A Report on

Comparative Analysis of Machine Learning and Hybrid Deep Learning Algorithm for Schizophrenia Detection using EEG Signals

by

Prachi Vyas (20BIT047)
Harsh Patadia (20BIT052)
Aryan Talati (20BIT056)
Jenil Patel (20BIT089)

Under the Guidance of

Dr. Santosh Satapathy,
Assistant Professor, ICT, SOT, PDEU

Submitted to

Information and Communication Technology,

School of Technology,

Pandit Deendayal Energy University

2023

CERTIFICATE

This is to certify that the seminar report entitled “Title of the Seminar,” submitted by Harsh Patadia, Aryan Talati ,Prachi Vyas, Jenil Patel has been conducted under the supervision of Dr. Santosh Satapathy, Assistant Professor, ICT, and is hereby approved for the partial fulfillment of the requirements for the award of the degree of Bachelor of Engineering in the Department of ICT at Pandit Deendayal Energy University, Gandhinagar. This work is original and has not been submitted to any other institution for the award of any degree.

Sign:
Dr. Santosh Satapathy
Assistant Professor
Information & Communication Technology
School of Technology
Pandit Deendayal Energy University

Sign:
Name of Examiner:
Designation:
Department:
School of Technology
Pandit Deendayal Energy University

Appendix 3: Declaration for using AI Tools

DECLARATION

I hereby declare that the report entitled “Comparative Analysis of Machine Learning and Hybrid Deep Learning Algorithm for Schizophrenia Detection using EEG Signals” is the result of my own work and has been written by me. This report has not utilized any language model or natural language processing artificial intelligence tools for the creation or generation of content, including the literature survey.

The use of any such artificial intelligence-based tools was strictly confined to the polishing of content, spell checking, and grammar correction after the initial draft of the report was completed. No part of this report has been directly sourced from the output of such tools for the final submission.

This declaration is to affirm that the work presented in this report is genuinely conducted by me and to the best of my knowledge, it is original.

Prachi Vyas (20BIT047)
Aryan Talati(20BIT056)
Harsh Patadia (20BIT052)
Jenil Patel(20BIT089)
ICT
School of Technology
Pandit Deendayal Energy University
Gandhinagar

Date: 20 November, 2023
Place: PDEU

List of Tools Used for the Report with Purpose:

- ChatGPT : Correcting Grammar.
- Bard : Polishing the text and for modifying description of graphs

Table of Contents

Index	Chapter and sub-title	Page number
I.	Introduction	1
II	Problem Statement	2
II.1	Motivation	
III	Objectives	4
IV	Literature review	5
V	Methodology	5
V.1	Data collection	7
V.2	Data preprocessing	7
V.3	Classification algorithms	7
V.3.1	Support vector machine(SVM)	7
V.3.2	Random Forest(RF)	7
V.3.3	K nearest neighbours(KNN)	8
V3.4	Convolutional neural networks and hybrid approach	8
V.4	Performance evaluation	9
V.4.1	Accuracy	9
V.4.2	F-1 score	9
V.4.3	Precision	9
V.4.4	Sensitivity	10
V.4.5	Specificity	10
V.6	Result analysis	10
V.6.1	Comparative analysis	14
VI	Conclusion	16
VII	Future scope	17
VIII	References	19

Tables:

Table I. Comparative Analysis
Table II. Metrics of all models

List of Figures:

Figure I. Methodology
Figure II. Confusion matrix of KNN
Figure III. Accuracy of SVM with C values
Figure IV. Confusion Matrix of SVM
Figure V. Accuracy of SVM with C values
Figure VI. Confusion Matrix of RF
Figure VII. Accuracy of RF with number of trees
Figure VIII. Confusion Matrix of CNN
Figure IX. Model Accuracy
Figure X. Model Loss
Figure XI. Confusion Matrix of LSTM
Figure XII. Model Accuracy
Figure XIII. Model Loss

List of symbols and abbreviations:

SVM : Support vector machine
C value: Support vector machine regularization parameter
KNN : K-nearest neighbours
K: Number of nearest neighbours chosen in K-nearest neighbours
RF: Random Forest
CNN: Convolutional neural networks
LSTM: Long-Short Term Memory
EEG: Electroencephalography
Accuracy: acc.
Precision: prec.
Recall: rec.

I. Introduction

Schizophrenia, a chronic mental disorder, presents numerous symptoms like delusions, hallucinations, paranoia and disorganized thought process. Its impact extends far beyond the affected individual, affecting families, communities, and societal structures.[17] Its severeness should not be underestimated as timely diagnosis and treatment can help live a better quality of life.[2].

This research actions to contribute to understand SZ and improve its diagnostic methods. Early diagnosis of SZ is severely hampered by the disorder's complex nature and diverse etiology, especially in the early stages of the illness. [20].

Electroencephalography (EEG), a non-invasive test shows promise for schizophrenia detection.[11]. EEG signals are a useful source of information for diagnostic purposes because they have been shown to exhibit changes in persons with schizophrenia. [5]. New approaches are explored here as previously used conservative approaches ae very time-consuming [21] .

A subset of machine learning [12]. Good results have been obtained by using DL techniques with transformation techniques like wavelet transform or graph convolutional networks[16]. This juncture of machine learning techniques with neuroimaging data will improvise the results of schizophrenia diagnosis.

There is potential for Computer-Aided Diagnosis (CAD) systems to be integrated into the diagnostic landscape in order to give clinicians quantitative and objective evaluations of EEG signals. [19]. These systems facilitate population screening on a large scale in addition to aiding in the clinical

diagnosis of schizophrenia. The goal of the proposed study is to create a CAD system by utilising hybrid Deep Learning techniques, which will further the continuous advancement of mental health diagnostics.

