STEWART, C., FUNG, W.K., FOUGH, N. and PRABHU, R. 2022. Automated tonic-clonic seizure detection using random forests and spectral analysis on electroencephalography data. In *Valle, M., Lehmhus, D., Gianoglio, C. et al. (eds.) Advances in system-integrated intelligence: proceedings of the 6th International conference on system-integrated intelligence 2022 (SysInt 2022), 7-9 September 2022, Genova, Italy*. Lecture notes in networks and systems (LNNS), 546. Cham: Springer [online], pages 679-688. Available from: https://doi.org/10.1007/978-3-031-16281-7_64

Automated tonic-clonic seizure detection using random forests and spectral analysis on electroencephalography data.

STEWART, C., FUNG, W.K., FOUGH, N. and PRABHU, R.

2022

Automated Tonic-Clonic Seizure Detection using Random Forests and Spectral Analysis on Electroencephalography Data

Craig Stewart[1], Wai Keung Fung[2,] Nazila Fough[1], Radhakrishna Prabhu[1]

[1] School of Engineering, Robert Gordon University, Aberdeen, UK
[2] Cardiff School of Technologies, Cardiff Metropolitan University, Cardiff, UK
C.stewart35@rgu.ac.uk

Abstract. Artificial intelligence (AI) has a potential for impact in the diagnosis of neurological conditions, the academic consensus generally has a positive outlook regarding how AI can improve the care of stroke victims and those who suffer from neuro-degenerative conditions such as dementia. These technologies have tangible applications in improving the welfare of epileptics, epilepsy being a common neurological condition that can result in premature death without a quick response. As such it is important for the system to avoid false negatives, ie. falsely diagnosing that the subject is fine when having a seizure. This investigation focused on how machine learning algorithms can be utilised to identify these events through Electroencephalography (EEG) data. The UCL/Bonn dataset, a classic benchmark for automated epilepsy detection systems was identified and utilised. This investigation focused on the random forest algorithm a popular technique to allow for pattern detection and classifcation. Given that EEG neurological data represents time series data and machine learning excels at this task, automation could be achievable. From there, Fast Fourier Transforms (FFT) were applied to identify if spectral features of EEG signals would aid identification of seizures. This method achieved an accuracy of 99%, precision of 99% and a recall of 100% in 12.2 milliseconds time to classify and one second of EEG data. These results show that random forests combined with FFT are a viable technique for attaining high recall, thus minimising false negatives specifically, when detecting grand mal epileptic seizures in short periods of time.

Keywords: Epilepsy, Seizure Detection, Random Forest, Fast Fourier Transfomrm (FFT), Artificial intelligence (AI)

1. Introduction

Epilepsy is a group of neurological conditions that result in recurrent seizures of various magnitudes. This condition displays heterogeneity in outcome, this is caused by sheer variation of source within the brain meaning that they can result in anything from relatively minor behavioural infringements to life threatening arrests. The cause usually is a form of neurological damage that can lead to either increased or reduced sensitivity to stimulus triggering a chain reaction that spreads throughout an area of the brain as discussed in [1] by Thjis *et-al*. The size of the area in the brain is related to the magnitude of the seizures as seen also in [1]. This family of conditions is relatively common throughout society with the World Health Organisation (WHO) estimating that 70 million people globally have some form as discussed in [2]. By in large, epileptic conditions are treatable with conventional medications, reducing the magnitude of the seizures or stopping them completely. However, there is a significant minority that will endure the condition throughout life experiencing a diminished quality of life with the treatment itself inducing side-effects as well as social barriers being imposed [1].

Epileptic seizures can be hazardous in of themselves; however, injuries can also be sustained from the uncontrolled nature of collapse, serious head injuries as a result can be sustained that prove fatal. Care is also hampered by the unpredictability of seizures with forewarning being limited often to the moments leading up to an event, often known as an “aura”, a grace period that allows for preparation. The seizures themselves, in nature, can be unpredictable, individuals prone to tonic-clonic seizures are not excluded from experiencing an entire spectrum of seizures from absence to full tonic-clonic arrests. Thus, there is vested research interest into technological solutions capable of predicting when a seizure is likely to occur or otherwise provide prompt warning to carers as to if one has occurred. There are also applications within the healthcare profession where automatic epileptic event detection can speed up identifying of activity allowing for quicker diagnosis. Fig 1 illustrates the diversity epileptic conditions.

