Machine Learning Methods and Schizophrenia Spectrum Disorders: A Scoping Review

Laura Rivera, Kenya Costa-Dookhan, Joel Joel Dissanayake, Ishraq Siddiqui, Martin Rotenberg, Margaret Hahn, Marta Maslej, George Foussias

Submitted to: JMIR Mental Health
on: August 15, 2025

Table of Contents

Machine Learning Methods and Schizophrenia Spectrum Disorders: A Scoping Review

Laura Rivera[1] MD, MPH; Kenya Costa-Dookhan[1] MD, MSc; Joel Joel Dissanayake[2] MSc; Ishraq Siddiqui[2] PhD; Martin Rotenberg[1, 2] MD, MSc; Margaret Hahn[1, 2] MD, PhD; Marta Maslej[2] PhD; George Foussias[2, 1] MD, PhD

[1] Department of Psychiatry University of Toronto Toronto, Ontario CA
[2] Centre for Addiction and Mental Health Toronto CA

Corresponding Author:
George Foussias MD, PhD

Department of Psychiatry
University of Toronto
250 College Street 8th floor416-979-6948
Toronto, Ontario
CA

Abstract

Schizophrenia spectrum disorders (SSD) pose significant challenges in terms of functional impairment, disability, and economic burden. The application of machine learning (ML) in mental health, particularly in SSD research, has grown rapidly in recent years. ML techniques offer the potential to analyze complex datasets, identify subtle patterns, and make predictions that could lead to more personalized and effective interventions for SSD. This scoping review aims to characterize how ML has been applied to SSD research, with a focus on determining clinical and functional outcome trajectories. We conducted a comprehensive scoping review following the methodology adapted from Arksey and O'Malley, adhering to the PRISMA extension for scoping reviews and following the five stages for scoping reviews. A health sciences librarian assisted in designing a search strategy that was implemented across seven bibliographic databases. Two independent reviewers screened articles at the title/abstract and full-text levels. Data extraction focused on study characteristics, machine learning techniques used, and key findings related to SSD outcomes. In total, 88 studies were included in the review which illustrated a diverse range of ML applications in SSD research, categorized into five primary outcome domains: clinical, treatment, functional, behavioral, and provider outcomes. Clinical outcome studies utilized various data types, including electronic medical records, clinical assessment scales, neuroimaging, and mobile sensing data, employing methods and algorithms such as support vector machines, random forests, and neural networks to predict symptom burden, remission, and relapse. Treatment outcome research focused on predicting response to specific interventions, treatment resistance, and adverse effects, primarily using supervised learning techniques. Functional outcome studies employed advanced ML methods, including deep learning, to analyze social cognition and functioning. Behavioral outcome research applied ML to predict risks of self-harm, aggression, and offending. Provider outcome studies mainly examined prescribing patterns using supervised learning approaches. Across all outcome domains, there was a trend towards integrating multiple data sources and applying increasingly sophisticated ML techniques to address the complex nature of SSD. There is a growing potential of ML in SSD research across various outcome domains. The integration of diverse data types and advanced ML techniques offers promising avenues for improving diagnosis, prognosis, and treatment planning in SSD. However, methodological challenges, ethical considerations, and the need for robust validation remain important issues to address. Future research should focus on refining ML models, addressing data quality and privacy concerns, and translating these computational approaches into clinical practice to enhance personalized care for individuals with SSD.

(JMIR Preprints 15/08/2025:82492)
DOI: https://doi.org/10.2196/preprints.82492

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

TITLE

Machine Learning Methods and Schizophrenia Spectrum Disorders: A Scoping Review

ABSTRACT

Schizophrenia spectrum disorders (SSD) pose significant challenges in terms of functional impairment, disability, and economic burden. The application of machine learning (ML) in mental health, particularly in SSD research, has grown rapidly in recent years. ML techniques offer the potential to analyze complex datasets, identify subtle patterns, and make predictions that could lead to more personalized and effective interventions for SSD. This scoping review aims to characterize how ML has been applied to SSD research, with a focus on determining clinical and functional outcome trajectories. We conducted a comprehensive scoping review following the methodology adapted from Arksey and O'Malley, adhering to the PRISMA extension for scoping reviews and following the five stages for scoping reviews. A health sciences librarian assisted in designing a search strategy that was implemented across seven bibliographic databases. Two independent reviewers screened articles at the title/abstract and full-text levels. Data extraction focused on study characteristics, machine learning techniques used, and key findings related to SSD outcomes. In total, 88 studies were included in the review which illustrated a diverse range of ML applications in SSD research, categorized into five primary outcome domains: clinical, treatment, functional, behavioral, and provider outcomes. Clinical outcome studies utilized various data types, including electronic medical records, clinical assessment scales, neuroimaging, and mobile sensing data, employing methods and algorithms such as support vector machines, random forests, and neural networks to predict symptom burden, remission, and relapse. Treatment outcome research focused on predicting response to specific interventions, treatment resistance, and adverse effects, primarily using supervised learning techniques. Functional outcome studies employed advanced ML methods, including deep learning, to analyze social cognition and functioning. Behavioral outcome research applied ML to predict risks of self-harm, aggression, and offending. Provider outcome studies mainly examined prescribing patterns using supervised learning approaches. Across all outcome domains, there was a trend towards integrating multiple data sources and applying increasingly sophisticated ML techniques to address the complex nature of SSD. There is a growing potential of ML in SSD research across various outcome domains. The integration of diverse data types and advanced ML techniques offers promising avenues for improving diagnosis, prognosis, and treatment planning in SSD. However, methodological challenges, ethical considerations, and the need for robust validation remain important issues to address. Future research should focus on refining ML models, addressing data quality and privacy concerns, and translating these computational approaches into clinical practice to enhance personalized care for individuals with SSD.

