Identification of Children at Risk of Schizophrenia Via Deep Learning

Dr. G Vijaya[1], Thirisha Visranthi S[2], Uthika C[3], Vaishnitha P[4]
[1,2,3,4]*Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, India*

Abstract- This study offers a novel approach for the early detection of children's schizophrenia risk that makes use of machine learning techniques, including Deep Learning, Voting Classifier, Naive Bayes, Random Forest, Decision Tree, and SVM. By using a comprehensive dataset that includes children's IDs, gender, age, migraine status, BPRS score, medication usage (including clozapine and traditional neuroleptics), and EEG data, the study aims to increase prediction accuracy. Through thorough analysis on a variety of sample populations, the study demonstrates the effectiveness of the proposed approach in accurately and consistently distinguishing between at-risk individuals. By offering targeted support and early intervention, this approach may enhance the timely identification and treatment of schizophrenia in children. For patients suffering from the illness, this could result in better long-term outcomes and a higher quality of life.

Keywords: Schizophrenia Risk Factors, Childhood Psychopathology, Electroencephalography (EEG)

I. INTRODUCTION

1.1 SCHIZOPHRENIA RISK FACTORS

Schizophrenia is a complex and multifaceted mental illness that affects approximately 1% of individuals globally yet is nonetheless a significant public health issue. While the exact etiology of schizophrenia remains unclear, a wealth of evidence suggests a complex interplay between genetic, environmental, and neurodevelopmental factors. Understanding the intricate pathophysiological mechanisms of schizophrenia and identifying high-risk individuals depend on identifying the risk factors associated with the onset of the disorder. This knowledge is crucial for implementing targeted prevention interventions and early therapeutic strategies. To fully comprehend the causes of schizophrenia and establish effective prevention measures in high-risk groups, it is imperative to look into the diverse range of risk factors, from genetic predisposition to environmental stressors. The intricate network of risk factors for schizophrenia is examined in this study, underscoring the illness's complexity and the importance of a comprehensive approach to detecting and lowering the risk of developing schizophrenia.

1.2 CHILDHOOD PSYCHOPATHOLOGY

The study of childhood psychopathology is highly valued in the field of mental health because it recognizes the crucial influence that early experiences have on a person's psychological development. "Childhood psychopathology" encompasses a broad spectrum of emotional, behavioural, and developmental issues that manifest in the early years of an individual's life. These conditions can cause acute challenges as well as long-term issues by affecting a child's cognitive, social, and emotional development well into adulthood. Examining childhood psychopathology is crucial to comprehending the complexities of mental health because it provides a unique perspective that allows researchers and therapists to identify early signs of distress and implement targeted interventions. In an attempt to shed light on the various causes of juvenile psychopathology, the broad spectrum of disorders it includes, and the importance of early detection and treatment in fostering the best possible mental health outcomes for children, this study delves into the intricate topic of childhood psychopathology.

1.3 ELECTROENCEPHALOGRAPHY (EEG)

Electroencephalography (EEG) is a vital instrument in the study of the human mind because it provides a unique window into the dynamic interaction of brain activity that underlies behaviour and cognition. EEG

is a non-invasive technique that records the electrical impulses generated by neurons and offers highly precise temporal monitoring of brain activity. Electroencephalography (EEG) has been a useful tool in clinical and academic settings since its inception, aiding in the resolution of issues pertaining to the intricate brain circuitry. Because EEG is utilized for everything from the diagnosis of neurological diseases to the study of cognitive processes, we have made great progress in our knowledge of the functional dynamics of the brain. With an explanation of its fundamental concepts, technological advancements, and crucial role in identifying neural correlates that have implications for numerous fields, including neurology, psychology, and the newly developing field of neuro informatics, this work initiates an exploration into the capabilities and applications of EEG.