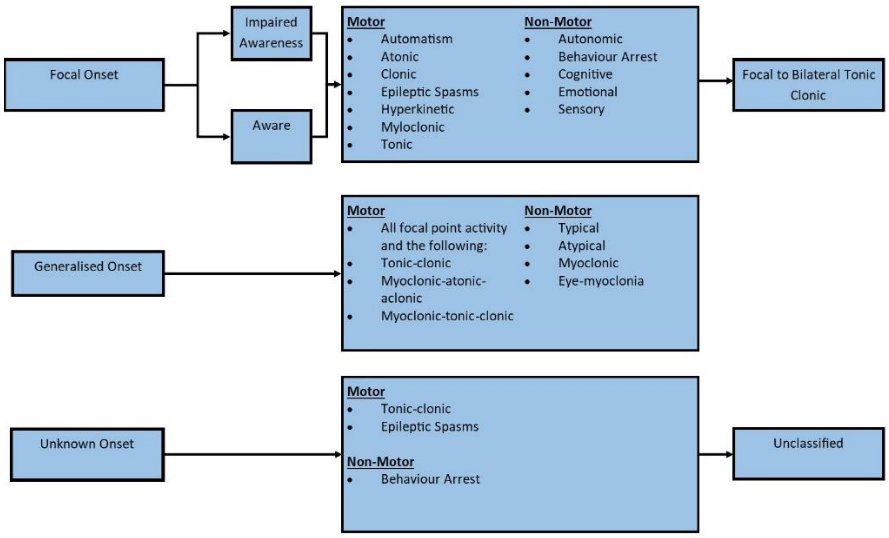

Fig 1. The International League Against Epilepsy framework for the classification of epileptic seizures [3]

There are numerous signal types that can either implicitly or explicitly suggest that a seizure event has occurred, some academics elect to follow implicit, indirect signals related to motion and temperature changes related to possible tonic-clonic activity, the embrace2 watch [4] takes such an approach whilst others work with explicit, direct signals from the brain itself utilising Magnetic Resonance Imaging (MRI) and Electroencephalography (EEG) [1]. Prediction and medication are typically done utilising EEG, this involves several electrodes being placed directly on to an individual’s scalp to monitor the electrical activity occurring. Identification is made possible through specific signal traces that correspond with an event occurring, helping to segregate signals that correspond with normal and abnormal functions. Seizures have four types defined by their amplitude levels, these are the preictal, postictal, ictal and interictal stages as discussed in Deivasigamani's paper [5]. Preictal stage is related to before the seizure occurs, the ictal stage relates to the onset of an attack and the postictal follows this stage as discussed. Interictal stage is referred to as the first ictal occurring stage. Currently it is understood that seizures can be identified as early at the preictal stage [5]. Prediction and searching for signals that anticipate seizures takes significant periods of expert time, an expert can often take days, weeks or even months to analyse the output of EEG signals. In addition, people can often overlook events by accident as well as due to the enormity of the data itself as seen in [6] by Gajic *et al*. Due to these factors, there has been a push to incorporate automation to effectively minimise the cost and human effort involved in the process. Another point of interest to epilepsy experts, as discussed above, is the location and then spread of an event. The location and spread defines what the outcome of an event will be, with localised seizures generally manifesting in events such as absences and small localised arrests, these are often referred to as focal-point seizures. Non-localised seizures can engulf entire portions of the brain and tend to result in tonic-clonic, otherwise known as "grand mal," seizures. These result in the typical rhythmic muscular driven convulsions. Beghi discussed this in [7].

Benefits of using EEG for epileptic seizures stem from their cost, portability and show clear interpretable patterns in the frequency domain according to [8] by Shoeibi *et-al*. Therefore, this is particularly relevant as pervasive computation and the Internet of Things (IoT) continues to proliferate society, offering new automation opportunities as well as those that can potentially benefit, not only epileptics, but others that fall prey to similar neurologically driven events. This paper is structured as such, with Section 2 discussing related works in the field, Section 3 discusses the methodology of the investigation, Section 4 discusses the simulation results and finally conclusion and future works is discussed in Section 5, Overall, this report shows and discusses the results from applying the Random Forest algorithm with spectral features extracted by Fast Fourier Transform (FFT) to a dataset of EEG signals derived from epileptic seizures through electrodes placed on the scalp.

