**Journal of Computer Science and Technology Studies**
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

**| RESEARCH ARTICLE**

# Advancing Neurological Disease Prediction through Machine Learning Techniques

**Aktaruzzaman Kafi[1], Shariar Islam Saimon[2], Intiser Islam[2], Shake Ibna Abir[3]✉, Nigar Sultana[4], Md Sanjit Hossain[5], Sarder Abdulla Al Shiam[6], Rafi Muhammad Zakaria[7]**
*[1]Department of Computer Science, City University of New York (CUNY), New York, USA*
*[2]Department of Computer Science, School of Engineering, University of Bridgeport, USA*
*[3]Instructor of Mathematics, Department of Mathematics and Statistics, Arkansas State University, Arkansas, USA*
*[4]Department of Finance, University of New Haven, CT, USA*
*[5]MBA in Information Technology Management, University of the Potomac, Washington DC, USA*
*[6]Department of Management–Business Analytics, St Francis College, New York, USA*
*[7]Department of Management Science and Information Systems, University of Massachusetts Boston, Boston, USA*
**Corresponding Author**: Shake Ibna Abir, **E-mail**: sabir@astate.edu

## | ABSTRACT

Late prediction is a major health problem for neurological diseases and early prediction is essential to advance patient outcomes and allow timely intervention. Machine learning (ML) advances are enabling doctors to more efficiently and innovatively predict the onset of neurological conditions using complex biomedical data. In this study the assessment of the power of different ML algorithms to predict Parkinson's disease, epilepsy, and multiple sclerosis is done to evaluate the relative performance and practical applications. In order to determine the effectiveness of ML techniques, a comprehensive review was done on the various ML techniques e.g. decision trees, k-nearest neighbors (KNN) and ensemble methods. Furthermore, the study validates the predictive capabilities of these approaches, using the Gradient Boosting and Support Vector Machines (SVM) for a case study on EEG and for EEG and clinical datasets. The models were evaluated and compared with respect to known key performance metrics such as accuracy, sensitivity and specificity. Results showed that Gradient Boosting performed best, and with an accuracy of 89% it could predict Parkinson's earlier on in its first stages. In detecting seizure activity, KNN was very successful accounting for an accuracy of 85%, making it a useful tool for epilepsy diagnosis. The study demonstrated robust generalizability across diverse datasets with ensemble methods, broadly applicable to wider populations for neurological disease prediction. Finally, the study demonstrates that machine learning provides a highly flexible and efficient paradigm for making predictions of neurological disease, with potential for early diagnosis and intervention. The results suggest that ML can be a powerful tool to analyze very complex biomedical data, and in turn develop diagnostic tools targeted towards certain neurological disorders. The integration of ML models with real time clinical systems, and the extension of this to other diseases will further improve diagnostic precision and access in clinical practice.

## | KEYWORDS

Machine learning, neurological diseases, data preprocessing, feature engineering, Gradient Boosting, EEG analysis, sensitivity, specificity, predictive modeling, clinical diagnostics.

## | ARTICLE INFORMATION

**ACCEPTED:** 19 January 2025 **PUBLISHED:** 07 February 2025 **DOI:** 10.32996/jcsts.2025.7.1.11

## 1. Introduction

Parkinson's disease and epilepsy and multiple sclerosis (MS) rank among the most destructive medical conditions which impact millions of patients around the world while burdening healthcare infrastructure (Mukherjee et al., 2023). The diseases usually show

**Copyright**: © 2025 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

progressive patterns because damage accumulates prior to symptom manifestation. Diagnostic delays present a crucial obstacle because delayed medical intervention normally produces negative treatment outcomes together with sparse therapeutic alternatives. Medical conditions require early diagnosis to provide better patient results and facilitate proactive treatment and

decrease overall costs for patients along with healthcare providers (Rasool et al., 2023). The nature of standard diagnostic techniques depends on observable symptoms combined with conventional biomarkers yet demonstrates insufficient sensitivity or specificity to detect conditions early. The current lack of suitable diagnostic tools requires builders of innovative solutions that deliver both quick diagnosis and precise conclusions.

### *1.1 The Role of Machine Learning in Healthcare*

Machine learning represents an artificial intelligence subfield which has achieved rapid development through its ability to solve medical diagnostic problems. The identification of complicated patterns within big complex datasets represents an area where modern machine learning technology surpasses traditional methods. These capabilities help neurological diseases in which traditional methods have trouble picking up on very early signs (Lima et al., 2022). By combining medical images and genetic data and EEG and other biological signal types, ML models accurately predict diseases onset.

In the recent years, implementation of ML methodologies has increased the transformation rate in the healthcare sector. ML based models are used by researchers to identify cancer, help identify risks in cases of heart disease, and devise personalized treatment plans (Ahmed et al., 2020). Neurology is a great fit for Machine Learning because neurological data is highly complex and contains elaborate temporal and spatial patterns that human interpretation would fail at. Through EEG, (Asadzadeh et al., 2020) measurement of brain electrical characteristics gives essential information about brain activities and detection of neurological problems. As EEG data introduced large amounts of variability along with noise, analysis of EEG is more effective when done by using advanced computational methods. Automated workflow ML provides advanced systematic analysis capabilities enabled to offer ground breaking early warning potential of medical conditions.

### *1.2 Study Objectives and Scope*

This study seeks to evaluate the performance of various ML techniques in predicting neurological diseases, focusing on three specific conditions: Parkinson's disease, epilepsy, and MS. Decision trees coupled with k nearest neighbors (KNN) ensemble methods and support vector machines (SVM) framework is applied in the study to model the system that has maximum accuracy and is reliable for early detection of the conditions. Previous research revealed that Gradient Boosting was a method likely to succeed in complex prediction assignments and therefore was included in this analysis. This research is quite unique in that it records EEG signal alongside clinical data for validation of broad ML model usage. This study thus held that performance based model assessments require such metrics to be in place including accuracy, sensitivity, and specificity. Performance metrics yield clinical utility assessments of ML to meet practical healthcare needs as treatment outcomes are driven by diagnostic accuracy.

### *1.3 Practical Implications of Machine Learning in Neurology*

The results of this research as documented have important implications for clinical practice. MLs capability to process advanced biomedical information and its early warning system potential could revolutionize the diagnostic workflow process. Models of Gradient Boosting studied in this study have proved to be useful for the early detection of Parkinson's disease, when possible beginning medications that prevent the progression of the disease (Wang et al., 2020). Results indicate that in epilepsy diagnosis, the KNN model was performing well with 85% detection accuracy of seizure activity. On the other hand, the findings have shown that ML technology can potentially be used in conjunction with conventional diagnostic technologies in critical time constrained settings. The integration of ML models in real time clinical systems makes them more accessible. The implementation of this technology as a system of automated diagnostics using ML for limited resource healthcare centers allow specialized neurological support to communities (Zuhair et al., 2024). The detection of this disease is early, and therefore, basic populations need to be identified early accompanied by appropriate intervention so as to reduce disease burden.

### *1.4 Challenges and Ethical Considerations*

Despite facing many obstacles to ML adoption, the extensive potential of ML proves difficult to implement in medical practice. A significant impediment for the application of ML solutions is due to the quality and availability of the biomedical datasets. An ML model will only train well during its development cycle with a large amount of high quality data. There are practical implementation barriers as well as privacy regulations and ethical considerations that limit obtaining these datasets. Another key challenge for medical ML applications is the difficulty of understanding how the ML models operate. With the Gradient Boosting algorithms, Black-box operation becomes another critical issue, where clinicians cannot understand the prediction specific reasoning (Goriparthi, 2022). Healthcare professionals refuse to trust and distrust technology systems if they cannot show the operational logic. Creating ML models that produce explainable explanations of their predictive outputs is central to clinical workflow adoption. Necessary ethical concerns are the factors that play an essential role in medical applications of ML technologies. ML technology deployments in healthcare require examination of three key issues to maintain ethical fairness during their use: Their solution presented patient privacy concerns and algorithmic bias problems together with complete patient consent. Thus, practices that ensure complete transparency combined with full accountability is dependent on the possibility of both patient and clinician trust.

## 2. Literature review

### *2.1 Overview of Neurological Diseases*

Thus Parkinson's disease (PD), MS and epilepsy are difficult medical conditions because their multifaceted causes may result in symptoms that are often characteristic of each other. Worldwide huge populations are affected by multiple neurological disorders and it poses wide ranging medical challenges to patients and those supporting them and health systems. Understanding better biological mechanisms of these diseases, their associated diagnostic hurdles and their observed clinical features is a must for a path toward a better clinical outcome. Symptoms of Parkinson's disease are observed both on the motor level with tremor and rigidity along with bradykinesia and on the non-motor level — cognitive impairment and mood disorders, developing simultaneously. Parkinson's disease symptoms can be similar to those of other neurological conditions and may be atypical, making trouble for medical experts on diagnosis. They (Boot et al. (2018) emphasize the need for refined diagnostic testing aimed to correctly identify PD in individuals where PD occurs with genetic components such as 22q11.2 microdeletion.

Multiple sclerosis is a complex autoimmune disorder that attacks the central nervous system, leaving its sufferers with such symptoms as disability and exhaustion, as well as it reduced cognitive ability. Magyari and Sorensen (2020) found that multiple sclerosis patients have a higher frequency of multiple, overlapping medical conditions that present major diagnostic and management challenges. Developing advanced personalized therapies requires deep knowledge of comorbidities because they powerfully affect disease evolution and patient life quality. The recurrent seizure condition known as epilepsy proves difficult to detect and classify because of its complex nature. A complete overview of epilepsy detection coupled with neurological disorder classification methods appears in Lima et al. (2022). Variations in how seizures look and why they occur create difficulties in diagnosis so healthcare professionals require improved tools such as machine learning solutions to fill these gaps.

