigm shift from model-centered strategies in isolation to system-level, integrated solutions. We have proposed and delineated a Hybrid ML framework as the most promising way of bridging the clinical-technical gap. This framework combines synergistically the predictive power of deep learning and interpretability of Explainable AI (XAI), surmounts data scarcity through privacy-preserving Federated Learning and synthetic data augmentation via GANs, and embraces multimodal data fusion to create a unified digital phenotype of the disorder. Most notably, we introduced the CTRL framework as a practical guide to evaluate and advance the maturity of AI tools from laboratory benchmarks to bedside application.

Yet technical innovation alone is insufficient. Ethical values of fairness, accountability, and equity must be embedded in the design process of these tools, adhering to recognized standards such as IEEE P7000. The ultimate success of ML for schizophrenia care hinges on a foundation of commitment to interdisciplinary co-design, uniting clinicians, patients, computer scientists, and ethicists in a shared mission to develop tools that are not only intelligent but also interpretable, accessible, and trustworthy.

Looking ahead, the potential of precision psychiatry is within reach. By staying on this dual path—where technical rigor is matched with clinical utility and ethical consideration—we can make machine learning a research oddity into a cornerstone of global mental health infrastructure. This will empower clinicians with robust, decision-support tools, enable earlier and more accurate diagnosis, personalize treatment plans, and ultimately reduce the global burden and diagnostic disparity of schizophrenia, paving the way toward a more effective and equitable future for psychiatric care.

Acknowledgement

We gratefully acknowledge the Department of Computer Science and Engineering at Pabna University of Science and Technology (PUST) for its continuous support and valuable guidance throughout this work.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Syed Mossabbir Hossain: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft. **Nitun Kumar Podder:** Conceptualization, Software, Validation, Resources, Writing – review & editing, Supervision, Project administration. **Md. Raihanul Haque:** Formal analysis, Data curation, Visualization. **Poly Akter:** Formal analysis, Investigation. **Tasfia Rahman Asma:** Formal analysis, Investigation.

References

[1] World Health Organization. (2022). *World mental health report: Transforming mental health for all*. Switzerland: World Health Organization.

[2] Charlson, F. J., Ferrari, A. J., Santomauro, D. F., Diminic, S., Stockings, E., Scott, J. G., . . . , & Whiteford, H. A. (2018). Global epidemiology and burden of schizophrenia: Findings from the global burden of disease study 2016. *Schizophrenia Bulletin*, *44*(6), 1195–1203. https://doi.org/10.1093/schbul/sby058

[3] Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., . . . , & Unützer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, *392*(10157), 1553–1598. https://doi.org/10.1016/S0140-6736(18)31612-X

[4] Tandon, R. (2024). Reinventing schizophrenia. Updating the construct a three-year international project. *Asian Journal of Psychiatry*, *97*, 104107. https://doi.org/10.1016/j.ajp.2024.104107

[5] Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine*, *49*(9), 1426–1448. https://doi.org/10.1017/S0033291719000151

[6] Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

[7] Norouzi, F., Machado, B. L. M. S., & Nematzadeh, S. (2024). Schizophrenia diagnosis and prediction with machine learning models. In*ternational Journal of Scientific and Applied Research (IJSAR)*, *4*(9), 113–122. https://doi.org/10.54756/IJSAR.2024.26

[8] Lai, J. W., Ang, C. K. E., Acharya, U. R., & Cheong, K. H. (2021). Schizophrenia: A survey of artificial intelligence techniques applied to detection and classification. *International Journal of Environmental Research and Public Health*, *18*(11), 6099. https://doi.org/10.3390/ijerph18116099

[9] Bocková, J., Jones, N. C., Topin, J., Hoffmann, S. V., & Meinert, C. (2023). Uncovering the chiral bias of meteoritic isovaline through asymmetric photochemistry. *Nature Communications*, *14*(1), 3381. https://doi.org/10.1038/s41467-023-39177-y

[10] Rodríguez Mallma, M. J., Zuloaga-Rotta, L., Borja-Rosales, R., Rodríguez Mallma, J. R., Vilca-Aguilar, M., Salas-Ojeda, M., & Mauricio, D. (2024). Explainable machine learning models for brain diseases: insights from a systematic review. *Neurology International*, *16*(6), 1285–1307. https://doi.org/10.3390/neurolint16060098

[11] Idowu, O., Aderinto, N., Olatunji, G., & Kokori, E. (2025). Machine learning in schizophrenia: A systematic review and meta-analysis of diagnostic and predictive models. *BJPsych Open*, *11*(S1), S44. https://doi.org/10.1192/bjo.2025.10148

[12] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195. https://doi.org/10.1186/s12916-019-1426-2

[13] Montano, I. H., Lafuente, E. P., Breñosa, J., Ortega-Mansilla, A., Díez, I. D. L. T., & Río-Solá, M. L. D. (2022). Correction to: Systematic review of telemedicine and ehealth systems applied to vascular surgery. *Journal of Medical Systems*, *47*(1), 15. http://doi.org/10.1007/s10916-022-01901-4

[14] Yang, H., Chang, F., Muroi, F., Liu, Z., Zhang, W., & Cai, J. (2026). Application of artificial intelligence in schizophrenia rehabilitation management: A systematic scoping review. *Translational Psychiatry*, *16*, 180. https://doi.org/10.1038/s41398-026-03872-3

[15] Coutts, F., Koutsouleris, N., & McGuire, P. (2023). Psychotic disorders as a framework for precision psychiatry. *Nature Reviews Neurology*, *19*(4), 221–234. https://doi.org/10.1038/s41582-023-00779-1