The process of taking an EEG involves the placement of conducting electrodes fixed with adhesive to the subject's scalp the number of which varies depending on how much area detail is necessary. Therefore, data can vary from set to set depending on the experiment carried out. Table 1 shows a selection of literature related to the study of random forests regarding this area, the Bonn dataset to be used and the implementation of Fast Fourier Transforms. The benefits of these implementations are that for a clinical situation involving considerable time to diagnose, this dataset uses 23 seconds of data in its unprocessed form from which classifcations can be made, accurate results can be obtained often as high as the late nineties or even 100% as can be seen in [11] and [13]. However, an epileptic going about their daily activity does not benefit themselves from this innovation. 23 seconds is a considerable period of elapsed time during a seizure event, given that time management is crucial, the air passage could be restricted due to the tongue being swallowed or bitten, and attention is deemed urgent so measures can be taken to reduce risk. The UCI/Bonn modified dataset considers a single second of time to make prompt decisions which is more relevant to the individual. Therefore, time to classify will be considered in this project in addition to accuracy, specificity etc.

Table 1. A selection of pertinent literature regarding this field

Classifier	Features	Performance (%)	Metric	Dataset
ELM and BPNN [9]	Sample Entropy	95.6	Accuracy	Bonn
Random Forest [10]	STFT, mean, energy and standard deviation	96.7	Accuracy	Bonn
LS-SVM [11]	FFT and DWT	~100	Accuracy	Bonn
Random Forest [12]	Intrinsic Mode Functions	98.4, 98.6, 96.4	Sensitivity, Specificity, Accuracy	Bonn
Random Forest [13]	L1-Penalised Robust Regression	100	Accuracy	Bonn

2. Methodology

In this paper, a detection mechanism is proposed that tends to focus on the diagnosis of seizure activity and discards classifying whether a seizure is tonic-clonic etc. The reasoning behind this is because an epileptic event, regardless of magnitude, tends to be significant event to the individual and the professional team in the greater picture. Given that a tonic-clonic seizure is a specifically profound event at any time during an epileptic's life with the condition, the other types of seizure are indeed just as relevant in long term analysis. The data regarding absences or otherwise can often indicate a change in status or a progression in the condition, if an epileptic has been on medication for a significant period, successfully managing the condition in doing so, then has even a light episode it could be cause for concern driving a shift in medication or otherwise. These seizures can also be a pre-emptive sign of a larger seizure incoming thus giving an opportunity for proactive care rather than reactive, this could be the difference between someone surviving and succumbing to the infliction. This simplifies the classification process in of itself in terms of functionality, rendering it from a multiclass to a binary classification problem.

The process of experimentation is as seen in fig 2.

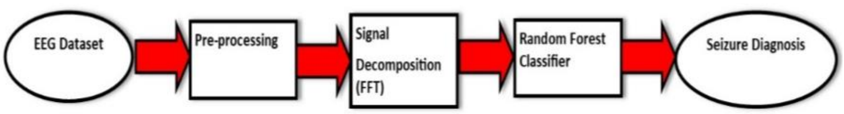

Fig 2. process diagram of the algorithm utilised

3.1. Dataset

Datasets are key in developing an adept learning system. An AI agent can only act as far as it has been trained properly to perceive. The demands of the dataset are dependent on several factors, application being one of them as the complexity of the system is directly related to this aspect, sitting down as still as possible in a clinical setting produces significantly different demands from a subject in motion, carrying out day to day activities. This simple caveat adds significant volumes of noise to a reading that can affect the machine learning process. Therefore, an application to be developed for this purpose of everyday diagnosis in life would need to have the relevant data, unfortunately this data is not readily available online with most readily available datasets focusing on a clinical environment application. A readily available dataset appropriate for basic seizure differentiation is the Bonn/UCI dataset as can be seen here [14]. Although it does not offer the movement data required to make a system suitable for everyday care it offers several advantages in that it has already seen data preparation to a degree for such training operations, this renders it time efficient as the pre-processing of data can take considerable time in of itself. Another aspect of simplicity is that it has been pre-rendered down to two-dimensional data, each line is a time-series datapoint which renders it simpler to interpret. The Bonn/UCI dataset also has five classes to differentiate from, these are as follows in table 2, whilst fig 3 illustrates the variation between signals of each class: -

Table 2. The five classifications in the Bonn Dataset [14]

Class Value	Description of Activity
1	Recording of seizure activity.
2	They record the EEG from the area where the tumour was located.
3	Yes, they identify where the region of the tumour was in the brain and recording the EEG activity from the healthy brain area.
4	Eyes closed, means when they were recording the EEG signal the patient had their eyes closed.
5	Eyes open, means when they were recording the EEG signal of the brain the patient had their eyes open.