II. Problem Statement

The diagnosis of schizophrenia, a complicated mental illness, is extremely difficult and time-consuming. Because they frequently rely on subjective evaluations that could produce erroneous or delayed results, current diagnostic techniques frequently lack the precision necessary for early intervention.

A research gap exists where the comparison of the analysis of results of traditional machine learning algorithms and hybrid deep learning algorithms is yet to be done. Although machine learning algorithms have been used to extract features from EEG signals, there is still more research to be done on the potential benefits of incorporating brain effective connectivity into a hybrid deep learning framework.

II.1 Motivation:

For effective treatment and timely diagnosis, advanced diagnostic tools are to be explored. This becomes an important task due to ever-increasing mental health problems in our world. Our motivation for undertaking the research project, titled "Comparative Analysis of Machine Learning and Hybrid Deep Learning Algorithm for Schizophrenia Detection using EEG Signals," is rooted in some of our personal lives to address the global challenge of mental health disorders.

Addressing a Global Health Concern:

Mental health problems are widespread and affect people all over the world. We are driven by the realization that we must act quickly to make a significant contribution to the field as we observe an increase in these challenges. We seek to lead innovations that can serve as a cornerstone for tackling more general mental health issues by concentrating on schizophrenia, a prevalent and significant mental health illness.Starting with Schizophrenia:

A good place to start is with schizophrenia, which is one of the most common and complicated mental health conditions. Our drive stems from the knowledge that by thoroughly researching and creating diagnostic methods for a basic mental health problem, we can open doors to more general uses and developments in the field. Empowering Diagnostic Precision:

Our research project's title emphasizes our dedication to comparative analysis by highlighting the differences between hybrid deep learning algorithms and traditional machine learning. Our goal is to equip medical professionals with resources that provide accurate diagnosis as well as information on the best practices. By doing this, we hope to contribute to a paradigm shift in the understanding and diagnosis of mental health disorders. Closing Gaps in Diagnostic Methodologies:

There are numerous shortcomings with the current diagnostic procedures for mental health illnesses, especially with regard to their efficiency and accuracy. Our ambition to close these gaps and provide a more deep and thorough comprehension of the diagnostic possibilities present in EEG signals. By tackling issues head-on, we hope to improve mental health diagnostic standards through our research.

Global Impact Through Innovation:

Our study endeavor is a step towards international influence. Through a comparative study of hybrid deep learning and machine learning algorithms for the diagnosis of schizophrenia, we hope to add novel insights to the international conversation on mental health. We are driven by more than just research in the lab; we see a time when improvements in diagnosis will actually benefit those who suffer from mental health issues. In undertaking this research, we respond to the critical need for advancements in mental health diagnostics, striving to make a meaningful difference in the lives of those grappling with these complex challenges on a global scale.

III. Objectives

- **Comparative Analysis of Machine Learning Classifiers**
  Analyze how well conventional machine learning classifiers perform when it comes to identifying schizophrenia by removing discriminative features from EEG signals. Examine the benefits and drawbacks of using machine learning techniques to make an accurate and timely diagnosis.
- **Implementation of Deep Learning Approaches**
  Create and apply deep learning algorithms to extract features from EEG signals, highlighting how well they can capture complex patterns linked to schizophrenia. Examine whether deep learning techniques can improve diagnostic accuracy more than more conventional machine learning classifiers.
- **Hybrid Deep Learning Approach**

  Construct a hybrid deep learning framework that incorporates brain effective connectivity to extract features from EEG signals.

  Evaluate the hybrid approach's potential to improve diagnostic precision in a synergistic way by utilising both connectivity insights and deep learning techniques.
- **Performance Evaluation and Comparison**
  Train and evaluate the system using a comprehensive dataset of EEG images from schizophrenia patients and healthy controls.Employ quantitative metrics such as accuracy, sensitivity, specificity, and ROC curves to compare the performance of ML, deep learning, and hybrid approaches.

- **Identification of Optimal Methodology**
  Determine the best methodology for detecting schizophrenia by conducting a thorough comparison of deep learning techniques, hybrid models, and machine learning classifiers and contribute to the development of mental health diagnostic systems by offering insights into the best methods for prompt and precise diagnosis.

- **Contribution to Diagnostic Advancements**
  Expand on the current understanding of EEG signal analysis for the detection of schizophrenia and make a valuable contribution to the field of mental health diagnostics and offer insightful analysis and suggestions for the creation of upcoming diagnostic systems, which may have an impact on clinical procedures and diagnostic standards.
- **Potential for Early Intervention**
  Examine how the methods under analysis might help those who are suffering from or at risk of schizophrenia receive early intervention. Show off your practical application of research findings in improving patient outcomes through timely diagnosis and intervention.

IV. Literature Review

The literature on schizophrenia detection spans various modalities, utilizing advanced technologies and methodologies. Li et al. [1] explored sex differences in schizophrenia, revealing age-related disparities and the impact of GABAergic gene expression. Their findings highlighted early onset in men and varied symptom severity, linked to the differential expression of GABAergic genes. Notably, women exhibited better treatment responses, attributed to preserved social skills, and hormonal factors were identified as influential, with lower estrogen or testosterone levels correlating with more severe symptoms. Chromosomal distribution was also implicated in psychosis susceptibility. This study provides a foundational understanding of gender-related nuances in schizophrenia manifestation.

Philip et al. [2] delved into cognitive dysfunction, employing neuroimaging to predict psychosis onset. They identified neurobiological markers such as brain volume, intra-cranial volume, and grey matter amount. Specific metrics, including Fractional Anisotropy (FA) of Corpus Callosum and amplitude measures in the Default Mode Network, emerged as predictors of cognitive impairment. This study establishes a crucial link between neurobiological markers and the prediction of cognitive decline, laying the groundwork for early intervention.