KEYWORDS

Schizophrenia; Psychotic Disorders; Machine Learning; Natural Language Processing

INTRODUCTION

Schizophrenia spectrum disorders (SSD) are associated with significant functional impairment, disability, and low rates of personal recovery [1,2]. These disorders pose an immense burden in terms of economic costs, generally due to lost productivity and premature mortality [3,4]. These challenges are compounded by societal stigma, leading to low rates of personal recovery and integration into the community [5,6]. The heterogeneous nature of SSD, characterized by varying symptom profiles and outcomes, underscores the urgent need to better understand long-term functional outcomes of people affected by SSDs [7,8].

The use of artificial intelligence, specifically the application of machine learning (ML), has become increasingly popular in mental health research and clinical practice, particularly concerning schizophrenia [9,10]. ML, a subset of artificial intelligence, involves the development of algorithms that can learn from and make predictions based on data[11]. In the context of schizophrenia, ML techniques have been employed to analyze complex datasets, ranging from neuroimaging scans to genetic information and electronic health records [12–14]. These advanced analytical methods enable the identification of patterns and correlations that may not be evident through traditional statistical approaches.

For instance, ML algorithms can predict disease onset, progression, and treatment response by integrating diverse data sources, potentially leading to more personalized and effective interventions for individuals with SSD [15]. Recent studies have demonstrated the potential of ML in early detection of psychosis risk [14], prediction of treatment outcomes [16], and identification of neuroimaging biomarkers associated with SSD [17]. Moreover, ML approaches have shown promise in analyzing language patterns and social media data to detect early signs of relapse or symptom exacerbation [18,19].

This scoping review aims to characterize how ML has been used for determining clinical and functional outcome trajectories in individuals with SSD, including psychotic affective disorders. By synthesizing the current state of ML applications in SSD research, this review seeks to identify promising avenues for future investigation, highlight methodological challenges, and explore the potential impact of ML on improving outcomes for individuals living with SSD.

METHODS

Given that ML applications within mental health are a novel and rapidly evolving field, few primary studies were expected. For this reason, a scoping review was chosen from the available methodologies to gain a comprehensive sense of existing evidence and to identify key knowledge gaps. Typically, scoping review stages involve determination of the research question(s); consultation with a librarian to identify key search terms; screening and abstracting identified articles; discussing and synthesizing the results; and finally creating a scientific manuscript. The proposed scoping review was conducted in five stages which have been adapted from Arksey and O'Malley as described below [20]. Additionally, the PRISMA extension for scoping reviews was used to ensure reporting and synthesis of evidence met current standards for reporting on knowledge synthesis research.

The primary question is: "How is ML being used for individuals with schizophrenia spectrum and psychotic affective disorders toward determining treatment and functional outcome trajectories?"

Search strategy

A comprehensive search strategy was designed by a health sciences librarian (TR) in consultation with the research team. The core strategy was designed, tested, and finalized in MEDLINE (Ovid), and then translated and run in the following seven bibliographic databases: MEDLINE (Ovid), APA PsycInfo (Ovid), Embase (Ovid), the Cumulative Index to Nursing & Allied Health (CINAHL) (EBSCO), Web of Science (Clarivate), Education Resources Information Center (ERIC) (Ovid), and Library, Information Science & Technology Abstracts (LISTA) (EBSCO). The search strategies used database-specific subject headings, keywords in natural language, and advanced search operators to capture two concepts: "artificial intelligence" and "schizophrenia spectrum disorders". Publication year limit for all original searches was 2023, with no earlier year limit; the initial search was completed July 28, 2023, and re-run on March 10, 2025. The following publication types were removed using the search strategy when possible: book chapters, dissertations, editorials, letters, commentaries, and conference abstracts. To supplement the database searches, reference lists of included studies were scanned for additional eligible articles.

Screening and selection

The authors discussed and determined the eligibility criteria which included: peer-reviewed studies investigating any use (e.g. predictive tool, treatment algorithm) of ML in determining outcomes in psychotic disorders. Exclusion criteria involved the following characteristics that disqualify prospective articles from being included: non-English studies, grey literature, systematic reviews, and conference abstracts. Two levels of screening took place: a title and abstract review for articles that fit predefined criteria and a full-text review of articles that meet those criteria. Two investigators reviewed each article and extracted data using Covidence, a systematic review management software. A 5-percent random sample of the results from the search strategy was used to calibrate and establish the reliability of the screening process applied. The review team met to compare their assessments and to modify the eligibility criteria as needed. There were two levels of screening: a title and abstract review to identify articles meeting criteria and a full-text review of those passing the first screen. For both levels, two independent reviewers (LR, KCD) performed the screen, met, and compared results, and, where needed, discussed with a third reviewer to resolve any differences.