Given that the latter groups 2-5 correspond with non-seizure data, converting them all into class 2 is enough to turn the problem into a binary issue. This has the effect of creating a large imbalance in the dataset, a ratio of 4:1 in favour of events unrelated to seizures, this was addressed by utilising oversampling of the minority class so that the split became a 1:1 ratio. Other caveats in terms of practicality in data is that epilepsy in of itself, is not merely driven by a given pattern throughout all epileptics, it is a diverse family of conditions, this means for an accurate system to be developed a large data set must be employed with as much seizure data as possible which represents are generalised tool which will not be applicable for all epileptics. Another option would be to train a highly specialised algorithm on as much individual data possible for a subject which would be a considerable project in labour and cost but would be bespoke to a given subject.

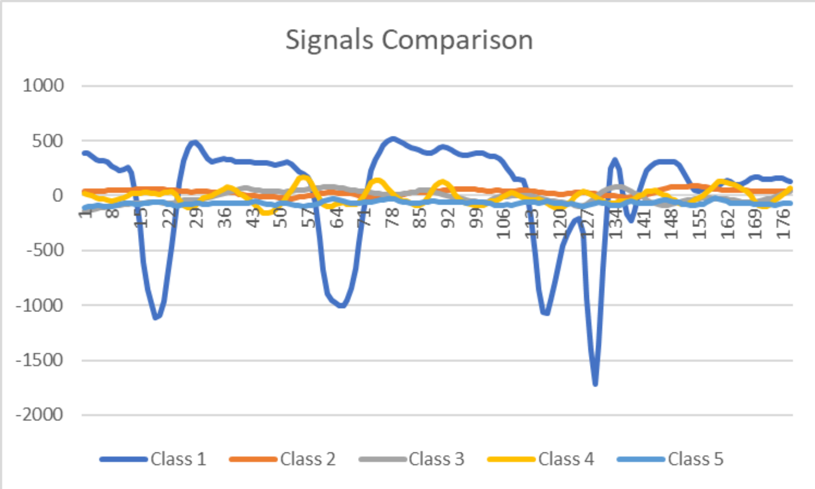

Fig 3. Illustration of signal from each class and the differences between each

3.2. Fast Fourier Transform (FFT)

Having pre-processed the data, Fast Fourier Transform (FFT), an algorithmic technique commonly utilised in various digital signal processing applications, is applied for signal decomposition. The core principle of an FFT is to implement a fast computation of Discrete Fourier Transform (DFT) coefficients, expressing a function as a sum of periodic components using the function stated in equation 1 as brought forward by Cooley and Tukey in [15].

$$y[k] = \sum_{n=0}^{N-1} e^{-2\pi j \frac{kn}{N} x[n]} \quad \textbf{[1]}$$

Where Y[k] is the FFT coefficients, N is the total number of input EEG samples, n is the signal sample index. Using this method, spectral features can be extracted from the time series data that allows for an alternate perspective to be taken on the EEG data. Transforms, in general, are widely implemented in the machine learning community Fourier and Wavelet transforms being readily available in addition within common libraries. Rationale for specifically choosing FFT is due to the fact it is a faster derivative of the DFT. The Discrete Wavelet Transform (DWT), an alternative algorithm utilised in this application, was not implemented because the lattice filter results in a loss of spectral resolution and, thus, can reduce the accuracy of the classification process, this can be seen in [16]. Also, given that FFT's time complexity is given by O(N*log2(N)), and DWT is given by O(N) this represent represents a compromise between accuracy and time, this could be justified by caveating that missing a severe seizure is riskier than the benefit from gain in processing time. The importance of a quick response within the epileptic community is well known as time can be relative to the harm caused because of seizure. The longer someone is arrested in seizure the higher the risk comes and therefore, the care needs change from something that can be handled privately to requiring an ambulance for emergency assistance. Therefore, given a second's worth of time has already been consumed to gather the necessary data for this algorithm, the actual classification process should be prompt to enable rapid attention, within seconds an individual could have collapsed and incurred concussive impact, as has been established, can result in a fatal outcome.

