School of Engineering, Computing and Mathematics

Programme

MSc Health Data Science and Statistics

Module: MATH516

Machine learning and Artificial Intelligence for Healthcare

Course work

Machine Learning-based Analysis of Real-World Biomedical Data:

EEG-based epilepsy detection

Student Name

Oluwabunmi Oluwayemisi Akintunde

Student ID

10938375

Table of Contents

1. Summary

The human brain encounters excessive electrical activity when someone has an epileptic seizure resulting in electroencephalography (EEG) alterations. This research investigates how machine learning operates on EEG data for detecting epilepsy with accuracy. The study implements a complete methodology for data preprocessing, model development and feature selection which uses data from 40 EEG features that analyse delta, theta, alpha, beta and gamma bands.

The research goal entails developing a classification system that accurately separates epilepsy patients from healthy controls while determining significant EEG features, the minimum operational features and assessing additional features for performance enhancement. The performance of various machine learning classifiers will be evaluated, and their strengths and weaknesses will be compared in terms of performance, optimization, robustness, interpretability, training time and limitations.

This study helps to advance automated epilepsy diagnosis technology through machine learning therefore enabling better treatment and detection outcomes for patients and healthcare

2. Introduction

Epilepsy, a neurological condition, causes repetitive seizures which affect numerous millions of people worldwide, according to An, Kang and Lee (2020). Accurate and timely diagnosis of epilepsy is crucial for effective management and treatment, which can significantly improve the quality of life for affected individuals (Miltiadous et al., 2023). The diagnosis and tracking of epilepsy strongly rely on electroencephalography, which allows scientists to capture brain electrical signals outside of the body without invasive procedures (Stafstrom and Carmant, 2015). Modern EEG data analysis through machine learning approaches shows potential to build better and accurate methods of detecting epilepsy (Supriya et al., 2023)

The application of machine learning with EEG analysis enables automatic detection of hidden signs that indicate epileptic seizure patterns, which may be challenging for human experts to discern. The training of machine learning algorithms enables them to detect distinct EEG patterns associated with epilepsy which leads to the development of automated detection systems (Jaafar, Kiat and Remli, 2021). Machine learning algorithms have the potential to transform raw EEG data into clinically relevant information, providing valuable insights for epilepsy diagnosis and management (Dasgupta et al., 2020). The development of robust and reliable machine learning models for epilepsy detection requires careful consideration of various factors, including data preprocessing, feature extraction, model selection, and performance evaluation. Although, the natural complexity and non-stationary nature of EEG signals present significant challenges for machine learning algorithms, necessitating the development of advanced techniques to extract meaningful features and improve classification accuracy.

This research develops and examines machine learning models intended for epilepsy detection using a dataset of EEG features, focusing on the delta, theta, alpha, beta, and gamma frequency bands. The study investigates the performance of various machine learning classifiers, including bagging classifier (BG), logistic regression (LR), support vector machines (SVM), and random forest (RF), in distinguishing between subjects with and without epilepsy based on their EEG features. The research also explores the importance of different EEG features in epilepsy detection and aims to identify the smallest subset of features that can achieve high accuracy, addressing critical questions related to the clinical utility and interpretability of machine learning models in epilepsy diagnosis.

3. Literature Review

3.1 Machine Learning and EEG in Epilepsy Diagnosis and Detection

The intersection of machine learning and EEG analysis has experienced major progress in recent years, leading to the development of advanced techniques for epilepsy diagnosis and detection. Machine learning algorithms have been successfully employed to analyse EEG data and identify patterns associated with epilepsy, offering a promising avenue for improving diagnostic accuracy and efficiency. Feature selection is an important factor in influencing classification accuracy, with studies emphasizing the optimization of parameters and feature subsets to enhance detection rates (Murugavel and Ramakrishnan, 2014). Machine learning models have shown potential in aiding diagnosis, prognosis, and treatment in psychiatric disorders (Nielsen et al., 2019). However, deciding the optimal machine learning model for a given problem can be challenging, requiring a better understanding of the various algorithms and their suitability for specific applications (Myszczynska et al., 2020).

Several research investigations have examined how machine learning algorithms operate on EEG signals to differentiate between subjects with and without epilepsy conditions. These researchers have demonstrated the potential of machine learning to identify tiny EEG patterns suggestive of epilepsy that may be missed by human experts, highlighting the benefits of automated detection systems. EEG functions as a common biomedical signal that aids in detecting diseases like Alzheimer’s and Parkinson’s, revealing mental states for psychological disease analysis (Wang et al., 2021). Research on physiological signals features has resulted in the development of waveform information and power spectral analysis and wavelet coefficients methods for features extraction (Chen et al., 2017). Additionally, various machine learning methods employing different architectures in EEG signal analysis provide essential information to develop advanced AI-based systems according to Praveena, Sarah and George (2020).

Machine learning techniques, including deep learning models, are used for processing unstructured data like free text or images, facilitating the extraction and structuring of relevant information. Various neuroimaging methods, combined with machine learning algorithms, have emerged as efficient tools for early Alzheimer's disease identification and diagnosis (Juganavar, Joshi and Shegekar, 2023). Computer-aided detection and diagnosis using machine learning algorithms can assist physicians in interpreting medical imaging findings and reduce interpretation times (Erickson et al., 2017).

3.2 Knowledge Gaps Identified

Despite the numerous advances in the application of machine learning to EEG-based epilepsy detection, several knowledge gaps remain that needs further investigation. One key challenge is the limited availability of properly tagged large EEG datasets, which are essential for training robust and generalizable machine learning models. The lack of interpretability of some machine learning models, particularly deep learning algorithms, poses another challenge, as clinicians need to understand the reasoning behind a model's predictions to trust and utilize it in clinical practice (Hicks et al., 2021). Research evidence shows traditional machine learning algorithms deliver better interpretability and stability for prognosis and diagnosis, so they get chosen instead of deep learning approaches (Karpov et al., 2023).

Addressing this variability and identifying consistent, reliable biomarkers for epilepsy detection is crucial for translating machine learning research into clinical practice. Another gap lies in the limited exploration of advanced feature extraction techniques that can capture subtle but important information from EEG signals. There is a need to develop machine learning models that are robust to noise and artifacts in EEG data, which are common in real-world clinical settings, suggesting that future research should focus on developing robust and interpretable machine learning models that can effectively handle the challenges of EEG data analysis and contribute to improved epilepsy diagnosis and management (Woodman and Mangoni, 2023).

The variety of epilepsy types and dissimilarities in EEG signs between patients also creates obstacles for machine learning model development and usage (Ogunpola et al., 2024). Future advancements in the field of epilepsy diagnosis and detection require the development of new solutions to address existing technique failures as well as the implementation of innovative approaches (Biswas, Chakraborty and Pramanik, 2020) (Ibrahim and Abdulazeez, 2021). Moreover, the translation of machine learning models into clinical practice requires careful consideration of ethical and regulatory issues, including data privacy, algorithmic bias, and the explainability of model predictions (Ngiam and Khor, 2019) (Buabbas et al., 2023). Models that learn to forecast diagnostic outcomes or therapeutic effects represent a potential solution for avoiding the restrictions imposed by models which detect biomarkers through visual analysis of human experts (Tveit et al., 2023).

Machine learning research focuses on explainable AI as a distinct field to interpret neural network operations because they are often regarded as black boxes (Wagner et al., 2024). Explainable AI is vital in surgery to address issues such as potential risks and uncertainty by making underlying models and derived results more interpretable. Addressing these challenges is crucial for realizing the full potential of machine learning in improving epilepsy diagnosis, treatment, and patient outcomes. Model biases and fairness represents a crucial problem since the training data becomes corrupted by bias during the data collection process (Weng, 2019). Health equity requires improving fairness in machine learning because it helps promote distributive justice between healthcare providers through responsible model development and implementation monitoring (Rajkomar et al., 2018).

The lack of interpretability is a key factor that limits wider adoption of machine learning in healthcare because healthcare workers often find it challenging to trust complex machine-learning models. Trust is often necessary for adoption to occur (Sendak et al., 2019). The interpretability and explainability of machine learning models are essential for applications in medicine and healthcare (Vellido, 2019) (Marcinkevičs and Vogt, 2020). The incorporation of techniques for uncertainty quantification and adversarial robustness can further enhance the reliability and trustworthiness of machine learning models in clinical settings, increasing their acceptance and adoption by healthcare professionals.

4. Methodology

This research analyses the developmental approach to building machine learning systems that classify epilepsy from electroencephalogram (EEG) data. This work seeks to evaluate the effectiveness of machine learning models in distinguishing between epileptic and non-epileptic patients while addressing key questions regarding diagnostic accuracy, critical EEG features, and the minimal feature set required for effective detection. The analysis starts by acquiring and exploring the dataset, then preprocessing the data to ensure quality and consistency, developing baseline models, optimising model performance through feature selection, and evaluating the models using multiple performance metrics. The following subsections provide a detailed description of each step, ensuring transparency and reproducibility of the analysis.