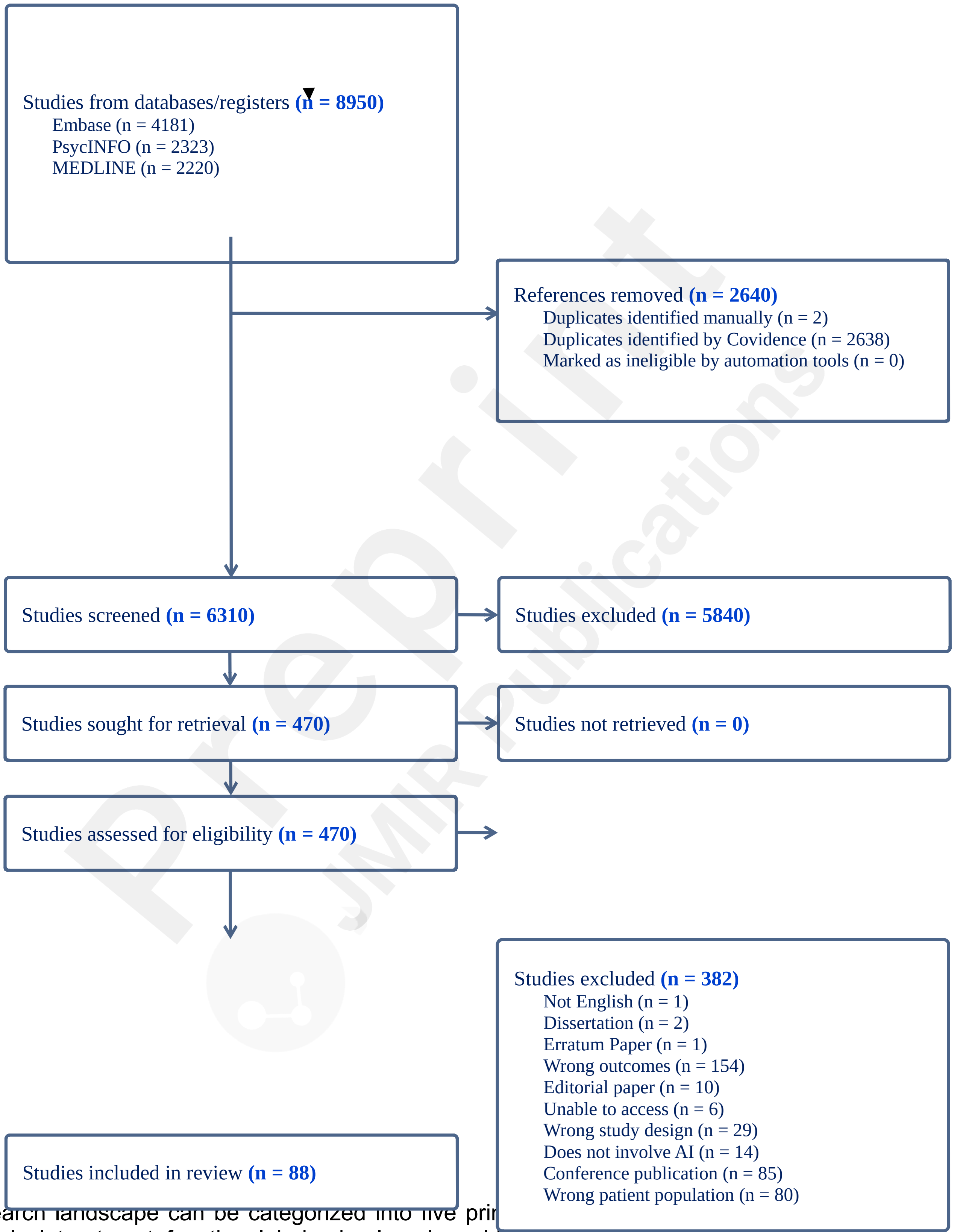
Data extraction and synthesis

As with screening, two independent investigators (LR, KCD) reviewed and extracted data from the selected articles. Data extraction was carried out using an extraction table developed by the authors on Covidence. Information extracted from studies included: author, year of study publication, study population and size, ML techniques used, study objectives, and summarized key findings. The literature was extracted by LR and KCD and then analyzed based on the study outcomes of focus (clinical vs behavioral vs treatment related outcomes).

RESULTS

Our scoping review revealed a diverse landscape of ML applications in the study of schizophrenia spectrum disorders (SSD). In total, 88 studies were analyzed and included in the final study. Included studies were cohort, cross-sectional, and evaluation studies. Sample sizes ranged between n=18 to n=24,614 study participants. Figure 1 illustrates the complete study PRISMA diagram. Further detailed descriptive information about included articles is contained in an additional table in the multimedia appendix.

Table 1. PRISMA diagram.



The research landscape can be categorized into five pri
focus: clinical, treatment, functional, behavioral, and provider outcomes (Table 2).

Table 2. Summary of types of outcomes and corresponding data types elicited via scoping review for individuals with SSD.