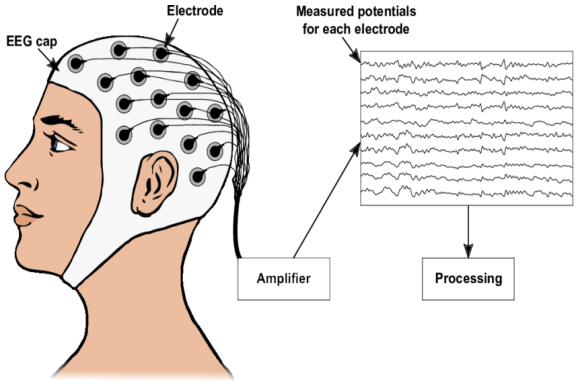

FIGURE 1. EEG

II.. LITERATURE REVIEW

As demonstrated in this work by Carla Barros [1] et al., the intricacy and diversity of symptoms associated with schizophrenia pose a challenge to an objective diagnosis, which is often based on behavioural and clinical manifestations. Additionally, it may be challenging to differentiate schizophrenia from other nosologic conditions, such as bipolar disorder. Early diagnosis and more effective therapy may improve the quality of life for those who suffer from schizophrenia. Over the past few decades, hundreds of studies have been carried out using techniques such as electroencephalography (EEG) in an attempt to identify the brain mechanisms behind the clinical manifestations of schizophrenia. Changes in the electroencephalogram's (EEG) event-related potentials have been proposed as markers of schizophrenia and connected to impairments in perception and cognition. In addition to helping with a more precise diagnosis, biomarkers can be extremely important in prognosticating and anticipating the onset of schizophrenia. But a lot more clinical research is required to show a biomarker's affordability and effectiveness. Thanks to developments in computational neuroscience, attempts have been made to automatically classify schizophrenia at various stages (prodromal, first episode, and chronic). Changes in functional brain activity have been documented using brain imaging pattern recognition tools. The investigation of advanced learning systems has shown encouraging results. The possibilities and constraints of new machine learning-based methods for categorizing schizophrenia from EEG data are covered in this research.

In this paper, Brian Roach [2] et al. proposed deep learning techniques for electroencephalogram (EEG) inputs, which have intriguing psychiatric implications. Schizophrenia, one of the most crippling neuropsychiatric disorders, is often associated with auditory hallucinations. Auditory processing impairments have been investigated and connected to cognitive failure and clinical symptoms in schizophrenia using EEG-derived event-related potentials. Due to their consistent amplitude swings, several ERP components—like the auditory N100—have been proposed as biomarkers of schizophrenia. In this study, we examine the potential use of distinct electrical brain activity patterns during auditory processing to differentiate between people with schizophrenia and healthy people. Using deep convolutional neural networks, we offer an architecture for categorizing multi-channel auditory-related EEG single-trials acquired during a passive listening job. We looked at the distribution and laterality of the electrical activity over the scalp, as well as the effect of the number of electrodes used. The results show that the proposed model can classify people with schizophrenia and healthy participants with an average accuracy of 78% using only 5 midline channels. The current study shows how deep learning methods can be used to study the poor auditory processing associated with schizophrenia and its implications for diagnosis. The proposed structure

might act as a fundamental model for future developments in the field of schizophrenia research.

According to Jeong-Youn Kim [3] and colleagues, a precise diagnosis and a comprehensive assessment of the severity of symptoms are important clinical considerations for people with schizophrenia (SZ). We examined the potential for effectively classifying SZ subtypes based on the intensity of their symptoms using electroencephalography (EEG) features from EEG source network investigations. Sixty-four electrode EEG signals were recorded from 119 SZ patients (53 males and 66 females) and 119 normal controls (NC, 51 men and 68 females) while they were at rest with their eyelids closed. Two characteristics of the brain network, the global path length and the global and local clustering coefficient, were ascertained using EEG source activity. Based on positive, negative, and cognitive/disorganization symptoms, the SZ patients were divided into two groups, high and low, using the positive and negative syndrome scale (PANSS). To pick features for classification, we used the sequential forward selection (SFS) method.

