A Comparative Analysis of Classifier Performance for Epileptic Seizure Detection Using EEG Signals

Nida Alyas, David P. Hastings and Abbas Mehrabidavoodabadi
Department of Computer & Information Sciences, Northumbria University, Newcastle-upon-Tyne, U.K.

Keywords: Machine Learning, Epilepsy, Seizure Detection, Signal Processing, EEG, Classification.

Abstract: In middle and low-income countries, epilepsy remains undiagnosed in many instances because of an insufficient number of medical specialists and expensive EEG recording devices. In previous studies, many machine learning (ML) based methods were proposed to investigate and classify the EEG signals. However, little work has been performed with EEG data recorded with consumer-grade devices. The extraction of the most discriminating set of features and high misclassification rate is another challenge. To address these problems, this study empirically investigates several data segment sizes and chooses the optimal window size to segment the Guinea-Bissau dataset. Several statistical and spectral feature extraction methods were investigated to obtain useful sets of features from segmented epochs in combination with conventional ML algorithms and ensemble methods. The proposed framework is then implemented on a comparable dataset collected from Nigeria to validate the reliability of the framework. A comparative analysis is performed with conventional ML models and with existing techniques to prove the effectiveness of the proposed methodology. The obtained results demonstrate that XGBoost and LightGBM achieved the highest levels of performance in terms of F1 score and AUC.

1 INTRODUCTION

Epilepsy is a chronic neurological illness that affects 1-2% of the population worldwide (Panayiotopoulos, 2010), with nearly 2.4 million people newly diagnosed per year (Megiddo, et al., 2016). Unfortunately, 30-40% of epileptic patients have uncontrolled seizures and are left without any proper treatment, or their seizures do not respond to medication (Cook, et al., 2013). A brain test called an electroencephalogram (EEG) is used to spot anomalies in the electrical brain signals. EEG signals are one of the frequently used methods to categorise and predict neurological diseases and disorders. Usually, the visual representation of EEG signals must be analysed and monitored by experienced medical professionals (Hu et al., 2020) a time-demanding process and one where a clinician's fatigue can cause less accurate outcomes (San-Segundo et al., 2019). Therefore, an automatic epilepsy seizure detection system that can assist healthcare experts and staff in analyzing the EEG signals quickly and efficiently would assist in providing more precise and reliable diagnoses.

Historically, epilepsy seizures are diagnosed by classifying electroencephalography (EEG) electrical signals into epilepsy or control classes and require expensive devices to record the EEG signals. The detection of epileptic seizures by classifying EEG signals is a demanding and challenging task, as it identifies the seizure and seizure-free states from non-linear and non-stationary data. Previous research has seen many machine learning based approaches introduced to analyze and interpret EEG signals for accurate classification. However, the nature of EEG data (non-linear and non-stationary) makes it difficult to extract proper information regarding these dynamic biomedical signals, while the extraction of the most relevant features set from EEG recordings is also challenging. Another issue is the potential for high misclassification rates due to the oscillatory and fractal characteristics of EEG signals (epilepsy and control) possessing a high resemblance. To address these problems, this study empirically investigates several data segment sizes and selects the optimal window size to segment the Guinea-Bissau dataset, (a dataset collected with consumer-grade devices, mimicking the conventional way to record these signals in many parts of the word). Several statistical and spectral feature extraction methods are investigated to obtain useful features from segmented epochs, in combination with conventional ML

Alyas, N., Hastings, D. and Mehrabidavoodabadi, A.
A Comparative Analysis of Classifier Performance for Epileptic Seizure Detection Using EEG Signals.
DOI: 10.5220/0011631600003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 237-244
ISBN: 978-989-758-626-2; ISSN: 2184-4313

algorithms and ensemble methods. The proposed framework is then implemented on a dataset of similar quality (the Nigeria dataset, collected with same protocol as Guinea-Bissau dataset), to validate the reliability of the framework. A comparative analysis is performed to demonstrate the effectiveness of the proposed framework. The obtained results demonstrate that XGBoost and LightGBM achieved the highest levels of performance, in terms of F1 score and AUC. The main contributions of this research project address relevant feature extraction problems, low performance levels, and detection using recordings from low-cost devices, are stated as follows:

- An effective ensemble method for epileptic seizure detection is implemented that classifies EEG signals obtained using a low-cost device into epilepsy and control classes with enhanced performance.
- The best window size is determined empirically to get the optimal segments of the long EEG signals for better interpretation.
- A combination of spectral and temporal feature extraction techniques is investigated and extracts the most useful set of features by implementing a TFD-based statistical analysis of segmented EEG signals.
- The comparison analysis of state-of-the-art and ensemble classifiers is performed using several evaluation metrics.

