COMPARATIVE ANALYSIS OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR EEG-BASED SCHIZOPHRENIA DETECTION

Faiyaz Bin Yousf, Asief Iqbal Dieyaz, Annajiat Alim Rasel,
Farah Binte Haque, Humaion Kabir Mehedi & Md. Sabbir Ahmed
Department of Computer Science and Engineering (CSE)
BRAC University
KHA 224, Pragati Sarani, Merul Badda, Dhaka-1212, Bangladesh
faiyaz.bin.yousuf@g.bracu.ac.bd, asief.iqbal.dieyaz@g.bracu.ac.bd

***Abstract*—Psychiatric disorders demand enhanced diagnostics. This research conducts a comparative analysis of machine learning and deep learning models for the detection of a specific mental disorder through EEG analysis. Leveraging a dataset of EEG signals from schizophrenic patients, we employ eight distinct models, including Random Forest, KNN (K-Nearest Neighbors), XGBoost, Logistic Regression, SVM, Decision Tree, and a Convolutional Neural Network (CNN). Our feature extraction process involves signal processing techniques, and the models are rigorously evaluated on their accuracy, precision, and recall. The study aims to provide comparative insights into the efficiency of different modeling approaches in the challenging task of mental disorder detection using EEG data.**

***Index Terms*—Mental Disorder, Schizophrenia, EEG Analysis, Classification, Machine learning Deep Learning, Logistic Regression, KNN, SVM, Random Forest, XGBoost, Decision Tree, Convolutional Neural Network, Feature Extraction, Signal Processing, Diagnostic Accuracy, Early Intervention**

I. Introduction

According to the World Health Organization (WHO), approximately 24 million individuals, constituting 0.32% of the global population or 1 in 300 people, are affected by Schizophrenia. Detecting mental disorders, especially through non-invasive methods like EEG analysis, has gained considerable attention in recent years. EEG signals offer valuable insights into the neural dynamics associated with various mental health conditions. In this context, our research focuses on a specific mental disorder, schizophrenia, utilizing EEG data from some of the patients.

The primary objective is to conduct a comparative study of machine learning and deep learning models for precisely detecting schizophrenia. Schizophrenia is one of the most excruciating mental disorders which significantly affects the mental health of a patient. The significance of this research lies in the diverse modeling techniques that are implemented, ranging from classifiers like Logistic Regression and SVM to more advanced approaches such as Random Forest, XGBoost, KNN, and the deep learning model of Convolutional Neural Network (CNN) architecture.

Our feature extraction process involves a series of signal processing steps, emphasizing the importance of extracting relevant information from EEG signals. The selected models are subjected to a rigorous evaluation process. The evaluation process also includes metrics like precision, accuracy, and recall.

By examining the performance of each model, we aim to provide valuable information and insights into the strong points, weak points, and limitations of different modeling approaches in the context of mental disorder detection. This research contributes to the ongoing discourse on leveraging machine learning techniques for mental health diagnosis, with potential implications for improving diagnostic accuracy and facilitating early intervention strategies.

II. Background and Literature Review

In recent years, there has been a growing interest in leveraging electroencephalography (EEG) as a powerful tool for the detection and diagnosis of mental disorders. Historically, mental health assessments have relied heavily on subjective methods such as interviews and observations. However, the limitations of these approaches have led researchers to explore objective and quantitative measures, with EEG emerging as a promising candidate for providing valuable insights into the neural dynamics associated with various mental health conditions.

Schizophrenia Detection Through EEG Analysis: The study by Devia et al. (2019) delves into the specific application of EEG for diagnosing schizophrenia, a complex mental disorder characterized by disturbances in thought processes, perception, and emotional responsiveness[3]. The conventional diagnostic methods for schizophrenia often face challenges in terms of subjectivity and variability. Devia et al. address this gap by recording EEG during free-viewing tasks, capturing distinct EEG activity patterns in response to natural scenes. Their findings highlight significant differences in occipital areas, indicating the potential of EEG signals as biomarkers for

schizophrenia. Moreover, the development of a classifier based on these differences demonstrates promising sensitivity and specificity for detecting schizophrenia.

Hybrid Models for Depression Detection: Depression, another prevalent mental health condition, has also been a focal point of research in the context of EEG-based detection. The work by Saidi et al. (2020) presents a hybrid method that integrates convolutional neural networks (CNN) and support vector machines (SVM) to enhance the efficiency of detecting depression [1].. This hybrid model demonstrates superior accuracy compared to a baseline CNN model, emphasizing the potential of combining the feature extraction capabilities of CNN with the classification strength of SVM. This work contributes to the broader exploration of hybrid models in EEG analysis, opening avenues for enhanced accuracy in detecting diverse mental health conditions.