4.1 Data Acquisition and Pre-processing

The dataset used in this study was sourced from Kaggle and comprises EEG signals measured from 198 subjects, with 99 diagnosed with epilepsy and 99 without. The EEG signals for each subject were pre-processed, and features associated with epilepsy were extracted. A total of 40 features were extracted for each subject from their EEG signals under five frequency bands (delta, theta, alpha, beta, gamma) that included power and fractal dimension values as well as six wavelet coefficients for each frequency band. Stat is the target variable that demonstrates epilepsy status, with 1 representing epilepsy and 2 as non-epileptic patients, but classifiers received 1 and 0 to indicate this binary format. An additional ID column uniquely identifies each subject but was excluded from modelling.

Data pre-processing was conducted to prepare the EEG dataset for machine learning analysis. Initially, the dataset was loaded into Python using Pandas' library, and basic exploratory analysis was performed to assess its structure and identify anomalies. Missing values were quantified as a fraction of total samples per feature, though no missingness was observed requiring imputation or column removal.

Feature variability was assessed by calculating the number of unique values per feature using Pandas' nunique() function, excluding ID and stat columns to identify no-variance and Low variance features. No-variance are features where all values are identical across the subjects. Features with no variance cannot distinguish between epilepsy and non-epilepsy because they're constant across all subjects. Low-variance features are features with very little variation. Low variance features contribute minimally to classification, as their small changes are often regarded as noise rather than signal.

Feature collinearity was evaluated to remove redundant EEG features, using a Pearson correlation matrix computed on 40 features (excluding ID and stat). A threshold of 0.80 was applied to the upper triangle of the matrix, identifying features with absolute correlations exceeding this value. This process identified four features for removal, reducing dimensionality to 36 features while retaining independent information for epilepsy classification.

4.2 Model Development and Investigation

The next stage involved the selection of appropriate machine learning models for epilepsy detection. Eleven models were initially evaluated using 10-fold cross-validation on the 36 remaining EEG features with accuracy as the performance measure. Features were normalized using StandardScaler to eliminate scale differences by setting means to zero and variances to one. The top five models, random forest (RF), bagging classifier (BG), support vector machine (SVM), logistic regression (LR), and extra trees (ET) were selected based on their mean accuracy and minimal variability (Figure A3, Appendix).

To assess the impact of feature selection, the five models were first evaluated on the full set of 36 features. Subsequently, Recursive Feature Elimination with Cross-Validation (RFECV) was used to identify the most relevant features. Out of the 36 remaining features, 13 were ranked 1, indicating their importance for epilepsy classification. These selected features, with the stat variable, were used to create a new data frame. This new data frame is then divided into training and test sets using 70:30 split, ensuring a balanced representation of epilepsy and non-epilepsy in both sets. To avoid data leakage the data sets were scaled after the split. The five models selected were then re-evaluated on the 13 features, and their performance was compared to the initial evaluation using multiple measures such as sensitivity, specificity, AUC, accuracy, precision and F1-score.

5. Results, Discussion and Conclusion

5.1 Results

The performance of the five selected machine learning models, support vector machine (SVM), random forest (RF), Extra trees (ET), bagging classifier (BG), and logistic regression (LR) was evaluated in two stages, first using all 36 EEG features, and then using the 13 features selected through Recursive Feature Elimination with Cross-Validation (RFECV). The test set comprised 30% of the dataset based on a 70:30 train-test split. The split was chosen to balance the need for sufficient training data, with a robust test set for evaluation, ensuring generalizability while maintaining the dataset’s balanced class distribution (99 epilepsy, 99 non-epilepsy).

Performance was evaluated using multiple measures: accuracy, sensitivity, specificity, precision, F1-score, and Area Under the Curve (AUC). Results for both stages are summarized in Tables 1 and 2, with confusion matrices and ROC curves providing additional insight into classification performance for the 13-feature evaluation.

Table 1: Model Performance Measures on the Test Set (All 36 Features)

Model Metrics Summary

	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
SVM	0.9833	1.0	0.9667	0.9677	0.9836	0.9978
Random Forest	0.9833	0.9667	1.0	1.0	0.9831	1.0
Extra Trees	0.9833	1.0	0.9667	0.9677	0.9836	1.0
Bagging	0.9833	0.9667	1.0	1.0	0.9831	1.0
Logistic Regression	0.9833	1.0	0.9667	0.9677	0.9836	0.9989

Table 1 shows that all models achieved an accuracy of 0.9833 when using all 36 features, with AUC scores ranging from 0.9978 (SVM) to 1.0 (RF, ET, BG). Sensitivity and specificity varied slightly, with RF and BG showing perfect specificity (1.0) but a sensitivity of 0.9667 (one false negative each), while SVM, ET, and LR had perfect sensitivity (1.0) but a specificity of 0.9667 (one false positive each).

Table 2: Model Performance Metrics on the Test Set (13 Selected Features)

Model Metrics Summary

	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
SVM	0.9833	1.0	0.9667	0.9677	0.9836	1.0
Random Forest	0.9833	1.0	0.9667	0.9677	0.9836	1.0
Extra Trees	1.0	1.0	1.0	1.0	1.0	1.0
Bagging	0.9833	0.9667	1.0	1.0	0.9831	1.0
Logistic Regression	0.9667	0.9667	0.9667	0.9667	0.9667	0.9978

After feature selection, Extra Trees achieved perfect performance across all the performance measures, correctly classifying all test subjects, as shown in its confusion matrix (Figure 1) and ROC curve (Figure 2). Random Forest and SVM both maintained their accuracy of 0.9833, with perfect sensitivity (1.0000) but a specificity of 0.9667, indicating one false positive each (one non-epilepsy subject misclassified as epilepsy), as seen in Figures A6 and A7 (Appendix) for RF and SVM, respectively. Their ROC curves (Figures A10 and A11, Appendix) confirm perfect AUC scores of 1.0000. The Bagging Classifier also scored 0.9833 in accuracy, with perfect specificity (1.0000) but a sensitivity of 0.9667, misclassifying one epilepsy subject as non-epilepsy (Figure A8, Appendix), yet achieving an AUC of 1.0000 (Figure A12, Appendix). Logistic Regression's performance decreased slightly worse to an accuracy of 0.9667, misclassifying one subject in each class (sensitivity and specificity both 0.9667, Figure A9, Appendix), and an AUC of 0.9978 (Figure A13, Appendix).

Feature selection via RFECV reduced the feature set from 36 to 13, focusing on the most predictive EEG features. The important features, identified using Random Forest's feature importance scores, are shown in Figure A4 (Appendix), with features like pow_a (power in the alpha band) and cd2_a (second coefficient wavelet in the alpha band) ranking highest. Conversely, Figure A5 (Appendix) highlights the least important features, such as cd5_g (fifth coefficient wavelet in the gamma band), which contributed minimally to classification.