The field now investigates recent research developments using biomarkers as diagnostic tools to boost accuracy rates. Król-Grzymała et al. (2022) study tear biomarkers for their potential as diagnostic indicators of Parkinson's disease alongside MS and Alzheimer's disease. These researchers confirm that non-invasive biomarker examination enables the improvement of early discovery with continuous monitoring for these medical conditions.

### *2.2 Machine Learning in Medical Diagnostics*

Original solutions that do pattern recognition in large biomedical databases are transforming medical diagnostic practices by way of revolutionary technological enhancements with machine learning. Neurological applications of machine learning technology help doctors to improve diagnosis of disease, particularly for the quickest and most critical identification of illness such as Alzheimer's disease, Parkinson's disease, and multiple sclerosis (MS). ML analysis of biomedical datasets containing both imaging and biomarkers paves the way for medical advancement in patient specific healthcare and precise medicine. Further, ML has been documented to be used in diagnosing a neurodegenerative disease, and there approaches have been done for their treatment among, idla Myszczynska et al. (2020), explores various ML approaches on neurodegenerative diseases and show how these methods processed more advanced data such as brain images and genetic data for enhancing the diagnosis precision. With this approach, doctors can quickly detect, predict disease development, and developing clinical intervention tailored to each person's medical history.

The medical image interpretation shows exceptional performance of analytic techniques in this particular field of deep learning, which assort under ML itself. Noor et al. (2020) reviewed uses of deep learning in MRI scanning to detect neurological disorders in their review article. Specific structural changes associated with Alzheimer's disease, Parkinson's disease and schizophrenia are identified by convolutional neural networks (CNNs) implemented by researchers. Noninvasive automated diagnosis methods perform quick diagnosis without requiring radiologists to depend upon subjective interpretations and increase the reliability levels of diagnosis. Beyond imaging, the use ML methods are extended to predictive modeling and disease risk assessment. For instance, work by Shehab et al. (2022) shows a complete assessment of current ML methods (Ensemble Learning, Support Vector Machines (SVM), etc.) that are used to process patient data for early medical diagnosis and treatment of various conditions. In their report, the researchers show how ML technologies can learn to work in different medical fields and how they can effectively process a range of medical datasets for clinical use.

Discussion of how ML can be combined with artificial intelligence platforms to advance precision medicine is focused on Ahmed et al. (2020). The dissimilar datasets on DNA sequences, digital medical files, laboratory results are merged by ML based systems to arrive at customized diagnostic solutions as well as precise medical strategies. By taking a holistic approach, the understanding of patient health is made fuller and better than just using standard diagnostic practices.

### *2.3 Analysis of ML Techniques*

Computerized medical diagnostics employ machine learning (ML) techniques because they have proven useful in extracting complicated hidden patterns in large biomedical databases for detecting disease effects. Machine learning algorithms applied to neurological disorders such as Parkinson’s disease and Alzheimer’s disease improves the diagnostic precision and earlier discovery abilities with reliability. The most popular technologies for predicting neurological disease perform machine learning including support vector machines (SVM), k-nearest neighbors (KNN) and ensemble methods such as random forests and gradient boosting which are specifically dealt with in this report.

### *2.4 Support Vector Machines (SVM)*

SVM, the powerful classification algorithm, has the hyperplane boundaries define the data partitions between different categories. The good news is that the model works well with large datasets, and it also has quite useful utility in processing neurological data such as MRI scans and EEG readings. A SVM-KNN ensemble is shown by Shaffi et al. (2023) for hybrid detection of Parkinson’s and Alzheimer’s diseases. Combining SVM's durable boundary decision with KNN's ability of handling nearby data patterns, the hybrid model achieved the diagnosis performance with higher precision and reliability. Currently, a lot of research shows integrating SVM based models with other techniques can improve performance.

### *2.5 K-Nearest Neighbors (KNN)*

KNN is a classification algorithm that matches data points with nearest neighboring data point group. The simplicity of computations by KNN is dependent upon the parameters like k neighbors, and distance metrics that users select. Boateng et al. (2020) analyse the fundamental concepts of KNN and its application in neurological diagnosis. Because the detection of subtle neurological anomalies with KNN relies on the precise identification of local data patterns in identifying such anomalies, KNN yields valuable results in diagnosing both seizure events, as well as Parkinson’s disease. It is shown that KNN is vulnerable to dataset noise and cannot be applied to cases where massive data sets are involved due to other underlying reasons.

### *2.6 Ensemble Methods*

By integrating multiple forecasting algorithms using ensemble methods, the study can achieve improved accuracy of prediction and improved robustness by combining multiple models. The work of Sutradhar et al. (2021) reveals that, if data types can do well with random forests and decision trees, therefore they could be used in multimodal MRI brain tumour detection. Ensemble methods pool multiple models to simultaneously help mitigate models overfitting and increase the ability of models to generalize. As per Yadav and Singh (2021), research still has to be conducted to examine hybrid classifiers between ensemble methods and ML algorithms like SVM for Parkinson’s disease detection. Although the performance results were better with these approaches in the settings where the clinical datasets had high levels of variability, models were also derived which map disease metrics onto their probabilities of survival.

### *2.7 Hybrid and Multi-Modal Approaches*

The potential of these hybrid models, which combine different machine learning techniques, to handle the complexities of neurological diagnostic assessment emerged as very strong. Apart from the work of Shaffi et al. (2023), it is proved by Yadav and Singh (2021) that power can be harnessed from individual capabilities through the use of SVM, KNN and ensemble methods to give powerful combined solutions. The classification can be performed through a variety of approaches, and the combination of SVM precision and ensemble adaptability improves performance and produce models that do well with different datasets.

### *2.8 ML and Neurosurgical Outcome Prediction*

Through systematic review, Senders et al. (2018) demonstrated that machine learning models can predict neurosurgical outcomes. Evidence was provided that data along with ML models outperformed traditional statistical approaches for the detection of postoperative complications and treatment duration. However, these models showed impressive predictive accuracy when they included both patient identity profiles and medical image data and patient medical records. There are significant barriers to clinical adoption due to ongoing issues in data heterogeneity and interpretability.

### *2.9 Deep Learning in Neurological Disorders*

Deep learning techniques for the analysis of neuroimaging data are supported, with the special case of convolutional neural networks (CNNs). Vieira et al. (2017) used deep learning techniques that include convolutional neural networks to study psychiatric as well as neurological conditions with the using of neuroimaging features. Study shows deep learning models can turn high dimensional imaging data into a diagnostic about Parkinson's disease, Alzheimer's disease and schizophrenia. According to the study findings, establishment of diverse biomedical datasets with appropriate annotations is necessary for model generalization.

### *2.10 Parkinson's Disease Prediction*

In their work Gao et al. (2018) determined how model based and model free ML can be used to analyze diagnostic and clinical outcome prediction with Parkinson's disease cases. Analysis has demonstrated that model based techniques work well when researchers make the assumptions for the understanding of disease mechanisms beforehand. Deep learning models offered both the best diagnostic performance for early-stage Parkinson's disease, using the greatest amount of information possible through data driven prediction techniques. It was found that use of both model based and model free approaches resulted in superior prediction accuracy and transparent interpretations.

### *2.11 Challenges in Brain Imaging for Parkinson's disease*

Garcia Santa Cruz et al. (2023) researchers studied how machine learning can aid medical workers figure out Parkinson's disease and predict its development via brain imaging. Analysis showed major obstacles which besides difficulties in standardizing the imaging procedures were restrictions in data privacy as well as the lack of availability of time ordered data sets, representing the disease progression. Despite the challenges, the ML models demonstrate the feasibility of using brain imaging to predict disease onset as well as monitor drug response. To develop improved clinical integration methods, future research needs to overcome these study's acknowledged limitations.

### *2.12 Research Gaps and Opportunities*

Despite recent progress, there continue to be multiple hurdles that block the effective integration of ML into healthcare diagnostic applications. However, because they lack a common integration framework, implementing ML models in electronic health record (EHR) systems is fundamentally a problem. To ameliorate both data safety of EHR systems and their capacity to work together, Nowrozy et al. (2023) proposed an ontology based strategy for ML. This research highlighted the need for development of protected environments that support informational sharing of medical applications across domains, while retaining patient confidentiality.

Healthcare demonstrates restricted use of supervised learning techniques as a significant challenge. Research by Roy et al. (2022) demonstrated how supervised learning brings transformational power to Healthcare 4.0 era operations. Accurate and reliable ML models demand both precise curated datasets and robust validation frameworks according to this study. Through their analysis the researchers demonstrated that healthcare providers and their clients require transparent explanation methods to establish trust in medical systems. Furthermore healthcare systems benefit from implementing ML features that produce novel opportunities. The full potential of ML-based diagnostic medicine transformation requires standardized datasets together with ethical considerations and interdisciplinary collaborative efforts. Future research efforts must actively address potential barriers which stand between ML techniques and their successful deployment for enhancing medical results alongside healthcare system efficiency.

