JOIV : Int. J. Inform. Visualization, 8(3-2): IT for Global Goals: Building a Sustainable Tomorrow - November 2024 1819-1826

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Systematic Literature Review: An Early Detection for Schizophrenia Classification Using Machine Learning Algorithms

Ainin Sofiya Azizi [a], Marnisha Mustafa Kamal [a], Nurzarifah Azizan [a], Rohaizaazira Mohd Zawawi [a], Noor Hidayah Zakaria [a,*], Mohamad Aizi Salamat [b], Yulherniwati [c]

[a] *Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia*
[b] *Soft Computing and Data Mining Centre (SMC), Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, Malaysia*
[c] *Department of Information Technology, Politeknik Negeri Padang, Padang, Indonesia*
*Corresponding author: *noorhidayah.z@utm.my*

Abstract*—Schizophrenia is a complex mental health disorder that poses significant challenges in diagnosis and treatment due to its multifaceted symptoms, such as hallucinations, delusions, and cognitive impairments. Early detection is crucial for effective intervention, yet traditional diagnostic methods often fail in precision and scalability. This systematic literature review investigates the application of machine learning (ML) algorithms in the early detection and classification of schizophrenia. By synthesizing findings from 40 primary studies, the review highlights the effectiveness of diverse ML models, including Random Forests, Support Vector Machines (SVM), and advanced deep learning techniques like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. Key datasets such as clinical records, EEG signals, and neuroimaging data were analyzed to evaluate model performance across metrics like accuracy, precision, and sensitivity. Studies demonstrated that hybrid approaches, integrating multiple data sources and deep learning architectures, achieved classification accuracies exceeding 90%, with notable advancements in early-stage diagnosis. However, the review identifies critical challenges, including data quality issues, biases, and limited external validation, which hinder the widespread clinical application of these models. Through a comparative analysis of ML methods and traditional supervised approaches, the study underscores the transformative potential of ML in enhancing diagnostic accuracy and facilitating personalized treatment plans. Addressing current limitations, such as expanding data diversity and improving model interpretability, is essential for translating these findings into practical healthcare solutions. This research contributes to the growing knowledge in ML-driven diagnostics, advocating for its integration into clinical workflows to optimize schizophrenia management.

***Keywords*—Schizophrenia; mental health; machine learning; classification; early detection.**

Manuscript received 11 Apr. 2024; revised 7 Jul. 2024; accepted 22 Sep. 2024. Date of publication 30 Nov. 2024.

I. INTRODUCTION

Schizophrenia is a severe mental illness that profoundly affects an individual's thinking, emotions, and behavior. People with schizophrenia often experience a combination of symptoms, including hallucinations, delusions, disorganization, and cognitive and motivational impairments. Despite the availability of various treatments and medications, some individuals do not benefit from them. However, advancements in machine learning techniques offer new possibilities for preventing and treating schizophrenia.

Recent studies have made significant progress using machine learning algorithms to detect schizophrenia. For example, Support Vector Machines (SVM) were used to differentiate between schizophrenia patients and healthy individuals based on structural MRI data, achieving promising results [1]. On the other hand, the eXtreme Gradient Boosting (XGB) algorithm was utilized to analyze EEG signals in schizophrenia patients. The study compared the XGB-based approach with four other supervised machine learning systems for classifying schizophrenia patients based on EEG recordings [2]. The XGB-based method showed superior performance, with an AUC and accuracy value of 0.94, outperforming the other methods. This system demonstrated high accuracy and robustness, indicating its potential as a valuable tool for hospital clinical use.

In contrast to previous approaches focusing on traditional methods [1], recent work introduced deep learning (DL) algorithms to improve the classification accuracy of EEG signals in schizophrenia research [3]. They segmented the

EEG signal into three bands and extracted distinct domain features from each band. These features were then converted into red-green-blue (RGB) images for input into their network. Additionally, they developed a hybrid DL network that combined a Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) for processing EEG-based schizophrenia classification, achieving high classification accuracy.

This journal explores the application of machine learning algorithms in the early detection classification of schizophrenia [4], [5], [6], [7], [8]. By leveraging diverse datasets and employing various machine learning models, the research aims to identify the most effective techniques for diagnosing schizophrenia at an early stage. The study utilizes several key datasets, each contributing unique insights into the clinical presentation and neuroimaging of schizophrenia.

The research employs several machines learning algorithms, including Random Forests (RF), Support Vector Machines (SVM), and Logistic Regression, each offering distinct advantages for classification tasks [9], [10], [11], [12], [13]. Random Forests, introduced by Breiman, are renowned for their robustness and accuracy in various domains, including neuroimaging and neurodegenerative diseases. Support Vector Machines are recognized for their effectiveness in handling complex decision boundaries and have been widely applied in fields ranging from cancer classification to remote sensing. Logistic Regression, a staple in statistical modeling, is valued for its simplicity and interpretability, making it a popular choice in healthcare and social sciences.

This study aims to comprehensively analyze these algorithms' performance in classifying schizophrenia based on clinical symptoms and neuroimaging data. By comparing the accuracy, precision, recall, and F1 scores of each model, the research seeks to determine the most effective machine-learning technique for the early detection of schizophrenia. Additionally, hyperparameter tuning, specifically through Grid Search, will optimize the models' performance, ensuring the best possible outcomes for schizophrenia classification.