Figure 1: Confusion Matrix for Extra Trees (ET) with 13 Features

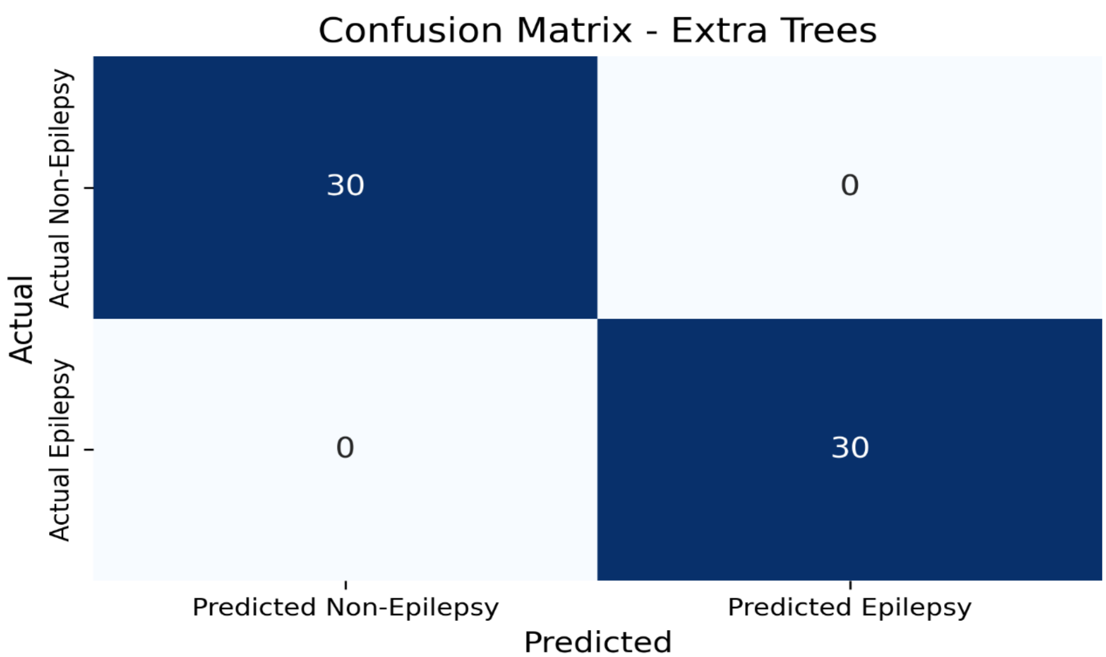
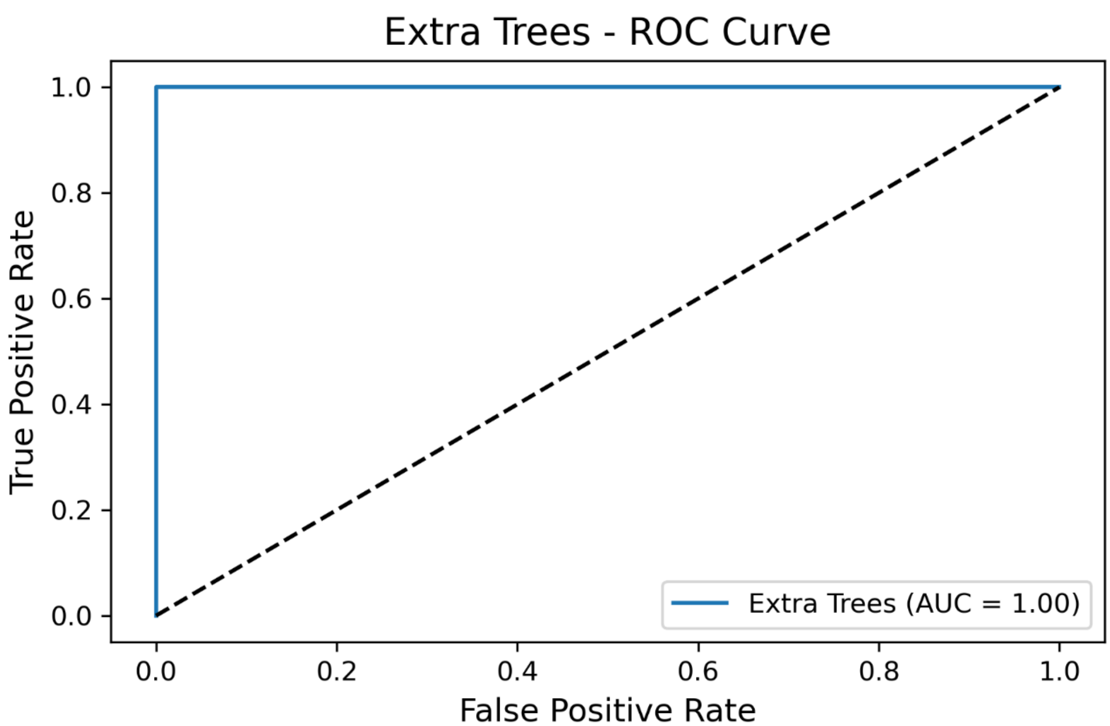


Figure 2: ROC Curve for Extra Trees (ET) with 13 Features



5.2 Discussion

The results demonstrate that machine learning models, when applied to EEG features, can effectively classify epilepsy with high accuracy, addressing the key questions posed in this study. The comparison between the 36-feature and 13-feature evaluations highlights the significance of feature selection on model performance.

Impact of Feature Selection

Feature selection improved overall model performance, as evidenced by the comparison between Tables 1 and 2. With all 36 features, all models achieved an accuracy of 0.9833, with AUC scores ranging from 0.9978 to 1.0000. However, after reducing to the 13 most important features, Extra Trees achieved perfect performance (accuracy, AUC, sensitivity, specificity, F1-score = 1.0000), and Logistic Regression's accuracy dropped slightly to 0.9667, while RF, SVM, and BG maintained their accuracy of 0.9833. The improvement in ET's performance (from 0.9833 to 1.0000) suggests that removing less relevant features (e.g., cd5_g, Figure A5, Appendix) reduced noise, allowing the model to focus on the most predictive EEG features like pow_a and cd2_a (Figure A4, Appendix). For LR, the slight decrease in performance (0.9833 to 0.9667) may indicate that it benefited from the additional features to capture linear relationships, but the reduced feature set introduced more classification errors (one false positive and one false negative, Figure A9, Appendix). This underscores the importance of feature selection in enhancing model performance, particularly for ensemble methods like ET, RF, and BG, which thrive on relevant features.

Can Machine Learning and EEG Be Used to Detect Epilepsy Accurately (AUC > 0.8)?

All five models exceeded the AUC threshold of 0.8 in both evaluations, with ET, RF, SVM, and BG achieving a perfect AUC of 1.0000 with 13 features and LR scoring 0.9978 (Figures 2, A10-A13, Appendix). Even with all 36 features, AUC scores ranged from 0.9978 to 1.0 confirming that EEG-based machine learning is highly effective for epilepsy detection, aligning with prior research (Miltiadous et al., 2023). The perfect AUC, visualised in the ROC curves, indicates that the models can perfectly distinguish between epilepsy and non-epilepsy subjects in the test set, likely due to the balanced dataset and strong predictive power of the selected features. The high sensitivity (0.9667-1.0000) and specificity (0.9667-1.0000), shown by the confusion matrices (Figures 1, A6-A9, Appendix), further demonstrate that the models are reliable for diagnosis. For instance, Figure 1 demonstrates that ET achieves perfect classification, while RF detects one false positive in Figure A6, highlighting their robustness and making them suitable for clinical applications where minimising both false positives and false negatives is critical.

Which EEG Features Are Most Indicative of Epilepsy?

The feature selection using RFECV identified 13 important features (Figure A4, Appendix) highlighting ‘pow_a’ and ‘cd2_a’, as the most indicative of epilepsy. The alpha band power is associated with relaxed wakefulness and is often suppressed during periods of mental activity or sensory exposure (Hsu, Cheng and Chiu, 2017). Alteration of alpha band power and reactivity can indicate underlying brain abnormalities, including those related to epilepsy (Jiang et al., 2023). The second coefficient of the wavelet transform applied to the alpha band holds particular significance in epilepsy diagnosis due to its sensitivity to subtle changes in alpha activity that may precede or accompany seizures. Wavelet coefficients capture the amplitude and phase information of the signal at different scales, providing a more detailed representation of the underlying neuronal dynamics than traditional frequency analysis methods. The second coefficient, specifically, may reflect alterations in the correlation and stability of alpha oscillations, which are known to be disrupted in epileptic conditions (Zhong et al., 2023).

Conversely, features like cd5_g (Figure A5, Appendix) were least important, possibly because high-frequency gamma activity is less specific to epilepsy in this dataset. The gamma band is associated with higher-level cognitive functions, sensory processing, and inter-neuronal communication. While gamma activity is known to be altered in various neurological disorders, including epilepsy, its role in seizure generation and propagation is less well defined than that of other frequency bands, such as alpha and beta (Guan et al., 2022). The correlation matrix (Figure A1, Appendix) supports the removal of collinear features (e.g., fr_g, cd5_t), ensuring that the selected features provide independent, epilepsy-specific information. The improved performance with the 13 features (Table 2) further validates their relevance for epilepsy detection.

What is the Smallest Set of EEG Features for Accurate Detection (AUC > 0.8)?

The 13 features selected by RFECV enabled all models to achieve AUC > 0.8, with ET achieving a perfect AUC of 1.0000 (Figure 2). To explore a smaller feature set, we can consider the top features from Figure A4 (e.g., pow_a, cd2_a, etc.), which likely account for most of the predictive power. However, further validation is needed to confirm generalisability, as reducing features risks overfitting on smaller datasets. The test results between 36-feature and 13-feature evaluations (Tables 1 and 2) reveal that the reduction in feature dimensions leads to improved performance for every predictive model, thus validating that a focused set of features can achieve accurate detection.

Comparison of Classifiers

The classifiers exhibited distinct advantages and disadvantages, as summarised below:

- **Performance:** With 13 features, Extra Trees outperformed all models, achieving perfect scores across all the performance measures (accuracy, AUC, sensitivity, specificity, F1-score = 1.0000), as seen in its ROC curve (Figure 2) and confusion matrix (Figure 1). RF, SVM, and BG were close behind (accuracy = 0.9833), with LR at 0.9667. With 36 features, all models performed equally (accuracy = 0.9833), but feature selection allowed ET to achieve perfection, highlighting its ability to leverage the most relevant features. The boxplot comparison from cross-validation (Figure A2, Appendix) aligns with the test set results, showing ET, RF, and BG consistently near 1.0, while SVM improved to 1.0 after tuning (from an initial 0.85 with a polynomial kernel) using mean accuracy as the performance measure.

- **Ease of Optimization:** RF, ET, and BG required minimal tuning, as their ensemble nature makes them robust to hyperparameter choices. SVM needed extensive tuning (from initial polynomial kernel to linear kernel) to improve from 0.85 to 0.98, making it more labourintensive. LR was easiest to optimize, requiring only max_iter adjustments for convergence.