EEG Connectivity Networks in Schizophrenia Classification: Bougou et al. (2019) propose an innovative architecture for classifying schizophrenia by focusing on the connectivity patterns extracted from EEG signals [2]. The study introduces the concept of effective connectivity measures, employing graph theory metrics to construct functional networks. Their findings underscore the significance of effective connectivity in providing directional information for improved schizophrenia classification. By utilizing the Random Forest classifier, the study achieves notable accuracy, emphasizing the potential of effective connectivity measures in enhancing the precision of schizophrenia detection.

Evaluation of Brain Connectivity using EEG: In Olejarczyk's (2017) research, the primary focus was on exploring the correlation between schizophrenia and brain connectivity by employing EEG graph-based analysis. The study involved the examination of electroencephalography data obtained from 14 individuals diagnosed with schizophrenia and 14 healthy counterparts. The results uncovered a decline in the strength of connectivity and a reduction in the clustering coefficient. Moreover, noticeable changes were observed in the measures of shorter characteristic path length in phase synchronization following the application of either CSD transformation or REST re-referencing. Particularly, the most prominent differences in connectivity were identified in the alpha band. The study suggests that a valuable approach to comprehending disorders characterized by disrupted connectivity, such as schizophrenia, involves scrutinizing various connectivity measures with graph-based indices for each frequency band separately.

While existing studies have made strides in EEG-based mental disorder detection, a comprehensive comparative analysis across various models is lacking. This paper aims to fill this gap by evaluating the performance of diverse models, including Random Forest, KNN (K-Nearest Neighbors), XGBoost, Logistic Regression, SVM, Decision Tree, and a Convolutional Neural Network. The objective is to provide insights into the strengths and limitations of these models for detecting a specific mental disorder through EEG analysis.

III. Dataset Collection and Analysis

To compile the dataset, the study utilized information from a peer-reviewed journal [4], incorporating two groups: a patient group comprising 14 individuals diagnosed with paranoid schizophrenia (7 males and 7 females) and a control group of 14 healthy individuals (7 males and 7 females). The age range for male patients was 24.6-31.2 years, and for female patients, it was 24.2-32.4 years, aligning with the age range of healthy subjects. All participants underwent a resting state with closed eyes, and EEG data were recorded for fifteen minutes using the standard 10–20 EEG montage. 19 EEG channels were placed on the participant's head. The EEG signals were segmented into thirty-second intervals without any external effects. A second-order Butterworth filter was applied to the EEG data to isolate five distinct frequency bands. The signals were sampled at a rate of 240 hertz.

IV. Methodology

This research endeavors to conduct a comprehensive analysis of Electroencephalogram (EEG) data using the MNE Library, with a specific focus on discerning distinctive patterns between healthy individuals and those diagnosed with mental disorders. The dataset undergoes meticulous categorization, delineating two discrete groups of healthy patients for subsequent supervised learning. To ensure temporal consistency, EEG signals undergo segmentation into fixed-length epochs of 5 seconds with a 1-second overlap, fostering temporal uniformity and comparability across segments.

The labeled dataset materializes from the classification of epochs into healthy and patient groups, laying the cornerstone for a supervised learning paradigm. For feature extraction, a set of 7201 features is derived from diverse mathematical functions. For robust model evaluation, an 80:20 train-test split is implemented, allocating 80% of the features for training and reserving the remaining 20% for testing which results in 5760 features for training and 1441 features for testing shown in Table I.

TABLE I: TOTAL DATA IN THE TRAINING AND TEST SETS

Data Split	**Total Data**
Training Set	5760
Test Set	1441

The primary objective of this study is an assessment and comparison of machine learning and deep learning models. Random Forest, KNN (K-Nearest Neighbors), XGBoost, Logistic Regression, SVM, and Decision Tree are used for ML models. StandardScaler pipeline is incorporated for consistent feature scaling, and hyperparameter tuning is meticulously conducted using GridSearchCV over a spectrum of regularization parameters. SVM acknowledged for its efficacy in neural network analysis, is specifically chosen, with an exploration of linear, polynomial, and Radial Basis Function (RBF) kernels. Random Forest, XGBoost, and Decision Tree models are

subjected to hyperparameter tuning, including considerations for maximum depth, number of estimators, and learning rate, to optimize overall model performance.

The Convolutional Neural Network (CNN) model introduces a sophisticated three-layered architecture culminating in a fully connected dense layer. This architecture comprises a 1D CNN layer with 32 kernels of size 5, with subsequent layers experiencing an augmentation in the number of filters. Batch normalization, LeakyReLU activation function, MaxPooling1D, and Dropout are strategically employed to enhance performance stability and alleviate concerns related to overfitting. The model concludes with a dense layer for binary classification using the Adam optimizer and Binary Cross entropy loss function.