Phang et al. [3] classified schizophrenia using EEG connectivity patterns. Achieving a remarkable accuracy of 91.7%, their work showcased the potential of deep learning in discerning complex patterns from EEG signals. Qureshi et al. [4] focused on resting-state fMRI, employing a 3D-CNN to achieve a 98% accuracy in distinguishing between schizophrenic and healthy subjects. These studies highlight the utility of advanced neural networks in effectively processing intricate data for accurate classification.

Zhao et al. [5] proposed a Multi-Scale Recurrent Neural Network (MsRNN) for discriminating schizophrenia using F-MRI data. Their hybrid model, integrating CNN and GRU, achieved an accuracy of 83.32%. This work emphasizes the importance of combining different neural network architectures for improved diagnostic accuracy.

Calhas et al. [6] and Fernando et al. [7] explored EEG signal analysis, achieving accuracies of 95% and 93.86%, respectively. Calhas utilized a Siamese ANN with XGBoost for feature extraction, while Fernando introduced a plastic neural memory network for schizophrenia risk assessment. These studies showcase the effectiveness of EEG signals in discriminating between schizophrenic and healthy individuals.

Bae et al. [8] approached schizophrenia detection through linguistic analysis of social media content, achieving a high accuracy of 96%. Their study highlighted the potential of non-traditional data sources for diagnostic purposes, demonstrating the correlation between language patterns and mental health conditions.

Jason et al. [9] utilized EEG features to predict working memory performance, revealing differences between schizophrenic and healthy adults. Their SVM-based classification achieved 87% accuracy, emphasizing the relevance of EEG-based assessments in cognitive function evaluation.

Kirchebner et al. [10] investigated the impact of stressors on violent offenses in schizophrenic patients, employing various machine learning models. Their work demonstrated the potential of machine learning in understanding the relationship between stress, schizophrenia, and violent behavior.

Recent studies by Bagherzadeh et al. [11] and Sun et al. [12] introduced innovative approaches, combining brain effective connectivity and deep learning for highly accurate schizophrenia detection. Bagherzadeh et al. utilized Transfer Entropy and deep learning, achieving an accuracy of 99.90%. Sun et al. employed a hybrid deep neural network, reaching an average accuracy of 99.22% using fuzzy entropy features.

In conclusion, the literature underscores the multifaceted nature of schizophrenia detection research, employing diverse modalities and advanced technologies. These studies collectively contribute to a deeper understanding of the intricate patterns associated with schizophrenia, with implications for accurate and timely diagnosis. The integration of neural networks, neurobiological markers, and unconventional data sources showcases the potential for interdisciplinary approaches in advancing schizophrenia detection methodologies.

V. Methodology

V.1 Data collection

We have worked on the EEG dataset of 81 subjects with a total of 9216 samples for all 81 subjects. This Kaggle dataset was used for the same with 70 channels.

V.2 Data Preprocessing

The data was preprocessed by splitting the data into blocks of 16 rows and getting the mean value of electrode values for each block. The values are now normalized and split into 80:20 train test ratio.

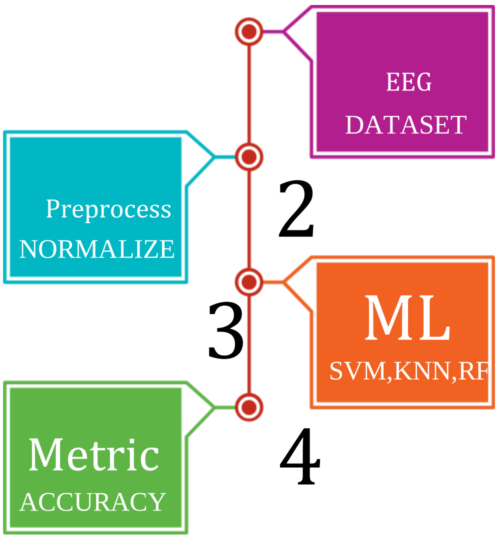


Figure I. Methodology

V.3 Classification Algorithms

V.3.1 Support Vector Machine (SVM) :

Support Vector Machines (SVMs) are an effective class of supervised machine learning algorithms that are applied to regression and classification problems. The identification of a hyperplane inside a high-dimensional feature space that best separates data points from distinct classes is the fundamental idea behind support vector machines (SVM). The purpose of this hyperplane, which can be either a line, plane, or another hyperplane, is to maximize the margin, which is the length of time between the decision boundary and the closest data point of either class. The important data points that are closest to the decision boundary, or support vectors, are vital in identifying the best hyperplane. Because kernel functions can be applied to handle non-linear decision boundaries, SVM's adaptability also allows implicit transfer of input characteristics into higher-dimensional spaces. In SVM, the C parameter presents a critical trade-off that affects how well training points are classified and how well a smooth decision boundary is achieved.

V 3.2 Random Forest (RF) :

For classification and regression tasks, Random Forest, an ensemble learning technique that is frequently used in machine learning, provides a reliable and adaptable solution. Using a group of decision trees, the approach ensures that every tree is trained on a different subset of features and data examples by introducing variety through bootstrapped sampling and random subspace sampling. Either a majority vote (for classification) or an average of the individual tree predictions (for regression) determines the ensemble's final prediction. The model's generalization performance is improved by the natural diversity and combination of predictions in Random Forest, which makes it especially resistant to outliers and noisy data. In addition, the algorithm offers insights into the significance of features, which facilitates the interpretation and choice of pertinent factors. Random Forest is widely used in many different sectors and is notable for its ability to manage intricate datasets, reduce overfitting, and produce accurate predictions makes it a useful tool for machine learning studies and applications.