- **Robustness to Noise or Missing Data:** Ensemble methods (RF, ET, BG) are less affected by noise due to their averaging mechanisms, as seen in their low variability. SVM and LR, while effective, are more sensitive to noisy features, though feature selection mitigated this. The dataset had no missing values, but RF and ET would likely handle missing data better by leveraging feature importance.
- **Interpretability:** LR offers the highest interpretability, with coefficients directly indicating feature impact on epilepsy likelihood. SVM and ensemble methods (RF, ET, BG) are less interpretable, though RF provides feature importance scores (Figures A4, A5). This makes LR valuable for clinical settings where understanding feature contributions is crucial, but its slight decrease in performance with 13 features suggests that it may rely on a broader feature set for linear modelling.
- **Training Time:** Training time was slightly reduced with 13 features compared to 36, as expected due to dimensionality. LR trained the fastest while SVM was the slowest because of its computational complexity.
- **Limitations:** SVM's initial poor performance (0.85) with a polynomial kernel highlights its sensitivity to kernel choice, requiring careful tuning. LR, while interpretable, struggles with non-linear relationships in EEG data, as reflected in its slightly lower AUC (0.9978, Figure A13) and reduced accuracy with 13 features. Ensemble methods, while high-performing, risk overfitting on small datasets, though the balanced test set performance suggests this was not an issue here.

Additional EEG Features for Improved Detection

Beyond the 40 features provided, incorporating **spectral entropy** and **phase synchronization** could enhance detection. To significantly improve detection performance, it is crucial to consider the use of advanced EEG features like spectral entropy and phase synchronization, which offer complementary perspectives on the underlying neural dynamics (Zhong et al., 2023). Spectral entropy, a measure rooted in information theory, quantifies the complexity and irregularity of the EEG signal's power spectrum, offering a richer characterisation of brain activity compared to traditional frequency band analysis (Chen et al., 2017). By quantifying the distribution of power across different frequency bands, spectral entropy can capture subtle changes in the EEG signal that may be indicative of various neurological conditions or cognitive states (Čukić, López and Pavón, 2020). Phase synchronization, on the other hand, assesses the degree of temporal coordination between different brain regions, providing insights into the functional connectivity and information flow within the brain (Yang et al., 2022). By examining the consistency of phase relationships between EEG signals recorded from different electrodes, phase synchronization can reveal patterns of neural communication that are not readily visible from amplitude-based measures alone. The features extracted from EEG signals, which include statistical measures, amplitude-based metrics, frequency domain characteristics, and entropy measures, serve as inputs for machine learning models designed to categorise EEG data into distinct classes (Li et al., 2018).

5.3 Conclusion

This study successfully developed and evaluated machine learning models to detect epilepsy through EEG features, achieving AUC scores well above the 0.8 threshold. Feature selection significantly improved performance, with Extra Trees emerging as the top performer (accuracy, AUC = 1.0000) using 13 features, compared to 0.9833 with 36 features. Random Forest, SVM, and Bagging Classifier maintained high performance (accuracy = 0.9833) while Logistic Regression experienced a slight decrease (0.9667). Feature selection identified 13 critical features, with pow_a and cd2_a being the most indicative of epilepsy. The comparison of classifiers highlights the trade-offs between performance, interpretability, and computational cost, with ensemble methods (RF, ET, BG) offering the best balance for this task. Future work could explore spectral entropy and phase synchronization to further enhance detection, particularly for complex and challenging epilepsy cases, and validate the models on larger, multi-channel EEG datasets.

6. References

An, S., Kang, C. and Lee, H.W. (2020). Artificial Intelligence and Computational Approaches for Epilepsy. *Journal of Epilepsy Research*, 10(1), pp.8–17. doi:https://doi.org/10.14581/jer.20003.

Biswas, S.D., Chakraborty, R. and Pramanik, A. (2020). A Brief Survey on Various Prediction Models for Detection of ADHD from Brain-MRI Images. *Proceedings of the 21st International Conference on Distributed Computing and Networking*, pp.1–5. doi:https://doi.org/10.1145/3369740.3372775.

Buabbas, A.J., Miskin, B., Alnaqi, A.A., Ayed, A.K., Shehab, A.A., Syed-Abdul, S. and Uddin, M. (2023). Investigating Students' Perceptions Towards Artificial Intelligence in Medical Education. *Healthcare*, 11(9), p.1298. doi:https://doi.org/10.3390/healthcare11091298.

Chen, L., Zhao, Y., Ye, P., Zhang, J. and Zou, J. (2017). Detecting Driving Stress in Physiological Signals Based on Multimodal Feature Analysis and Kernel Classifiers. *Expert Systems with Applications*, 85, pp.279–291. doi:https://doi.org/10.1016/j.eswa.2017.01.040.

Čukić, M., López, V. and Pavón, J. (2020). Classification of Depression through Resting-State Electroencephalogram as a Novel Practice in Psychiatry: Review. *Journal of Medical Internet Research*, 22(11), p.e19548. doi:https://doi.org/10.2196/19548.

D. Merlin Praveena, D. Angelin Sarah and George, S. (2020). Deep Learning Techniques for EEG Signal Applications – a Review. *Iete Journal of Research*, 68(4), pp.3030–3037. doi:https://doi.org/10.1080/03772063.2020.1749143.

Dasgupta, A., Nayak, L., Das, R., Basu, D., Chandra, P. and De, R.K. (2020). Pattern and Rule Mining for Identifying Signatures of Epileptic Patients from Clinical EEG Data. *Fundamenta Informaticae*, 176(2), pp.141–166. doi:https://doi.org/10.3233/fi-2020-1968.

Erickson, B.J., Korfiatis, P., Akkus, Z. and Kline, T.L. (2017). Machine Learning for Medical Imaging. *RadioGraphics*, 37(2), pp.505–515. doi:https://doi.org/10.1148/rg.2017160130.

Guan, A., Wang, S., Huang, A., Qiu, C., Li, Y., Li, X., Wang, J., Wang, Q. and Deng, B. (2022). The Role of Gamma Oscillations in Central Nervous System diseases: Mechanism and Treatment. *Frontiers in Cellular Neuroscience*, 16. doi:https://doi.org/10.3389/fncel.2022.962957.

Hicks, S.A., Isaksen, J.L., Thambawita, V., Ghouse, J., Ahlberg, G., Linneberg, A., Grarup, N., Strümke, I., Ellervik, C., Olesen, M.S., Hansen, T., Graff, C., Holstein-Rathlou, N.-H., Halvorsen, P., Maleckar, M.M., Riegler, M.A. and Kanters, J.K. (2021). Explaining Deep Neural Networks for Knowledge Discovery in Electrocardiogram Analysis. *Scientific Reports*, [online] 11(1), p.10949. doi:https://doi.org/10.1038/s41598-021-90285-5.

Hsu, C.-C., Cheng, C.-W. and Chiu, Y.-S. (2017). Analyze the Beta Waves of Electroencephalogram Signals from Young Musicians and non-musicians in Major Scale Working Memory Task. *Neuroscience Letters*, 640, pp.42–46. doi:https://doi.org/10.1016/j.neulet.2017.01.022.

Ibrahim, I. and Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, 2(01), pp.10–19. doi:https://doi.org/10.38094/jastt20179.

Jaafar, R., Chua Soon Kiat and Rabani Remli (2021). An ocular artefacts removal technique for interictal epileptiform discharges detection in epilepsy. *AIP conference proceedings*, [online] 2355, pp.020001–020001. doi:https://doi.org/10.1063/5.0053216.

Jiang, S., Chen, W., Ren, Z. and Zhu, H. (2023). EEG-based Analysis for Pilots' at-risk Cognitive Competency Identification Using RF-CNN Algorithm. *Frontiers in Neuroscience*, 17. doi:https://doi.org/10.3389/fnins.2023.1172103.

Juganavar, A., Joshi, A. and Shegekar, T. (2023). Navigating Early Alzheimer's Diagnosis: a Comprehensive Review of Diagnostic Innovations. *Cureus*, [online] 15(9). doi:https://doi.org/10.7759/cureus.44937.

Karpov, O.E., Pitsik, E., Kurkin, S.A., Maksimenko, V.A., A.V. Gusev, Shusharina, N. and Hramov, A.E. (2023). Analysis of Publication Activity and Research Trends in the Field of AI Medical Applications: Network Approach. *Environmental Research and Public Health*, 20(7), pp.5335–5335. doi:https://doi.org/10.3390/ijerph20075335.

Li, X., Song, D., Zhang, P., Zhang, Y., Hou, Y. and Hu, B. (2018). Exploring EEG Features in Cross-Subject Emotion Recognition. *Frontiers in Neuroscience*, 12. doi:https://doi.org/10.3389/fnins.2018.00162.