Mateo de Bardeci [4] et al. demonstrated in this paper that deep learning (DL) techniques have made significant strides in recent medical research. This article thoroughly examines how deep learning (DL) techniques have been applied to electroencephalogram (EEG) data for diagnostic and prognosis purposes in order to conduct research on mental illnesses. EEG investigations on mental illnesses classified according to the DSM-V or ICD-10 classifications employed convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. The quality of the information found in these investigations was analysed in three areas: deep learning, EEG data processing, and clinical. Many of the studies lacked a thorough definition of clinical features, even if most of them provided a fair description of EEG recording and pre-processing. Additionally, several studies used subpar testing techniques or incorrect model selection procedures. It has been argued for several years that the identification, treatment, and management of mental diseases will soon undergo a dramatic shift due to advancements in psychiatric research. Future psychiatric disorders studies using DL are encouraged to adhere to state-of-the-art model selection and testing techniques and enhance the quality of clinical data in order to achieve better research standards and get closer to therapeutic importance.

The idea that schizophrenia is a severe mental disorder associated with a range of cognitive and neurophysiological dysfunctions was presented by J. Ruiz de Miras [5] et al. in this work. Since an early diagnosis is dependent on the condition's manifestation, it remains difficult. We have evaluated whether machine learning techniques may help diagnose schizophrenia and presented a processing pipeline to develop machine learning classifiers of schizophrenia based on resting state EEG data. In order to compute well-known linear and non-linear metrics, we used sliding windows of the EEG data. Then, using principal component analysis, we combined the metrics that best differentiated between sick and healthy controls. These elements were then utilized as features in five popular machine learning algorithms: logistic regression (LR), decision trees (DT), random forest (RF), support vector machines (SVM), and k-nearest neighbours (kNN). Complexity measures demonstrated a significant degree of ability to differentiate between healthy controls and people with schizophrenia. These group differences were primarily found in the opercula region and the temporal pole, which correspond to a bounded zone of the right cerebral hemisphere.

III.. RELATED WORK

The prospective identification of children who are at risk of developing schizophrenia is a crucial method to support early interventions that may lessen the possibility of development to clinical psychosis. Electroencephalogram (EEG) patterns from brain activity and deep learning techniques are helpful tools for this identification. We propose automated approaches for processing raw EEG waveforms to detect children who could be at risk for schizophrenia, in contrast to generally developing children. Additionally, we look at deviant traits that continue during a four-year developmental follow-up period in children whose susceptibility to schizophrenia was initially assessed at the ages of nine or twelve. A passive auditory oddball paradigm was recorded while the participants' EEGs were being recorded. We carry out a thorough investigation to find brain abnormalities. First, we use classification approaches

to hand-engineered features (potential components connected with events) in order to study classical machine learning algorithms. We next use unprocessed data to compare these methods' performance to end-to-end deep learning techniques. Using average cross-validation performance metrics, we demonstrate that recurrent deep convolutional neural networks can outperform traditional machine learning methods for sequence modelling.

IV. METHODOLOGY

The suggested method aims to establish a strong basis for the early detection of children's schizophrenia risk by utilizing state-of-the-art machine learning techniques. Among the essential elements are data loading, dataset splitting, and the application of multiple classification algorithms, such as deep learning and ensemble methods like the voting classifier. In order to accurately identify children who may be at risk of developing schizophrenia, the technique uses a comprehensive dataset that includes demographic information, clinical markers, medication history, and EEG data. By providing physicians with a reliable tool for early intervention and individualized support through comprehensive evaluation and testing, the suggested method aims to enhance the quality of life and long-term results for those who are affected.

