

Article

EEG Oscillatory Power and Complexity for Epileptic Seizure Detection

Lina Abou-Abbas ^{1,2,*} , Imene Jemal ^{2,3}, Khadidja Henni ^{1,2}, Youssef Ouakrim ^{1,2}, Amar Mitiche ^{2,3} and Neila Mezghani ^{1,2} 

- ¹ Imaging and Orthopedics' Research Laboratory, The CHUM Research Center, Montreal, QC H2X 0A9, Canada; khadidja.henni@teluq.ca (K.H.); oua.youssef@gmail.com (Y.O.); neila.mezghani@teluq.ca (N.M.)
- ² Research Center LICEF, Teluq University, Montreal, QC H2S 3L4, Canada; imen.djmal@gmail.com (I.J.); amar.mitiche@inrs.ca (A.M.)
- ³ INRS-Centre Énergie, Matériaux et Télécommunications, Montreal, QC H5A 1K6, Canada
- * Correspondence: lina.abou-abbas@teluq.ca

Abstract: Monitoring patients at risk of epileptic seizure is critical for optimal treatment and ensuing the reduction of seizure risk and complications. In general, seizure detection is done manually in hospitals and involves time-consuming visual inspection and interpretation by experts of electroencephalography (EEG) recordings. The purpose of this study is to investigate the pertinence of band-limited spectral power and signal complexity in order to discriminate between seizure and seizure-free EEG brain activity. The signal complexity and spectral power are evaluated in five frequency intervals, namely, the delta, theta, alpha, beta, and gamma bands, to be used as EEG signal feature representation. Classification of seizure and seizure-free data was performed by prevalent potent classifiers. Substantial comparative performance evaluation experiments were performed on a large EEG data record of 341 patients in the Temple University Hospital EEG seizure database. Based on statistically validated criteria, results show the efficiency of band-limited spectral power and signal complexity when using random forest and gradient-boosting decision tree classifiers (95% of the area under the curve (AUC) and 91% for both F-measure and accuracy). These results support the use of these automatic classification schemes to assist the practicing neurologist interpret EEG records more accurately and without tedious visual inspection.

Keywords: epileptic seizure; entropy; spectral power; random forest; gradient-boosting decision tree; support vector machine; k-nearest neighbors



Citation: Abou-Abbas, L.; Jemal, I.; Henni, K.; Ouakrim, Y.; Mitiche, A.; Mezghani, N. EEG Oscillatory Power and Complexity for Epileptic Seizure Detection. *Appl. Sci.* **2022**, *12*, 4181. <https://doi.org/10.3390/app12094181>

Academic Editors: Fabio La Foresta and Serena Dattola

Received: 9 March 2022

Accepted: 19 April 2022

Published: 21 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Epilepsy is a neurological disorder characterized by recurrent epileptic seizures. It affects as many as 50 million people of all ages worldwide. Uncontrolled seizures can lead to a disruption of the nervous system and physical risks such as injury and even death. It is generally accepted that better seizure control is essential for better brain health in the long term. Brain activity can be recorded by one of several means, such as electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI). Of these, EEG is the most prevalent because it is simple to effect, low cost, and has high temporal resolution. Its interpretation is commonly used in the study of various brain disorders such as epilepsy, autism, and attention deficit hyperactivity. It is also the most common method in seizure detection and prediction [1–18].

Visual inspection and interpretation of EEG signals are tedious and complex tasks done by experts, which justify research and development of computer-processing methods of EEG that use EEG data representation and machine-learning technics.

Investigations of EEG patterns of seizure have shown that, in general, seizure events have been observed within a wide range of EEG bandwidths. Likewise, their amplitude

and frequency in the delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (13–30 Hz), and gamma (30–70 Hz) bands are significantly different than normal [17,18]. In line with this observation, several studies have used EEG spectral power density (PSD) for seizure detection [3,7,8,19]. PSD is commonly used to evaluate the power of each observed frequency in different neurological brain states. According to previous research, PSDs during seizure events differ significantly from the PSD during rest or sleep states [20], and EEG PSD features have been shown to be able to distinguish between seizure and seizure-free events in the frequency domain [3,7,8,19]. Using the absolute and relative band power of the PSD in 13 frequency sub-bands [21], the authors achieved a 91.36% AUC in a patient-specific study. In another study [22], they found that relative spectral power features led to improved seizure prediction and an average fault rate per hour of 0.1. In a recent study, a comparison of power spectrum analysis and time-frequency analysis was performed. The results revealed that power spectrum features are likely to be seizure markers, and there was a significant difference between the distribution of the power spectrum in seizure segments versus segments in which there were no seizures [23]. Time and frequency domain signal complexity measures have also been used to describe EEG data and serve for the classification of normal and seizure EEG data patterns [2,5,7,11–13,17–19,24]. Other EEG data descriptions and properties have been used as well—for instance, permutation entropy, which decreases significantly in the transition from a seizure-free phase to a seizure phase [16]; Shannon and logarithmic energy entropy, which decrease during epileptic seizures [4]; spectral entropy, which increases for seizure segments [14]; and wavelet entropy [5,9]. Although some studies have considered single complexity features in their proposed approaches [5,16], others have combined several such measures [12,14].