1) CLINICAL OUTCOMES -symptom burden --depression --psychosis --OCD --sleep -symptom severity -symptom remission -relapse prediction -suicide risk -disease progression -hospital readmission -length of hospitalization -laboratory values	Data types: • Electronic medical records • Clinical assessment scales (e.g., PANSS, BPRS, CGI) • Neuroimaging data (MRI, EEG) • Speech and language data • Mobile sensing data Machine learning methods: • Supervised Learning: Support Vector Machines (SVM), Random Forests, Gradient Boosting, Logistic Regression • Unsupervised Learning: Topic Modeling, Latent Class Growth Analysis • Deep Learning: Neural Networks • Regularization Methods: LASSO
2) TREATMENT OUTCOMES -treatment response --clozapine --ECT -treatment resistance -treatment discontinuation -treatment adherence -adverse treatment effects including toxicity	Data types: • Electronic medical records • Clinical assessment scales • EEG data Machine learning methods: • Supervised Learning: Support Vector Machines, Random Forests, Artificial Neural Networks, Kernel Partial Least Squares Regression
3) FUNCTIONAL OUTCOMES -social cognition -social functioning -social integration -social isolation -educational/academic functioning -quality of life -resilience	Data types: • Clinical assessment scales • Speech data • Mobile sensing data Machine learning methods: • Supervised Learning: Random Forests Deep Learning: Neural Network Analysis, Long Short-Term Memory (LSTM) neural networks • Probabilistic Methods: Bayes Network
4) BEHAVIOURAL OUTCOMES -self injurious behaviour -aggressive behaviours -violent offending -completed homicide -recidivism	Data types: • Electronic medical records • Clinical assessment scales • Neuroimaging data Machine learning methods: • Supervised Machine Learning using various algorithms, Support Vector Machines, Random Forests
5) PROVIDER OUTCOMES -prescribing patterns	Data types: • Electronic medical records Machine learning methods: • Supervised Machine Learning: Random Forests, Support Vector Machines, Ordinary Least Square Regression

Clinical outcomes

A wide range of data sources have been uses to investigate clinical outcomes. Electronic medical records with both structured and unstructured data have been a rich source of information [21,22], complemented by standardized clinical assessment scales such as Positive and Negative Syndrome Scale (PANSS), Brief Psychiatric Rating Scale (BPRS), and Clinical Global Impression (CGI) [23,24]. Additionally, researchers have utilized neuroimaging data from magnetic resonance imaging (MRI) [25–27], electroencephalogram (EEG), and speech analysis and mobile sensing technologies such as smartphones and videos [28–30]. This diverse data ecosystem has enabled the application of sophisticated machine learning techniques, ranging from traditional supervised learning methods like Support Vector Machines [19,26,31] and Random Forests [32,33] as well as regularization techniques like LASSO [30,34], to more advanced approaches such as deep learning. These tools have been applied to a broad spectrum of clinical challenges, including predicting symptom burden across various domains (depression, psychosis, OCD, sleep disturbances), assessment and diagnostic aims, assessing symptom severity and remission, forecasting relapses, evaluating suicide risk, modeling disease progression, and predicting hospital readmissions and length of stay [35–38].

Treatment outcomes

There has been focused application of machine learning, primarily drawing from electronic medical records, clinical assessment scales, and EEG data on treatment outcomes. Researchers have employed supervised learning techniques, including Support Vector Machines, Random Forests, Artificial Neural Networks, and Kernel Partial Least Squares Regression. These methods have been used to predict treatment response, with particular attention to specific interventions such as clozapine pharmacotherapy [39–41] and electroconvulsive therapy [42,43]. Other important areas of investigation include treatment resistance, factors influencing treatment discontinuation and adherence, and the prediction of adverse treatment effects [44–47].

Functional outcomes

A combination of clinical assessment scales, speech data, and mobile sensing technologies have been used to investigate functional outcomes. This domain has seen a particular emphasis on advanced ML techniques, including deep learning approaches such as neural network analysis and Long Short-Term Memory (LSTM) neural networks, as well as probabilistic methods like Bayes Networks. These tools have been applied to study various aspects of social and cognitive functioning, including social cognition, social integration, educational and academic performance, and overall quality of life [30,48–50].

Behavioral outcomes

Researchers have primarily relied on electronic medical records, clinical assessment scales, and neuroimaging data to study behavioural outcomes. The ML approaches in this domain have largely focused on supervised learning techniques, including Support Vector Machines and Random Forests. These methods have been applied to study and predict a range of challenging behaviors associated with SSD, including self-injurious behavior [33,51,52], aggression, violent offending, homicide, and recidivism [53,54].

Provider outcomes

A small but significant body of research has focused on provider outcomes, particularly prescribing patterns [55,56]. This work has primarily utilized patient data from hospital visits and healthcare encounters via electronic medical records and employed supervised machine learning methods such as Random Forests, Support Vector Machines, and Ordinary Least Square Regression to analyze and predict physician behavior in managing SSD.

Most studies either did not address generalizability at all or only mentioned it as a limitation without substantive discussion. When generalizability was actively addressed, researchers typically used cross-validation techniques, particularly leave-one-site-out and 10-fold cross-validation; testing models on unseen data/test sets; and geographical validation across different recruiting centers. For studies that did address generalizability in their limitations, the common theme was acknowledging uncertainty about the models' generalizability to broader populations or different clinical settings.