2 LITERATURE REVIEW

Data can be collected using an EEG monitoring device that records the brain activity in form of brain signals through different channels/electrodes connected to the scalp. It records the signals with voltage and spatial information. EEGs are non-stationary data, as the statistical characteristics of data change over time (Azami et al., 2011). To allow for the data to be used by a range of classifiers, data segmentation divides the signals into segments that are expected to contain the same temporal and spectral features (Hassanpour & Shahiri 2007). Dividing the data in such a way helps to generate more training samples and a more appropriate way to train classification algorithms. Feature extraction is a fundamental component of developing an efficient epileptic seizure detection system. The most frequently used feature extraction techniques for EEG data include Time-domain (TD), Frequency-domain (FD), Time-Frequency domain (TFD), Fourier transform (FT), Short-time Fourier Transform (STFT) and Continuous wavelet transform (CWT) (Usman et al., 2019). Moreover, Empirical mode decomposition (EMD), Kalman filter (KF), Singular spectrum analysis (SSA), Discrete wavelet transform (DWT), and Savitzgy-Golay (SG) filtering are also used to enhance the performance of ML and deep learning (DL) techniques (Azami & Sanei 2014). Regarding the algorithmic approach, different classifiers such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbour (KNN) have been seen to produce high levels of performance, especially in brain signal processing, and many previous studies preferred hybrid models for automatic seizure detection systems. Shoeb & Guttag (2010) presented a machine learning-based classifier to develop a patient-specific model for onset detection of an epileptic seizure and took a step to explore and solve the automatic epileptic seizure detection problem. They used the data from the CHB-MIT database that contains a total of 163 seizure episodes and EEG signals recordings of 844 hours. They used SVM with feature extraction, using time and frequency domains (FD). The evaluation was performed using performance measures such as sensitivity, detection delay, and false alarms per hour. The model achieved good results with an average sensitivity of around 96%. However, the study highlights the main challenges that are intrinsic to this problem, including data quality issues. Wang et al., (2015) implemented and compared ML algorithms such as DT based algorithm named C4.5, RF, SVM, and SVM-based random forest (SVM+RF), and DT-based SVM (SVM+C4.5) for seizure detection. A RF outperformed all other implemented models in this study, yielding the highest accuracy among all algorithms (Wang et al., 2015). Dash et al., (2020) proposed a novel approach that extracted sub-components from EEG signals using an iterative filtering decomposition approach and implemented Hidden Markov Model (HMM) as a classifier. They evaluated the models using a private EEG dataset from the All India Institute of Medical Science (AIIMS), Patna, and a publicly available database (CHB-MIT). The final class was decided based on the maximum score from HMM classifier. The proposed novel approach using decomposition of EEG signals and HMM attained 99.60% and 99.74% accuracy for the online CHB-MIT and AIIMS Patna EEG datasets, respectively. ML was found effective for epilepsy detection by Kavitha et al., (2022), who implemented KNN, DT, Naive Bayes (NB), and SVM for EEG signal classifications. A University of Bonn database and real-time private dataset obtained from the Senthil Multispecialty Hospital, India, were used. The EEG signals were extracted from both datasets and broken up into six frequency sub-bands using DWT and six statistical features were extracted, allowing for the combining of different features and classifiers. Van Hees et al. (2018) collected EEG signals for 5-minutes with an EMOTIVE device for epilepsy identification from two low-income countries in rural areas: Guinea-Bissau and Nigeria. They achieved an accuracy of 83% and 70% in Guinea-Bissau and Nigeria respectively. The key results from the a number of key studies in this domain are shown in Table 1.

Table 1: Existing machine learning approaches for seizure detection.