In summary, this reflective journal documents the research process, challenges encountered, and insights gained while exploring the application of machine learning algorithms in the early detection classification of schizophrenia. Through this exploration, the study contributes to the growing body of knowledge in medical diagnostics and machine learning, aiming to improve early diagnosis and treatment strategies for schizophrenia.

II. Material and Method

The methodology for early detection and classification of schizophrenia using machine learning algorithms is meticulously devised. This section explains how we used machine learning algorithms to conduct a systematic literature review on schizophrenia disease. We implemented specific suggestions to achieve this systematic literature review's objectives. Our method aligns with the framework presented in Figure 1.

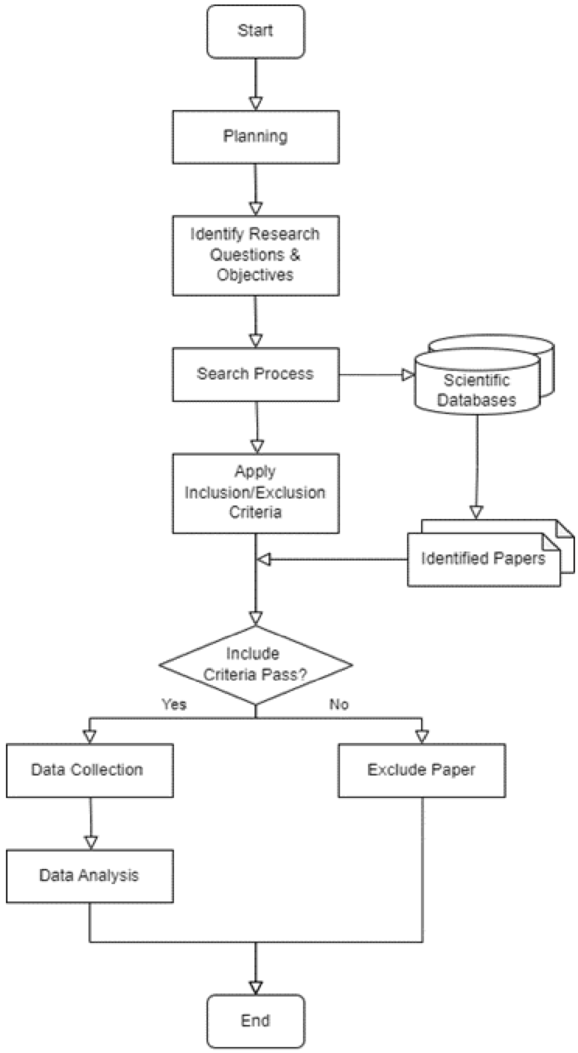

Fig. 1 Search process flowchart

The essential steps are:

1) Planning Stage: Creating a methodical approach for carrying out the study

2) Identifying research questions and motivations: Identifying the specific questions and objectives the research attempts to solve.

3) Searching process: thorough searching for appropriate literature papers and data sources.

4) Applying inclusion and exclusion criteria: Filtering throughout the search results to find solid and relevant information using predefined criteria.

5) Quality assessment: Evaluating the chosen sources' reliability and approach

6) Collecting data: Extracting relevant data and insight from the sources.

7) Analyzing data: Collecting and interpreting the information to the findings that address the research questions.

A. Planning Stage

In the preliminary stage of our study, we identified the crucial steps to achieve our research objectives. Our study centers on the impact of Machine Learning (ML) algorithms in the Schizophrenia domain. We ensured that we established and created a combination of strategic and technical approaches to guarantee that the rest of our proposed technique could be carried out methodically and coherently. The planning stage was the foundation for effectively implementing our Systematic Literature Review (SLR) methodology.

B. Research Question and Motivation

In this section, we discuss and explain the research topics addressed in the current systematic literature review study. The goals derive from the important results of the study implementing machine learning algorithms for the categorization of schizophrenia. Several studies suggest that using different machine learning methods for early identification of schizophrenia can increase performance. Thus, Table 1 contains the research questions (RQs) investigated in this study.

TABLE I
RESEARCH QUESTIONS AND THE MOTIVATIONS

No	Research Questions	Motivations
RQ1	What machine learning algorithms can be effectively utilized for the early detection and classification of schizophrenia?	RQ1 was created by the requirement to identify various machine learning algorithms capable of classifying schizophrenic diseases.
RQ2	How does the performance of machine learning algorithms compare with other supervised methods in the context of early schizophrenia detection and classification?	RQ2 is motivated to evaluate the models by comparing the effectiveness of machine learning algorithms against other supervised methods in early schizophrenia detection and classification.
RQ3	What are the differences in performance metrics such as accuracy, precision, and sensitivity when employing machine learning to detect early signs of schizophrenia compared to alternative supervised machine learning approaches?	RQ3 is motivated by the need to thoroughly compare machine learning approaches to other supervised methods in order to assess their effectiveness in identifying early indicators of schizophrenia.