The clinical utility of automated classification methods that involve neuroimaging techniques have become increasingly important in many areas of health care due to their ability to identify atypical neural activity, such as seizure episodes, without visual inspection. Accordingly, research has addressed the design of end-to-end machine learning (ML) systems to detect seizure segments and determine seizure type. ML offers the opportunity to automatically distinguish between patterns in seizure and seizure-free segments. Automatic methods that use ML classification have been investigated for their potential for classification in several healthcare domains, including epilepsy detection, seizure detection, seizure localization, and seizure type identification [25].

There have been several studies of low false alarm detection of epileptic seizure episodes in EEG recordings by ML [4,8,9,14,18]. Along this vein, various classifiers have been used, such as the k-nearest neighbors algorithm (k-NN) [14], support vector machines (SVM) [15,26,27], neural networks (NN) [1,28], and decision trees (DT) [7]. In [27], for neonatal seizure detection, the authors used 55 features with SVM and reported an overall performance of 89% with only one false seizure detection per hour. The system was trained and tested using EEG data collected from 17 newborns with seizures. The study [29] focused on EEG feature selection by relevance and redundancy analysis and used a back propagation neural network to evaluate the effectiveness of feature selection. Results showed an average seizure detection rate of 91% with a false detection rate of 1.17 per hour. Using the Children’s Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) dataset, the study [30] reported an accuracy of 96% with a false-positive rate of 0.08 per hour. Using the same database, which consists of only 23 subjects, a recent study [31] showed an overall accuracy of 91.8% by deploying linear discriminant analysis as a classifier validated on only five subjects. With the European Epilepsy Database, results achieved include an average sensitivity of 91.72% by neural network classification [32], 90.8% by SVM (false-positive rate of 0.094 per hour) [33], and 93.8% by a feed-forward back propagation artificial neural network [34]. These results all relate to patient-specific experimentation and have not been confirmed on across-patient data.

Although there are many studies that focus on seizure detection based on EEG, most of them have some limitations, mainly including experimentation on data from few patients

and using of invasive intracranial EEG records. Furthermore, previous research has used a patient-specific problem formulation and experimental test in which the model is trained and tested on a single patient. Despite attempts to create algorithms to obtain values as accurate as possible, no single one has yet gained widespread acceptance in clinical practice, and the results of studies are far from substituting manual interpretation of the EEG [35,36]. Indeed, despite the numerous commercially available seizure detection devices, there are several issues that must be taken into consideration before such devices can be used in everyday clinical practice. According to a study published in 2021 [37], a comparative study was conducted between 23 devices on the market for seizure detection/alerting. This study showed that these devices rely largely on movement detectors, autonomic change detectors, heart rates, or eye tracking to detect seizures. According to the study's authors, almost no commercially available seizure detection methods use EEG except for Epihunter, a wearable headband device that is connected to a smartphone and is specifically designed for automatic absence seizure detection, which is a specific type of seizure. Thus, the authors concluded that commercially available seizure detection devices do not have the capability of detecting multiple seizure types. Another recent study [38] compared the sensitivity of three well-known commercial seizure detection softwares (Besa, Encevis, and Persyst) to determine whether they correctly detected seizures over long-term video-electroencephalography monitoring (VEM). Based on results from 81 unseen patients, the researchers found a sensitivity of 67.6% for Besa, 77.8% for Encevis, and 81.6% for Persyst on a patient-by-patient basis. They suggested that the false alarm rate needs to be improved. Furthermore, in another study published in 2020 [39], the authors concluded that most commercially available methods focus on using non-electroencephalography EEG signals, which is perceived as a major limitation given the inability of these devices to detect all seizures types. Therefore, it is crucial to develop an automated system for the detection of epileptic seizures based on EEG signals. To develop such an automated system, significant further research and experimentation are necessary so as to develop seizure detection methods that are clinically relevant, high-performing, and statistically validated. Therefore, testing the clinical relevance of state-of-the-art findings through the analysis of a large number of EEG data collated from an accurate representation of clinical situations and collected from multiple sites will be necessary.