Author	Method	Features	Dataset	Accuracy (%)
Birjandtalab, Pouyana, Cogana, Nourani, & Harvey (2017)	Random Forest	TD, PS	Epilepsy	93.8
Donos, Dumpel-mann, & Schulze-Bonhage (2015)	Random Forest, k-Nearest Neighbour	FD, PS	CHB-MIT	80.9
Pinto-Orellana & Cerqueira (2016)	Random Forest	STFT, PCA, MMF	CHB-MIT	97.1
Wang, Gong, Li, & Qiu (2019)	Random Forest	TFD	BONN, CHB-MIT	100

3 PROPOSED FRAMEWORK

This study proposes an improved framework to automate epileptic seizure detection with the aim of classifying the input EEG signals into epilepsy and control classes, as shown in Figure 1. The proposed framework consists of 6 sub-modules: (1) EEG dataset acquisition using the consumer-grade device, that describes the EEG data and its parameters in detail; (2) pre-processing of data, which deals with the cleaning of data from noise and artefacts; (3) signal segmentation, related to dividing the long signal into segments using a window size; (4) feature extraction to extract the most relevant features from the data; (5) classifier building, including building a classifier using ensemble method to compare with conventional ML models; and (6) evaluation, which deals with the performance evaluation of models using different metrics.

3.1 Data Collection

The datasets used in this research were acquired by van Hees et al. (2018), and are available for public access. They collected EEG signals from epileptic (N=51) and healthy individuals (N=46) in the low-income country of Guinea-Bissau. A low-cost, portable, and consumer-grade device (EMOTIVE) was used to acquire the 5-minutes of 14 channels includes: AF3, AF4, F3, F4, F7, F8, FC5, FC6, O1, O2, P7, P8, T7, T8) resting-state EEG data from rural areas with the 128 Hz sampling frequency.

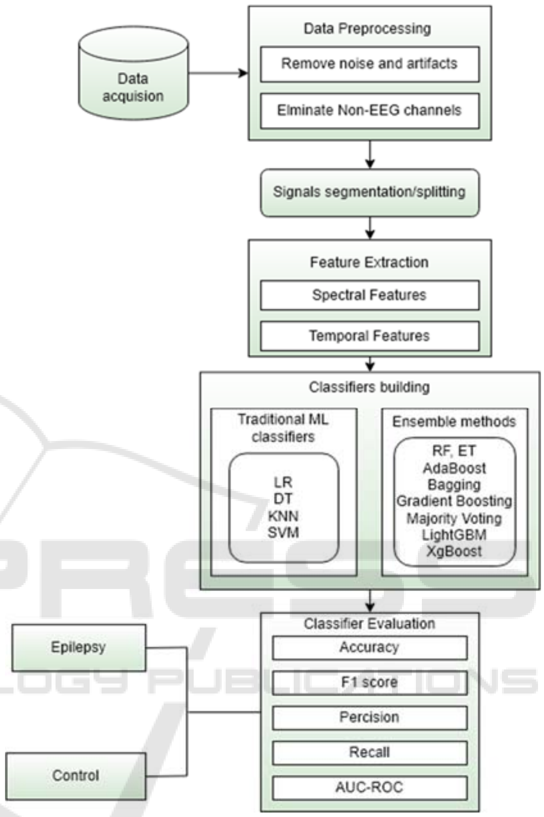

Figure 1: Proposed framework for epilepsy detection.

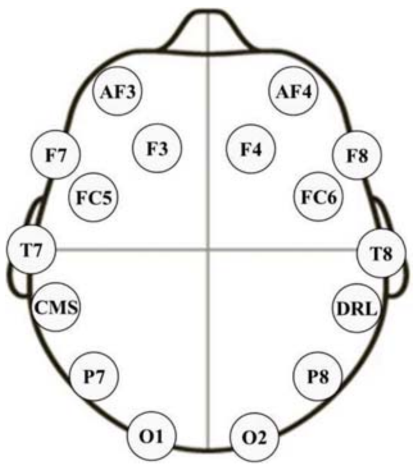

Figure 2: Emotive-EPOC device/headset 14 channels scalp placement (Mehmoud & Lee, 2016).

Table 2: Description of Dataset.

Country	Guines-Bissau		Nigeria	
	Epilepsy	Control	Epilepsy	Control
N	51	46	112	92
Gender (M/F)	31/20	41/5	67/45	56/36
Age (Mean)	25	25	21	20
Age (S.D.)	13	8	12	8

The electrodes were placed on the scalp at anterofrontal (AF3, AF4, F3, F4, F7, F8), parietal (P7, P8), occipital (O1, O2), frontocentral (FC5, FC6), and temporal sites (T7, T8), according to the International standard 1020 as shown in Table 2. The Guinea-Bissau dataset consisted of 97 subjects in total while Nigeria dataset consist of 204 subjects.

