STEWART, C., FUNG, WAI KEUNG, FOUGH, N. and PRABHU, R. 2023. Extremely random forest based automatic tonic-clonic seizure detection using spectral analysis on electroencephalography data. In *Proceedings of the 21st IEEE (Institute of Electrical and Electronics Engineers) Interregional NEWCAS conference 2023 (NEWCAS 2023), 26-28 June 2023, Edinburgh, UK*. Piscataway: IEEE [online], article 10198101. Available from: https://doi.org/10.1109/NEWCAS57931.2023.10198101

Extremely random forest based automatic tonic-clonic seizure detection using spectral analysis on electroencephalography data.

STEWART, C., FUNG, WAI KEUNG, FOUGH, N. and PRABHU, R.

2023

Extremely Random Forest based Automatic Tonic-Clonic Seizure Detection using Spectral Analysis on Electroencephalography Data

Craig Stewart,
School of Engineering,
Robert Gordon University, Aberdeen, UK

Wai Keung Fung,
Cardiff School of Technologies,
Cardiff Metropolitan University, Cardiff, UK

Nazila Fough,
School of Engineering,
Robert Gordon University, Aberdeen, UK

Radhakrishna Prabhu,
School of Engineering,
Robert Gordon University, Aberdeen, UK

***Abstract*— Machine learning proliferates society and has begun changing medicine. This report covers an investigation into how Extremely Random Forests combined with Fast Fourier Transform feature extraction performed on two-dimensional time-series Epileptic Seizure data from the Bonn/UCI dataset. It found that robust classification can take place with lower channel counts, achieving 99.81% recall, 98.8% precision and 99.35% accuracy, outperforming previous works carried into this scenario.**

***Keywords*— epilepsy, extremely random forest, electroencephalography, Fourier Transform**

I. INTRODUCTION

Epilepsies are a series of neurological conditions that are common throughout the populace.. The diversity in epilepsies and effects means it is challenging to know the truth of how a given epileptic experiences seizures daily and their frequency as they can often result in loss of consciousness, given that missing a Tonic-clonic seizure can possibly result in the death of a subject, an automated system could allow for improved, timely and efficient care and treatment.

Electroencephalography (EEG) is a method of analysing brain activity generally. This involves reading directly the bio-electrical signals occurring in the brain of an individual through a series of electrode being placed on the scalp. These translate the electrical activity of interest into an interpretable signal for analysis utilising Machine Learning algorithms [1].Building upon the work carried out in [2], where it was found Random Forests (RF) combined with Fast Fourier Transforms (FFT) produced timely, accurate diagnosis of seizure activity in this scenario, Extremely Random Forests (ERF) and FFT were utilised to improve the results. This work showed that even with low-channel count data streams reduced from clinical systems epileptic signals can be accurately diagnosed. This report shall introduce some of the previous works carried out into the field in Section 2. Section 3 shall discuss the methodology utilised in the investigation and section 4 shall interpret the results and thus, the conclusions can be discussed.

II. METHODOLOGY

The methodology has been specifically designed to study how epilepsy can be detected quickly and effectively in low channel devices. Given the selected dataset containing corresponding signals, the FFT algorithm was used to extract Fourier coefficients, the ERF algorithm was trained on this information in order to improve accuracy, reduce overfitting then performance evaluated using confusion matrix and derived performance metrics.

A. The Bonn and CHB-MIT Datasets

The Bonn dataset [3] is frequently used in the academic literature for the purpose of automated diagnosis of seizure activity, the UCI/Bonn version is a subset of the Bonn dataset that has been pre-processed for the purposes of machine learning driven analysis. Essentially, this takes the form of two-dimensional data, 178 samples representing a time-series EEG signal over the course of a second.

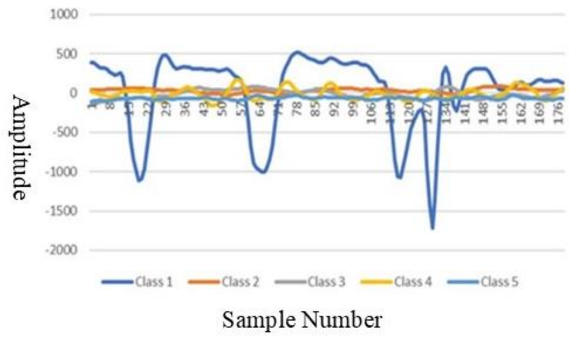

FIG 1. Examples of EEG signals contained in Bonn dataset

This dataset is upsampled by quadrupling the sample frequency to balance the disproportionality between seizure to non-seizure instances and then reduced to a binary classification problem, solving for seizures versus non-seizures. [4] discusses the placement of the electrodes on the brain in this dataset, utilising the 10-20 standard electrode placement and figure 1 shows the five classes of signals present in the dataset, class 1 corresponds with seizure data, the others being non-seizure. In addition, to analyse for robust classification, the classifier was also tested on a variation of the CHB-MIT dataset [5].