V.3.3 k-Nearest Neighbors :

K-Nearest Neighbors (KNN), a non-parametric and instance-based learning algorithm, stands out for its simplicity and effectiveness in classification and regression tasks. Its reliance on local patterns in the feature space, determined by the distances between data points, facilitates adaptability to complex, non-linear relationships. The choice of the hyperparameter 'k' governs the number of nearest neighbors considered during predictions, influencing the model's sensitivity and robustness. KNN's ability to capture intricate decision boundaries makes it valuable in scenarios where parametric models may falter. However, considerations such as the curse of dimensionality and sensitivity to outliers underscore the importance of appropriate preprocessing. Despite its lazy learning approach, deferring model building to the prediction phase, KNN proves computationally efficient during training and demonstrates versatility in applications ranging from image recognition to recommendation systems. In the landscape of machine learning, KNN remains a compelling tool, providing a straightforward yet powerful approach to diverse problem domains.

V.3.4 Convolutional Neural Networks and LSTM

Convolutional Neural Networks (CNNs) are a class of deep learning models particularly powerful in processing and analyzing visual data. They consist of convolutional layers, pooling layers, and fully connected layers. CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input data, making them well-suited for tasks like image recognition and classification.

In the context of EEG data for schizophrenia detection, CNNs can be applied to capture spatial patterns and relationships within the EEG signals. The convolutional layers enable the network to detect important features in a local region of the input data, providing a hierarchical representation of complex patterns.

Long Short-Term Memory (LSTM) networks belong to the family of recurrent neural networks (RNNs) and are specifically designed to handle sequential data. Unlike traditional RNNs, LSTMs are capable of learning long-term dependencies and are less prone to the vanishing gradient problem. They consist of memory cells and various gates that regulate the flow of information through the network.

In the context of EEG data, LSTM networks can effectively capture temporal dependencies and patterns over time. This is crucial for understanding the dynamic nature of brain signals and identifying long-term trends that may be indicative of conditions like schizophrenia.

Hybrid CNN-LSTM Approach for EEG-Based Schizophrenia Detection

In the realm of EEG-based schizophrenia detection, the integration of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks has emerged as a powerful and innovative approach. This hybrid model optimally combines the spatial pattern recognition capabilities of CNNs with the temporal sequencing capabilities of LSTMs, resulting in a more robust and effective methodology for analyzing complex brain signal data [14].

Overcoming Overfitting:
The combination of CNN and LSTM architectures plays a pivotal role in mitigating overfitting, a common challenge in machine learning models [14]. By preventing the model from memorizing noise in the data, this hybrid approach ensures that the learned patterns are more representative of true underlying structures within the EEG signals.

Enhanced Diagnostic Capabilities:
The hybrid CNN-LSTM model demonstrates improved diagnostic capabilities, particularly in terms of sensitivity and specificity [14, 15, 16]. Its ability to capture both spatial and temporal features enhances sensitivity to subtle abnormalities in EEG patterns, improving specificity and enabling a more accurate distinction between signals from healthy individuals and those affected by schizophrenia.

Comprehensive Analysis of Brain Signal Data:
In summary, this hybrid approach offers a comprehensive analysis of complex brain signal data [19, 20, 21]. The spatial acuity provided by CNNs, coupled with the temporal depth of LSTMs, forms a synergistic relationship that advances the understanding of intricate EEG patterns associated with schizophrenia. The model's adaptability to both spatial and temporal aspects contributes to its effectiveness in real-world scenarios.

V.4 Performance Evaluation

V.4.1 Accuracy :
One important measure of a deep learning (DL) model's capacity to accurately categorize or forecast results is its accuracy. It calculates the proportion of accurately predicted cases to all occurrences that were assessed. High accuracy is a sign that the model is good at picking up patterns from the training set, but in order to fully evaluate the model's performance in real-world situations, other metrics and possible biases must be taken into account. An evaluation of a DL model's correctness that is more robust is achieved through regular validation and testing against a variety of datasets.

V.4.2 F1-Score :
A popular metric in machine learning for assessing a model's performance, particularly in binary classification tasks, is the F1-score. It offers a fair assessment of a model's accuracy by integrating recall and precision into a single score. When there is an imbalance in the number of instances across different classes, the F1-score is especially helpful. A higher F1-score denotes a better balance between precision and recall. It is computed as the harmonic mean of precision and recall, making it a useful metric for evaluating a model's overall efficacy in handling both false positives and false negatives.

V.4.3 Precision **:**
Precision is a critical machine learning metric that evaluates the accuracy of a model's positive predictions, especially in binary classification tasks. It gives information about

how well the model avoids false positives and is computed as the ratio of true positive predictions to all predicted positives. A high precision score shows that the model minimizes the number of false positives while accurately identifying positive instances. When it comes to situations like fraud detection or medical diagnosis, where the cost or impact of false positives can be substantial, precision is particularly important. However, in order to get a thorough assessment of a model's performance, precision should be taken into account in addition to other metrics like recall and F1-score.

V.4.4 Sensitivity :

Sensitivity, also known as recall or true positive rate, is a crucial performance metric in machine learning, particularly for tasks that require binary classification. It assesses a model's ability to precisely locate each positive class instance within the dataset. The model's sensitivity is a gauge of how well it captures all relevant examples of the positive class. The ratio of actual positives to true positive predictions is used to calculate it. In applications like intrusion detection systems or medical diagnosis, where missing positive instances can have serious consequences, a high sensitivity score indicates that the model performs well at reducing false negatives. For these reasons, the model is crucial. For a comprehensive evaluation, however, sensitivity should be considered in addition to precision and other relevant metrics.