DISCUSSION

Our scoping review has revealed a diverse and rapidly evolving landscape of ML applications in SSD research. The main areas where ML is being applied can be categorized into five primary domains: clinical outcomes; treatment outcomes; functional outcomes; behavioural outcomes; and provider outcomes. Across these categories, we observed a trend towards integrating multiple data types, including electronic medical records, neuroimaging data, clinical assessment scales, and novel data streams such as mobile sensing and speech analysis. The most common ML methods used include supervised learning techniques like Support Vector Machines (SVM) and Random Forests, as well as more advanced approaches such as deep learning neural networks. We also noted a lack of standardization in how generalizability is addressed in this field.

The use of multiple data types and application of sophisticated ML techniques reflects the complex, multifaceted nature of SSDs and highlights the potential of ML to uncover novel insights and improve patient outcomes. Notably, there has been a shift towards more sophisticated ML techniques, particularly in functional and clinical outcome prediction. For instance, deep learning approaches have shown promise in analyzing complex, high-dimensional data such as neuroimaging and speech patterns [57]. This trend reflects the growing recognition of the complex, multifaceted nature of SSD and the potential of ML to uncover novel insights that may not be apparent through traditional statistical methods [9].

The application of ML in SSD research holds significant potential for improving clinical practice across several domains. Regarding diagnosis, ML algorithms have demonstrated the ability to identify subtle patterns in text, neuroimaging, and clinical data that may aid in early detection and differential diagnosis of SSD [12,14]. Predictive models developed using ML techniques have shown promise in forecasting disease trajectories, relapse risk, and long-term functional outcomes [16]. In the realm of treatment planning, ML-based approaches have been used to predict treatment response and adverse effects, potentially enabling more personalized and effective treatment strategies [15]. These advancements offer opportunities for personalized medicine in SSD, moving towards a more individualized approach to patient care. For example, ML models could potentially be used to identify patients at high risk of treatment resistance, allowing for earlier intervention with clozapine or other targeted therapies. Similarly, ML-driven prediction of functional outcomes could inform rehabilitation strategies and social support interventions. However, it is crucial to note that while these applications show promise, most are still in the development and research phase and require further validation before widespread clinical implementation [58].

While ML approaches offer powerful tools for analyzing complex datasets in SSD research, several methodological considerations must be addressed. The performance of ML models is heavily dependent on the quality and quantity of available data. Issues such as missing data, small sample sizes, and lack of standardization across datasets pose significant challenges [59]. Many advanced ML techniques, particularly deep learning models, operate as "black boxes," making it difficult to understand the reasoning behind their predictions. This lack of interpretability can be a significant barrier to clinical adoption [60]. Regarding generalizability, ML models developed on specific patient populations or datasets may not generalize well to other contexts or populations. More specifically, performance disparities in ML predictions for socially marginalized and equity seeking groups emerging in other areas of healthcare highlight the importance of considering the potential for ML to contribute to health inequities[61]. Especially seeing as many patients with SSDs face social disadvantages, barriers to accessing care, and concerns about inequity may be particularly relevant in this context, since many patients with SSDs already face social and structural disadvantages and barriers to accessing care. It is therefore critical to consider how biases in training data could contribute to disproportionate negative impacts of ML deployment on specific populations, especially when it comes to prediction of behavioural outcomes, like violence or recidivism.

Based on our review, there are several promising areas for future research and development in ML applications for SSD. This includes potential advancements from combining diverse data types (e.g., genomics, neuroimaging, clinical assessments, and real-world functioning data) to provide a more comprehensive understanding of SSD [62]. Developing ML approaches that can effectively model the dynamic, time-varying nature of SSD symptoms and functioning can also be clinically useful [9]. Finally, advancing techniques for making complex ML models more accessible and user friendly could assist in promoting more inclusive and interdisciplinary research as well as trust among care providers.

While this scoping review provides a comprehensive overview of ML applications in SSD research, several limitations should be noted. Given the fast-evolving nature of ML research, some very recent developments may not be captured in this review such as ongoing developments and use of generative artificial intelligence tools (GenAI). This review primarily captured research applications of ML in SSD, with limited information on real-world clinical implementation. Our review may be influenced by publication bias, potentially overrepresenting positive findings, inclusion of only English language studies, and no-grey literature. As a scoping review, we did not conduct a formal quality assessment of included studies, which could be valuable for future systematic reviews in this area.

ML applications in SSD research offer promising avenues for advancing our understanding of these complex disorders and improving patient outcomes. The integration of diverse data types and increasingly sophisticated ML techniques has the potential to enable more personalized and effective approaches to diagnosis, prognosis, and treatment in SSD. However, realizing this potential will require addressing significant methodological, ethical, and implementation challenges. As the field continues to evolve, close collaboration between data scientists, clinicians, ethicists, and individuals with lived experience of SSD will be crucial to ensure that ML-driven innovations translate into meaningful improvements in clinical care and quality of life for individuals with SSD.

CONFLICTS OF INTEREST
The authors have no conflicts of interest to declare.