TABLE II
REPOSITORIES AND THE CORRESPONDING SEARCH STRINGS

Digital database	Field	Search strings	No.
IEEEXplore	All	(Mental Health) OR (Schizophrenia) AND (Machine Learning) AND (Classification)	153
SpringerLink	Artificial Intelligence,	Schizophrenia classification using Computational Intelligence, Bioinformatics, Computational Biology/Bioinformatics machine learning	415
Scopus	All	Machine Learning AND Schizophrenia AND Mental Disease AND Mental Health AND Classification AND Deep Learning AND Classification Algorithm And Machine Learning	8
Web of Science	All	Mental health and Schizophrenia and Machine Learning and Classification	155
ScienceDirect	All	Machine Learning AND Schizophrenia AND Mental Disease AND Mental Health AND Classification AND Deep Learning AND Classification Algorithm And Machine Learning	513
Wiley Online Library	All	Mental Health, Schizophrenia, Machine Learning, Classification	329
Overall			**1,573**

C. Searching Process

This section describes how articles were chosen for the study. It involved using machine learning to search various digital scientific databases for recent studies on the early detection of schizophrenia. Table 1 presents the databases, their fields, search terms, and the initial number of studies found. Each article was extracted from these databases manually, focusing on journal articles and conference proceedings.

D. Inclusion and Exclusion Criteria

After the searching process, in order to ensure that only relevant studies were included in the literature review, specific criteria were established to assess the eligibility of each study. For inclusion in the review, articles needed to meet various conditions, such as must be related to the topic of mental health specifically schizophrenia and if the articles written is not focused on the topic and other than the chosen keywords will be excluded from consideration due to potential complicated challenges. A comprehensive list of the inclusion and exclusion criteria is provided in Table 3.

TABLE III
INCLUSION AND EXCLUSION CRITERIA

Inclusion Criteria	Exclusion Criteria
The articles must be focused on and related to mental health, schizophrenia applied in machine learning, or deep learning based on classification models.	Studies that are not related to the topic or only related to any one subtopic, like schizophrenia, machine learning, deep learning, or classification, but not all of the above.
Each article must be written in clear and comprehensible English and published in an	Articles written in languages other than English or those that are inaccessible.

Inclusion Criteria	Exclusion Criteria
accessible format.	
Each article should present empirical evidence that directly addresses the research questions relevant to the topic.	Literature reviews, opinion pieces, editorials, and non-peer-reviewed sources that do not directly contribute to answering the research questions are not accepted
The publication year of each article must be within 2019-2024.	Article published outside of the period specified

E. Quality Assessment

Alongside implementing inclusion and exclusion criteria, a thorough quality assessment (QA) was carried out to refine the data collection and analysis scope. The primary objective of the QA was to assess the quality of the author's responses to the research questions posed in the SLR. This assessment facilitated the precise extraction of pertinent information while eliminating irrelevant research studies. Table 4 provides a detailed depiction of the QA standards.

TABLE IV
QA QUESTIONNAIRE

No	QA Question	Relevant to the RQ
QA1	What appropriate machine learning algorithms can be chosen for the task of early detection and classification of schizophrenia?	RQ1
QA2	How are the alternative supervised machine learning methods chosen for comparison appropriate for the task and widely recognized field of schizophrenia?	RQ2
QA3	What are the appropriate performance metrics to evaluate the effectiveness of machine learning in detecting early signs of schizophrenia?	RQ3

TABLE V
QA SCORING MATRIX

QA Question	Score
The author(s) have presented a comprehensive, clear, and unambiguous explanation of the responses to the particular RQ.	High = H = 1
The author(s) have provided some elaboration, but it is not specific or detailed and lacks clarity regarding the particular RQ.	Medium = M = 0.5
The author(s) have provided minimal to no technical information addressing the particular RQ.	Low = L = 0

All the research papers underwent an evaluation procedure based on the quality assessment questionnaire outlined previously in Table 5. A scoring matrix in Table 5 was used to assign and evaluate each paper's score. Only papers with a score greater than three were deemed acceptable and inclusive in this research review, while those with a lower score would be considered exclusive.

F. Data Collection

After evaluating the quality of the research papers, any unrelated to the study will be excluded. The next step is to extract the data, which involves thoroughly analyzing and gathering significant information from each research paper that has passed the quality assessment. This process creates a list of papers with specific details that can be used for paper classification and further analysis. The information and details extracted from the primary research studies are summarized in Table 6.

TABLE VI
DATA COLLECTION

Data Fields	Description	Research Questions
Reference ID	Unique ID for each primary study for documentation purposes	RQ1
No. of Primary Studies	Number of studies qualified after a quality scoring scheme	RQ1
Year	Year of publication	RQ1
Publication Source	Type of study and publication name where the primary study was conducted	RQ1
Databases	Name of the digital database for each primary study	RQ1
Objectives	Major objectives of the study	RQ1
Machine Learning Algorithms	Machine learning algorithms used in the study	RQ2
Comparison with Other Methods	Comparison of machine learning algorithms with other methods	RQ2
Limitations	Limitations of the study	RQ3
Future Works	Future works suggested by the study	RQ3