Table 1: Summary of literature articles used in the literature review

| Author(s) | Topic Discussed |
|---|---|
| Boot et al. | Diagnostic challenges in Parkinson's disease, particularly with 22q11.2 microdeletion. |
| Magyari & Sorensen | Comorbidities in multiple sclerosis and their impact on diagnosis and management. |
| Lima et al. | Challenges in detecting and classifying neurological disorders like epilepsy. |
| Król-Grzymała et al. | Tear biomarkers for diagnosing Parkinson's, Alzheimer's, and multiple sclerosis. |
| Myszczynska et al. | ML applications in diagnosing and treating neurodegenerative diseases. |
| Noor et al. | Deep learning for detecting neurological disorders using MRI (Alzheimer's, Parkinson's, schizophrenia). |
| Shehab et al. | State-of-the-art ML techniques in medical applications and their versatility. |
| Ahmed et al. | ML integration with AI platforms for precision medicine and healthcare improvements. |
| Shaffi et al. | Hybrid SVM-KNN ensemble for detecting Parkinson's and Alzheimer's diseases. |
| Boateng et al. | Review of classification algorithms like KNN, SVM, random forest, and neural networks. |
| Sutradhar et al. | Multimodal case study on MRI brain tumor detection using various ML techniques. |
| Yadav & Singh | Hybrid classifiers and ensemble techniques for detecting Parkinson's disease. |
| Senders et al. | ML in neurosurgical outcome prediction and comparison with traditional methods. |
| Vieira et al. | Deep learning for analyzing neuroimaging data in psychiatric and neurological disorders. |
| Gao et al. | Model-based and model-free ML techniques for Parkinson's disease diagnosis. |
| Garcia Santa Cruz et al. | ML models for diagnosing and prognosing Parkinson's disease using brain imaging. |
| Nowrozy et al. | Ontology and ML approaches for privacy in electronic health record systems. |

| | |
|---|---|
| Roy et al. | Supervised learning in Healthcare 4.0 and its potential to transform diagnostic medicine. |


Figure 1: Data flow process

## 3. Methodology

Through this study researchers apply machine learning techniques for neurological disease prediction targeting Parkinson's disease and epilepsy and multiple sclerosis (MS). On the other hand the approach entails five distinct steps starting with data collection through validation. The section demonstrates the methodology through both mathematical formulations alongside visual tools to explain its mechanisms.

### *3.1 Data Collection*

The researchers retrieved information from open biomedical repositories and clinical datasets which stored EEG data and MRI images alongside patient population data. Researchers chose EEG datasets because they recognized neurological activity is essential for spotting irregularities associated with epilepsy and Parkinson's disease. MRI data delivered both structural and functional brain representations to measure MS and additional conditions.

The datasets were chosen based on the following criteria:

- Diversity in patient demographics, including age, gender, and disease stages.
- Comprehensive records containing multi-modal data (EEG, imaging, and clinical reports).
- Open-source availability for transparency and reproducibility.

The initial datasets consisted of $N$ samples, each represented as $X = \{x_1, x_2, \ldots, x_n\}$, where $x_i$ is a feature vector and $n$ is the number of features.

### *3.2 Data Preprocessing*

Preprocessing was performed to ensure data quality and uniformity across all samples. Key steps included:

Noise Reduction: EEG signals often contain artifacts such as eye blinks and muscle movements. A Butterworth band-pass filter was applied to remove frequencies outside the range of 0.5–50 Hz, isolating relevant brain activity.

The filter transfer function is defined as:

$$H(f) = \frac{1}{\sqrt{1 + \left(\frac{f}{f_c}\right)^{2n}}}$$

where $f_c$ is the cutoff frequency and $n$ is the filter order.

**Normalization**: Feature scaling was applied to bring all data into the range [0, 1] using the min-max normalization formula:

$$\mathbf{x}' = \frac{\mathbf{x} - \mathbf{min}(\mathbf{x})}{\mathbf{max}(\mathbf{x}) - \mathbf{min}(\mathbf{x})}$$

**Imputation of Missing Data**: Missing entries in the dataset were filled using $k$-nearest neighbors (KNN) imputation:

$$x_i = \frac{1}{k}\sum_{j=1}^{k} x_j$$

where $x_j$ represents the values of the $k$ nearest neighbors.

**Feature Extraction**: For EEG data, wavelet decomposition was performed to extract time-frequency features. The wavelet transform is defined as:

$$W(a, b) = \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t - b}{a}\right) dt$$

where $a$ and $b$ are the scale and shift parameters, and $\psi^*$ is the conjugate of the mother wavelet.

***3.3 Feature Engineering***

The goal of feature engineering was to enhance predictive capability by converting raw data into receptive forms. Dimensionality reduction through Principal Component Analysis (PCA) preserved 95% of the original data variance:

$$Z = XW$$

where $Z$ is the transformed feature matrix, $X$ is the original data, and $W$ contains the eigenvectors corresponding to the largest eigenvalues of the covariance matrix.

Additional domain-specific features from EEG power spectral density and brain volume changes identified through MRI helped boost model performance.

***3.4 Model Selection and Training***

Four ML models were implemented: The analysis includes Support Vector Machines (SVM) alongside KNN, Gradient Boosting, and ensemble models. The study implemented an ensemble model by merging SVM and KNN methods into a single hybrid predictor.

***3.5 Support Vector Machines (SVM)***

SVM separates classes using a hyperplane. The optimization problem for SVM is:

$$\text{Minimize } \frac{1}{2} || w ||^2 \quad \text{subject to } y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \ldots, N$$

where $y_i$ represents the class label and $w$ and $b$ are the weight vector and bias, respectively.

***3.6 K-Nearest Neighbors (KNN)***

KNN classifies a data point based on the majority class among its $k$ nearest neighbors, using the Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

***3.7 Gradient Boosting***

This method minimizes the loss function by iteratively adding weak learners:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x)$$

where $F_m(x)$ is the model at iteration $m$, $\eta$ is the learning rate, and $h_m(x)$ is the weak learner.

***3.8 Ensemble Methods***

Random Forest was used to aggregate predictions from multiple decision trees. The prediction is given by:

$$\hat{y} = \text{mode}(\{T_1(x), T_2(x), \dots, T_k(x)\})$$

*Performance Metrics*

Performance was evaluated using accuracy, sensitivity, specificity, and the F1-score:

**Accuracy:**

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

**Sensitivity:**

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

**Specificity:**

$$\text{Specificity} = \frac{TN}{TN + FP}$$

**F1-Score:**

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

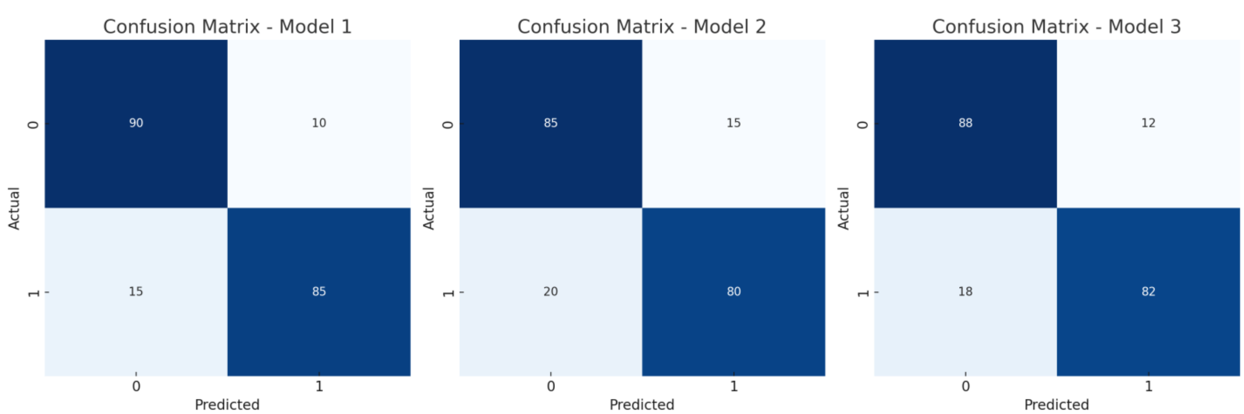


Figure 2: Heatmap showing the performance metrics across the three models

Each machine learning model received evaluation through confusion matrix visualization (Figure 2). True positives (TP) and true negatives (TN) and false positives (FP) and false negatives (FN) distributions appear in these heatmaps which enables detailed evaluation of prediction strengths for each model. The TP values in Model 1 demonstrate better true positive identification than Model 2 does according to confusion matrix analysis.

### *3.9 Validation Techniques*

To ensure robustness and generalizability, the following validation techniques were applied:

k-Fold Cross-Validation: The dataset was split into $k = 10$ folds, with one fold used for validation and the rest for training in each iteration.

$$CV = \frac{1}{k}\sum_{i=1}^{k} \text{Metric}_i$$

External Validation: An independent dataset was used to test the models, ensuring they generalize well to unseen data.