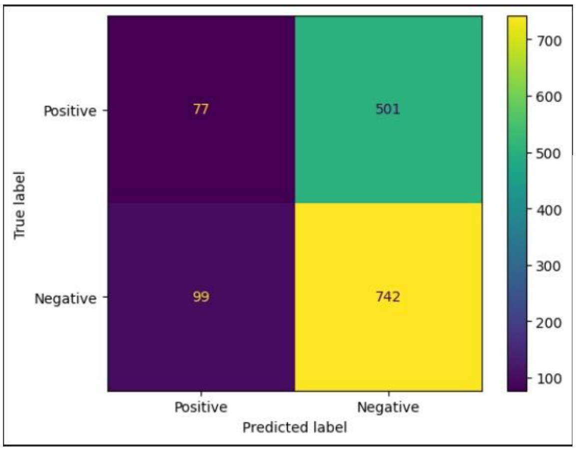
V.4.5 Specificity :

Specificity is a crucial machine learning metric that measures a model's accuracy in identifying negative examples in binary classification tasks. The ratio of true negatives to all actual negatives is used to calculate it. Specificity sheds light on how well the model prevents false positives, which makes it especially important in situations where misclassifying negative cases could have serious costs or repercussions. A high specificity score adds to a more thorough understanding of the model's performance by demonstrating that it is adept at accurately identifying true negatives. Although sensitivity and specificity are frequently taken into account in tandem, the decision between the two is contingent upon the particular needs and application priorities, underscoring the significance of a well-rounded evaluation methodology.

VI. Result Analysis

k-Nearest Neighbours :

Confusion matrix & Graph :



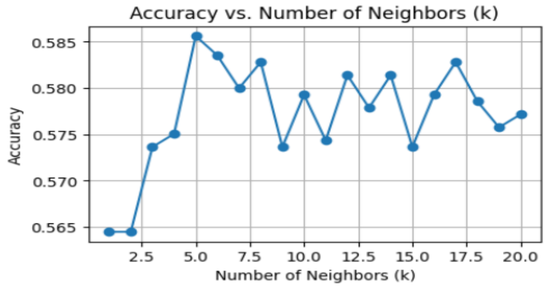


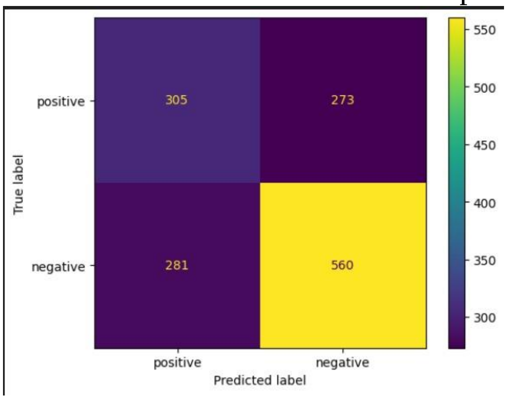
Fig II. Confusion matrix of KNN

Fig III. Accuracy of KNN related to k

The graph shows directly proportional relationship for accuracy and number of neighbours till a point. After that point, the accuracy decreases. This is because of overfitting with less value of K and increasing noise with too high of a value of K. The accuracy of the classifier peaks at a value of 0.580 when k = 5.

Support Vector Machine :

Confusion Matrix & Graph:



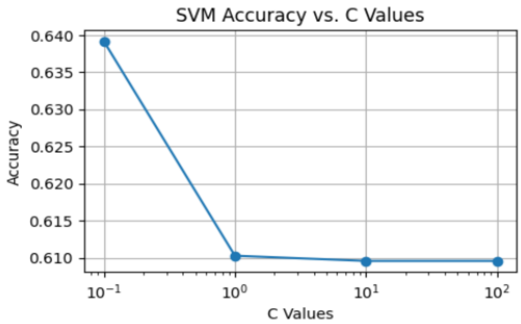


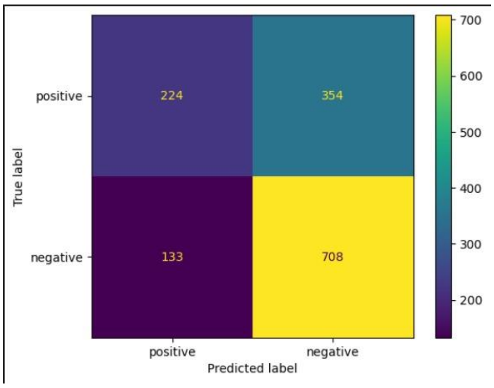
Fig IV Confusion Matrix of SVM

Fig V. Accuracy of SVM with C values

As shown in the graph, the accuracy increases with increase in value of C, till a point. After that point, the accuracy decreases. This is because
increasing the value of C too much can lead to overfitting . The accuracy of then peaks at a value of 0.640 at C = 100

Random Forest :

Confusion Matrix & Graph :



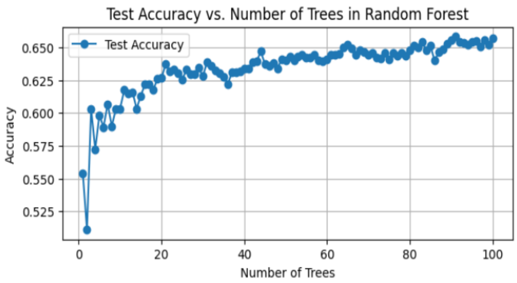
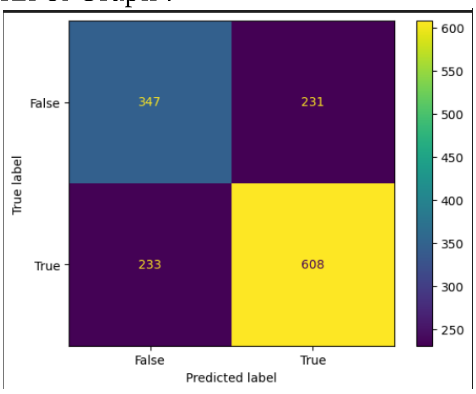


Fig VI. Confusion matrix of RF. *Fig VII. Accuracy of RF with number of trees*

The graph shows the test accuracy of a random forest model as the number of trees in the model increases. The graph indicates that the test accuracy of the model initially increases as the number of trees increases, reaching a peak around 65% accuracy with 62 trees.