ABBREVIATIONS
EEG, electroencephalopgram
ML, machine learning
MRI, magnetic resonance imaging
PANSS, positive and negative symptom scale
SSD, schizophrenia spectrum disorders
SVM, support vector machines

REFERENCES

1. Jääskeläinen E, Juola P, Hirvonen N, et al. A systematic review and meta-analysis of recovery in schizophrenia. *Schizophr Bull*. 2013;39(6):1296-1306. doi:10.1093/schbul/sbs130

2. Owen MJ, Sawa A, Mortensen PB. Schizophrenia. *The Lancet*. 2016;388(10039):86-97. doi:10.1016/S0140-6736(15)01121-6

3. Chong HY, Teoh SL, Wu DBC, Kotirum S, Chiou CF, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. *Neuropsychiatr Dis Treat*. 2016;12:357-373. doi:10.2147/NDT.S96649

4. Olfson M, Gerhard T, Huang C, Crystal S, Stroup TS. Premature Mortality Among Adults With Schizophrenia in the United States. *JAMA Psychiatry*. 2015;72(12):1172-1181. doi:10.1001/jamapsychiatry.2015.1737

5. Corrigan PW, Watson AC. Understanding the impact of stigma on people with mental illness. *World Psychiatry*. 2002;1(1):16.

6. Yanos PT, Roe D, Lysaker PH. The Impact of Illness Identity on Recovery from Severe Mental Illness. *American journal of psychiatric rehabilitation*. 2010;13(2):73. doi:10.1080/15487761003756860

7. an der Heiden W, Häfner H. Course and outcome. In: *Schizophrenia, 3rd Ed*. Wiley Blackwell; 2011:104-141. doi:10.1002/9781444327298.ch7

8. Tandon R, Gaebel W, Barch DM, et al. Definition and description of schizophrenia in the DSM-5. *Schizophr Res*. 2013;150(1):3-10. doi:10.1016/j.schres.2013.05.028

9. Bzdok D, Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2018;3(3):223-230. doi:10.1016/j.bpsc.2017.11.007

10. Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu Rev Clin Psychol*. 2018;14:91-118. doi:10.1146/annurev-clinpsy-032816-045037

11. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science*. 2015;349(6245):255-260. doi:10.1126/science.aaa8415

12. Doan NT, Kaufmann T, Bettella F, et al. Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders. *Neuroimage Clin*. 2017;15:719-731. doi:10.1016/j.nicl.2017.06.014

13. Kambeitz J, Kambeitz-Ilankovic L, Leucht S, et al. Detecting neuroimaging biomarkers for schizophrenia: a meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology*. 2015;40(7):1742-1751. doi:10.1038/npp.2015.22

14. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al. Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or With Recent-Onset Depression: A Multimodal, Multisite Machine Learning Analysis. *JAMA Psychiatry*. 2018;75(11):1156-1172. doi:10.1001/jamapsychiatry.2018.2165

15. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry*. 2016;3(3):243-250. doi:10.1016/S2215-0366(15)00471-X

16. Fond G, Bulzacka E, Boucekine M, et al. Machine learning for predicting psychotic relapse at 2 years in schizophrenia in the national FACE-SZ cohort. *Prog Neuropsychopharmacol Biol Psychiatry*. 2019;92:8-18. doi:10.1016/j.pnpbp.2018.12.005

17. Winterburn JL, Voineskos AN, Devenyi GA, et al. Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multi-method and multi-dataset study. *Schizophr Res*. 2019;214:3-10. doi:10.1016/j.schres.2017.11.038

18. Bedi G, Carrillo F, Cecchi GA, et al. Automated analysis of free speech predicts psychosis onset

in high-risk youths. *npj Schizophr*. 2015;1(1):1-7. doi:10.1038/npjschz.2015.30

19. Birnbaum ML, Ernala SK, Rizvi AF, et al. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. *npj Schizophr*. 2019;5(1):1-9. doi:10.1038/s41537-019-0085-9

20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*. 2005;8(1):19-32. doi:10.1080/1364557032000119616

21. Chandran D, Robbins DA, Chang CK, et al. Use of Natural Language Processing to identify Obsessive Compulsive Symptoms in patients with schizophrenia, schizoaffective disorder or bipolar disorder. *Sci Rep*. 2019;9(1):14146. doi:10.1038/s41598-019-49165-2

22. Han X, Jiang F, Needleman J, Zhou H, Yao C, Tang YL. Comorbidity combinations in schizophrenia inpatients and their associations with service utilization: A medical record-based analysis using association rule mining. *Asian J Psychiatr*. 2022;67:102927. doi:10.1016/j.ajp.2021.102927

23. Ambrosen KS, Skjerbæk MW, Foldager J, et al. A machine-learning framework for robust and reliable prediction of short- and long-term treatment response in initially antipsychotic-naïve schizophrenia patients based on multimodal neuropsychiatric data. *Transl Psychiatry*. 2020;10(1):276. doi:10.1038/s41398-020-00962-8

24. Amoretti S, Verdolini N, Mezquida G, et al. Identifying clinical clusters with distinct trajectories in first-episode psychosis through an unsupervised machine learning technique. *Eur Neuropsychopharmacol*. 2021;47:112-129. doi:10.1016/j.euroneuro.2021.01.095