3.2 Data Pre-processing

Data processing was performed to prepare the data for use. The datasets contain non-EEG parameters: for this research they can be considered superfluous, hence they are removed, leaving the 14 EEG channels. Data is filtered using a band-pass filter range of 0.1Hz-45Hz: removal of the high frequencies eliminates the effects of a few artifacts as well as line noise, while the suppression of low frequencies below 0.1 Hz eliminates the effects of slow voltage shifts due to the skin potentials. Re-referencing was then used to normalize the signal.

3.3 Signal Segmentation

To segment the signals, the full EEG signal of around 300 seconds time series is segmented into small epochs to make a better interpretation and classification. The appropriate segment size is selected by evaluating different split sizes. Firstly, different splits were investigated with overlapping of 1-s. The selected EEG epochs are of 8-s window size with 1-s overlapping.

3.4 Feature Extraction

3.4.1 Power Spectral

Power spectral density using the Welch method is performed to calculate the spectral features. The EEG signal Y is decomposed to evaluate the power distribution across the neurological frequency spectrum. The Welch method is a Power Spectral Density (PSD) estimation technique that uses a STFT to calculate the periodogram for segmenting the EEG data (TD signal to FD). Overlapping segments are windowed with a discrete Fourier transform applied to calculate the periodogram, then the data is squared and each periodogram is averaged to obtain the power measure.

3.4.2 Statistical Methods

Statistical methods were used to extract the different temporal features such as average, standard deviation (SD), peak to peak (PTP), variance, minimum, maximum, index of minimum value, index of maximum value, root mean square, and absolute difference of signal, skewness, and kurtosis.

3.5 Classifier Building and Evaluation

In this study, four conventional machine learning (ML) algorithms, Logistic Regression (LR), K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM), and ensemble methods as Random Forest (RF), Bagging, Majority Voting, Extra Tree (ET), AdaBoost, Gradient Boosting, XGBoost, LightGBM are implemented to build classifiers for epileptic seizure detection using EEG.

For the experiments, the chosen dataset consists of a total of 4245 EEG epochs. For each classifier, the dataset is spliced into train, validation, and test subsets as shown in Table 3. Grid search is used for hyper-parameters tuning and to overcome the overfitting problem using k-fold cross validation. The training subset is used to develop the classifier, the validation set to tune the algorithms and test subsets are used to evaluate the performance of classifiers.

Table 3: Train Test Split.

Splits	No. of Epochs
Training Data	3438
Validation Data	382
Test Data	425

The classifiers are evaluated through a range of performance metrics, namely accuracy, precision, recall (both used to calculate the F1 score), and AUC, along with the macro and weighted averages, to ensure robustness in the evaluative process.

4 RESULTS AND DISCUSSION

This subsection provides an analysis and discussion of the results obtained by implementing conventional ML and ensemble methods with a different set of features. All the stated results were achieved after model-specific hyperparameter tuning to optimise model performance. F1 score and AUC-ROC have been chosen as the principal methods of evaluation, as in combination they provide a robust means evaluating predictive performance.

4.1 Results of Conventional ML Algorithms

Tables 4 and 5 show the F1 scores for the conventional ML algorithms. The results demonstrate that highest level of performance among these models comes from k-NN using spectral features (with a weighted average F1 score of 0.833). SVM achieves a similar level of performance, with a weighted average F1 score of 0.830.

The experimental results are also evaluated in terms of AUC-ROC, presented in Figures 3 and 4, which demonstrate the performance of conventional ML algorithms with spectral features achieve better performance than those using temporal features.

Table 4: F1 scores for conventional ML algorithms (ML + statistical features).

Classifier	Control	Epilepsy	Macro Avg.	Weighted Avg.
Decision Tree	0.787	0.744	0.765	0.766
k-Nearest Neighbour	0.698	0.525	0.612	0.617
Logistic Regression	0.790	0.769	0.780	0.780
Support Vector Machine	0.746	0.730	0.738	0.738

Table 5: F1 scores for conventional ML algorithms (ML + spectral features).

Classifier	Control	Epilepsy	Macro Avg.	Weighted Avg.
Decision Tree	0.694	0.648	0.671	0.673
k-Nearest Neighbour	0.839	0.825	0.832	0.833
Logistic Regression	0.771	0.757	0.764	0.765
Support Vector Machine	0.842	0.816	0.829	0.830

The AUC plots confirm the performance discrepancy between models, with K-NN and SVM considerably outperforming the decision tree and logistic regression models.