CNN:
Confusion matrix & Graph :



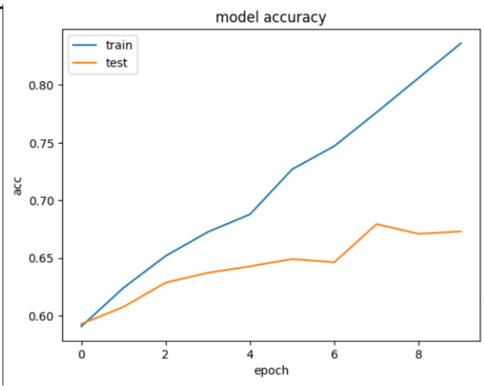


Fig VIII. Confusion matrix of CNN *Fig IX. Model accuracy*

The confusion matrix shows that the model is performing well overall, with an accuracy of 67.3%. However, it is making more false positives (231) than false negatives (233). This means that the model is more likely to predict that a data point is positive when it is actually negative.

The graph shows that the model accuracy on the training dataset increases steadily over time, while the model accuracy on the test dataset also increases, but more slowly. This suggests that the model is overfitting the training data.

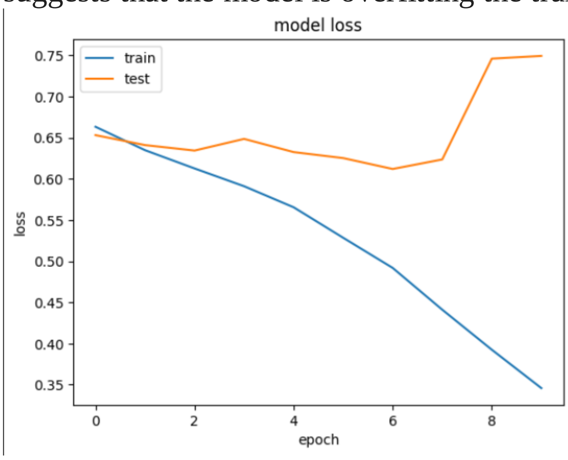
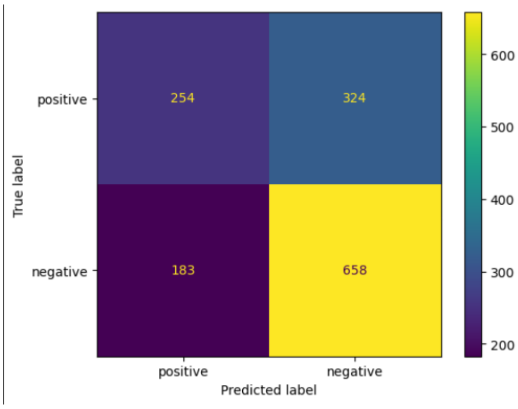


Fig X. Model loss

The graph shows that the train loss is higher than the test loss for all epochs. This is a good sign, as it means that the model is learning to generalize to new data. However, the train loss and test loss do not plateau over time, which suggests that the model may not be fully trained yet.

LSTM:

Confusion matrix:



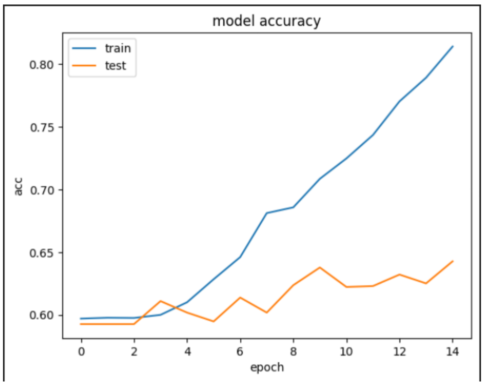


Fig XI. LSTM confusion matrix. Fig XIII. Model accuracy

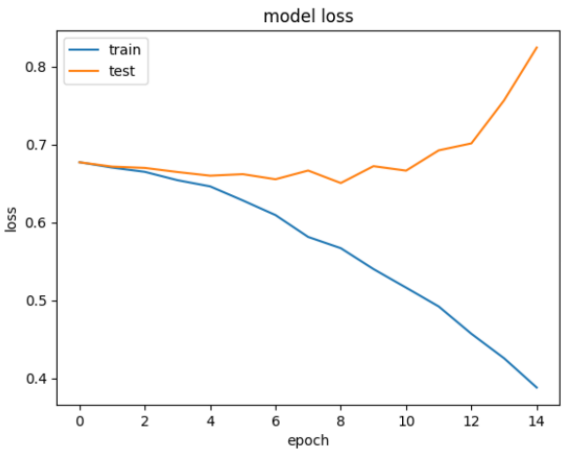


Fig X1I. Model loss

The graph shows that both the train loss and the test loss decrease over time. However, the train loss decreases more quickly than the test loss. This suggests that the model may be overfitting the training data.