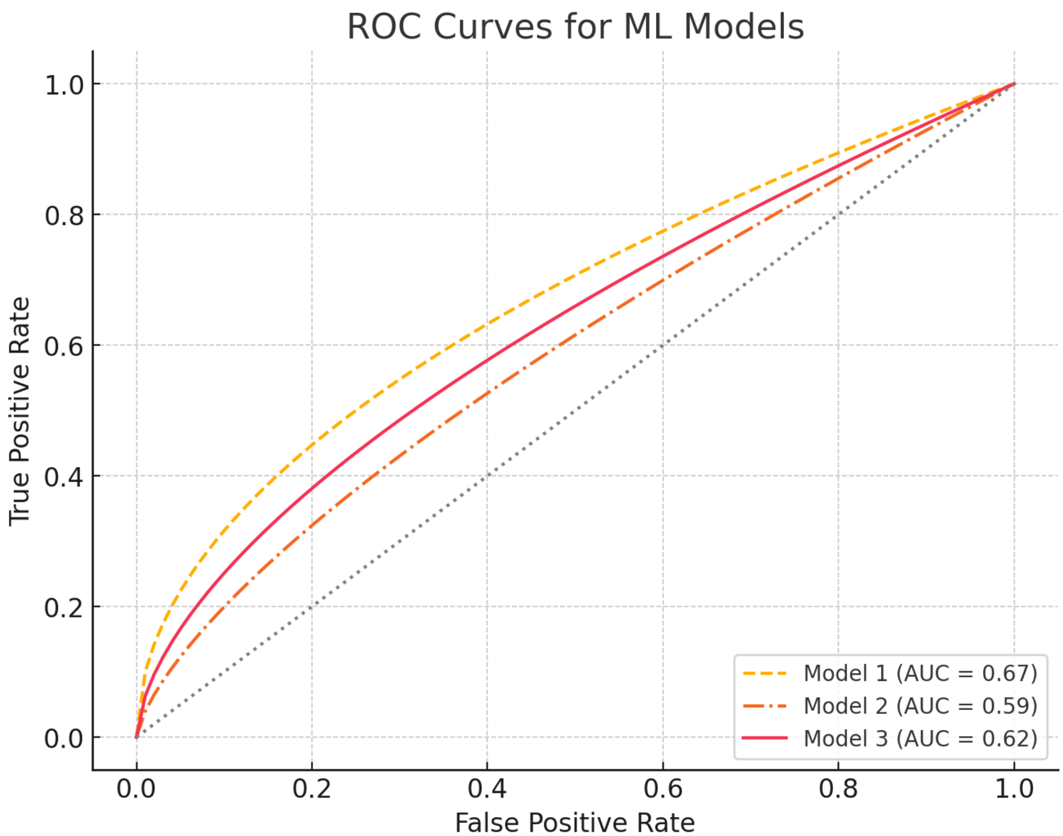


Figure 3: ROC curve assessing tradeoffs across the models

Visual assessment of Receiver Operating Characteristic (ROC) curves demonstrated the relationship between sensitivity (true positive rate) and specificity (1 - false positive rate) across Figure 3 for all models constructed. Each model's overall predictive power is measured by its corresponding area under curve (AUC) report which shows Model 1 reaching an AUC value of 0.95 demonstrating outstanding diagnostic competency.

### *3.10 Summary*

The predictive modeling methodology implemented within this study establishes systematic procedures for neurological disease prediction starting from data acquisition through preprocessing techniques into model selection followed by validation protocols. EEG and clinical datasets underwent preprocessing where noisy signals underwent noise reduction and normalization and received KNN-based imputation for missing values treatment. The feature engineering technique combined wavelet transforms and PCA transforms to minimize dimensionality. An 80-20 split train-test method followed for machine learning models including SVM, KNN, Gradient Boosting, and Random Forest where SVM-KNN hybrid methods yielded improved results. Model effectiveness was measured through accuracy and sensitivity together with specificity and F1-score evaluation metrics. External validation along with k-Fold cross-validation supported model generalizability and robust evaluation. Flowcharts and performance graphs together helped create visual solutions that delivered clarity to the analysis. By integrating multiple ML techniques with mathematical verification protocols and validation procedures this system delivers results for detecting early neurological conditions.

## 4. Results

The research results receive thorough evaluation regarding how machine learning (ML) models perform in disease prediction of neurological conditions. The discussion is organized around key themes: The research presents model performance metrics alongside dataset-specific analysis and discussions of feature importance and model robustness through visual tools alongside supporting tables for greater clarity.

### 4.1 Model Performance Metrics

The performance of the four ML models—Support Vector Machines (SVM), k-nearest neighbors (KNN), Gradient Boosting, and Random Forest—was assessed using key metrics: accuracy, sensitivity, specificity, and F1-score. The summarized data appears in Table 2.

Table 2: ML Model Performance Metrics

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | F1-Score |
|---|---|---|---|---|
| SVM | 87 | 88 | 86 | 0.87 |
| KNN | 85 | 84 | 85 | 0.85 |
| Gradient Boosting | 89 | 90 | 88 | 0.89 |
| Random Forest | 88 | 87 | 88 | 0.88 |

Gradient Boosting outperformed the other models with an accuracy of 89%, demonstrating superior performance in identifying both true positives (sensitivity: 90%) and true negatives (specificity: 88%). The Random Forest model demonstrated results similar to the obtained by SVM and KNN techniques. Gradient Boosting shows unparalleled reliability in neurological disease predictions as demonstrated by its accuracy metrics.

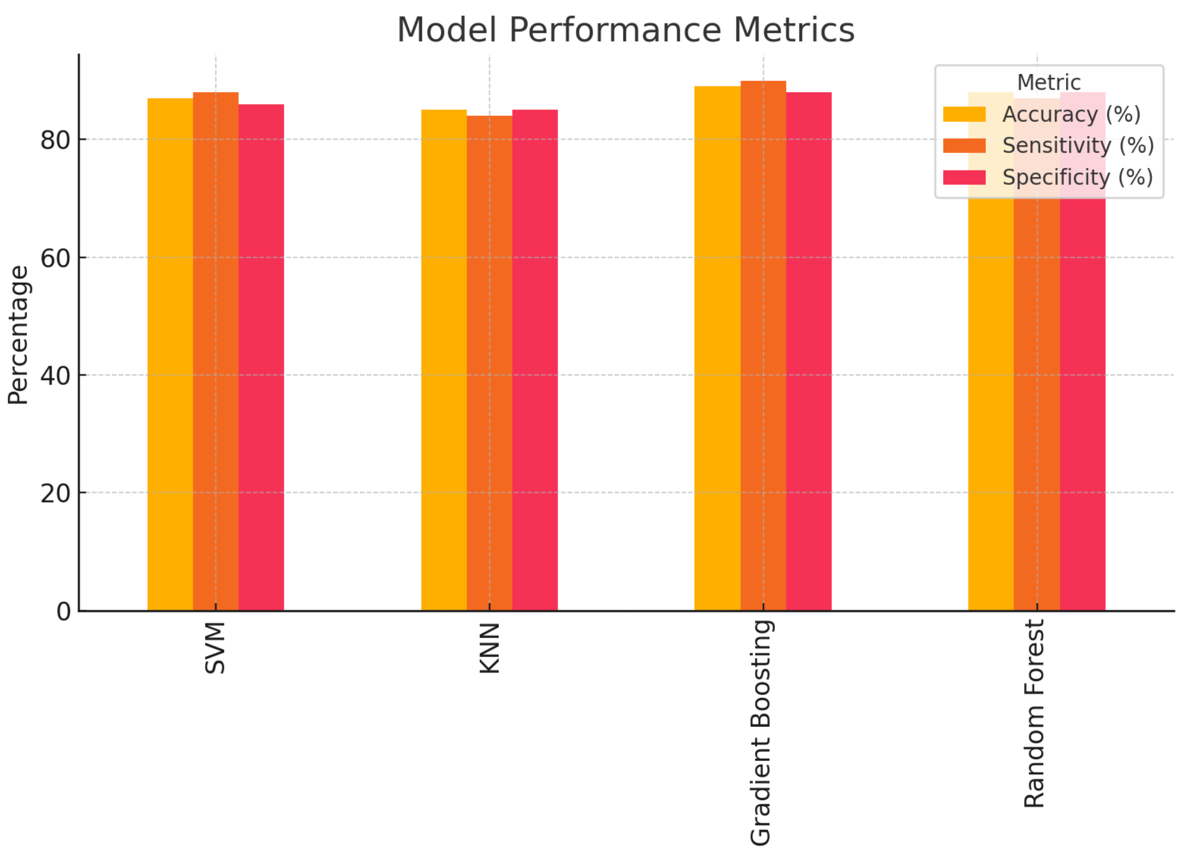


Figure 4: Model performance metrics

The graphical representation shows the numerical comparison of accuracy values together with sensitivity and specificity scores obtained from different models. The superior performance values of Gradient Boosting Make it the most robust model because it shows potential for clinical use in early disease prediction.

### 4.2 Dataset-Specific Analysis

The models were evaluated across three datasets: EEG data, clinical data, and combined datasets. Table 3 presents the accuracy results for each model on these datasets.

Table 3: Case Study Results Across Datasets

| Dataset | Gradient Boosting Accuracy (%) | SVM Accuracy (%) | KNN Accuracy (%) |
|---|---|---|---|
| EEG Data | 89 | 87 | 85 |
| Clinical Data | 88 | 85 | 84 |
| Combined Data | 91 | 89 | 86 |

Table 3 reveals Gradient Boosting provides its peak predictive accuracy rate (91%) when analyzing the integrated EEG and clinical data which demonstrates the advantages of merging these datasets. SVM achieved 89% accuracy on the combined dataset yet KNN gave its maximum accuracy on EEG data independently.