[16] Del Fabro, L., Bondi, E., Serio, F., Maggioni, E., D'Agostino, A., & Brambilla, P. (2023). Machine learning methods to predict outcomes of pharmacological treatment in psychosis. *Translational Psychiatry*, *13*(1), 75. https://doi.org/10.1038/s41398-023-02371-z

[17] Verma, S., Goel, T., Tanveer, M., Ding, W., Sharma, R., & Murugan, R. (2023). Machine learning techniques for the schizophrenia diagnosis: A comprehensive review and future research directions. *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), 4795–4807. https://doi.org/10.1007/s12652-023-04536-6

[18] Sharma, M., Patel, R. K., Garg, A., SanTan, R., & Acharya, U. R. (2023). Automated detection of schizophrenia using deep learning: A review for the last decade. *Physiological Measurement*, *44*(3), 03TR01. https://doi.org/10.1088/1361-6579/acb495

[19] McCutcheon, R. A., Keefe, R. S., & McGuire, P. K. (2023). Cognitive impairment in schizophrenia: Aetiology, pathophysiology, and treatment. *Molecular psychiatry*, *28*(5), 1902–1918. https://doi.org/10.1038/s41380-023-01949-9

[20] van Dee, V., Schnack, H. G., & Cahn, W. (2023). Systematic review and meta-analysis on predictors of prognosis in patients with schizophrenia spectrum disorders: An overview of current evidence and a call for prospective research and open access to datasets. *Schizophrenia Research*, *254*, 133–142. https://doi.org/10.1016/j.schres.2023.02.024

[21] Orouskhani, M., Zhu, C., Rostamian, S., Zadeh, F. S., Shafiei, M., & Orouskhani, Y. (2022). Alzheimer's disease detection from structural MRI using conditional deep triplet

network. *Neuroscience Informatics*, *2*(4), 100066. https://doi.org/10.1016/j.neuri.2022.100066

[22] Haug, C. J., & Drazen, J. M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023. *New England Journal of Medicine*, *388*(13), 1201–1208. https://doi.org/10.1056/NEJMra2302038

[23] Gupta, N., Gupta, M., & Esang, M. (2023). Lost in translation: Challenges in the diagnosis and treatment of Early-Onset schizophrenia. *Cureus*, *15*(5), 1–11. https://doi.org/10.7759/cureus.38567

[24] Ibrahim, T., Gebril, A., Nasr, M. K., Samad, A., Zaki, H. A., & Nasr Sr, M. (2023). Exploring the mental health challenges of emergency medicine and critical care professionals: A comprehensive review and meta-analysis. *Cureus*, *15*(7), e42345. https://doi.org/10.7759/cureus.42345

[25] Raj, P., Rauniyar, S., & Sapkale, B. (2023). Psychedelic drugs or hallucinogens: Exploring their medicinal potential. *Cureus*, *15*(11), 1–8. https://doi.org/10.7759/cureus.48345

[26] Wang, Y., Xu, Y., Wu, P., Zhou, Y., Zhang, H., Li, Z., & Tang, Y. (2024). Exploring the interplay between core and mood symptoms in schizophrenia: A network analysis. *Schizophrenia Research*, *269*, 28–35. https://doi.org/10.1016/j.schres.2024.04.016

[27] Lee, S., Cho, Y., Ji, Y., Jeon, M., Kim, A., Ham, B. J., & Joo, Y. Y. (2024). Multimodal integration of neuroimaging and genetic data for the diagnosis of mood disorders based on computer vision models. *Journal of Psychiatric Research*, *172*, 144–155. https://doi.org/10.1016/j.jpsychires.2024.02.036

[28] Chen, R., Liu, Y., Djekidel, M. N., Chen, W., Bhattacherjee, A., Chen, Z., . . . , & Zhang, Y. (2022). Cell type–Specific mechanism of Setd1a heterozygosity in schizophrenia pathogenesis. *Science Advances*, *8*(9), eabm1077. https://doi.org/10.1126/sciadv.abm1077

[29] Smith, K. A., Hardy, A., Vinnikova, A., Blease, C., Milligan, L., Hidalgo-Mazzei, D., . . . , & Cipriani, A. (2024). Digital mental health for schizophrenia and other severe mental illnesses: An international consensus on current challenges and potential solutions. *JMIR Mental Health*, *11*, e57155. https://doi.org/10.2196/57155

[30] Senthil Kumar, S., Venmathi, A. R., Thangavel, Y., & Raja, L. (2025). ResDense fusion: Enhancing schizophrenia disorder detection in EEG data through ensemble fusion of deep learning models. *Neural Computing and Applications*, *37*(4), 2411–2433. https://doi.org/10.1007/s00521-024-10701-5

[31] Walker, E., Rummel, N., & Koedinger, K. R. (2009). CTRL: A research framework for providing adaptive collaborative learning support. *User Modeling and User-Adapted Interaction*, *19*(5), 387–431. https://doi.org/10.1007/s11257-009-9069-1

[32] Rojas-Pérez, F. J., Conde-Sánchez, J. R., Morlett-Paredes, A., Moreno-Barbosa, F., Ramos-Fernández, J. C., Luna-Muñoz, J., . . . , & Pérez-Pérez, E. G. (2026). Exploratory study on hybrid systems performance: A first approach to hybrid ML models in breast cancer cassification. *AI*, *7*(1), 29. https://doi.org/10.3390/ai7010029

How to Cite: Hossain, S. M., Podder, N. K., Haque, M. R., Akter, P., & Asma, T. R. (2026). A Review for Bridging Clinical and Technical Gaps with Hybrid ML in Schizophrenia Diagnosis. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62028369