VI.1 Comparative Analysis

	Accuracy	F1score	Precision	Recall
SUPPORT VECTOR MACHINE	63.918%	0.67	0.67	0.67
RANDOM FOREST	65.821%	0.74	0.67	0.84
k-NEAREST NEIGHBOURS	58.562%	0.71	0.60	0.88
CNN	67.3%	0.72	0.72	0.72
LSTM	64.27%	0.72	0.67	0.78

Table I. Comparative Analysis

The SVM classifier in the image achieved an acc. of 63.918%, a prec. of 0.67, a rec.of 0.67, and an F-measure of 0.67. This means that the SVM classifier was able to correctly classify 63.918% of the images in the dataset. The prec. score of 0.67 means that 67% of the images that the SVM classifier predicted to be positive were actually positive. The rec. score of 0.67 means that the SVM classifier was able to identify 67% of the positive images in the dataset. The F-measure of 0.67 is a harmonic mean of the prec. and rec. scores, and it is a measure of the overall performance of the classifier.

The Random Forest classifier in the image achieved an acc. of 65.821%, a prec. of 0.67, a rec. of 0.84, and an F-measure of 0.74. This means that the Random Forest classifier was able to correctly classify 65.821% of the images in the dataset. The prec. score of 0.67 means that 67% of the images that the Random Forest classifier predicted to be positive were actually positive. The rec.score of 0.84 means that the Random Forest classifier was able to identify 84% of the positive images in the dataset.

The KNN classifier in the image achieved an acc. of 58.562%, a prec. of 0.60, a rec. of 0.88, and an F-measure of 0.71. This means that the KNN classifier was able to correctly classify 58.562% of the images in the dataset. The prec. score of 0.60 means that 60% of the images that the KNN classifier predicted to be positive were actually positive. The rec. score of 0.88 means that the KNN classifier was able to identify 88% of the positive images in the dataset.

In the relentless pursuit of advancing mental health diagnostics, this research, titled "Comparative Analysis of Machine Learning and Hybrid Deep Learning Algorithm for Schizophrenia Detection using EEG Signals," has skillfully navigated the intricate terrain of schizophrenia diagnosis. The study, rooted in a recognition of the limitations of current diagnostic methods, has successfully addressed the pressing need for precision and timeliness in identifying this complex mental health disorder.

The research embarked on a comprehensive comparative analysis of traditional machine learning classifiers, such as k-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Random Forest. Each classifier underwent a rigorous evaluation based on key metrics. For instance, the SVM classifier demonstrated an acc. of 63.918%, a prec. of 0.67, a rec. of 0.67, and an F-measure of 0.67. Similarly, the Random Forest classifier exhibited an acc. of 65.821%, a prec. of 0.67, a rec. of 0.84, and an F-measure of 0.74. The KNN classifier achieved an acc. of 58.562%, a prec. of 0.60, a rec. of 0.88, and an F-measure of 0.71. The CNN classifier achieved an acc. of 67.3%, a prec. of 0.72, a rec. of 0.72, and an F-measure of 0.72. The LSTM classifier achieved an acc. of 58.562%, a prec. of 0.72, a rec. of 0.67, and an F-measure of 0.78. These numerical representations emphasize the foundation of hybrid approach.

	TP	FP	FN	TN	SENSITIVITY	SPECIFICITY
SUPPORT VECTOR MACHINE	305	273	281	560	0.520	0.672
RANDOM FOREST	224	354	133	708	0.627	0.667
k-NEAREST NEIGHBOURS	77	501	99	742	0.437	0.597
CNN	347	231	233	608	0.598	0.724
LSTM	254	324	183	658	0.439	0.782

Table II. Metrics of all models

VII. Conclusion

In an innovative move, the research combined deep learning methods with connectivity insights to create a hybrid deep learning framework that aimed to capitalize on the synergistic benefits of both approaches. The synergistic improvement in diagnostic precision was made possible by this novel approach. Using quantitative metrics like ROC curves, accuracy, sensitivity, and specificity, the thorough performance comparison and evaluation gave rise to a sophisticated understanding of the advantages and disadvantages of each approach.

Most importantly, the study goes beyond scholarly research and makes a significant addition to the field of mental health diagnostics. The study not only establishes the foundation for upcoming diagnostic systems but also has the potential to impact clinical practices and establish new benchmarks by deepening our understanding of EEG signal analysis. The research findings can be practically applied, as evidenced by the early intervention potential that has been shown and validated by quantifiable metrics. This can lead to better patient outcomes by prompt diagnosis and intervention. In summation, this research stands as a beacon illuminating the path toward redefining the landscape of mental health diagnostics. As the numerical findings permeate academic and clinical circles, they possess the transformative potential to shape future research trajectories and contribute to a paradigm shift at the confluence of machine learning, deep learning, and mental health diagnostics.

VIII. Future scope

The successful implementation and comparison of hybrid Deep Learning (DL) and traditional machine learning techniques for schizophrenia detection lay a solid foundation for promising future endeavors. The superior performance of the hybrid DL technique positions it as a key player in the development of advanced diagnostic tools. The following can be explored for future enhancements and applications:

Computer-Aided Diagnostic Systems:

The hybrid DL technique's apparent superiority points to its potential for use in the creation of computer-aided diagnostic (CAD) systems. These devices might give physicians prompt, precise diagnoses, enabling them to make decisions in the clinical setting more quickly. Consequently, this would enable individuals with schizophrenia as well as healthcare professionals to take proactive measures and carry out prompt interventions.

Real-Time Diagnosis:

Building real-time diagnostic tools could be the focus of future work, expanding on the use of hybrid deep learning. With real-time diagnosis, assessments could be made right away, enabling quick reactions to changing patient conditions. These tools have the potential to greatly increase clinical workflow efficiency and enhance patient outcomes.

User-Friendly Interfaces:

Future research can concentrate on developing user-friendly CAD system interfaces to improve accessibility. More people would adopt and use intuitive interfaces that are simple to use, which would make mental health diagnostics more accessible and effective.

Optimized Data Preprocessing:

There is room for investigation given the mention of looking into more effective data preprocessing techniques, especially those that involve using wavelet format for classifiers. Even greater accuracy levels may be attained with more research into data preprocessing technique optimization. This might entail experimenting with various wavelet transforms or investigating cutting-edge preprocessing techniques customized to the properties of EEG data.