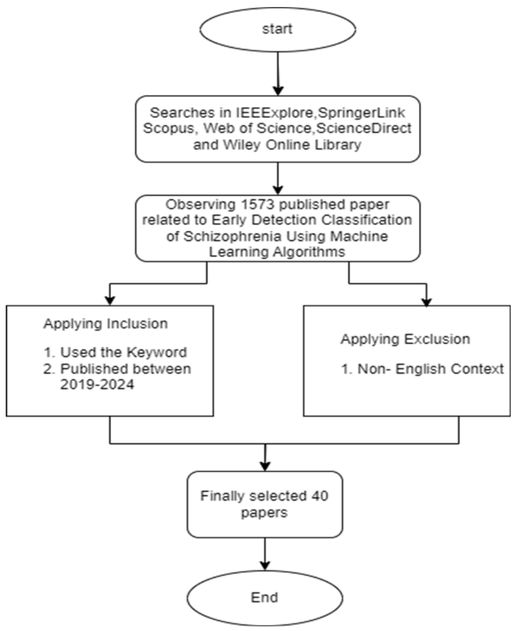

Fig. 2 PRISMA diagram

G. Data Analysis

This section marks the concluding phase of the systematic literature review methodology. It involves analyzing the primary studies and the extracted data from each study to address the three main research questions posed in this study. The quality assessment report, containing details of the primary studies, their main concepts, and their respective quality scores, is discussed.

III. Results and Discussion

In this SLR, a comprehensive search was conducted across numerous scientific databases, including Web of Science, IEEE Xplore, Scopus, ScienceDirect, ACM, SpringerLink, and Wiley Online Library, finding 1,573 results. Following the PRISMA procedure, we narrowed our selection to 40 papers that addressed our study questions. Therefore, the brief background and methods utilized in the primary research are valid. The results demonstrated that the primary studies are topic-focused and the review is valid.

A. *RQ1: How can machine learning algorithms be effectively utilized to develop predictive models for the early detection and classification of schizophrenia?*

Machine learning algorithms have shown significant potential in the early detection and classification of schizophrenia, leveraging various techniques and data sources. Hybrid deep learning models integrating fMRI, retinal imaging, EEG, and biomarkers have achieved up to 98% detection accuracy, with retinal imaging alone providing an 86% detection rate [4]. Deep learning, particularly convolutional neural networks (CNNs), has proven effective in processing complex image data such as retinal and OCT scans. Traditional machine learning models like SVM have also excelled, achieving a 100% precision rate in diagnosing schizophrenia from EEG data [14]. The predictive power of ML models for clinical features such as insomnia, depression, and anxiety, contributing to personalized care by anticipating symptoms and treatment outcomes [15]. Furthermore, the efficacy of Random Forest algorithms applied to event-related potentials (ERP), achieving 96.4% accuracy and underscoring the importance of sensor placement and detailed feature extraction [16]. Despite these advancements, challenges remain regarding data quality, biases, and the external validity of models in diverse clinical settings. Addressing these issues is crucial for enhancing the reliability and applicability of machine learning in schizophrenia diagnosis and early intervention.

The study investigated the use of various machine-learning algorithms to detect and classify schizophrenia. The algorithms examined included Decision Tree, Gaussian Naive Bayes, k-Nearest Neighbors, Support Vector Machine, and XGBoost [17]. Using Matlab's statistical tools with five-fold cross-validation to avoid overfitting, they discovered that XGBoost was the most effective algorithm. Similarly, Support Vector Machine, Random Forest, and Convolutional Neural Network were utilized on MRI data, finding that SVM achieved an accuracy of 75% [18]. SVM, AdaBoost, and XGBoost were employed on EEG data, highlighting XGBoost as the top performer after balancing the classes with SMOTE [19]. SVM, Multinomial Naive Bayes, Random Forest, and XGBoost were applied on brain imaging data, showing promising results across all models [20]. Additionally, Radial Basis Function neural network was used on EEG recordings, which outperformed other traditional algorithms [21].

Another study has demonstrated the potential of structural MRI, fMRI, ERP, and facial expression data in distinguishing schizophrenia patients from healthy controls with significant accuracy. For instance, deep capsule networks and ensemble techniques have improved classification accuracy to 82.83% and AUC values up to 0.9141 [22], while CNN models based on facial expressions achieved 95.18% accuracy [23]. Moreover, Random Forest and SVM models have shown promise in schizophrenia diagnosis and predicting aggressive behaviors in hospitalized patients, with sensitivity and specificity values reaching up to 91.7 [24], [25]. However, the impact of comorbidities such as antisocial personality disorder and substance use disorder on model performance, the complexity of interpreting nonlinear models, and the need for external validation highlight areas for further research [26]. Overall, integrating diverse machine-learning approaches enhances the accuracy and robustness of schizophrenia prediction, paving the way for precise, data-driven clinical applications.

Machine learning algorithms such as logistic regression, deep neural networks, decision trees, support vector machines (SVMs), and k-nearest neighbors (k-NNs) have been effectively used for early detection and classification of schizophrenia using peripheral inflammatory biomarkers [27]. Techniques such as feature extraction and sequential feature selection with grid search for optimal hyperparameters have also proven effective for early detection and classification based on event-related potentials (ERPs) [28]. Furthermore, an ensemble learning method combining SVM and PAM algorithms is successful in classifying schizophrenia using peripheral blood gene expression profiles [29]. Multiple kernel learning (MKL) classifiers have also been shown to be effective for early detection using ERPs [30]. Various classification algorithms such as SVM, KRR, TWSVM, TBSVM, LSTWSVM, and RELSTSVM have been used for detection using MRI [31].