Marcinkevičs, R. and Vogt, J.E. (2020). Interpretability and Explainability: a Machine Learning Zoo Mini-tour. *arXiv:2012.01805 [cs]*. [online] Available at: https://arxiv.org/abs/2012.01805.

Miltiadous, A., Tzimourta, K.D., Giannakeas, N., Tsipouras, M.G., Glavas, E., Kalafatakis, K. and Tzallas, A.T. (2023). Machine Learning Algorithms for Epilepsy Detection Based on Published EEG Databases: A Systematic Review. *IEEE Access*, 11, pp.564–594. doi:https://doi.org/10.1109/access.2022.3232563.

Murugavel, A.S.M. and Ramakrishnan, S. (2014). Optimal Feature Selection Using PSO with SVM for Epileptic EEG Classification. *International Journal of Biomedical Engineering and Technology*, 16(4), p.343. doi:https://doi.org/10.1504/ijbet.2014.066229.

Myszczynska, M.A., Ojamies, P.N., Lacoste, A.M.B., Neil, D., Saffari, A., Mead, R., Hautbergue, G.M., Holbrook, J.D. and Ferraiuolo, L. (2020). Applications of Machine Learning to Diagnosis and Treatment of Neurodegenerative Diseases. *Nature Reviews Neurology*, [online] 16(8), pp.440–456. doi:https://doi.org/10.1038/s41582-020-0377-8.

Ngiam, K.Y. and Khor, I.W. (2019). Big Data and Machine Learning Algorithms for health-care Delivery. *The Lancet Oncology*, 20(5), pp.262–273. doi:https://doi.org/10.1016/s1470-2045(19)30149-4.

Nielsen, A.N., Barch, D.M., Petersen, S.E., Schlaggar, B.L. and Greene, D.J. (2019). Machine Learning with Neuroimaging: Evaluating Its Applications in Psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. doi:https://doi.org/10.1016/j.bpsc.2019.11.007.

Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A.M. and Qasem, S.N. (2024). Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases. *Diagnostics*, [online] 14(2), p.144. doi:https://doi.org/10.3390/diagnostics14020144.

Rajkomar, A., Hardt, M., Howell, M.D., Corrado, G. and Chin, M.H. (2018). Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of Internal Medicine*, [online] 169(12), p.866. doi:https://doi.org/10.7326/m18-1990.

Sendak, M., Elish, M., Gao, M., Futoma, J., Ratliff, W., Nichols, M., Bedoya, A., Balu, S. and O'Brien, C. (2019). *'The Human Body is a Black Box': Supporting Clinical Decision-Making with Deep Learning*. [online] arXiv.org. Available at: https://arxiv.org/abs/1911.08089.

Stafstrom, C.E. and Carmant, L. (2015). Seizures and Epilepsy: An Overview for Neuroscientists. *Cold Spring Harbor Perspectives in Medicine*, 5(6), pp.a022426–a022426. doi:https://doi.org/10.1101/cshperspect.a022426.

Supriya, S., Siuly, S., Wang, H. and Zhang, Y. (2023). Epilepsy Detection from EEG Using Complex Network Techniques: A Review. *IEEE Reviews in Biomedical Engineering*, 16, pp.292–306. doi:https://doi.org/10.1109/rbme.2021.3055956.

Tveit, J., Harald Aurlien, Plis, S., Calhoun, V.D., Tatum, W.O., Schomer, D.L., Arntsen, V., Cox, F., Firas Fahoum, Gallentine, W.B., Gardella, E., Hahn, C.D., Husain, A.M., Kessler, S., Mustafa Aykut Kural, Nascimento, F.A., Hatice Tankisi, Ulvin, L.B., Wennberg, R. and Sándor Beniczky (2023). Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence. *JAMA Neurology*. doi:https://doi.org/10.1001/jamaneurol.2023.1645.

Vellido, A. (2019). The Importance of Interpretability and Visualization in Machine Learning for Applications in Medicine and Health Care. *Neural Computing and Applications*, 32. doi:https://doi.org/10.1007/s00521-019-04051-w.

Wagner, P., Mehari, T., Haverkamp, W. and Nils Strodthoff (2024). Explaining Deep Learning for ECG analysis: Building Blocks for Auditing and Knowledge Discovery. *Computers in Biology and Medicine*, 176, pp.108525–108525. doi:https://doi.org/10.1016/j.compbiomed.2024.108525.

Wang, L., Liu, H., Zhou, T., Liang, W. and Shan, M. (2021). Multidimensional Emotion Recognition Based on Semantic Analysis of Biomedical EEG Signal for Knowledge Discovery in Psychological Healthcare. *Applied Sciences*, [online] 11(3), pp.1338–1338. doi:https://doi.org/10.3390/app11031338.

Weng, W.-H. (2019). *Machine Learning for Clinical Predictive Analytics*. [online] arXiv.org. Available at: https://arxiv.org/abs/1909.09246 [Accessed 4 Apr. 2025].

Woodman, R. and Mangoni, A.A. (2023). A Comprehensive Review of Machine Learning Algorithms and Their Application in Geriatric medicine: Present and Future. *Aging Clinical and Experimental Research*. doi:https://doi.org/10.1007/s40520-023-02552-2.

Yang, K., Tong, L., Zeng, Y., Lu, R., Zhang, R., Gao, Y. and Yan, B. (2022). Exploration of Effective Electroencephalography Features for the Recognition of Different Valence Emotions. *Frontiers in Neuroscience*, 16. doi:https://doi.org/10.3389/fnins.2022.1010951.

Zhong, L., Wan, J., Yi, F., He, S., Wu, J., Huang, Z., Lu, Y., Yang, J. and Li, Z. (2023). Epileptic Prediction Using Spatiotemporal Information Combined with Optimal Features Strategy on EEG. *Frontiers in Neuroscience*, 17. doi:https://doi.org/10.3389/fnins.2023.1174005.

7. Appendix

7.1 GitHub Repository for Script

The full Jupyter Notebook used in this study for analysis is at the following GitHub repository.

GitHub Link: https://github.com/Oluwabunmi-akintunde/EEG-Based-Epilepsy-Detection.git

The notebook includes the following steps:

1. Data preparation and exploratory analysis.
2. Feature analysis and dimensionality reduction.
3. Model selection and ranking.
4. Feature selection using RFECV.
5. Training and evaluating selected models using all features.
6. Training and evaluating selected models using selected features.

Each section is clearly marked within the notebook and all relevant outputs including confusion matrices, correlation matrix, ROC curves, are generated within the same file.