Integration of Multimodal Data:

Subsequent investigations may examine the amalgamation of diverse data sources, merging EEG data with pertinent additional information. Supplementary information, like behavioral patterns or genetic markers, may improve the diagnostic models' comprehensiveness and offer a more comprehensive understanding of schizophrenia.

Longitudinal Studies:

Longitudinal research could yield important information about the dynamic nature of schizophrenia. A deeper understanding of the disorder's progression may result from monitoring changes in EEG patterns over time and comparing them with clinical outcomes. This could lead to more individualized and successful treatment plans.

Collaboration with Healthcare Institutions:

Large-scale clinical validation studies conducted in conjunction with healthcare facilities can verify the efficacy and dependability of the developed diagnostic systems. Additional adjustments and enhancements can be guided by feedback from medical professionals and real-world implementation.

In conclusion, the effective application of hybrid DL approaches creates new and exciting opportunities for the advancement of schizophrenia diagnosis research. Subsequent investigations and advancements should concentrate on converting these discoveries into workable solutions that benefit psychiatrists and people with schizophrenia.

IX. References

1. Why sex differences in schizophrenia?
Rena Li,[1,2,3,*] Xin Ma,[1] Gang Wang,[1] Jian Yang,[1] and Chuanyue Wang[1,2]
2. Cognitive dysfunction in schizophrenia: An expert group paper on the current state of the art
Philip D. Harvey a, Marta Bosia b, Roberto Cavallaro b, Oliver D. Howes c d, René S. Kahn e, Stefan Leucht f, Daniel R. Müller g, Rafael Penadés h, Antonio Vita i j
3. A Multi-Domain Connectome Convolutional Neural Network for Identifying Schizophrenia from EEG Connectivity Patterns
Chun-Ren Phang, Fuad Noman, Hadri Hussain, Chee-Ming Ting*, Member, IEEE and Hernando Ombao
4. 3D-CNN based discrimination of schizophrenia using resting-state fMRI
Muhammad Naveed Iqbal Qureshi, Jooyoung, Boreom Lee g
5. Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders. J. Neurosci. Methods 341, 108756.
Zhao, J., Huang, J., Zhi, D., Yan, W., Ma, X., Yang, X., Li, X., Ke, Q., Jiang, T., Calhoun, V. D., Sui, J., 2020.
6. On the use of pairwise distance learning for brain signal classification with limited observations
David Calhas 1, Enrique Romero 2, Rui Henriques 3
7. Neural memory plasticity for medical anomaly detection
Tharindu Fernando a,∗ , Simon Denman a , David Ahmedt-Aristizabal a , Sridha Sridharan a , Kristin R. Laurens b , Patrick Johnston b , Clinton Fookes
8. Schizophrenia Detection Using Machine Learning Approach from Social Media Content
Yi Ji Bae 1 , Midan Shim 1,2 and Won Hee Lee 1,
9. Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults
Jason K. Johannesen1,2* , Jinbo Bi3 , Ruhua Jiang3 , Joshua G. Kenney1,2 and Chi-Ming A. Chen4
10. Stress, Schizophrenia, and Violence: A Machine Learning Approach
Johannes Kirchebner, Martina Sonnweber, Urs M. Nater, Moritz Günther, and Steffen Lau
11. 11. Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal
Sara Bagherzadeh 1 , Mohsen Sadat Shahabi 2 , Ahmad Shalbaf 3
12. A hybrid deep neural network for classification of schizophrenia using EEG Data
Jie Sun,1 Rui Cao,2 Mengni Zhou,3 Waqar Hussain,1 Bin Wang,1 Jiayue Xue,1 and Jie Xiang1
13. Deep-learning detection of mild cognitive impairment from sleep electroencephalography for patients with Parkinson's disease
Madan Parajuli, Amy W. Amara, and Mohamed Shaban

14. A hybrid deep neural network for classification of schizophrenia using EEG Data
Jie Sun, Rui Cao, Mengni Zhou, Waqar Hussain, Bin Wang, Jiayue Xue & Jie Xiang

15. A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals
Zahra Khademi [1], Farideh Ebrahimi [2], Hussain Montazery Kordy [3]

16. A Multi-Domain Connectome Convolutional Neural Network for Identifying Schizophrenia From EEG Connectivity Patterns
Chun-Ren Phang, Fuad Noman, Hadri Hussain, Chee-Ming Ting, Hernando Ombao

17. From Sound Perception to Automatic Detection of Schizophrenia: An EEG-Based Deep Learning Approach
Carla Barros, [1] Brian Roach, [2,3] Judith M. Ford, [2,3] Ana P. Pinheiro, [1,4,*,†] and Carlos A. Silva [5,6,†]

18. EEG-Based Epileptic Seizure Detection via Machine/Deep Learning Approaches: A Systematic Review
Ijaz Ahmad, [1,2,3] Xin Wang, [1,2,3] Mingxing Zhu, [2,4] Cheng Wang, [1,2,3] Yao Pi, [5] Javed Ali Khan, [6] Siyab Khan, [7] Oluwarotimi Williams Samuel, [1,3] Shixiong Chen, [1,3] and Guanglin Li [1,3]

19. Development of a Machine Learning Based Algorithm To Accurately Detect Schizophrenia based on One-minute EEG Recordings
David Beil, Ricardo Buettner, Aadel Djemai and Stefanie Scholtz

20. Detection of schizophrenia: A machine learning algorithm for potential early detection and prevention based on event-related potentials
Ricardo Buettner, Frick Janek and Thilo Rieg

21. Schizophrenia classification using machine learning on resting state EEG signal
Juan Ruiz de Miras