B. *RQ2: How does the performance of machine learning algorithms compare with other supervised methods in the context of early schizophrenia detection and classification?*

Machine learning algorithms significantly outperform traditional supervised methods for early schizophrenia identification and categorization. SVM's demonstrated better precision, attaining 100% accuracy in diagnosing schizophrenia from EEG signals, with SVM and CNN models surpassing Logistic Regression, Random Forest, and XGBoost [5]. CNNs have been shown to be effective at classifying both first-episode psychosis (FEP) and chronic schizophrenia patients [4]. 96.4% accuracy utilizing Random Forest algorithms on event-related potentials (ERP), claimed to be outperforming earlier models [16], [28], [31]. Researchers' structural MRI investigation found that an SVM and RFE framework delivers over 85% classification accuracy, outperforming PCA, ICA, and TBFS approaches. Combining neurocognitive and electrophysiological features with XGBoost results in 93.28% accuracy and 97.91% AUC, outperforming logistic regression and random forest. These studies indicate that machine learning algorithms, particularly SVM, CNN, Random Forest, and XGBoost, offer higher accuracy and precision compared to traditional methods, although challenges like data availability and external validity remain [32], [33].

Multiple machine learning methods were compared and determined that XGBoost was the most effective, excelling in both accuracy and the area under the curve (AUC) [18]. Similarly, XGBoost surpassed SVM and AdaBoost in performance, especially after applying SMOTE for class balancing, proving highly effective in detecting schizophrenia [20]. The Random Forest algorithm achieved the highest accuracy and AUC among the models they examined [21]. Additionally, the study discovered that their Radial Basis Function neural network method outperformed SVM, BLDA, GNB, KNN, and AdaBoost, showing significant improvements in various performance metrics [22].

These machine-learning approaches were compared to other supervised techniques, and the results show that ensemble methods, such as decision forests with multiple trees, enhance classification accuracy. Neural networks have demonstrated great sensitivity in diagnosing schizophrenia [28]. The SVM model attained 91% sensitivity and 90.8% specificity, demonstrating its superior early detection capabilities [29]. Ensemble models accurately categorized schizophrenia samples [30]. The MKL classifier used ERPs to classify patients from controls with 86% accuracy [31].

The performance of machine learning algorithms in the context of early schizophrenia detection and classification demonstrates a notable improvement over the traditional supervised method. Studies highlight that deep learning approaches, such as deep capsule networks and ensemble techniques, significantly enhance classification accuracy. Capsule network ensemble approach outperformed other methods, achieving an 82.83% classification accuracy and an AUC value of up to 0.9141, thus surpassing conventional classifiers like SVM, ELM, and CNN [23]. Similarly, the machine learning models, including Random Forest, Multi-Layer Perceptron, Lasso, and SVM, effectively predicted aggressive behaviors in hospitalized schizophrenia patients, with the Random Forest model demonstrating superior predictive value [26]. This indicates that machine learning not only improves accuracy but also aids in managing clinical outcomes more precisely. However, limitations such as the instability of shallow classifiers, the need for external validation, and the challenge of interpreting complex models suggest areas for further development. Overall, machine learning algorithms, particularly when integrated with advanced techniques like deep learning, offer a more robust and accurate approach to schizophrenia detection and classification compared to traditional supervised methods, paving the way for more effective clinical applications [23], [26].

C. *RQ3: What are the differences in performance metrics such as accuracy, precision, and sensitivity when employing machine learning for detecting early signs of schizophrenia compared to alternative supervised machine learning approaches?*

When it comes to recognizing early indicators of schizophrenia, machine learning techniques outperform traditional supervised methods in terms of accuracy, precision, and sensitivity. PCA, ICA, and TBFS in feature selection outperformed by achieving over 85% accuracy with SVM and RFE [34]. A 100% accuracy rate was claimed for detecting schizophrenia using EEG signals, with SVM and CNN models outperforming Logistic Regression, Random Forest, and XGBoost [5]. CNNs were discovered to be effective at classifying FEP and chronic schizophrenia patients using retinal pictures, even though OCT metrics underperformed [4]. 96.4% accuracy was achieved with Random Forest on ERP data, emphasizing the necessity of thorough feature extraction and sensor location [16]. Machine learning has been noted to effectively predict schizophrenia symptoms, although data quality and bias remain concerns [17]. These studies collectively suggest that SVM, CNN, and Random Forest algorithms offer higher accuracy, precision, and sensitivity than traditional methods, though improvements in data quality and external validity are needed.

The study highlighted the impressive performance metrics of XGBoost, which achieved a recall of 94.51%, an accuracy of 94.25%, an F1 score of 94.92%, and a precision of 94.62%, making it the top performer among the algorithms tested [18]. The study focused on accuracy, reporting that SVM reached 75% but did not provide detailed information on other metrics like precision and sensitivity [19]. XGBoost had the highest accuracy at 93% and superior precision and sensitivity compared to other methods [20]. The Random Forest algorithm achieved an accuracy of 68.6% and an AUC of 0.680 but did not provide precision and sensitivity values [21]. The study noted that their Radial Basis Function neural network achieved a balanced accuracy of around 93%, with high precision and recall values also around 93%, outperforming other algorithms across several performance metrics [22].