7.2 Figures

Figure A 1: Correlation Matrix of All 40 EEG Features

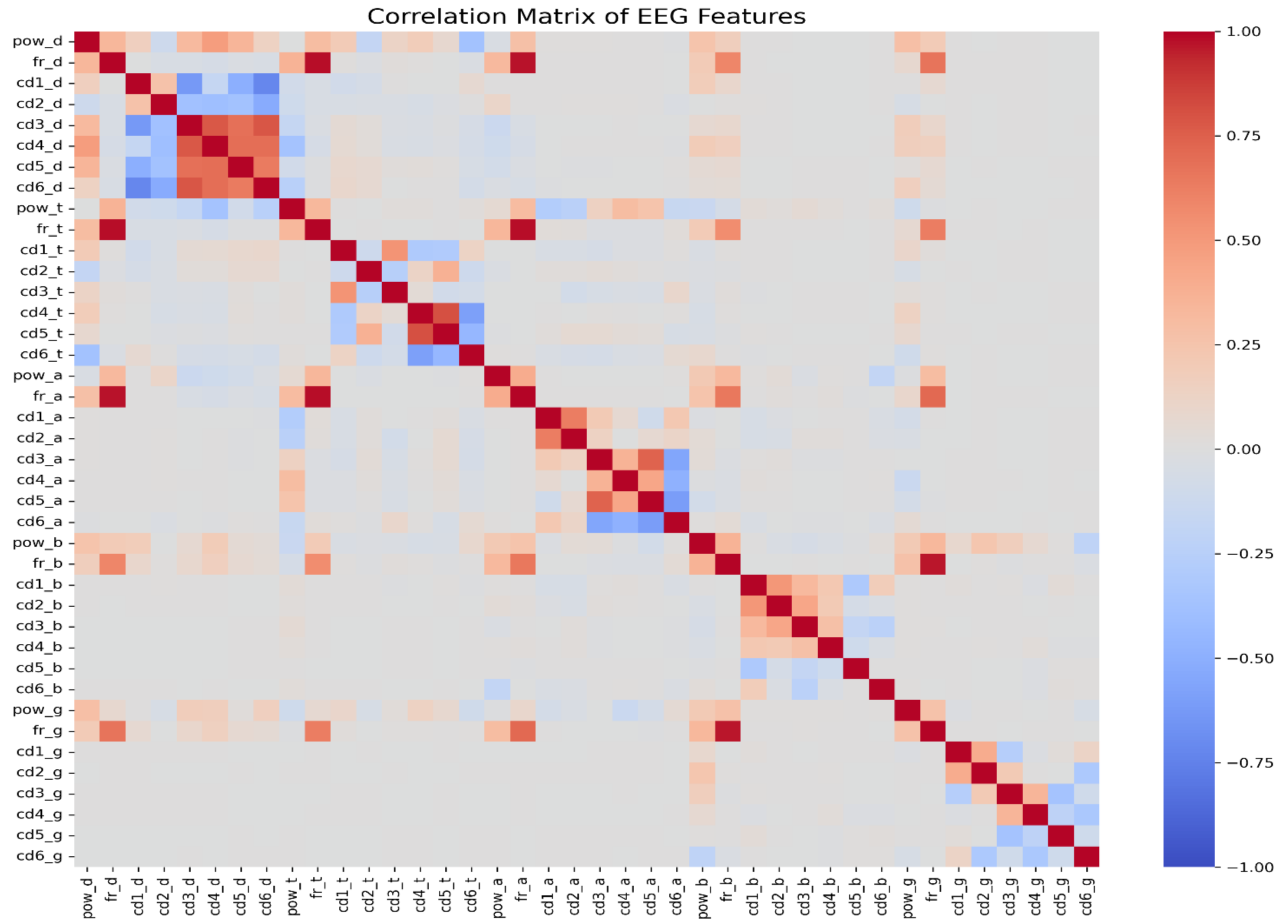


Figure A 2: Boxplot Comparison of Model Accuracy

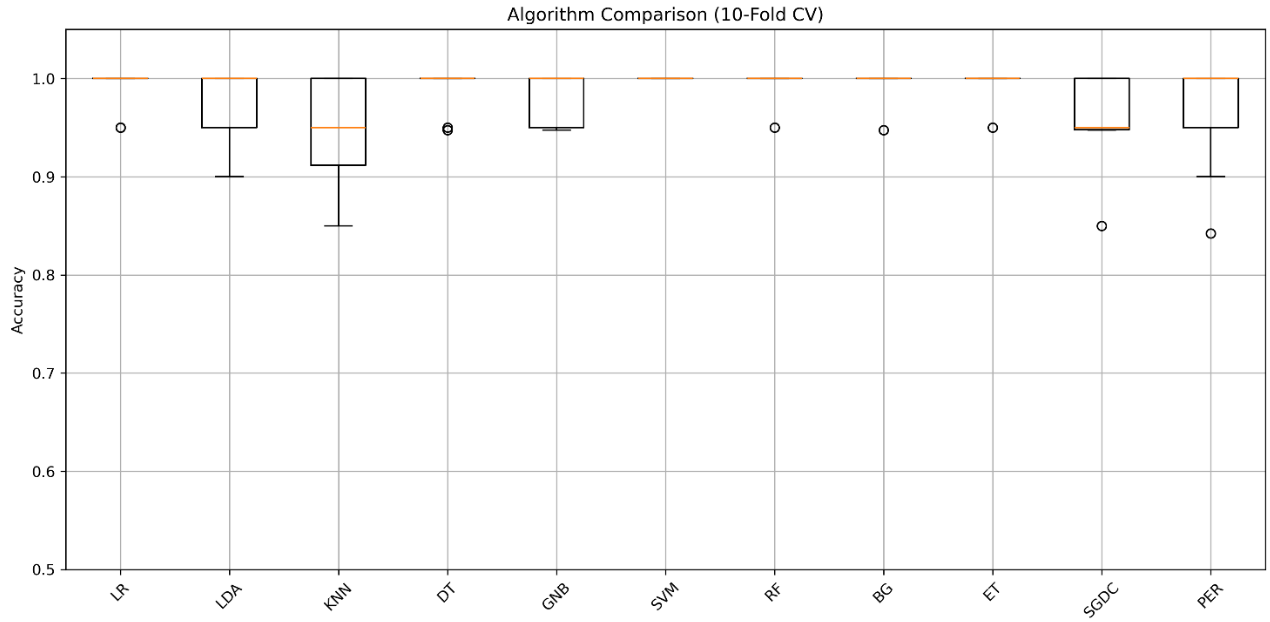


Figure A 3: Ranked Models Based on Mean Accuracy

Ranked Models by Mean Accuracy (10-Fold CV)

Rank	Model	Mean Accuracy	Std Dev
1	SVM	1.0000	0.0000
2	RF	0.9950	0.0150
3	ET	0.9950	0.0150
4	BG	0.9947	0.0158
5	LR	0.9900	0.0200
6	DT	0.9897	0.0205
7	GNB	0.9795	0.0252
8	LDA	0.9750	0.0335
9	PER	0.9642	0.0521
10	SGDC	0.9595	0.0437
11	KNN	0.9497	0.0500