In summary, the methodology is meticulously designed for an in-depth comparative analysis of machine learning and deep learning models. It employs a pragmatic test-train split strategy for robust model evaluation and incorporates sophisticated techniques to address overfitting concerns. The feature extraction process is conducted with meticulous precision, ensuring a comprehensive representation of EEG data and paving the way for a nuanced understanding of mental health patterns.

TABLE II: CONFUSION MATRIX AND EVALUATION METRICS FOR DIFFERENT MODELS

Model	**TP**	**FN**	**TN**	**FP**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	701	78	582	80	0.8976	0.8999	0.8987
K-Nearest Neighbors	693	86	533	129	0.8431	0.8896	0.8657
SVM	717	62	591	71	0.9099	0.9204	0.9151
Random Forest	753	26	500	162	0.8230	0.9666	0.8890
XGBoost	748	31	559	103	0.8790	0.9602	0.9178
Decision Tree	676	103	589	73	0.9025	0.8678	0.8848
CNN	582	197	658	4	0.9932	0.7471	0.8527

V. Result and Discussion

The assessment of model performance hinges on the efficacy of prediction and classification within the test dataset keeping precision and recall in mind. Following comprehensive testing and evaluation, SVM & XGboost emerged as the standout model, exhibiting an impressive accuracy of 90.77% and 90.7% respectively. This is closely followed by Logistic Regression, attaining a noteworthy accuracy of 89.03%. Notably, the Decision Tree & Random Forest model achieved an accuracy of 87.79% & 86.95% respectively. Subsequently, the only deep learning model CNN demonstrated substantial performance with a commendable accuracy of 86.05%. The remaining model, Logistic Regression also impressed by achieving an accuracy of 85.07%.

The SVM model's exceptional performance underscores the potency of EEG analysis for schizophrenia detection. Its ability to capture intricate patterns within the data contributes to its superior accuracy. CNN model follows closely, leveraging its multi-layer neural architecture for effective classification. XGBoost showcases its adaptability and competitiveness in this domain. The results underscore the diversity of models applicable to EEG analysis for schizophrenia detection. While SVM excels, other models, particularly XGBoost, Decision Tree & Random Forest, also exhibit considerable efficacy. CNN showcases much better performance due to its optimized architecture. A better and more complex architecture would result in higher accuracy. The strengths of each model offer flexibility in choosing the most suitable approach based on specific requirements and considerations. This comprehensive evaluation provides valuable insights into the comparative performance of diverse models, paving the way for informed decision-making in future EEG analysis studies.

TABLE III: ACCURACY Table

Model	**Accuracy (%)**
Logistic Regression	89.03
K-Nearest Neighbors	85.07
SVM	90.77
Random Forest	86.95
XGBoost	90.7
Decision Tree	87.79
CNN	86.05

VI. Conclusion

In EEG analysis for schizophrenia detection, our comparative study across diverse models reveals distinct strengths and weaknesses. SVM and XGboost emerge as the top-performing model, showcasing superior accuracy, precision, recall, and F1 score. The Convolutional Neural Network (CNN) also stands out, displaying exceptional accuracy and precision. Other ML models performed admirably, underscoring their potential in schizophrenia detection.

While each model demonstrates unique advantages, the choice of the optimal model depends on the specific requirements of the application, balancing considerations of accuracy, precision, and recall. Future research could explore ensemble methods or hybrid models to leverage the strengths of multiple algorithms, fostering even more robust and accurate schizophrenia detection systems. This could potentially bring improvements for faster EEG analysis for future applicability

especially in on-time detection using wearables and smartphones.

REFERENCES

[1] Saidi, A., Othman, S. B., & Saoud, S. B. (2020). Hybrid CNN-SVM classifier for efficient depression detection system. 2020 4th International Conference on Advanced Systems and Emergent Technologies (IC-ASET). doi:10.1109/ic-aset49463.2020.9318302.

[2] Bougou, V., Mporas, I., Schirmer, P., & Ganchev, T. (2019). Evaluation of EEG Connectivity Network Measures based Features in Schizophrenia Classification. 2019 International Conference on Biomedical Innovations and Applications (BIA). doi:10.1109/bia48344.2019.8967453.

[3] Devia, C., Mayol-Troncoso, R., Parrini, J., Orellana, G., A, A. R., Maldonado, P. E., & Egana, J. I. (2019). EEG Classification during Scene Free-Viewing for Schizophrenia Detection. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 1–1. doi:10.1109/tnsre.2019.2913799.

[4] Olejarczyk, E., & Jernajczyk, W. (2017). Graph-based analysis of brain connectivity in schizophrenia. PLOS ONE, 12(11), e0188629. doi:10.1371/journal.pone.0188629