This study explicitly highlights differences in performance metrics such as accuracy, precision, and sensitivity when using machine learning to detect early signs of schizophrenia compared to alternative supervised machine learning approaches [28]. This study does not directly compare these performance metrics with other supervised machine learning methods [29]. Similarly, this study did not explicitly compare machine learning performance metrics with other supervised approaches for early detection of schizophrenia [20]. This analysis did not directly compare performance metrics for using machine learning to detect early signs of schizophrenia compared to other methods [21]. This study did not explicitly compare performance metrics and alternative supervised machine-learning approaches for detecting early signs of schizophrenia [22].

Another study reported that a deep capsule network ensemble achieved 82.83% accuracy and an AUC value of up to 0.9141, outperforming traditional classifiers like SVM, ELM, and CNN [24]. Similarly, it has been found that a CNN-LSTM model achieved a remarkable 99.25% accuracy in diagnosing schizophrenia using EEG signals, significantly surpassing conventional machine learning methods such as SVM, KNN, and Random Forest [25]. A hybrid deep learning technique with Mayfly optimization has been highlighted for improving model performance in EEG signal classification, achieving lower loss and higher accuracy compared to other classifiers [35][36]. Additionally, it was demonstrated that a CNN model based on facial expression analysis achieved 95.18% accuracy, indicating a robust ability to distinguish schizophrenia patients from healthy controls [34]. These findings underscore the enhanced performance of deep learning models in terms of accuracy, precision, and sensitivity, making them more effective for early detection compared to traditional supervised methods [23], [24], [35], [36], [37], [38].

D. Limitation of Study

The research on early schizophrenia identification and classification using machine learning algorithms reveals numerous constraints that impact their usefulness and dependability. While noninvasive technologies such as fMRI and CNN are promising, their breadth is limited [3]. Their emphasis on retinal imaging and related metrics may overlook other important criteria for a thorough diagnosis. This narrow focus may limit the strength and generalizability of their findings across groups and environments.

The difficulties created by the restricted use of EEG in detecting neurodegenerative illnesses, particularly in generating valid data from case-control groups, were discussed [5]. Similarly, OCT machine metrics were noted to have limited classification performance for schizophrenia patients, and deep features near the output layer may be ineffective [4]. These limitations suggest that, while these machine learning approaches show promise, their current implementations may only be trustworthy and accurate enough for widespread clinical usage if further refined and validated.

Further constraints are identified throughout these machine-learning models' evaluation and testing stages. The method for early diagnosis of schizophrenia has not been verified in clinical settings, raising questions regarding its external validity [16]. Furthermore, the effect of medicine on model performance was not tested, which could affect accuracy. Broader issues such as data quality, bias, and patient confidentiality, which can affect the reliability and usefulness of machine learning models in various clinical scenarios, have been addressed [17]. These limitations highlight the importance of continued research and development in improving machine learning algorithms' practical utility and accuracy in the early diagnosis and classification of schizophrenia.

IV. Conclusion

The systematic literature review (SLR) comprehensively examined the use of machine learning (ML) algorithms for the early detection and classification of schizophrenia, revealing their substantial advantages over traditional supervised methods. Across 40 selected studies, various ML techniques demonstrated superior accuracy, precision, and sensitivity. Studies highlighted that deep learning models, such as CNN-LSTM, capsule networks, and hybrid deep learning approaches, consistently outperformed traditional classifiers like SVM, ELM, and Random Forest. The integration of advanced optimization techniques further enhanced the performance of these models, indicating the potential of ML in effectively handling complex and varied data types for schizophrenia detection.

However, the review also identified several limitations impacting the practical application of these ML models in clinical settings. Significant concerns were a narrow focus on specific data types, challenges with data quality and bias, and the lack of external validation. Some studies emphasized the need for comprehensive diagnostic features and valid data from diverse case-control groups. Additionally, the impact of comorbidities and medication on model performance and issues related to patient confidentiality were noted as critical areas needing further research and development. These challenges highlight the necessity for ongoing efforts to enhance data quality, address biases, ensure external validation, and improve model interpretability to optimize ML algorithms for clinical use in schizophrenia diagnosis.

Acknowledgment

We want to express our sincere gratitude to UTM for their generous financial support through the UTM Fundamental Grant (PY/2023/01586) with Cost Centre No. Q.J130000.3828.23H44. This funding has been instrumental in supporting this research, and without it, this work would not have been possible.

References

[1] B. Rashid et al., “Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity,” NeuroImage, vol. 134, pp. 645–657, Jul. 2016, doi:10.1016/j.neuroimage.2016.04.051.

[2] L. Dipietro et al., “The evolution of Big Data in neuroscience and neurology,” Journal of Big Data, vol. 10, no. 1, Jul. 2023, doi:10.1186/s40537-023-00751-2.

[3] S. Maneesha, A.S. Pillai, “Schizophrenia Detection Using Deep Learning Techniques,” International Journal of Advanced Research in Science, Communication and Technology, pp. 616–619, Mar. 2023, doi:10.48175/ijarsct-8647.