The purpose of this study is to investigate a new data representation model that simultaneously exploits signal oscillatory power in frequency bands and signal complexity as EEG. This representation serves prevailing potent ML algorithms for seizure vs. seizure-free classification. We will rigorously evaluate the proposed EEG data representation features on the publicly available large dataset of Temple University Hospital (TUSZ) collected in clinical settings [40].

The main contribution of this paper can be summarized as follows: (1) An analytical framework for seizure and seizure-free EEG classification is proposed and validated on a new large EEG seizure corpus. (2) The feature significance level is investigated by a univariate data analysis. (3) The performance of complexity measures and oscillatory power is analyzed individually and then combined, as EEG signal representation features, to classify normal and seizure data. (4) The results of four supervised methods are compared: RF, GBDT, SVM, and k-NN. (5) The impact of adding a feature selection step on the performance of classifiers is investigated. (6) The contribution of each channel to the performance of classifiers is examined.

2. Materials and Methods

2.1. Dataset

EEG data were drawn from the TUH EEG seizure database (TUSZ), which is a part of Temple University Hospital EEG corpus, the largest open-source EEG corpus. The latter comprises more than 16,986 sessions of EEG recordings collected from 10,874 unique subjects [40,41]. The version of the database used in this work was v1.5.1, released in March 2020. In this study, the dataset consists of EEG signals collected from 341 patients, providing

a substantial amount of data for seizure detection investigation. The EEG segments were labelled by a team of neuroscience and bioengineering students who underwent several months of intensive training to gain annotation skills. They have worked closely with a team of neurologists at Temple Hospital to understand their workflow and their clinical needs while labeling EEG segments. There have been numerous revisions of the corpus during the annotation process (two or three annotators per file system), each time applying more refinement and criteria to enhance the clarity and accuracy of the data. Annotations are created by viewing files with open-source software known as an EDF viewer or by using a customized annotation tool developed by the team, which provides time and frequency domain visualizations. The annotation was made available publicly, except for an evaluation corpus not released to the public, to be allocated for research competition purposes. The dataset used contains 886 sessions that were broken down into 7634 files, of which 1780 are seizure activities of different lengths, in seconds, for a total of 40.40 h. The data were collected in real-time clinical environments, including an intensive care unit, an epilepsy monitoring unit, emergency rooms, cardiac intensive care, and surgical and respiratory intensive care units [41]. Most of the EEG recordings have at least 19 electrodes corresponding to the international standard 10/20 system and range from one second to one hour in duration. The sampling rate varied between 250 Hz and 500 Hz. We resampled all data at 256 Hz. In this study, we discarded seizure events that lasted less than our sliding window of 20 s. Eight types of seizures were present in the TUSZ database: focal non-specific seizure, generalized non-specific seizure, complex partial seizure, simple partial seizure, tonic-clonic seizure, absence seizure, tonic seizure, and myoclonic seizure. Since our goal was not to detect the type of seizure but to determine the presence or absence of seizures, we combined all these types under one label, “seizure.”

The dataset description is summarized in Table 1, and Figure 1 shows an example of seizure and seizure-free EEG epochs. Detailed information about the database annotation can be found in the following reference [42].

Table 1. Overview of the subset of the TUSZ EEG corpus used in this study for seizure detection.

	Number
Total patients (female)	341 (188 F)
Patients with seizure (female)	133 (72 F)
Sessions	886
Files	7634
Seizure files	1780
Seizure-free files	5854
Total duration in hours	655.36