WEKA (Waikato Environment for Knowledge Analysis) is utilized for its powerful machine learning capabilities, specifically in handling tabular data for predictive modelling. WEKA's built-in classifiers, such as Naïve Bayes, Random Forest, and Decision Trees, are leveraged within a Voting Classifier to improve prediction accuracy by aggregating multiple models' decisions. Instead of using CNN (Convolutional Neural Networks) and RNN (Recurrent Neural Networks), which are typically more suited for image and sequential data—this project employs Deep Neural Networks (DNNs) and traditional ensemble learning approaches like Voting Classifiers. This choice is driven by the nature of the dataset, which consists of structured, tabular data where DNNs can effectively learn deep feature representations, while the Voting Classifier ensures robustness by combining diverse machine learning models. By utilizing WEKA alongside deep learning, a balance in interpretability and accuracy, optimizing the detection of children at risk of schizophrenia based on historical and behavioural data is achieved.

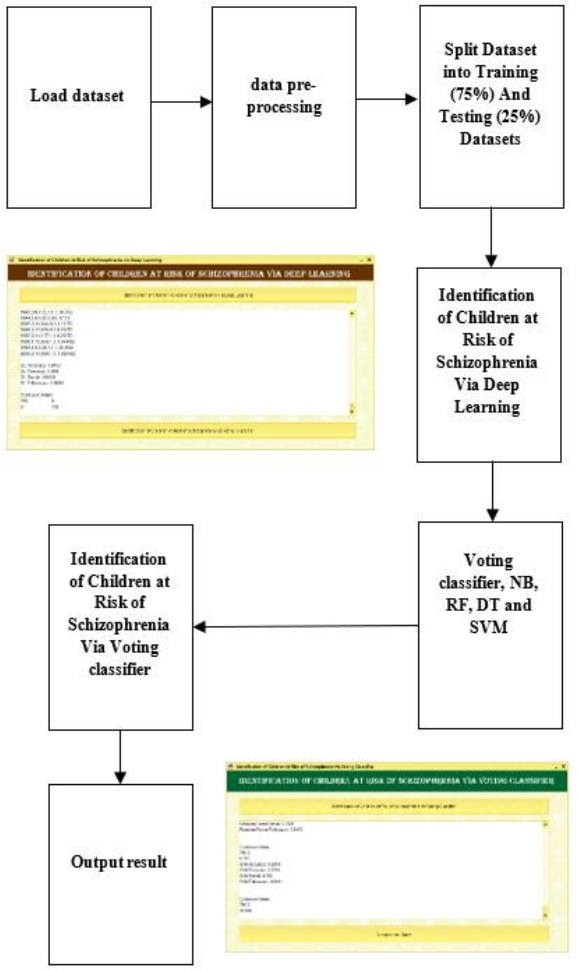

FIGURE 2. PROPOSED BLOCK DIAGRAM

A. Load Dataset

Importing the dataset, which includes a range of child-related variables such demographic data, clinical indicators, medication history, and EEG data, is the initial step in using the Load dataset module. This module ensures that the dataset is correctly read into memory from its storage format, such as CSV files. To prepare the dataset for further analysis, this module may also cover data pre-processing and appropriate handling of missing values.

B. Split Dataset into Training (75%) And Testing (25%) Datasets

This module separates the supplied dataset into training and testing datasets. The training dataset, which includes 75% of the original data, is used to train the machine learning models. The testing dataset, which makes up the remaining 25% of the dataset, is used to evaluate the performance of the trained models. This step is crucial to ensure an unbiased assessment of the models' predictive ability on unseen data.

C. Identification of Children at Risk of Schizophrenia Via Deep Learning

Finding Children at Risk of Schizophrenia using Deep Learning In this module, children who may be at risk of developing schizophrenia are identified using a deep learning algorithm, such as a neural network, based on the data extracted from the dataset. Because deep learning models can automatically find complex patterns and connections in the data, they are suitable for challenging jobs like risk prediction. The model's reported accuracy of 0.90% demonstrates its high accuracy in identifying youngsters at risk of schizophrenia.