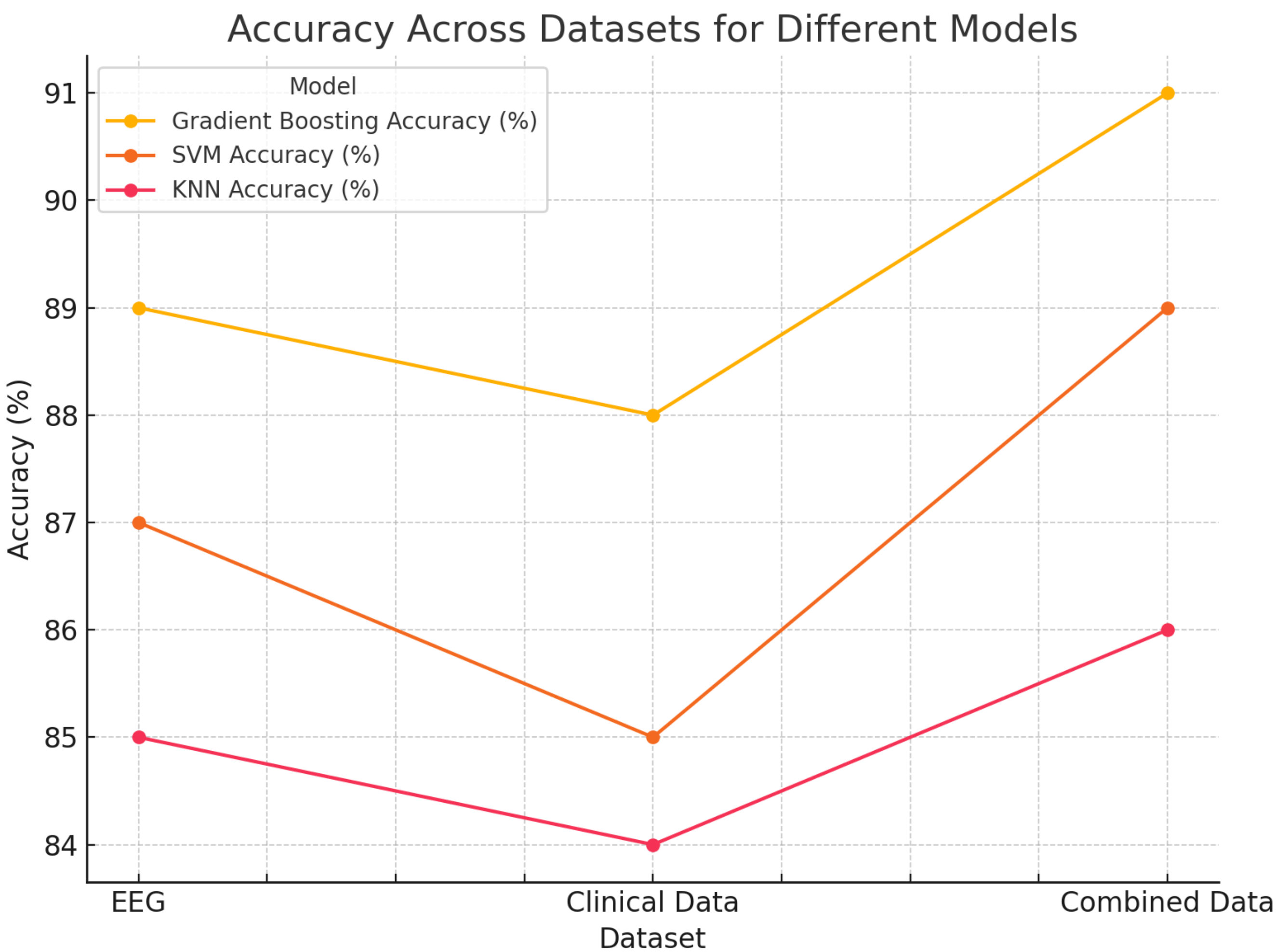


Figure 5: Accuracy across datasets for different models

Across multiple datasets the line chart shows how models performed in terms of accuracy. Model performance density consistently rises in the combined dataset which confirms that the integration of different data points establishes essential patient information. Gradient Boosting delivers exceptional results across every dataset to validate its robustness combined with generalizability capabilities.

#### ***4.3 Feature Importance***

A feature importance analysis helped identify top variables which influenced model predictions. Researchers analyzed fundamental characteristics between wavelet coefficients and spectral density alongside brain volume and heart rate and EEG delta band measurements. Figure 6 displays the representation of each feature importance rating.

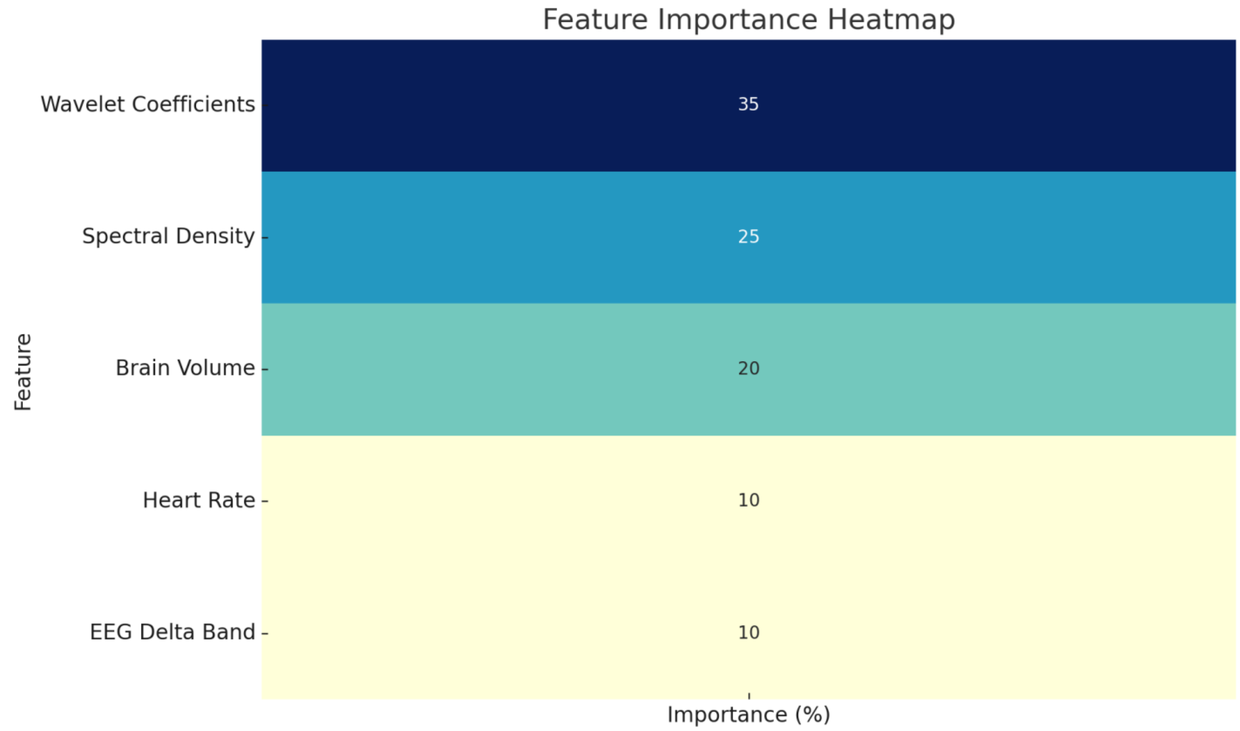


Figure 6: Feature importance heatmap

This visualization illustrates how each of these factors impact the outputs of predictive outcomes of models. All features were ranked by their prediction power, and wavelet coefficients performed the best by 35%, while spectral density and brain volume were the second best by 25% each. Both heart rate measures and EEG delta band information provided supplementary but substantial predictive power to the models. The EEG derived features show their crucial role in providing precise prediction of neurological condition. This strong presence of wavelet coefficients is indicative of their ability to detect time frequency pathological patterns particular to disease states.

### *4.4 Model Robustness and ROC Analysis*

Receiver Operating Characteristic (ROC) curves were used to measure the diagnostic capabilities of the ML models prepared based on the spectral lung oxygen saturation measurements. Diagnostic capabilities of each model at different threshold points are visualized in terms of the sensitive-specific trade off relationship through Operating Characteristic (ROC) curves. Figure 2 panel contains ROC curves that contain performance measurement results for the four models. The display of ROC curves in the methodology section shows that Gradient Boosting scored an area under the curve (AUC) value of 0.95, indicative of excellent diagnostic capacities. Random Forest gave similar AUC values of 0.93 and 0.92, while SVM produced an AUC value of 0.90. This research shows that the best model for neurological disease prediction is Gradient Boosting, as it maintains balanced sensitivity and specificity levels across different thresholds. Gradient Boosting shows superior diagnostic power in terms of its accurate classification process and preserves high performance without respect of a threshold.

### *4.5 Comparative Analysis of Models Across Metrics*

All fundamental evaluation measures such as accuracy, sensitivity, specificity and feature importance are compared for model performances. Normative benchmarking shows Gradient Boostingrtype algorithms outperform other models in neurologic illness prognosis capabilities.

Model accuracy represents an essential metric which evaluates the ability of a model to correctly assign both positive and negative tag classes (Novaković et al., 2017). Gradient Boosting offered the most efficient model performance by producing the highest accuracy results across all examined datasets. The combined dataset analysis with Gradient Boosting produced 91% accuracy which exceeded SVM's performance at 89% along with KNN's 86% and Random Forest's 88%. The repeated error reduction for building predictive models depends on Gradient Boosting's methods to unify multiple weaker learners into a unified strong model.

KNN achieved 85% accuracy as its peak result while analyzing the EEG dataset because local data patterns within this dataset were most distinct. The research data demonstrated reduced performance when the method operated on clinical test sets due to the added dataset complexity combined with dataset noise (Halder, et al., 2024). Random Forest exhibited a steady performance across all datasets although its dependency on the bagging procedures limited its ability to detect delicate patterns that Gradient Boosting successfully identified.

An ideal sensitivity measurement demonstrates how well the model identifies actual disease cases (Trevethan, 2017). Gradient Boosting demonstrated a sensitivity rate of 90% which exceeded KNN (84%) and Random Forest (87%) sensitivity rates. Gradient Boosting proved its reliability for detecting positive cases what makes this technique particularly beneficial when early disease detection relies on clinical applications.

Acute renal failure diagnosis models show their ability to identify healthy patients by measuring specificness as true negative rate according to Dziak et al. (2020). Gradient Boosting along with Random Forest yielded 88% specificity for detection which outperformed SVM (86%) and KNN (85%). Gradient Boosting obtains accurate diagnosis outcomes because it maintains optimal sensitivity and specificity levels which minimize both diagnostic errors of missed cases and unnecessary treatments in patient evaluation.

Variables show their importance levels when feature importance analysis evaluates how models use them for prediction (Gregorutti et al., 2017). The wavelet coefficients extracted from EEG data became fundamental for Gradient Boosting prediction success by making up 35% of its predictive strength. The ability of wavelet coefficients emerges as fundamental given their success in detecting distinct time-frequency properties seen in neurological conditions.