3.2. Random Forest Algorithm

The random forest algorithm is a form of ensemble learning that utilises several individual learners, in this case decision trees. As Safavian discusses in [17] decision tree classifiers (DTC) are pervasive in machine learning systems used in a broad range of scenarios from radar signal classification to speech recognition. The essence of DTCs is in the algorithm's ability to classify complex signals and data into simple rules, this represents a solution that is often easier to interpret. Decision Trees in of themselves can find success in time series signal data such as seizure detection with EEG signals. Random forests can offer advantages over individual learners.

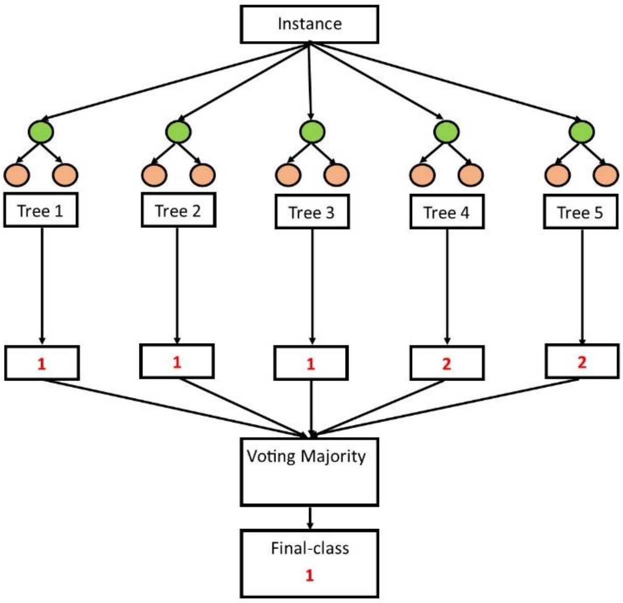

Fig 4. A general Random Forest process

As Biau *et al.* discussed in paper [18], a random forest is a general-purpose regression and classification algorithm that makes use of several decision trees arranged into an "ensemble." Fig 4 shows a simple illustration of a random forest process. This technique involves randomising these individual trees and then aggregates their individual predictions into an average score. This is a useful technique especially in areas where the variables are great in proportion to observations. Random forest has been already utilised successfully in studies regarding the analysis of seizure data. Therefore, there is already an element of verifiable proof available that this method is relatively successful within this context. In this study, 100 individual classifiers were utilised for this analysis using the standard framework given within the Scikit learn library with none of the other parameters tuned. Analytics were then carried out using a confusion matrix to understand overall performance.

3. Results

This proposed methodology was carried out in code using Python 3.6. Within Python several libraries were utilised including SciPy for its transform suite and Scikit Learn for simple practical access to common machine learning algorithms. In addition, some pre-processing was achieved using pandas and numpy. In this experiment, the splitting and shuffling was carried out using Scikit Learn's K-fold cross-validation with shuffling engaged, and 10-fold splits being carried out amongst the data. The benefits of K-fold cross validation being a reduction in bias during learning, rendering the learner a superior generaliser. Performance was ascertained in utilisation of typical metrics in recall, precision and accuracy. As discussed, the key metric is sensitivity for an epileptic seizure classifier. This operation was repeated six times to analyse if the results were consistent through multiple iterations of learning. Table 3 shows the performance of the random forest and FFT algorithm over 6 random trials of learning and testing.

Table 3 shows the results of training the random forest with the FFT enabled

Iteration	Recall	Precision	Accuracy
1	99.78%	98.39%	99.08%
2	100%	97.97%	98.97%
3	100%	99.73%	99.36%
4	99.78%	98.09%	98.91%
5	99.57%	97.67%	98.61%
6	99.89%	97.71%	98.79%
Average	99.84%	98.03%	98.95%

Average time to classify was 12.2 ms on 1 second of EEG data using GPU on Google Collab Cloud service. These results suggest that this methodology potentially, could lead to fast classification of epilepsy signals in a fast and timely manner when implemented on devices with low channel numbers. Of particular interest was that false negatives were minimised to the point where they were in the single digits achieving almost 100% accuracy in specificity. This means that false negatives were effectively decreased whilst training, therefore, the chance hazardous events being missed is reduced. This outperformed [9], [10] and [12] in the literature using techniques such as STFT and IMF but did not outperform [11] and [13], this could be because the dataset had more information to work with, given that these operations were based on analysing a considerable period worth of information simultaneously (23 seconds). Naturally, the extra information given on such a set-up would allow for alternative conclusions to be developed, possibly more accurate. These scores were achieved with one second worth of data window and 12.2 milliseconds worth of classification time, meaning that this can detect seizures fast in the context of this dataset. Given that this was done using GPU on the Google Colab Cloud Service, it is prudent to say that if this was running on a local device with limited resources this would run slower, further experiments could reveal the extent to which this is the case. This is far more practical for an epileptic having a seizure, reducing the time delay for a carer to act in case of an emergency.