4.2 Results of Ensemble Methods

Nine ensemble methods were used in this study: bagging, RF, extra tree (ET), hard majority voting, soft majority voting, AdaBoost, gradient boosting, (XGBoost), and LightBGM.

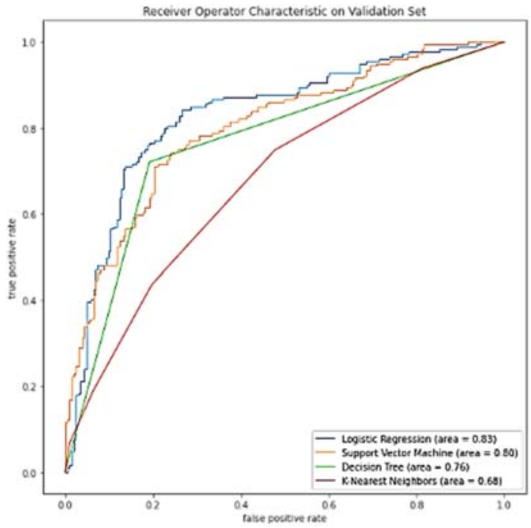

Figure 3: AUC plot for conventional ML models with statistical features.

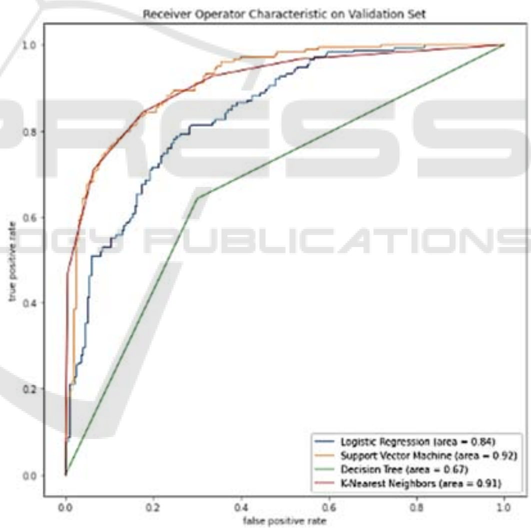

Figure 4: AUC plot for conventional ML models with spectral features.

The F1 score results of the ensemble classifiers using statistical features are shown in Table 6. In these results, XGBoost demonstrates the best performance with a weighted average of 0.921. LightGBM also performed well when compared to with the other methods, with a weighted average of 0.900.

Table 6: F1 scores for ensemble methods + statistical features.

Classifier	Control	Epilepsy	Macro Avg.	Weighted Avg.
Adaboost	0.772	0.714	0.743	0.745
Bagging	0.688	0.388	0.534	0.547
Extra Tree	0.888	0.849	0.869	0.870
Gradient Boosting	0.801	0.754	0.778	0.779
Hard Majority Voting	0.820	0.778	0.799	0.800
LightGBM	0.909	0.891	0.900	0.900
Random Forest	0.880	0.842	0.861	0.862
Soft Majority Voting	0.821	0.770	0.795	0.797
XGBoost	0.928	0.914	0.921	0.921

Table 7: F1 scores for ensemble methods + spectral features.

Classifier	Control	Epilepsy	Macro Avg.	Weighted Avg.
Adaboost	0.770	0.703	0.737	0.739
Bagging	0.814	0.738	0.776	0.779
Extra Tree	0.890	0.867	0.879	0.879
Gradient Boosting	0.725	0.655	0.690	0.692
Hard Majority Voting	0.824	0.785	0.804	0.806
LightGBM	0.861	0.850	0.856	0.856
Random Forest	0.851	0.815	0.833	0.834
Soft Majority Voting	0.792	0.767	0.779	0.780
XGBoost	0.864	0.847	0.855	0.856

Table 7 shows the performance of ensemble methods with the set of spectral features. The results demonstrate that ET achieved the highest levels of performance when compared to other ensemble algorithms, with a weighted average of 0.879.