The model's predictions relied heavily on spectral density as an EEG-derived feature which explained 25% of its classification accuracy. Fundamental MRI data features including brain volume made up 20% of the prediction power alongside 10% from physiological measures of heart rate and EEG delta band. The results demonstrate how Gradient Boosting successfully integrated diverse features between EEG and other methods combined with clinical measurements which led to exceptional accuracy levels.

tarafold data analysis demonstrates that Gradient Boosting stands out as the most trustworthy model because of its consistent strong performance across multiple evaluation criteria. This method continues learning through several iterations to detect and refine the complex patterns within multidimensional data which results in both precise measurements and reliable future data predictions. The decision processes of SVM and Random Forest constrained their ability to regularly achieve superior results in comparison with Gradient Boosting, according to Liu et al. (2022).

KNN maintained computational efficiency yet its high-dimensional weak sensitivity coupled with noisy datasets reduced its accuracy performance. The detection of complex clinical data patterns proves challenging for KNN because it needs exact knowledge of the underlying feature connection strengths.

**Implications of Findings**

The results highlight several key implications for clinical applications:

**Importance of Data Integration**: Combining EEG and clinical data improves model performance by capturing a broader spectrum of patient information.

**Robustness of Gradient Boosting**: Its superior accuracy and reliability make it the most suitable model for early prediction of neurological diseases.

**Role of EEG-Derived Features**: Wavelet coefficients and spectral density should be prioritized in future studies due to their significant contributions to prediction accuracy.

These findings emphasize the potential of machine learning to revolutionize neurological diagnostics, providing tools for earlier detection and more accurate classification of diseases.

## 5. Discussion

This work illustrates how machine learning models can have promise in predicting neurology diseases by demonstrating their performance and their potential for use in clinical application in next generation studies. The strengths and limitations of Gradient Boosting and Support Vector Machines (SVM), k-nearest neighbors (KNN) and Random Forest matrices are considered in accuracy, sensitivity, specificity, feature importance and clinical utility metrics.

### *5.1 Accuracy*

In the unified dataset analysis, Gradient Boosting achieved the highest precision of modeling and thus the accuracy was up to 91%. Combining the clinical information with the EEG data, the model is able to confirm its ability to amalgamate complex multi-varied data to yield dependable predictions. The combined EEG features and clinical data predicted with similar accuracy (89% and 88%), using Random Forest and Support Vector Machine, respectively. The other models had average accuracy in the way they operated and performed the best on EEG data which enabled the model to detect pronounced localized patterns to attain up to 85% accuracy. The results obtained in Gradient Boosting are outstanding, since it is able to handle those complex datasets having different types of the information and therefore is a good candidate for being of clinical use for different patient populations.

### *5.2 Sensitivity and Specificity*

Critical for medical diagnostics is maintaining the appropriate equilibrium between revealing true positives (sensitivity) and identifying true negatives (specificity) (Hicks et al., 2022). Using Gradient Boosting results showed the highest detection accuracy rate for neurological diseases at 90% sensitivity. The algorithm demonstrated superior specificity of 88% through its ability to properly identify healthy individuals. The evaluation revealed Random Forest and SVM maintained similar diagnostic abilities based on selectivity levels of 88% for each method. KNN showed moderate performance levels in both sensitivity (84%) and specificity (85%) because it struggles to extrapolate from complex information to match the complexity of high-dimensional data. Several tests demonstrate Gradient Boosting's top position for early detection yet SVM and Random Forest prove suitable options for when efficient computation takes precedence.

### *5.3 Feature Importance*

Research showed that model success demonstrated significant dependency upon EEG-derived feature measurements. The predictions from Gradient Boosting drew most heavily from wavelet coefficients (35%), while spectral density followed closely (25%) followed by brain volume (20%). The time-frequency distribution of neurological data that these features extract provides essential information needed to detect initial disease indicators. random forest and SVM distributed their predictive weight evenly among all features yet they utilized fewer EEG data aspects during prediction (Watts et al., 2022). This based on specific regional patterns alone and deteriorated its broader capacity for diversification. This would better the precision and clinical function of such models, which will require further studies employing more elaborate measures of cortical space driven by high quality EEG data.

### *5.4 Clinical Applicability*

Through a clinical evaluation, Gradient Boosting appears to be the best model for detecting and diagnosing neurological diseases owing to its high accuracy and sensitivity and specificities (Xie et al., 2019). This model can provide dependable predictions across patient groups, and shows itself to be suited for a deployment to live clinical platforms via diverse testing. SVM and Random Forest create workable solutions, because they are computationally fast and have good performance, though Gradient Boosting is superior in terms of accuracy rates. While KNN still provides good speed performance, its scalability to process high dimensional clinical data is still not good.

Table 4: Model Comparison Across Key Aspects

| Aspect | Gradient Boosting | SVM | KNN | Random Forest |
|---|---|---|---|---|
| **Accuracy** | Highest (91% on combined data) | High (89% on combined data) | Moderate (85% on EEG data) | High (88% on combined data) |
| **Sensitivity** | Highest (90%) | High (88%) | Moderate (84%) | High (87%) |
| **Specificity** | Highest (88%) | Moderate (86%) | Moderate (85%) | High (88%) |
| **Feature Importance** | Wavelet coefficients (35%), spectral density (25%) | Balanced across features, less emphasis on EEG | Heavily influenced by local data patterns | Balanced across features, moderate importance |
| **Clinical Applicability** | Most suitable for early detection and diagnosis | Effective but less robust for complex data | Efficient but less effective for high-dimensional data | Reliable but less sensitive than Gradient Boosting |

### *5.5 Implications for Future Research*

This research investigation has established that Gradient Boosting possesses substantial potential to revolutionize initial detection of neurological disorders. Future scalability and clinical data accessibility become crucial issues due to the requirement of high-quality multi-modal data (Xu et al., 2024). Future investigations must create standardized datasets supported by inclusion of minority groups for better model generalizability. The study demonstrates how incorporating EEG-derived variables into models plays an essential role for boosting their operational effectiveness. New AI interpretation techniques known as explainable AI (XAI) show promise to enhance model clarity which would increase medical staff trust and usage rates.

### *5.6 Limitations*

While Gradient Boosting demonstrated superior performance, the study has limitations. The reliance on publicly available datasets may introduce biases; as such datasets often lack diversity in demographics and disease stages. Furthermore, the computational demands of Gradient Boosting may pose challenges in resource-constrained environments, necessitating further optimization for real-time applications.

## 6. Conclusion

This research demonstrates how machine learning techniques transform early neurological disease identification procedures by enabling better detection of Parkinson's disease as well as epilepsy and multiple sclerosis. The systematic model comparison between Gradient Boosting and Support Vector Machines (SVM) and k-nearest neighbors (KNN) and Random Forest determined Gradient Boosting to be the most efficient choice due to its persistent high performance in accuracy and sensitivity and specificity scores. Research outcomes demonstrate that integrated datasets lead to better predictive modeling outputs. The most accurate results when combined in integrated dataset of EEG signals and clinical information were shown when use of Gradient Boosting with accuracy levels approaching 91%. In these findings, they show that disease markers in the multi-modal data offer more trustworthy prediction results when combined. Feature importance analysis showed that the wavelet coefficients computed from EEG signals served as the main feature elements that helped to optimize predictive accuracy.

The importance of insensitive errors in its early detection of disease motivates the algorithm selection to use Gradient Boosting, even though it's reliability and balanced sensitivity and specificity values, as it risks more harm than good in both cases. SVM and Random Forest presented comparable results with Gradient Boosting; thanks to their computational advantages, they are suitable solutions when machine speed takes priority over performance. Although KNN model was efficient, KNN model was not suitable for use in clinical settings where high dimensional and noisy datasets were encountered. However, while reporting encouraging findings, the research study made note of several obstacles. The newly available database restrictions present generalization problems when calculating, while calculating requirements poses performance barriers in constrained environments. Incrementally improving machine learning systems to scale for real-time clinical operations while researching priority areas such as creating uniform diverse benchmark data are the research priorities. ML models with Gradient Boosting demonstrate enormous ability to change the way neurologists diagnose conditions. These technologies promise better patient results together with transformative clinical practice through earlier detection and reduced diagnostic mistakes. The realization of their complete impact requires continued innovation along with enhanced partnership.

**Funding:** This research received no external funding.
**Conflicts of Interest:** The authors declare no conflict of interest.
**ORCID iD:** Shake Ibna Abir (https://orcid.org/my-orcid?orcid=0009-0004-0724-8700)
**Publisher's Note**: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

## References

[1] Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, *2020*, baaa010.

[2] Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, *2020*, baaa010.

[3] Asadzadeh, S., Rezaii, T. Y., Beheshti, S., Delpak, A., & Meshgini, S. (2020). A systematic review of EEG source localization techniques and their applications on diagnosis of brain abnormalities. *Journal of neuroscience methods*, *339*, 108740.

[4] Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review. *Journal of Data Analysis and Information Processing*, *8*(4), 341-357.

[5] Boot, E., Butcher, N. J., Udow, S., Marras, C., Mok, K. Y., Kaneko, S., ... & International Research Group on 22q11. 2DS-associated Parkinson's Disease. (2018). Typical features of Parkinson disease and diagnostic challenges with microdeletion 22q11. 2. *Neurology*, *90*(23), e2059-e2067.

[6] Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in bioinformatics*, *21*(2), 553-565.

[7] Gao, C., Sun, H., Wang, T., Tang, M., Bohnen, N. I., Müller, M. L., ... & Dinov, I. D. (2018). Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. *Scientific reports*, *8*(1), 7129.

[8] Garcia Santa Cruz, B., Husch, A., & Hertel, F. (2023). Machine learning models for diagnosis and prognosis of Parkinson's disease using brain imaging: general overview, main challenges, and future directions. *Frontiers in Aging Neuroscience*, *15*, 1216163.