25. Bohaterewicz B, Sobczak AM, Podolak I, et al. Machine Learning-Based Identification of Suicidal Risk in Patients With Schizophrenia Using Multi-Level Resting-State fMRI Features. *Front Neurosci*. 2020;14:605697. doi:10.3389/fnins.2020.605697

26. Buciuman MO, Oeztuerk OF, Popovic D, et al. Structural and Functional Brain Patterns Predict Formal Thought Disorder's Severity and Its Persistence in Recent-Onset Psychosis: Results From the PRONIA Study. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2023;8(12):1207-1217. doi:10.1016/j.bpsc.2023.06.001

27. Kottaram A, Johnston LA, Tian Y, et al. Predicting individual improvement in schizophrenia symptom severity at 1-year follow-up: Comparison of connectomic, structural, and clinical predictors. *Hum Brain Mapp*. 2020;41(12):3342-3357. doi:10.1002/hbm.25020

28. Abbas A, Hansen BJ, Koesmahargyo V, et al. Facial and Vocal Markers of Schizophrenia Measured Using Remote Smartphone Assessments: Observational Study. *JMIR Formative Research*. 2022;6(1):e26276. doi:10.2196/26276

29. Bishay M, Palasek P, Priebe S, Patras I. SchiNet: Automatic Estimation of Symptoms of Schizophrenia from Facial Behaviour Analysis. Published online August 7, 2018. doi:10.48550/arXiv.1808.02531

30. Lamichhane B, Zhou J, Sano A. Psychotic Relapse Prediction in Schizophrenia Patients Using A Personalized Mobile Sensing-Based Supervised Deep Learning Model. *IEEE J Biomed Health Inform*. 2023;27(7):3246-3257. doi:10.1109/JBHI.2023.3265684

31. Hofmann LA, Lau S, Kirchebner J. Maintaining social capital in offenders with schizophrenia spectrum disorder—An explorative analysis of influential factors. *Front Psychiatry*. 2022;13. doi:10.3389/fpsyt.2022.945732

32. Barone A, De Prisco M, Altavilla B, et al. Disorganization domain as a putative predictor of Treatment Resistant Schizophrenia (TRS) diagnosis: A machine learning approach. *J Psychiatr Res*. 2022;155:572-578. doi:10.1016/j.jpsychires.2022.09.044

33. Cheng N, Guo M, Yan F, et al. Application of machine learning in predicting aggressive behaviors from hospitalized patients with schizophrenia. *Frontiers in Psychiatry*. 2023;14:1016586. doi:10.3389/fpsyt.2023.1016586

34. Bosia M, Bechi M, Bosinelli F, et al. From cognitive and clinical substrates to functional profiles: Disentangling heterogeneity in schizophrenia. *Psychiatry Res*. 2019;271:446-453. doi:10.1016/j.psychres.2018.12.026

35. Bao Y, Wang W, Liu Z, et al. Leveraging deep neural network and language models for predicting long-term hospitalization risk in schizophrenia. *Schizophrenia (Heidelb)*. 2025;11(1):35. doi:10.1038/s41537-025-00585-2

36. Patel R, Dembek C, Won Y, et al. A real-world data analysis of electronic health records to investigate the associations of predominant negative symptoms with healthcare resource utilisation, costs and treatment patterns among patients with schizophrenia. *BMJ Open*. 2024;14(7):e084613. doi:10.1136/bmjopen-2024-084613

37. Góngora Alonso S, Herrera Montano I, De la Torre Díez I, et al. Predictive Modeling of Hospital Readmission of Schizophrenic Patients in a Spanish Region Combining Particle Swarm Optimization and Machine Learning Algorithms. *Biomimetics (Basel)*. 2024;9(12):752. doi:10.3390/biomimetics9120752

38. Jeong JH, Kim J, Kang N, et al. Modeling the Determinants of Subjective Well-Being in Schizophrenia. *Schizophr Bull*. Published online September 10, 2024:sbae156. doi:10.1093/schbul/sbae156

39. Fonseca de Freitas D, Kadra-Scalzo G, Agbedjro D, et al. Using a statistical learning approach to identify sociodemographic and clinical predictors of response to clozapine. *J Psychopharmacol*. 2022;36(4):498-506. doi:10.1177/02698811221078746

40. Khodayari-Rostamabad A, Hasey GM, Maccrimmon DJ, Reilly JP, de Bruin H. A pilot study to determine whether machine learning methodologies using pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. *Clin Neurophysiol*. 2010;121(12):1998-2006. doi:10.1016/j.clinph.2010.05.009

41. Lin CC, Wang YC, Chen JY, et al. Artificial neural network prediction of clozapine response with combined pharmacogenetic and clinical data. *Comput Methods Programs Biomed*. 2008;91(2):91-99. doi:10.1016/j.cmpb.2008.02.004

42. Min B, Kim M, Lee J, et al. Prediction of individual responses to electroconvulsive therapy in patients with schizophrenia: Machine learning analysis of resting-state electroencephalography. *Schizophr Res*. 2020;216:147-153. doi:10.1016/j.schres.2019.12.012