The experimental results of the ensemble methods are evaluated in terms of AUC-ROC, the graphical demonstration of which is in Figures 5 and 6. When using the statistical features, XGBoost, LightGBM and ET all demonstrate high levels of discriminant ability, with AUC values of 0.98, 0.97 and 0.95 respectively, reinforcing the findings generated through the F1 score. The same models achieve the highest levels of performance when using the spectral features, each achieving AUC values of 0.93; lower than those achieved when using the statistical features. This pattern is repeated for the majority of ensemble models, which differs from the conventional ML models, where the individual highest levels of performance were achieved by using the spectral features.

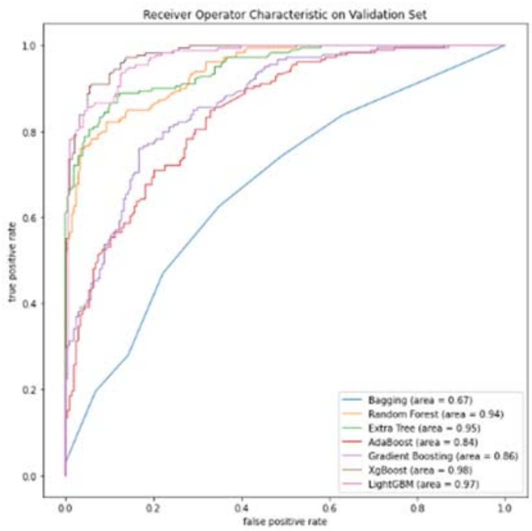

Figure 5: AUC plot for ensemble methods with statistical features.

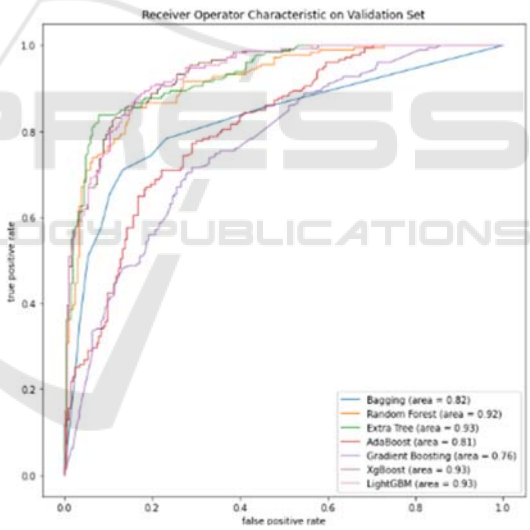

Figure 6: AUC plot for ensemble methods with spectral features

From the comparative analysis of classification results obtained from the Guinea-Bissau dataset it can be seen that the classification accuracies of KNN and SVM with spectral features are the best among all the conventional ML algorithms implemented in this study. Equally, for ensemble classifiers, XGBoost with statistical features achieved the best overall performance: this combination of model and features was also the optimal approach when considering all other approaches used in the study.

4.3 Results with Nigeria Dataset

We used the Nigeria EEG dataset that was recorded using the same protocols and standards as the Guinea-Bissau dataset to validate the performance of the proposed framework. The results achieved with this dataset were satisfactory and prove the reliability of the model. All the conventional ML algorithms and ensemble methods along with feature extraction techniques were implemented. The results gathered from this data demonstrate similar outcomes to those achieved using the Guinea-Bissau data: the highest performing model was XGBoost with a set of statistical features, with 79.45% accuracy and a weighted F1 score of 0.793. While the results for the Nigeria dataset are lower than those achieved when using the Guinea-Bissau dataset, this mirrors the findings of van Hees et al. (2018) and Anwar et al., (2021), who also document reduced levels of performance when using the data collected from Nigeria.

5 CONCLUSION

Epileptic seizures cause abnormalities of the brain and physical activities of epileptic patients, considered a chronic disease with an increased number of patients and sudden deaths every year. As earlier indicated, a better approach for epilepsy detection uses EEG data recorded using a consumer-grade device, and this study demonstrates that the optimal performance for an epilepsy detection model using such data can be achieved through ensemble machine learning methods using statistical features derived from the data. Accommodating the low-quality data using low-cost devices has not frequently been an approach used in previous research. However, the use of such data in the development of a system to detect epileptic seizures is better able to replicate the real-world data that can be collected from patients in much of the world and opens an avenue to increase the diagnosis rate of this disorder in low-income countries. However, additional factors may be considered that remain unaddressed within the study, such as geographical location of the patients and patient genetics that may affect the results. Further work will address this limitation to aid in the development of more generalisable findings. Moreover, when building the automatic seizure detection system, the potential effectiveness of deep learning methods should be investigated. Future work will identify whether deep learning algorithms can be implemented to further improve the development of accurate and reliable detection systems, along with attempting to optimise the datasets themselves, through the use of combined statistical and spectral features.