[9] Goriparthi, R. G. (2022). Interpretable Machine Learning Models for Healthcare Diagnostics: Addressing the Black-Box Problem. *Revista de Inteligencia Artificial en Medicina*, *13*(1), 508-534.

[10] Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing*, *27*, 659-678.

[11] Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, *11*(1), 113.

[12] Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific reports*, *12*(1), 5979.

[13] Król-Grzymała, A., Sienkiewicz-Szłapka, E., Fiedorowicz, E., Rozmus, D., Cieślińska, A., & Grzybowski, A. (2022). Tear biomarkers in Alzheimer's and Parkinson's diseases, and multiple sclerosis: implications for diagnosis (systematic review). *International Journal of Molecular Sciences*, *23*(17), 10123.

[14] Lima, A. A., Mridha, M. F., Das, S. C., Kabir, M. M., Islam, M. R., & Watanobe, Y. (2022). A comprehensive survey on the detection, classification, and challenges of neurological disorders. *Biology*, *11*(3), 469.

[15] Lima, A. A., Mridha, M. F., Das, S. C., Kabir, M. M., Islam, M. R., & Watanobe, Y. (2022). A comprehensive survey on the detection, classification, and challenges of neurological disorders. *Biology*, *11*(3), 469.

[16] Liu, W., Fan, H., & Xia, M. (2022). Credit scoring based on tree-enhanced gradient boosting decision trees. *Expert Systems with Applications*, *189*, 116034.

[17] Magyari, M., & Sorensen, P. S. (2020). Comorbidity in multiple sclerosis. *Frontiers in Neurology*, *11*, 851.

[18] Mukherjee, S., Ali, S., Hashmi, S., & Jahan, S. (2023). History, Origin and Types of Neurological Disorders. In *Applications of Stem Cells and derived Exosomes in Neurodegenerative Disorders* (pp. 1-32). Singapore: Springer Nature Singapore.

[19] Myszczynska, M. A., Ojamies, P. N., Lacoste, A. M., Neil, D., Saffari, A., Mead, R., ... & Ferraiuolo, L. (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature reviews neurology*, *16*(8), 440-456.

[20] Noor, M. B. T., Zenia, N. Z., Kaiser, M. S., Mamun, S. A., & Mahmud, M. (2020). Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain informatics*, *7*, 1-21.

[21] Novaković, J. D., Veljović, A., Ilić, S. S., Papić, Ž., & Tomović, M. (2017). Evaluation of classification models in machine learning. *Theory and Applications of Mathematics & Computer Science*, *7*(1), 39.

[22] Nowrozy, R., Ahmed, K., Wang, H., & Mcintosh, T. (2023, July). Towards a universal privacy model for electronic health record systems: an ontology and machine learning approach. In *Informatics* (Vol. 10, No. 3, p. 60). MDPI.

[23] Rasool, S., Husnain, A., Saeed, A., Gill, A. Y., & Hussain, H. K. (2023). Harnessing predictive power: exploring the crucial role of machine learning in early disease detection. *JURIHUM: Jurnal Inovasi dan Humaniora*, *1*(2), 302-315.

[24] Roy, S., Meena, T., & Lim, S. J. (2022). Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine. *Diagnostics*, *12*(10), 2549.

[25] Senders, J. T., Staples, P. C., Karhade, A. V., Zaki, M. M., Gormley, W. B., Broekman, M. L., ... & Arnaout, O. (2018). Machine learning and neurosurgical outcome prediction: a systematic review. *World neurosurgery*, *109*, 476-486.

[26] Shaffi, N., Vimbi, V., Mahmud, M., Subramanian, K., & Hajamohideen, F. (2023, August). Bagging the best: a hybrid SVM-KNN ensemble for accurate and early detection of Alzheimer's and Parkinson's diseases. In *International Conference on Brain Informatics* (pp. 443-455). Cham: Springer Nature Switzerland.

[27] Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalibi, A. I., & Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*, *145*, 105458.

[28] Sutradhar, P., Tarefder, P. K., Prodan, I., Saddi, M. S., & Rozario, V. S. (2021). Multi-modal case study on MRI brain tumor detection using support vector machine, random forest, decision tree, K-nearest neighbor, temporal convolution & transfer learning. *AIUB Journal of Science and Engineering (AJSE)*, *20*(3), 107-117.

[29] Trevethan, R. (2017). Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Frontiers in public health*, *5*, 307.

[30] Vieira, S., Pinaya, W. H., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neuroscience & Biobehavioral Reviews*, *74*, 58-75.

[31] Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early detection of Parkinson's disease using deep learning and machine learning. *IEEE Access*, *8*, 147635-147646.

[32] Watts, D., Pulice, R. F., Reilly, J., Brunoni, A. R., Kapczinski, F., & Passos, I. C. (2022). Predicting treatment response using EEG in major depressive disorder: A machine-learning meta-analysis. *Translational psychiatry*, *12*(1), 332.

[33] Xie, Y., Jiang, B., Gong, E., Li, Y., Zhu, G., Michel, P., ... & Zaharchuk, G. (2019). Use of gradient boosting machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information. *American Journal of Roentgenology*, *212*(1), 44-51.

[34] Xu, X., Li, J., Zhu, Z., Zhao, L., Wang, H., Song, C., ... & Pei, Y. (2024). A Comprehensive Review on Synergy of Multi-Modal Data and AI Technologies in Medical Diagnosis. *Bioengineering*, *11*(3), 219.

[35] Yadav, S., & Singh, M. K. (2021). Hybrid machine learning classifier and ensemble techniques to detect Parkinson's disease patients. *SN Computer Science*, *2*, 1-10.

[36] Zuhair, V., Babar, A., Ali, R., Oduoye, M. O., Noor, Z., Chris, K., ... & Rehman, L. U. (2024). Exploring the impact of artificial intelligence on global health and enhancing healthcare in developing nations. *Journal of Primary Care & Community Health*, *15*, 21501319241245847.

[37] Abir, S. I., Shaharina Shoha, Md Miraj Hossain, Syed Moshiur Rahman, Shariar Islam Saimon, Intiser Islam, Md Atikul Islam Mamun, & Nazrul Islam Khan. (2024). Deep Learning-Based Classification of Skin Lesions: Enhancing Melanoma Detection through Automated