B. *Fast Fourier Transform*

FFT [6] is a digital signal processing technique used to efficiently perform a Discrete Fourier Transform (DFT) on the data obtaining the corresponding spectral coefficients. It is frequently used as a method of signal decomposition in machine learning algorithms as seen in [7], this constitutes “feature extraction.” FFTs are regularly utilised over DFT typically as they are a faster algorithm, given DFT’s time complexity is O (n^2) whilst FFT’s O (n log n) [8]. Discrete Wavelet Transform (DWT) were not implemented as the lattice filter can reduce the accuracy in a classification application by reducing spectral resolution despite being a faster to implement algorithm. This represents a caveat between the speed and accuracy of the model, on balance the loss of a few milliseconds is not comparable to missing a Tonic-clonic seizure, therefore, accuracy was prioritised in this scenario.

C. *Extremely Random Forest*

RF [8] algorithms are a form of Ensemble Learning that utilises a series of individual learners, in this case, Decision Trees [9], to learn a series of rules regarding a given dataset, allowing for classification to take place. An ERF [10] is a variation on the RF algorithm. A standard RF selects a random subset of features and then attempts to find the most discriminative threshold for each subset whereas an ERF not only randomly selects the features in a subset, it randomises the thresholding process also. This additional randomisation stage can boost accuracy and reduce the effect of over-fitting relative to the RF algorithm. Thus, the inclusion of this algorithm should see a boost in performance theoretically. For this experiment, 100 learners were enabled on the dataset. They are also much faster to train than RF due to this process.

III. RESULTS

This experiment was carried out using SciKit learn, a common Python ML Library and SciPy for its signal decomposition suite. The dataset was shuffled and splitted using SciKit’s K-Fold Cross Validation Function, 10-fold splits were selected in this instance. Table I are the results of the experiment.

TABLE I. Results of the Extremely Random Forest Experiment

Algo	Recall	Precision	Accuracy
ERF (w/FFT)	99.81%	98.9%	99.35%
RF (w/FFT)	99.84%	98.03%	98.95%

The training session was re-run six times to get precision, recall and accuracy scores and these were then aggregated and averaged to get a final score. It was found that the ERF performed better consistently than using just RF with FFT in the work [1] according to the standard ML metrics. There was a slight deviation in speed with ERF taking two millisecond more to diagnose a seizure than the RF algorithm whilst running on a Google Colab GPU, (14.5ms vs 12.2ms). However, the diagnosis precision was boosted by nearly a whole percent rendering the process more robust consequently. Implementation on CHB-MIT resulted in a score of 97.8% accuracy, 98% recall and 97.9% precision. This is also a gain despite the slight drop compared to that found in the Bonn dataset, this is most likely for the same reasons discussed in [2].

IV. CONCLUSION AND FUTURE WORK

In conclusion, it was found that the ERF algorithm improved on the RF by raising the precision considerably at the sacrifice of two milliseconds worth of time and gain in accuracy on the Bonn dataset. In addition, there were general all-round improvements on the CHB-MIT dataset performance, the more complicated dataset. This suggests that the algorithm would improve significantly for the scenario laid out in standard metrics other than speed and possibly recall which is a concern, seeing that false negatives are the worst-case scenario. This represents a step forward towards developing a portable EEG based detection system. Future work would be to do further studies to analyse how seizures can be differentiated from other types of movement, more studies to augment this algorithm, minimize false negatives particularly and to look at developing a device ideal for this scenario and then running the experiments with that.

REFERENCES

[1] Y. U. Khan, N. Rafiuddin and O. Farooq, "Automated seizure detection in scalp EEG using multiple wavelet scales," 2012 IEEE International Conference on Signal Processing, Computing and Control, Solan, India, 2012, pp. 1-5, doi: 10.1109/ISPCC.2012.6224361.

[2] C.Stewart, WK Fung, N. Fough and R. Prabhu “Automated Tonic-Clonic Seizure Detection using Random Forests and Spectral Analysis on Electroencephalography Data,” The 6th International Conference on System-Integrated Intelligence (SysInt 2022)

[3] UCI Machine Learning Repository (2017, May 24), Epileptic Seizure Recognition Data Set [Online] Available: https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition

[4] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, & C. E. Elger, , "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Physical Review E, vol. 64, (6), pp. 061907, 2001.

[5] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, & H. E. Stanley, (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220

[6] H. R. Al Ghayab, Y. Li, S. Siuly, & S. Abdulla, "Epileptic seizures detection in EEGs blending frequency domain with information gain technique," Soft Computing, vol. 23, (1), pp. 227-239, 2019.

[7] J. Cooley and J. Tukey, "An algorithm for the machine calculation of complex Fourier series," Mathematics of Computation, vol. 19, (90), pp. 297-301, 1965

[8] G. Biau and E. Scornet, "A random forest guided tour," Test, vol. 25, (2), pp. 197-227, 2016.

[9] S. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," IEEE Trans. Syst. Man Cybern., vol. 21, (3), pp. 660-674, 1991.

[10] P. Geurts, D. Ernst and L. Wehenkel, "Extremely randomized trees," Mach. Learning, vol. 63, (1), pp. 3-42, 2006.