REFERENCES

Megiddo, I. et al., 2016. Health and economic benefits of public financing of epilepsy treatment in India: An agent-based simulation. *Epilepsia,* 57(3), pp. 1-11.

Panayiotopoulos, C. P., 2010. *A Clinical Guide to Epileptic Syndromes and their Treatment.* NewYork: Springer.

Cook, M. J. et al., 2013. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. *Lancet Neurol,* 12(6), pp. 563-571.

Hu, X. et al., 2020. Scalp EEG classification using deep Bi-LSTM network for seizure detection. *Computers in Biology and Medicine,* Volume 124.

San-Segundo, R. et al., 2019. Classification of epileptic EEG recordings using signal transforms and convolutional neural networks. *Computers in Biology and Medicine,* Volume 109, pp. 148-158.

Azami, H., Mohammadi, K. & Hassanpour, H., 2011. An Improved Signal Segmentation Method using Genetic Algorithm. *International Journal of Computer Applications,* 29(8).

Hassanpour, H. & Shahiri, M., 2007. *Adaptive Segmentation Using Wavelet Transform.* Lahore, Pakistan, IEEE.

Rashed-Al-Mahfuz, M. et al., 2021. A Deep Convolutional Neural Network Method to Detect Seizures and Characteristic Frequencies Using Epileptic Electroencephalogram (EEG) Data. *IEEE Journal of Translational Engineering in Health and Medicine,* Volume 9.

Usman, S. M., Latif, S. & Beg, A., 2019. Principle components analysis for seizures prediction using wavelet transform. *International Journal of Advances in Applied Sciences,* 6(3), p. 50–55.

Azami, H. & Sanei, S., 2014. Spike detection approaches for noisy neuronal data: Assessment and comparison. *Neuro computing,* Volume 133, pp. 491-506.

Shoeb, A. H. & Guttag, J. V., 2010. *Application of Machine Learning To Epileptic Seizure Detection.* s.l., ICML 2010.

Wang, G., Deng, Z. & Choi, K.-S., 2015. *Detection of epileptic seizures in EEG signals with rule-based interpretation by random forest approach.* s.l., Advanced Intelligent Computing Theories and Applications.

Dash, D. P., Kolekar, M. H. & Jha, K., 2020. Multi-channel EEG based automatic epileptic seizure detection using iterative filtering decomposition and Hidden Markov Model. *Computers in Biology and Medicine.*

Kavitha, K. N. et al., 2022. On the Use of Wavelet Domain and Machine Learning for the Analysis of Epileptic

Seizure Detection from EEG Signals. *Journal of Healthcare Engineering.*

Anwar, T., Rehmat, N., & Naveed, H. (2021). A Generic Approach for Classification of Psychological Disorders Diagnosis using EEG. Mexico: IEEE.

Birjandtalab, J., Pouyana, M. B., Cogana, D., Nourani, M., & Harvey, J. (2017). Automated seizure detection using limited-channel EEG and non-linear dimension reduction. *Computers in Biology and Medicine, 82*, 49-58.

Donos, C., Dümpelmann, M., & Schulze-Bonhage, A. (2015). Early Seizure Detection Algorithm Based on Intracranial EEG and Random Forest Classification. *International Journal of Neural Systems, 25*(5).

Hees, V. T., Diessen, E. v., Sinke, M. R., Buitenhuis, J. W., Maas, F. v., Ridder, L., & Otte, W. M. (2018). Reliable and automatic epilepsy classification with affordable, consumer-grade electroencephalography in rural sub-Saharan Africa. *bioRxiv* .

Mehmood, R. M. & Lee, H. J., 2016. *Towards human brain signal preprocessing and artifact rejection methods.* Lasvegas, USA, Int'l Conf. Biomedical Engineering and Sciences (BIOENG'16).

Pinto-Orellana, M. A., & Cerqueira, F. R. (2016). Patient-dependent epilepsy seizure detection using random forest classification over one-dimension transformed EEG data. *Elsevier*.

Wang, X., Gong, G., Li, N., & Qiu, S. (2019). Detection Analysis of Epileptic EEG Using a Novel Random Forest Model Combined With Grid Search Optimization. *Frontiers in Human Neuroscience, 13*, 12.