4. Conclusion

In conclusion, using the proposed process the system was relatively successful in diagnosing epileptic seizures from the status-quo utilising random forests and FFT algorithm achieving 99% average specificity score with a time to classify of 12.2ms on one second of single channel data. This shows that fast, accurate seizure detection is possible. In terms of future work, it would be compelling to get more diverse data to understand if it is possible to separate epilepsy signals from movement and refine algorithms for every use by epileptic individuals. Another avenue would be to attempt proactive performance rather than reactive, to identify the seizure from the initial aura rather than when the seizure has begun. Attempting to minimize the time needed to make an accurate detection would be another experiment worth attempting.

References

[1] Thijs, R. *et al*, "Epilepsy in adults," *The Lancet,* vol. 393, *(10172),* pp. 689-701, 2019.

[2] England M. *et al*, "Epilepsy across the spectrum: Promoting health and understanding.: A summary of the Institute of Medicine report," Epilepsy & Behavior, vol. 25, (2), pp. 266-276, 2012.

[3] Fisher, R. Cross, J. French, J. *et al*., "Operational classification of seizure types by the International League Against Epilepsy: position paper of the ILAE Commission for classification and terminology." Epilepsia 2017; 58: 522–30.

[4] Empatica (2022, April 13) [Online] Available: https://www.empatica.com/en-gb/index.html

[5] Deivasigamani, S. Senthilpari, C. and Yong W., "Machine learning method-based detection and diagnosis for epilepsy in EEG signal," Journal of Ambient Intelligence and Humanized Computing, vol. 12, (3), pp. 4215-4221, 2021.

[6] Gajic D. *et al*, "Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition," Biomedical Engineering: Applications, Basis and Communications, vol. 26, (02), pp. 1450021, 2014.

[7] Beghi, E., "The epidemiology of epilepsy," Neuroepidemiology, vol. 54, (2), pp. 185-191, 2020.

[8] Shoeibi A., *et al*, "Epileptic seizures detection using deep learning techniques: a review," International Journal of Environmental Research and Public Health, vol. 18, (11), pp. 5780, 2021.

[9] Logesparan, L. Rodriguez-Villegas E. and Casson A., "The impact of signal normalization on seizure detection using line length features," Med. Biol. Eng. Comput., vol. 53, (10), pp. 929-942, 2015.

[10] Wang, X. *et al*, "Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization," Frontiers in Human Neuroscience, vol. 13, pp. 52, 2019.

[11] Al Ghayab H. *et al*, "Epileptic seizures detection in EEGs blending frequency domain with information gain technique," Soft Computing, vol. 23, (1), pp. 227-239, 2019.

[12] Sharma, R. *et al*, "Automated system for epileptic EEG detection using iterative filtering," IEEE Sensors Letters, vol. 2, (4), pp. 1-4, 2018.

[13] Hussein R. *et al*, "Robust detection of epileptic seizures based on L1-penalized robust regression of EEG signals," Expert Syst. Appl., vol. 104, pp. 153-167, 2018.

[14] UCI Machine Learning Repository (2017, May 24), Epileptic Seizure Recognition Data Set [Online] Available: https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition
[15] Cooley J. and Tukey J., "An algorithm for the machine calculation of complex Fourier series," Mathematics of Computation, vol. 19, (90), pp. 297-301, 1965.
[16] T. Edwards, "Discrete wavelet transforms: Theory and implementation," Universidad De, pp. 28-35, 1991.
[17] Safavian S. and Landgrebe D., "A survey of decision tree classifier methodology," IEEE Trans. Syst. Man Cybern., vol. 21, (3), pp. 660-674, 1991.
[18] Biau G. and Scornet E., "A random forest guided tour," Test, vol. 25, (2), pp. 197-227, 2016.