Figure A 4: Top Important EEG Features Based on Random Forest

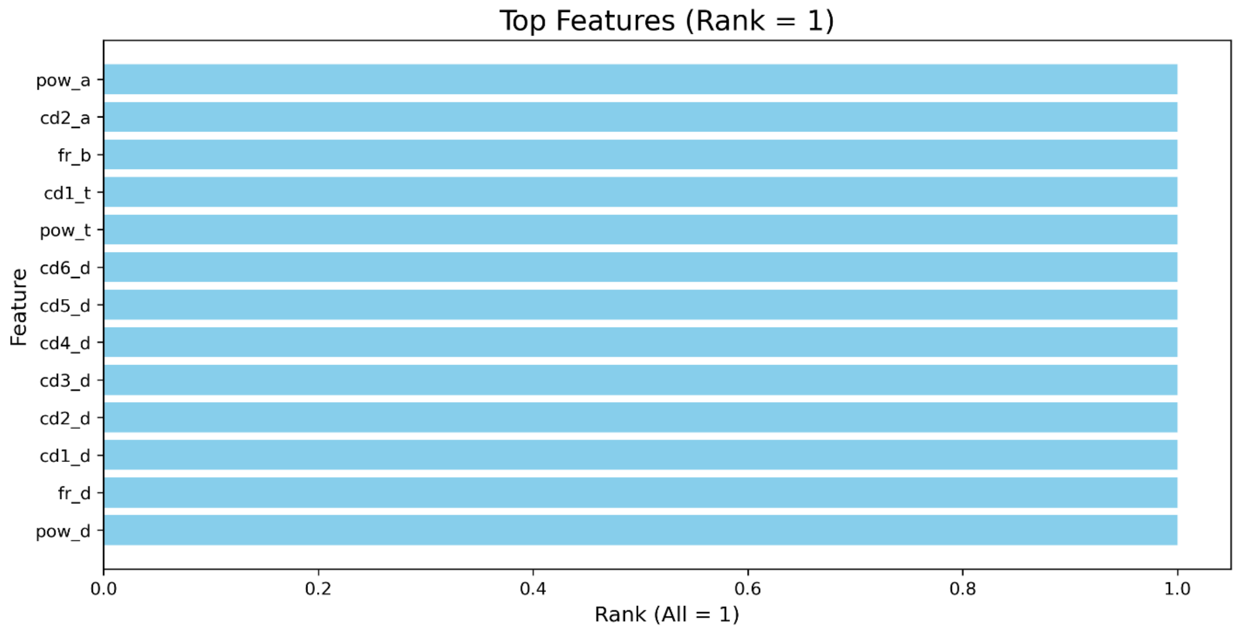


Figure A 5: Bottom 5 Least Important EEG features based on Random Forest

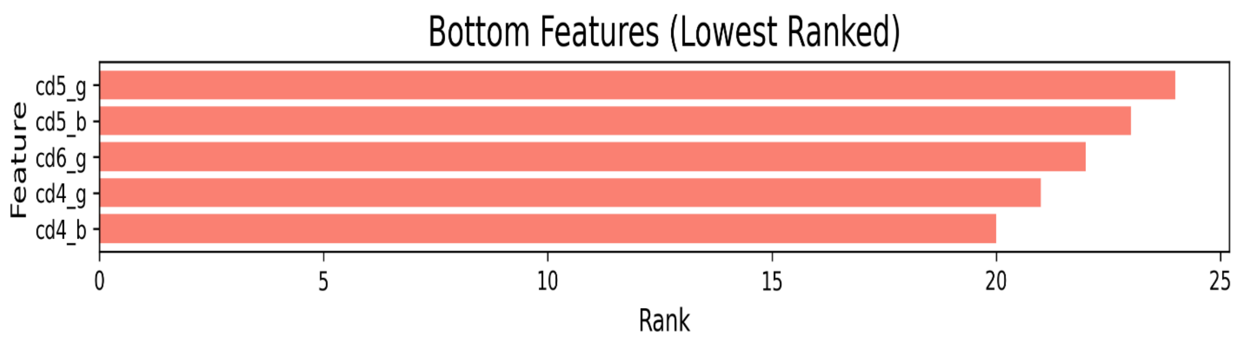


Figure A 6: Confusion Matrix for Random Forest (RF) with 13 Features

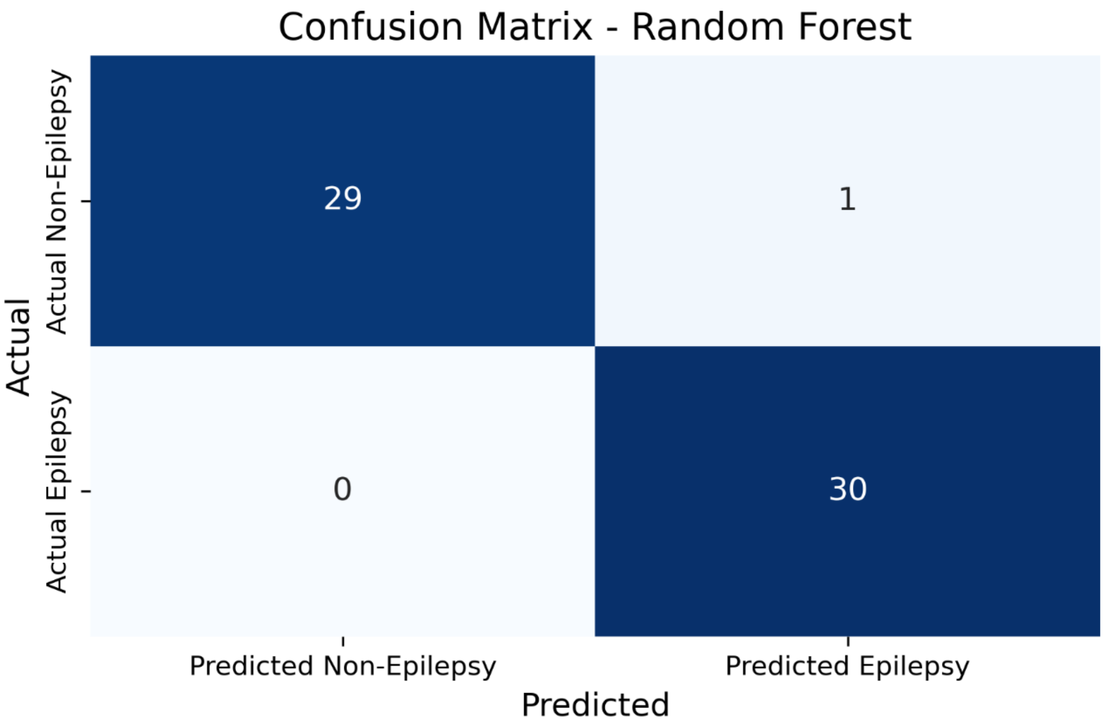


Figure A 7:Cofusion Matrix for Support Vector Machine (SVM) with 13 Features

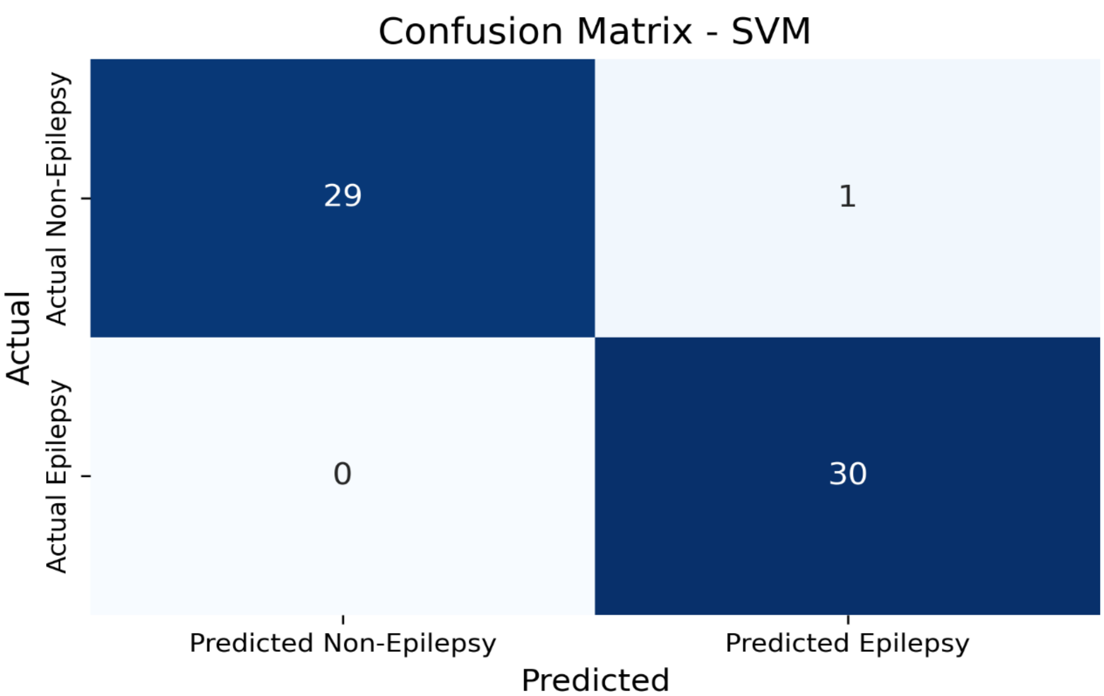


Figure A 8: Confusion Matrix for Bagging Classifier (BG) with 13 Features