D. Identification of Children at Risk of Schizophrenia Via Voting Classifier

The identification modules use a range of machine learning techniques to assess children's risk of developing schizophrenia. By aggregating the predictions of numerous classifiers, the voting classifier module achieves a high accuracy of 0.96%. At 0.81%, Naïve Bayes has a respectable accuracy rate even with its simpler assumptions. Using decision tree classification, which separates the dataset based on feature attributes, an accuracy of 0.83% is achieved. Random Forest achieves an accuracy of 0.75% by combining the output of numerous decision trees via ensemble learning. Finally, the support vector machine (SVM) module correctly classifies data points into groups of 0.76% using a hyperplane. These lessons together show how machine learning may be applied to detect children who might be at risk for schizophrenia; interestingly, the vote classifier outperforms the other techniques in this situation.

V ALGORITHM DETAILS

The method efficiently classifies data and finds patterns by following a set of stages. In order to ensure that essential characteristics like demographic information, clinical markers, and medication records are formatted correctly, data must first be collected and organized. After then, the dataset is split into two subsets: 25% is utilized for testing, and 75% is used for training. An objective evaluation of the classification models is made possible by this separation. The dataset is analyzed using a variety of machine learning techniques. To enable precise categorization, deep learning (Neural Networks) is used to find complex patterns and relationships in the data. Furthermore, to increase accuracy, the ensemble learning technique Voting Classifier aggregates the predictions of several models. A probabilistic classifier called Naïve Bayes uses past knowledge to estimate the likelihood of various classifications. While Random Forest, an extension of Decision Tree, constructs many trees and aggregates their outputs for a more stable prediction, Decision Tree uses a hierarchical structure to segment the dataset based on important attributes. By efficiently distinguishing data points in a high-dimensional space, the Support Vector Machine (SVM) is used to create an ideal decision boundary. The accuracy and efficacy of the categorization models are assessed through the use of conventional performance criteria. The ensemble-based Voting Classifier uses the advantages of several models to achieve higher accuracy. The outcomes demonstrate how well several classification methods may be combined for structured data analysis.

VI. RESULT ANALYSIS

According to the results analysis, the voting classifier module outperforms other identification modules in predicting children's risk of schizophrenia, achieving an astounding accuracy of 0.96%. This illustrates how well ensemble learning techniques leverage the benefits of multiple classifiers to provide accurate predictions. The Naïve Bayes module has an impressive accuracy of 0.81%, while the decision tree, random forest, and SVM modules offer accuracy values ranging from 0.75% to 0.83%. These accuracy levels show the potential of various machine learning algorithms in identifying individuals who are in danger, even though they are not as high as those of the voting classifier. Further refinement and

optimization of these models could improve their performance.

Algorithm	Accuracy	Precision	Recall	F-Measure
DL	0.9047	0.8287	0.8999	0.8628
Voting	0.9698	0.9746	0.9478	0.961
Naive Bayes	0.8105	0.7609	0.8177	0.7883
Decision Tree	0.8381	0.7996	0.7812	0.7903
Random Forest	0.7528	0.8121	0.8235	0.8178
SVM	0.7644	0.7554	0.7854	0.7701