43. Ying YB, Jia LN, Wang ZY, et al. Electroconvulsive therapy is associated with lower readmission rates in patients with schizophrenia. *Brain Stimul*. 2021;14(4):913-921. doi:10.1016/j.brs.2021.05.010

44. Zhu Z, Roy D, Feng S, Vogler B. AI-based medication adherence prediction in patients with schizophrenia and attenuated psychotic disorders. *Schizophr Res*. 2025;275:42-51. doi:10.1016/j.schres.2024.11.006

45. Yee JY, Phua SX, See YM, Andiappan AK, Goh WWB, Lee J. Predicting antipsychotic responsiveness using a machine learning classifier trained on plasma levels of inflammatory markers in schizophrenia. *Transl Psychiatry*. 2025;15(1):51. doi:10.1038/s41398-025-03264-z

46. van Hooijdonk CFM, van der Pluijm M, de Vries BM, et al. The association between clinical, sociodemographic, familial, and environmental factors and treatment resistance in schizophrenia: A machine-learning-based approach. *Schizophr Res*. 2023;262:132-141. doi:10.1016/j.schres.2023.10.030

47. Guo X, Zhou E, Wang X, et al. Machine learning-based prediction of antipsychotic efficacy from brain gray matter structure in drug-naive first-episode schizophrenia. *Schizophrenia (Heidelb)*. 2025;11(1):11. doi:10.1038/s41537-025-00557-6

48. Antonucci LA, Raio A, Pergola G, et al. Machine learning-based ability to classify psychosis and early stages of disease through parenting and attachment-related variables is associated with social cognition. *BMC Psychol*. 2021;9(1):47. doi:10.1186/s40359-021-00552-3

49. Mackinley M, Limongi R, Silva AM, et al. More than words: Speech production in first-episode psychosis predicts later social and vocational functioning. *Front Psychiatry*. 2023;14:1144281. doi:10.3389/fpsyt.2023.1144281

50. Antonucci LA, Pergola G, Rampino A, et al. Clinical and psychological factors associated with resilience in patients with schizophrenia: data from the Italian network for research on psychoses using machine learning. *Psychol Med*. 2023;53(12):5717-5728. doi:10.1017/S003329172200294X

51. Kappes JR, Huber DA, Kirchebner J, Sonnweber M, Günther MP, Lau S. Self-Harm Among Forensic Psychiatric Inpatients With Schizophrenia Spectrum Disorders: An Explorative Analysis. *Int J Offender Ther Comp Criminol*. 2023;67(4):352-372. doi:10.1177/0306624X211062139

52. Watts D, Mamak M, Moulden H, et al. The HARM models: Predicting longitudinal physical aggression in patients with schizophrenia at an individual level. *Journal of Psychiatric Research*. 2023;161:91-98. doi:10.1016/j.jpsychires.2023.02.030

53. Kirchebner J, Günther MP, Lau S. Identifying influential factors distinguishing recidivists among offender patients with a diagnosis of schizophrenia via machine learning algorithms. *Forensic Sci Int*. 2020;315:110435. doi:10.1016/j.forsciint.2020.110435

54. Sonnweber M, Lau S, Kirchebner J. Exploring Characteristics of Homicide Offenders With Schizophrenia Spectrum Disorders Via Machine Learning. *Int J Offender Ther Comp Criminol*. 2024;68(6-7):713-732. doi:10.1177/0306624X221102799

55. Machetanz L, Huber D, Lau S, Kirchebner J. Model Building in Forensic Psychiatry: A Machine Learning Approach to Screening Offender Patients with SSD. *Diagnostics (Basel)*. 2022;12(10):2509. doi:10.3390/diagnostics12102509

56. Marchi M, Galli G, Fiore G, et al. Machine-Learning for Prescription Patterns: Random Forest in the Prediction of Dose and Number of Antipsychotics Prescribed to People with Schizophrenia. *Clinical Psychopharmacology and Neuroscience*. 2022;20(3):450. doi:10.9758/cpn.2022.20.3.450

57. Corcoran CM, Carrillo F, Fernández-Slezak D, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry*. 2018;17(1):67-75. doi:10.1002/wps.20491

58. Torous J, Baker JT. Why Psychiatry Needs Data Science and Data Science Needs Psychiatry: Connecting With Technology. *JAMA Psychiatry*. 2016;73(1):3-4. doi:10.1001/jamapsychiatry.2015.2622

59. Cearns M, Hahn T, Baune BT. Recommendations and future directions for supervised machine learning in psychiatry. *Transl Psychiatry*. 2019;9(1):1-12. doi:10.1038/s41398-019-0607-2

60. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. Published online February 3, 2019. doi:10.48550/arXiv.1806.00069

61. Sikstrom L, Maslej MM, Hui K, Findlay Z, Buchman DZ, Hill SL. Conceptualising fairness: three pillars for medical algorithms and health equity. *BMJ Health Care Inform*. 2022;29(1):e100459. doi:10.1136/bmjhci-2021-100459

62. Sui J, Jiang R, Bustillo J, Calhoun V. Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. *Biol Psychiatry*. 2020;88(11):818-828. doi:10.1016/j.biopsych.2020.02.016