[4] D. Joseph, A. Lai, S. Silverstein, R. Ramchandran, and E. A. Bernal, “Using Machine Learning to Classify Schizophrenia Based on Retinal Images,” Apr. 2021, doi: 10.1101/2021.04.04.21254893.

[5] F. M. Bosco, L. Colle, R. Salvini, and I. Gabbatore, “A machine-learning approach to investigating the complexity of theory of mind in individuals with schizophrenia,” Heliyon, vol. 10, no. 9, p. e30693, May 2024, doi:10.1016/j.heliyon.2024.e30693.

[6] S. Góngora Alonso, G. Marques, D. Agarwal, I. De la Torre Díez, and M. Franco-Martín, “Comparison of Machine Learning Algorithms in the Prediction of Hospitalized Patients with Schizophrenia,” Sensors, vol. 22, no. 7, p. 2517, Mar. 2022, doi: 10.3390/s22072517.

[7] S.-L. Jeng et al., “Machine learning for prediction of schizophrenia based on identifying the primary and interaction effects of minor physical anomalies,” Journal of Psychiatric Research, vol. 172, pp. 108–118, Apr. 2024, doi: 10.1016/j.jpsychires.2024.02.032.

[8] S. Kumar and A. Saxena, “A Machine Learning Method for Predictive Detection of Depression in Men with Schizophrenia,” 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), pp. 1–6, Feb. 2023, doi:10.1109/icicacs57338.2023.10099887.

[9] J. W. Lai, C. K. E. Ang, U. R. Acharya, and K. H. Cheong, “Schizophrenia: A Survey of Artificial Intelligence Techniques Applied to Detection and Classification,” International Journal of Environmental Research and Public Health, vol. 18, no. 11, p. 6099, Jun. 2021, doi:10.3390/ijerph18116099.

[10] V. C. Nguyen, N. Lu, J. M. Kane, M. L. Birnbaum, and M. De Choudhury, “Cross-Platform Detection of Psychiatric Hospitalization via Social Media Data: Comparison Study,” JMIR Mental Health, vol. 9, no. 12, p. e39747, Dec. 2022, doi: 10.2196/39747.

[11] Q. Yang, Q. Xing, Q. Yang, and Y. Gong, “Classification for psychiatric disorders including schizophrenia, bipolar disorder, and major depressive disorder using machine learning,” Computational and Structural Biotechnology Journal, vol. 20, pp. 5054–5064, 2022, doi:10.1016/j.csbj.2022.09.014.

[12] N. Ramesh, Y. Ghodsi and H. Bolhasani, "A Survey on the Role of Artificial Intelligence in the Prediction and Diagnosis of Schizophrenia", 2023, doi: 10.48550/arxiv.2305.14370.

[13] C. Soria, Y. Arroyo, A. M. Torres, M. Á. Redondo, C. Basar, and J. Mateo, “Method for Classifying Schizophrenia Patients Based on Machine Learning,” Journal of Clinical Medicine, vol. 12, no. 13, p. 4375, Jun. 2023, doi: 10.3390/jcm12134375.

[14] S. Bhadra, C. J. Kumar, and D. K. Bhattacharyya, “Multiview EEG signal analysis for diagnosis of schizophrenia: an optimized deep learning approach,” Multimedia Tools and Applications, Sep. 2024, doi:10.1007/s11042-024-20205-y.

[15] M. A. Vázquez, A. Maghsoudi, and I. P. Mariño, “An Interpretable Machine Learning Method for the Detection of Schizophrenia Using EEG Signals,” Frontiers in Systems Neuroscience, vol. 15, May 2021, doi: 10.3389/fnsys.2021.652662.

[16] F. Janek, T. Rieg, and R. Buettner, “Detection of schizophrenia: A machine learning algorithm for potential early detection and prevention based on event-related potentials,” Proceedings of the 54th Hawaii International Conference on System Sciences, 2021, doi:10.24251/hicss.2021.460.

[17] V. R. Gashkarimov, R. I. Sultanova, I. S. Efremov, and A. R. Asadullin, “Machine learning techniques in diagnostics and prediction of the clinical features of schizophrenia: a narrative review,” Consortium Psychiatricum, vol. 4, no. 3, pp. 43–53, Sep. 2023, doi:10.17816/cp11030.

[18] J. Ruiz de Miras, A. J. Ibáñez-Molina, M. F. Soriano, and S. Iglesias-Parro, “Schizophrenia classification using machine learning on resting state EEG signal,” Biomedical Signal Processing and Control, vol. 79, p. 104233, Jan. 2023, doi: 10.1016/j.bspc.2022.104233.

[19] W. Zhu, S. Shen, and Z. Zhang, “Improved Multiclassification of Schizophrenia Based on Xgboost and Information Fusion for Small Datasets,” Computational and Mathematical Methods in Medicine, vol. 2022, pp. 1–11, Jul. 2022, doi: 10.1155/2022/1581958.

[20] J. T. Arias and C. A. Astudillo, “Enhancing Schizophrenia Prediction Using Class Balancing and SHAP Explainability Techniques on EEG Data,” 2023 IEEE 13th International Conference on Pattern Recognition Systems (ICPRS), pp. 1–5, Jul. 2023, doi:10.1109/icprs58416.2023.10179002.