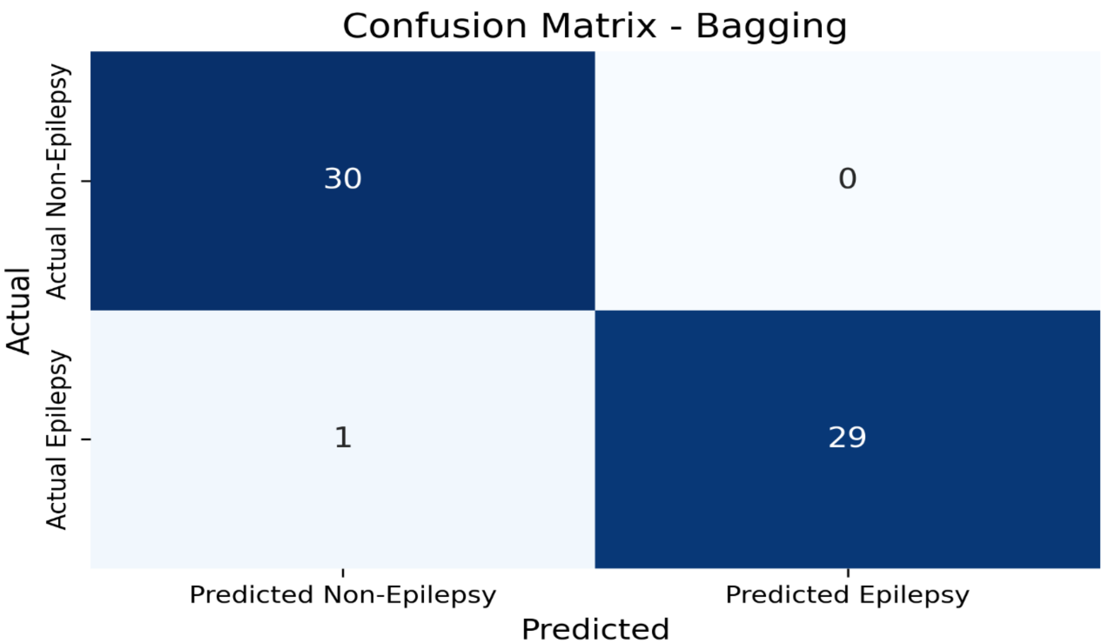


Figure A 9: Confusion Matrix for Logistic Regression (LR) with 13 Features

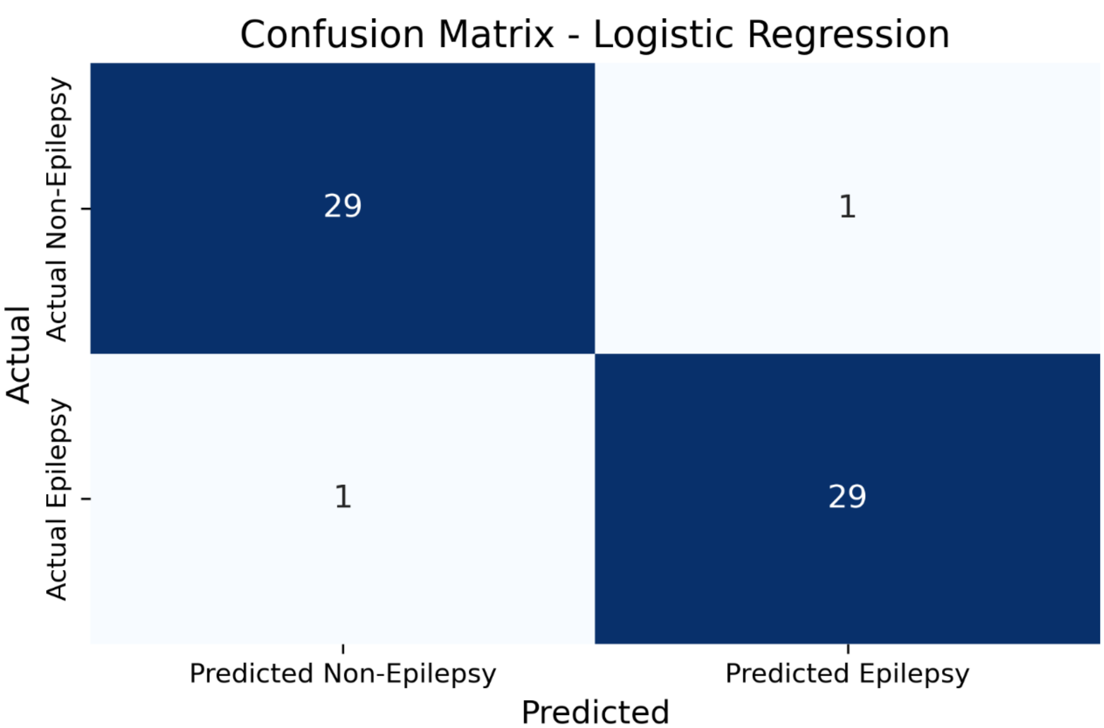


Figure A 10: ROC Curve for Random Forest (RF) with 13 Features

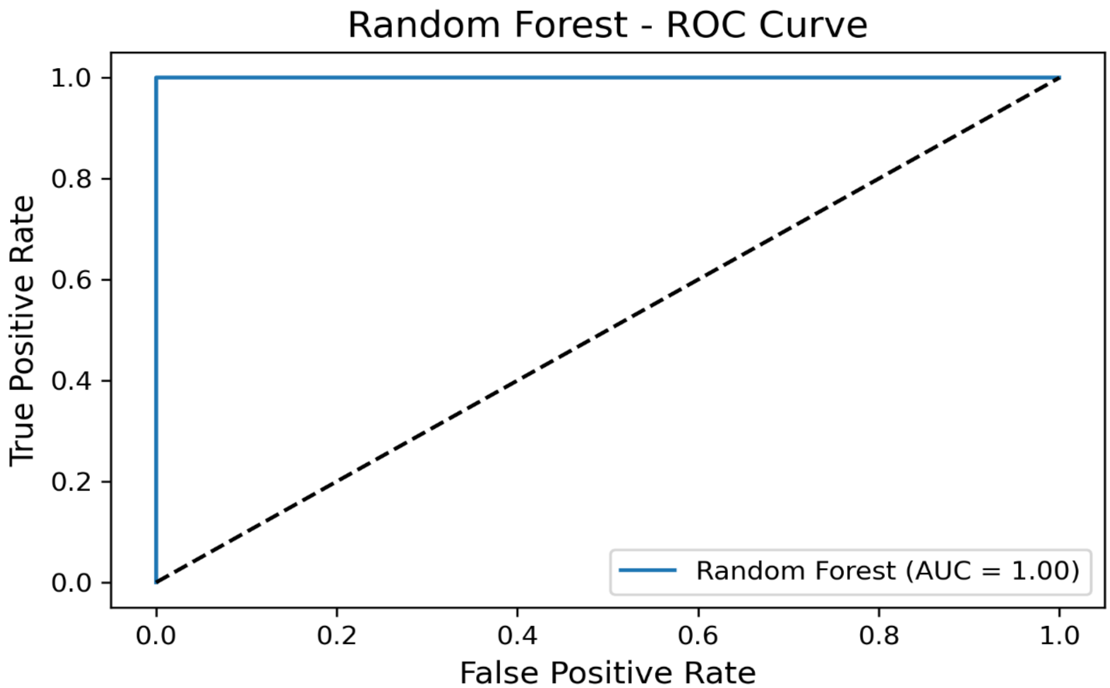


Figure A 11: ROC Curve for Support Vector Machine (SVM) with 13 Features

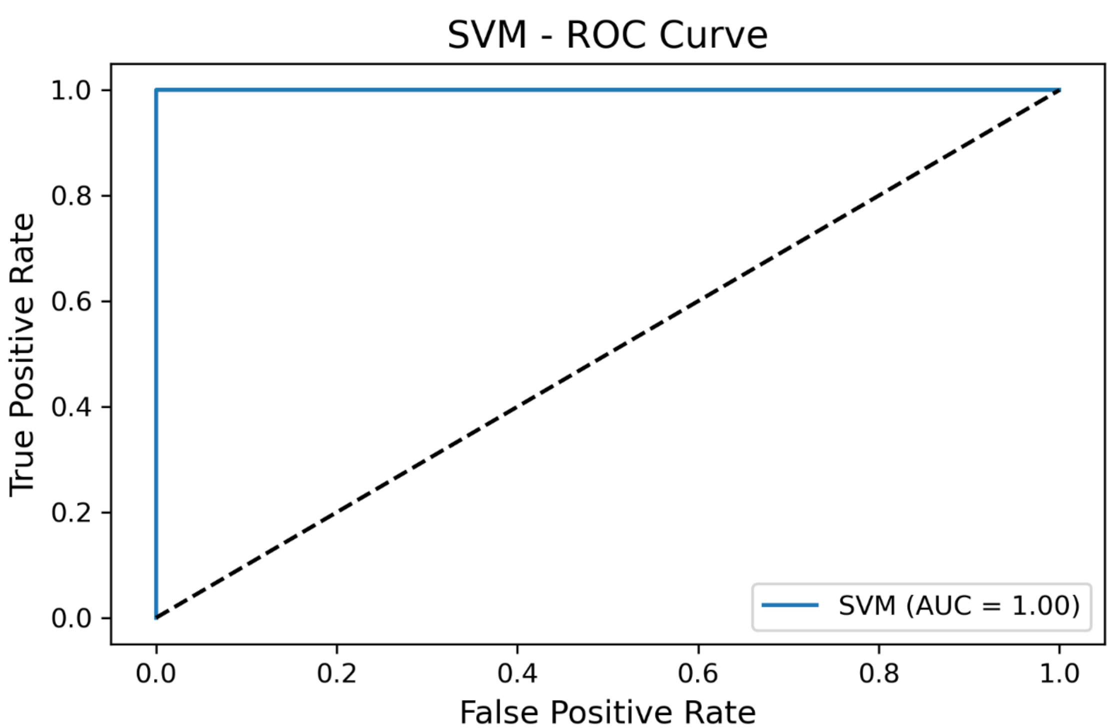


Figure A 12: ROC Curve for Bagging Classifier (BG) with 13 Features

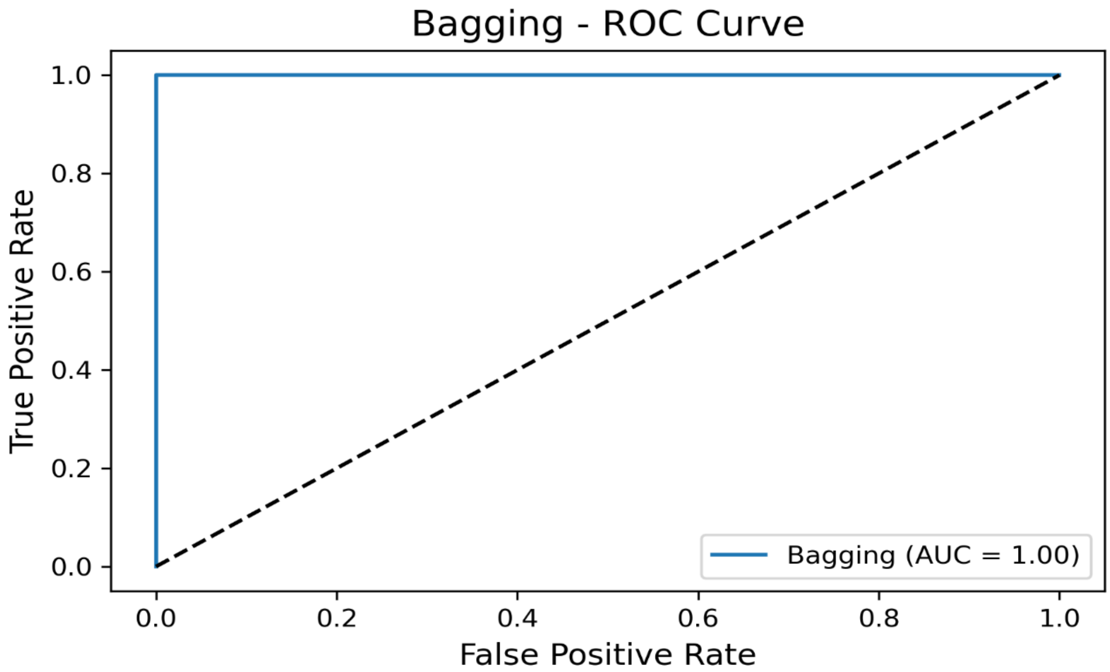


Figure A 13:ROC Curve for Logistic Regression (LR) with 13 Features