TABLE 1. COMPARISON TABLE

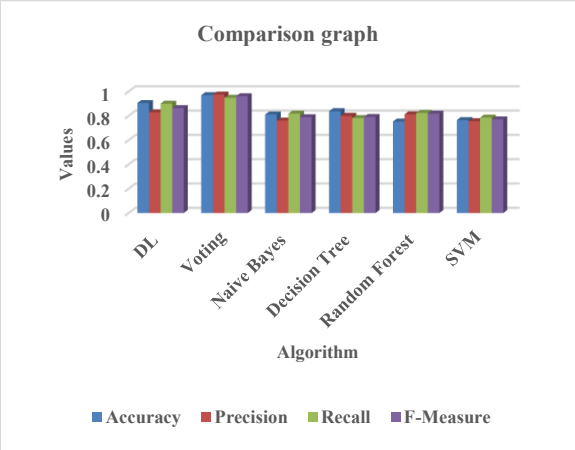

FIGURE 3. COMPARISON GRAPH

VII.CONCLUSION

In the study's conclusion, a novel paradigm for the early detection of children's schizophrenia risk using state-of-the-art machine learning algorithms is presented. By using a large dataset that contains clinical indicators, demographic information, medication history, and EEG data, the proposed method demonstrates encouraging results in accurately identifying individuals who are at-risk. When combined, deep learning and ensemble methods—such as voting classifiers—improve the models' prediction accuracy and enable targeted support and timely intervention for affected kids. Through careful testing and evaluation, the approach has the potential to greatly enhance quality of life and long-term results for individuals who are at risk of developing schizophrenia, underscoring the importance of early diagnosis and management in lessening the impact of this debilitating illness.

8.FUTURE WORK

To improve its effectiveness in detecting children at early risk for schizophrenia, the proposed method may be refined and enhanced in future studies. Adding new features or data sources may be necessary to increase the models' predictive ability. Furthermore, examining state-of-the-art machine learning techniques like deep reinforcement learning and recurrent neural networks may offer new insights into the complex patterns underlying the risk factors for schizophrenia. In order to assess the long-term effectiveness of the system's suggested intervention options, longitudinal studies may also be conducted. These studies could provide crucial information about how children develop schizophrenia and the effectiveness of early therapies.

REFERENCE

[1] Singh along with an in "Promises and pitfalls of advanced eeg-based learning approaches to predict schizophrenia," inf. Procedure, K. Misra, March 2017, Agriculture, vol. 4, pp. 41–49.

[2] "EEG-based deep learning approach for automatic detection of schizophrenia: moving from sound perception," Pattern Recognit., vol. 71, Nov. 2017, pp. 1–13, by H. Lee, c. S. Chan, S. J. Mayo, and p. Remagnino

[3] According to Hillel and Alchanatis, "a machine learning approach for the diagnosis of schizophrenia and the identification of subtypes based on symptom severity using an eEG source network," August 2019, Agriculture, vol. 21, pp. 1–19.

[4] P. Mohanty, D. P. Hughes, and m. Salathé, "A systematic review of deep learning applied to electroencephalogram data in mental disorders," Frontiers Plant Sci., vol. 7, September 2016, p. 1419

[5] F. Ren, w. Liu, and g. Wu, "Classifying schizophrenia using machine learning on resting state electroencephalogram signals," IEEE Access, vol. 7, 2019, pp. 122758–122768

[6] Proc. Adv. Neural inf. Procedure, "Biomarkers for schizophrenia prediction: insights from resting-state electroencephalogram microstates,"

by a. Krizhevsky, I. Sutskever, and g. E. Hinton. System, number 3, December 25, 2012, pp. 1097-1105 (nips).

[7] c. Szegedy, w. Liu, y. Jia, p. Sermanet, s. Reed, d. Anguelov, D. Vanhoucke, v. Erhan, and a. Rabinovich, "CNN-LSTM models for automatic diagnosis of schizophrenia in electroencephalogram (EEG) signals," in Proceedings of the IEEE International Computer Engineering Conference. Look for patterns. June 2015, pages 1–9 (CVR Press).

[8] The paper "Decomposition and nonlinear techniques with eEG signals for automated detection of conduct disorder and attention deficit hyperactivity disorder," by K. Simonyan and A. Zisserman, was published in 2014.

[9] "Abnormal feature extraction from eeg of person with SWT who has schizophrenia," in proceedings of the IEEEE conference, Computer, by K. He, X. Zhang, S. Ren, and J. Sun. (cvpr), June 2016; pages 770–778

[10] Huang, Van der Maaten, Z. Liu, and L. K. "EEG time-frequency analysis during verbal fluency and motor function tests: a preclinical diagnosis of schizophrenia," by Q. Weinberger, in Proceedings of the IEEE conference. Computer. Look for patterns. (cvpr), July 2017, pages 4700–4708.