[21] Y. T. Jo, S. W. Joo, S. Shon, H. Kim, Y. Kim, and J. Lee, “Diagnosing schizophrenia with network analysis and a machine learning method,” International Journal of Methods in Psychiatric Research, vol. 29, no. 1, Feb. 2020, doi: 10.1002/mpr.1818.

[22] G. S. Chilla, L. Y. Yeow, Q. H. Chew, K. Sim, and K. N. B. Prakash, “Machine learning classification of schizophrenia patients and healthy controls using diverse neuroanatomical markers and Ensemble methods,” Scientific Reports, vol. 12, no. 1, Feb. 2022, doi:10.1038/s41598-022-06651-4.

[23] B. Yang et al., “Schizophrenia Classification Using fMRI Data Based on a Multiple Feature Image Capsule Network Ensemble,” IEEE Access, vol. 7, pp. 109956–109968, 2019, doi:10.1109/access.2019.2933550.

[24] J. Ker, L. Wang, J. Rao, and T. Lim, “Deep Learning Applications in Medical Image Analysis,” IEEE Access, vol. 6, pp. 9375–9389, 2018, doi: 10.1109/access.2017.2788044.

[25] N. Shanarova, M. Pronina, M. Lipkovich, and J. Kropotov, “Machine learning based diagnostics of schizophrenia patients,” 2022 6th Scientific School Dynamics of Complex Networks and their Applications (DCNA), pp. 252–255, Sep. 2022, doi:10.1109/dcna56428.2022.9923292.

[26] N. Cheng et al., “Application of machine learning in predicting aggressive behaviors from hospitalized patients with schizophrenia,” Frontiers in Psychiatry, vol. 14, Mar. 2023, doi:10.3389/fpsyt.2023.1016586.

[27] M. Taipale et al., “Effects of Substance Use and Antisocial Personality on Neuroimaging-Based Machine Learning Prediction of Schizophrenia,” Schizophrenia Bulletin, vol. 49, no. 6, pp. 1568–1578, Jul. 2023, doi: 10.1093/schbul/sbad103.

[28] N. Shanarova, M. Pronina, M. Lipkovich, V. Ponomarev, A. Müller, and J. Kropotov, “Application of Machine Learning to Diagnostics of Schizophrenia Patients Based on Event-Related Potentials,” Diagnostics, vol. 13, no. 3, p. 509, Jan. 2023, doi:10.3390/diagnostics13030509.

[29] Y. J. Bae, M. Shim, and W. H. Lee, “Schizophrenia Detection Using Machine Learning Approach from Social Media Content,” Sensors, vol. 21, no. 17, p. 5924, Sep. 2021, doi: 10.3390/s21175924.

[30] H.Y. Raji-Lawal et al., “Ensemble of Machine Learning Classifiers for Schizophrenia Detection,” Caleb International Journal of Development Studies, vol. 05, no. 02, pp. 386–405, Dec. 2022, doi:10.26772/cijds-2022-05-02-20.

[31] E. Santos Febles, M. Ontivero Ortega, M. Valdés Sosa, and H. Sahli, “Machine Learning Techniques for the Diagnosis of Schizophrenia Based on Event-Related Potentials,” Frontiers in Neuroinformatics, vol. 16, Jul. 2022, doi: 10.3389/fninf.2022.893788.

[32] Tanveer, M., Jangir, J., Ganaie, M. A., Beheshti, I., Tabish, M., & Chhabra, N. “Diagnosis of schizophrenia: A comprehensive evaluation”. arXiv.org. 2022, doi:.10.48550/arXiv.2203.11610.

[33] Q. Tian et al., “Detection of Schizophrenia Cases From Healthy Controls With Combination of Neurocognitive and Electrophysiological Features,” Frontiers in Psychiatry, vol. 13, Apr. 2022, doi:10.3389/fpsyt.2022.810362.

[34] Z. Chen et al., “Detecting Abnormal Brain Regions in Schizophrenia Using Structural MRI via Machine Learning,” Computational Intelligence and Neuroscience, vol. 2020, pp. 1–13, Mar. 2020, doi:10.1155/2020/6405930.

[35] A. Shoeibi et al., “Automatic Diagnosis of Schizophrenia in EEG Signals Using CNN-LSTM Models,” Frontiers in Neuroinformatics, vol. 15, Nov. 2021, doi: 10.3389/fninf.2021.777977.

[36] V. Divya, S. Sendil Kumar, V. Gokula Krishnan, and M. Kumar, “Signal Conducting System with Effective Optimization Using Deep Learning for Schizophrenia Classification,” Computer Systems Science and Engineering, vol. 45, no. 2, pp. 1869–1886, 2023, doi:10.32604/csse.2023.029762.

[37] J. Zhang et al., “Detecting schizophrenia with 3D structural brain MRI using deep learning,” Scientific Reports, vol. 13, no. 1, Sep. 2023, doi:10.1038/s41598-023-41359-z.

[38] G. Sunil, S. Gowtham, A. Bose, S. Harish, and G. Srinivasa, “Graph neural network and machine learning analysis of functional neuroimaging for understanding schizophrenia,” BMC Neuroscience, vol. 25, no. 1, Jan. 2024, doi: 10.1186/s12868-023-00841-0.