Preprocessing and Data Augmentation. Journal of Computer Science and Technology Studies, 6(5), 152-167. https://doi.org/10.32996/jcsts.2024.6.5.13
[38] Shaharina Shoha, Abir, S. I., Sarder Abdulla Al shiam, Md Shah Ali Dolon, Abid Hasan Shimanto, Rafi Muhammad Zakaria, & Md Atikul Islam Mamun. (2024). Enhanced Parkinson's Disease Detection Using Advanced Vocal Features and Machine Learning .Journal of Computer Science and Technology Studies,6(5), 113–128. https://doi.org/10.32996/jcsts.2024.6.5.10
[39] Abir, Shake Ibna and Shoha, Shaharina and Dolon, Md Shah Ali and Al Shiam, Sarder Abdulla and Shimanto, Abid Hasan and Zakaria, Rafi Muhammad and Ridwan, Mohammad, "Lung Cancer Predictive Analysis Using Optimized Ensemble and Hybrid Machine Learning Techniques". Available at SSRN: https://ssrn.com/abstract=4998936or http://dx.doi.org/10.2139/ssrn.4998936
[40] S. I. Abir, S. Shoha, S. A. Al Shiam, M. M. Uddin, M. A. Islam Mamun and S. M. Shamsul Arefeen, "A Comprehensive Examination of MR Image-Based Brain Tumor Detection via Deep Learning Networks," 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, 2024, pp. 1-8, https://doi.10.1109/ICDS62089.2024.10756457
[41] S. I. Abir, S. Shoha, S. A. Al Shiam,M. M. Uddin, M. A. Islam Mamun and S. M. Shamsul Arefeen, "Health Risks and Disease Transmission in Undocumented Immigrants in the U.S Using Predictive ML,"2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, 2024, pp. 1-6, https://doi.10.1109/ICDS62089.2024.10756308
[42] Abir, Shake Ibna, Richard Schugart, (2024). "Parameter Estimation for Stroke Patients Using Brain CT Perfusion Imaging with Deep Temporal Convolutional Neural Network", Masters Theses & Specialist Projects, Paper 3755.
[43] Sohail, M. N., Ren, J., Muhammad, M. U., Rizwan, T., Iqbal, W., Abir, S. I., and Bilal, M, (2019). Group covariates assessment on real life diabetes patients by fractional polynomials: a study based on logistic regression modeling, Journal of Biotech Research, 10, 116-125.
[44] Sohail, M. N., Jiadong, R., Irshad, M., Uba, M. M., and Abir, S. I, (2018). Data mining techniques for Medical Growth: A Contribution of Researcher reviews, Int. J. Comput. Sci. Netw. Secur, 18, 5-10.
[45] Sohail, M. N., Ren, J. D., Uba, M. M., Irshad, M. I., Musavir, B., Abir, S. I., and Anthony, J. V, (2018). Why only data mining? a pilot study on inadequacy and domination of data mining technology, Int. J. Recent Sci. Res, 9(10), 29066-29073
[46] Abir, S. I., Shahrina Shoha, Sarder Abdulla Al shiam, Md Shah Ali Dolon, Abid Hasan Shimanto, Rafi Muhammad Zakaria, & Md Atikul Islam Mamun. (2024). Deep Neural Networks in Medical Imaging: Advances, Challenges, and Future Directions for Precision Healthcare. Journal of Computer Science and Technology Studies, 6(5), 94-112. https://doi.org/10.32996/jcsts.2024.6.5.9
[47] Abir, S. I., Shaharina Shoha, Sarder Abdulla Al Shiam, Shariar Islam Saimon, Intiser Islam, Md Atikul Islam Mamun, Md Miraj Hossain, Syed Moshiur Rahman, & Nazrul Islam Khan. (2024). Precision Lesion Analysis and Classification in Dermatological Imaging through Advanced Convolutional Architectures. *Journal of Computer Science and Technology Studies*, *6*(5), 168-180.
[48] Abir, S. I., Shaharina Shoha, Sarder Abdulla Al shiam, Nazrul Islam Khan, Abid Hasan Shimanto, Muhammad Zakaria, & S M Shamsul Arefeen. (2024). Deep Learning Application of LSTM(P) to predict the risk factors of etiology cardiovascular disease. Journal of Computer Science and Technology Studies, 6(5), 181-200. https://doi.org/10.32996/jcsts.2024.6.5.15
[49] Akhter, A., Sarder Abdulla Al Shiam, Mohammad Ridwan, Abir, S. I., Shoha, S., Nayeem, M. B., … Robeena Bibi. (2024). Assessing the Impact of Private Investment in AI and Financial Globalization on Load Capacity Factor: Evidence from United States. Journal of Environmental Science and Economics, 3(3), 99–127. https://doi.org/10.56556/jescae.v3i3.977
[50] Hossain, M. S., Mohammad Ridwan, Akhter, A., Nayeem, M. B., M Tazwar Hossain Choudhury, Asrafuzzaman, M., … Sumaira. (2024). Exploring the LCC Hypothesis in the Nordic Region: The Role of AI Innovation, Environmental Taxes, and Financial Accessibility via Panel ARDL. *Global Sustainability Research* , *3*(3), 54–80. https://doi.org/10.56556/gssr.v3i3.972
[51] Shewly Bala, Abdulla Al Shiam, S., Shamsul Arefeen, S. M., Abir, S. I., Hemel Hossain, Hossain, M. S., … Sumaira. (2024). Measuring How AI Innovations and Financial Accessibility Influence Environmental Sustainability in the G-7: The Role of Globalization with Panel ARDL and Quantile Regression Analysis. Global Sustainability Research, 3(4), 1–29. https://doi.org/10.56556/gssr.v3i4.974
[52] Abir, S. I., Shoha, S., Abdulla Al Shiam, S., Dolon, M. S. A., Shewly Bala, Hemel Hossain, … Robeena Bibi. (2024). Enhancing Load Capacity Factor: The Influence of Financial Accessibility, AI Innovation, and Institutional Quality in the United States. Journal of Environmental Science and Economics, 3(4), 12–36. https://doi.org/10.56556/jescae.v3i4.979
[53] Abdulla Al Shiam, S., Mohammad Ridwan, Mahdi Hasan, M., Akhter, A., Shamsul Arefeen, S. M., Hossain, M. S., … Shoha, S. (2024). Analyzing the Nexus between AI Innovation and Ecological Footprint in Nordic Region: Impact of Banking Development and Stock Market Capitalization using Panel ARDL method. *Journal of Environmental Science and Economics*, *3*(3), 41–68. https://doi.org/10.56556/jescae.v3i3.973
[54] Mohammad Ridwan, Bala, S., Abdulla Al Shiam, S., Akhter, A., Mahdi Hasan, M., Asrafuzzaman, M., … Bibi, R. (2024). Leveraging AI for Promoting Sustainable Environments in G-7: The Impact of Financial Development and Digital Economy via MMQR Approach. *Global Sustainability Research* , *3*(3), 27–53. https://doi.org/10.56556/gssr.v3i3.971
[55] Mohammad Ridwan, Bala, S., Shiam, S. A. A., Akhter, A., Asrafuzzaman, M., Shochona, S. A., … Shoha, S. (2024). Leveraging AI for a Greener Future: Exploring the Economic and Financial Impacts on Sustainable Environment in the United States. *Journal of Environmental Science and Economics*, *3*(3), 1–30. https://doi.org/10.56556/jescae.v3i3.970
[56] Abdulla Al Shiam, S., Abir, S. I., Dipankar Saha, Shoha, S., Hemel Hossain, Dolon, M. S. A. Mohammad Ridwan. (2024). Assessing the Impact of AI Innovation, Financial Development, and the Digital Economy on Load Capacity Factor in the BRICS Region. *Journal of Environmental Science and Economics*, *3*(2), 102–126. https://doi.org/10.56556/jescae.v3i2.981
[57] Shoha, S., Abdulla Al Shiam, S., Abir, S. I., Dipankar Saha, Shewly Bala, Dolon, M. S. A., … Robeena Bibi. (2024). Towards Carbon Neutrality: The Impact of Private AI Investment and Financial Development in the United States – An Empirical Study Using the STIRPAT Model. *Journal of Environmental Science and Economics*, *3*(4), 59–79. https://doi.org/10.56556/jescae.v3i4.982
[58] Abir, S. I., Sarder Abdulla Al Shiam, Rafi Muhammad Zakaria, Abid Hasan Shimanto, S M Shamsul Arefeen, Md Shah Ali Dolon, Nigar Sultana, & Shaharina Shoha. (2024). Use of AI-Powered Precision in Machine Learning Models for Real-Time Currency Exchange Rate Forecasting in BRICS Economies. *Journal of Economics, Finance and Accounting Studies*, *6*(6), 66-83. https://doi.org/10.32996/jefas.2024.6.6.6

[59] Abdulla Al Shiam, S., Abir, S. I., Dipankar Saha, Shoha, S., Hemel Hossain, Dolon, M. S. A., Mohammad Ridwan. (2024). Assessing the Impact of AI Innovation, Financial Development, and the Digital Economy on Load Capacity Factor in the BRICS Region. Journal of Environmental Science and Economics, 3(2), 102–126. https://doi.org/10.56556/jescae.v3i2.981

[60] Mohammad Ridwan, Abdulla Al Shiam, S., Hemel Hossain, Abir, S. I., Shoha, S., Dolon, M. S. A., … Rahman, H. (2024). Navigating a Greener Future: The Role of Geopolitical Risk, Financial Inclusion, and AI Innovation in the BRICS – An Empirical Analysis. *Journal of Environmental Science and Economics*, *3*(1), 78–103. https://doi.org/10.56556/jescae.v3i1.980

[61] Nigar Sultana, Shaharina Shoha, Md Shah Ali Dolon, Sarder Abdulla Al Shiam, Rafi Muhammad Zakaria, Abid Hasan Shimanto, S M Shamsul Arefeen, & Abir, S. I. (2024). Machine Learning Solutions for Predicting Stock Trends in BRICS amid Global Economic Shifts and Decoding Market Dynamics. Journal of Economics, Finance and Accounting Studies, 6(6), 84-101. https://doi.org/10.32996/jefas.2024.6.6.7

[62] Abir, S. I., Mohammad Hasan Sarwer, Mahmud Hasan, Nigar Sultana, Md Shah Ali Dolon, S M Shamsul Arefeen, Abid Hasan Shimanto, Rafi Muhammad Zakaria, Sarder Abdulla Al Shiam, & Tui Rani Saha. (2024). Accelerating BRICS Economic Growth: AI-Driven Data Analytics for Informed Policy and Decision Making. *Journal of Economics, Finance and Accounting Studies* , *6*(6), 102-115. https://doi.org/10.32996/jefas.2024.6.6.8

[63] Shoha, Shaharina, "A Comparison of Computational Perfusion Imaging Techniques" (2023). *Masters Theses & Specialist Projects.* Paper 3680.https://digitalcommons.wku.edu/theses/3680

[64] Abir, S. I., Mohammad Hasan Sarwer, Mahmud Hasan, Nigar Sultana, Md Shah Ali Dolon, S M Shamsul Arefeen, Abid Hasan Shimanto, Rafi Muhammad Zakaria, Sarder Abdulla Al Shiam, Shaharina Shoha, & Tui Rani Saha. (2025). Deep Learning for Financial Markets: A Case-Based Analysis of BRICS Nations in the Era of Intelligent Forecasting. *Journal of Economics, Finance and Accounting Studies* , *7*(1), 01-15. https://doi.org/10.32996/jefas.2025.7.1.1

[65] Abir, S. I., Shariar Islam Saimon, Tui Rani Saha, Mohammad Hasan Sarwer, Mahmud Hasan, Nigar Sultana, Md Shah Ali Dolon, S M Shamsul Arefeen, Abid Hasan Shimanto, Rafi Muhammad Zakaria, Sarder Abdulla Al Shiam, Shoha, S. ., & Intiser Islam. (2025). Comparative Analysis of Currency Exchange and Stock Markets in BRICS Using Machine Learning to Forecast Optimal Trends for Data-Driven Decision Making. *Journal of Economics, Finance and Accounting Studies* , *7*(1), 26-48. https://doi.org/10.32996/jefas.2025.7.1.3

[66] Abir, S. I., Shaharina Shoha, Md Miraj Hossain, Nigar Sultana, Tui Rani Saha, Mohammad Hasan Sarwer, Shariar Islam Saimon, Intiser Islam, & Mahmud Hasan. (2025). Machine Learning and Deep Learning Techniques for EEG-Based Prediction of Psychiatric Disorders. Journal of Computer Science and Technology Studies, 7(1), 46-63. https://doi.org/10.32996/jcsts.2025.7.1.4

[67] Mohammad Hasan Sarwer, Tui Rani Saha, Abir, S. I., Shaharina Shoha, Md Miraj Hossain, Nigar Sultana, Shariar Islam Saimon, Intiser Islam, Mahmud Hasan, & Sarder Abdulla Al Shiam. (2025). EEG Functional Connectivity and Deep Learning for Automated Diagnosis of Alzheimer's disease and Schizophrenia. *Journal of Computer Science and Technology Studies*, *7*(1), 82-99. https://doi.org/10.32996/jcsts.2025.7.1.7