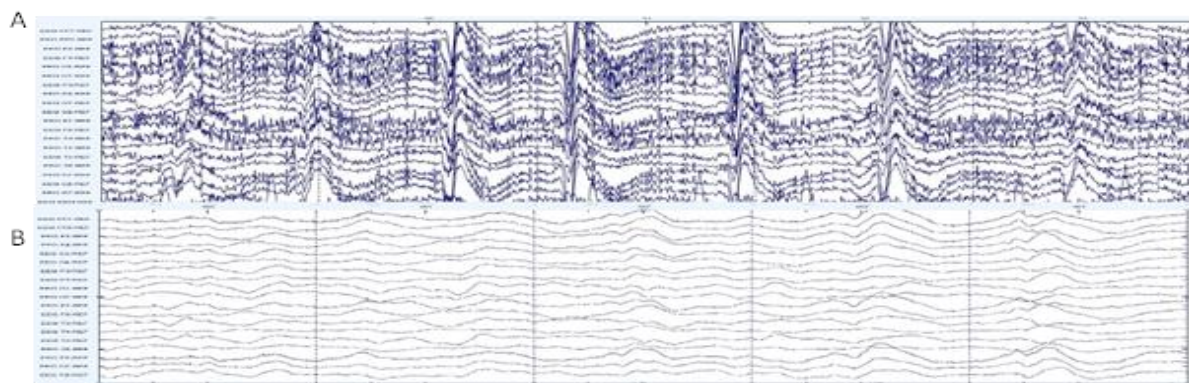


Figure 1. (A) An example of raw EEG including seizure epochs. (B) An example of normal EEG—seizure-free epochs.

2.2. Methodology

An overview of this study's generalized ML-processing strategy is shown in Figure 2. It gives the main steps for data pre-processing, feature extraction, and, finally, classification, which includes training and evaluation. The TUSZ EEG input data consist of pruned EEG recordings, and all uninteresting portions of the data (muscle artifacts, noise, electrode movements, and eye blinks) were purged from the cortical EEG signals. The steps involved in our proposed approach for the classification of epileptic EEG seizures are as follows:

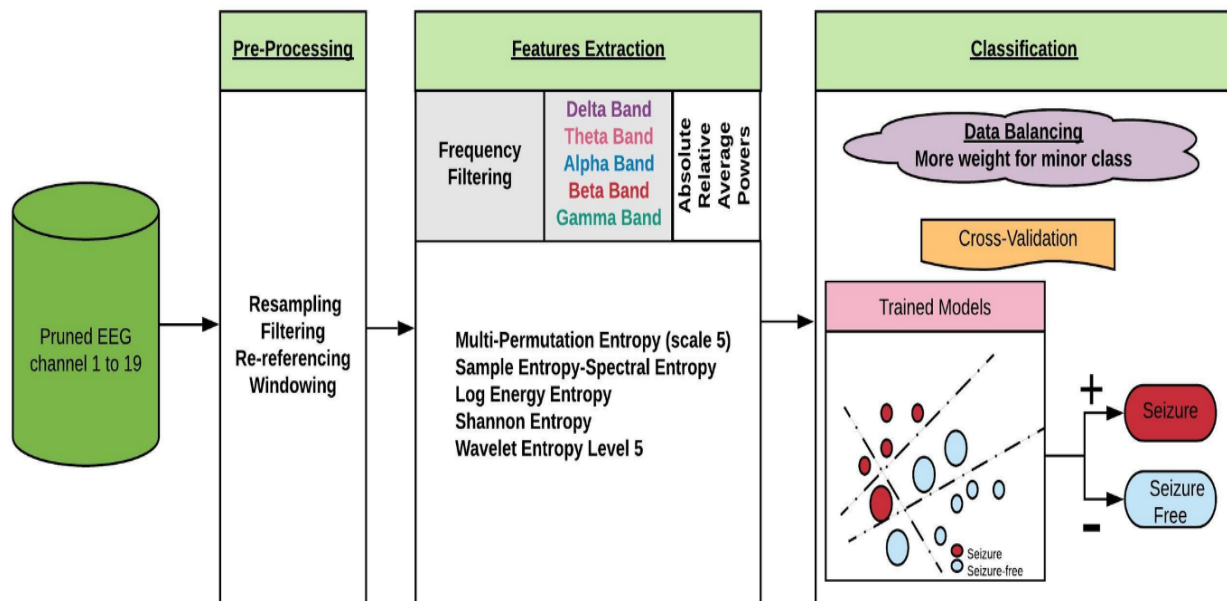


Figure 2. A block diagram of this study's automatic seizure detection method. A total of 19 pruned EEG channels from the TUSZ corpus are resampled, filtered, re-referenced, and segmented. Features are subsequently extracted and used as input for modelling and testing four classifiers separately (RF, GB, k-NN, and SVM).

2.2.1. Pre-Processing

In the staging phase, a 60 Hz infinite impulse response notch filter was applied to attenuate the power line, followed by a bandpass filter with lower and higher cut-off frequencies of 0.5 Hz and 75 Hz, respectively. This was followed by resampling to 256 Hz and re-referencing to the average of all electrodes. From the remaining EEG signal, 20 s fixed-length epochs were segmented.

2.2.2. Feature Extraction

Nine signal complexity measures were derived from the pre-processed EEG: multi-permutation entropy (4 levels), sample entropy, wavelet entropies, logarithmic entropy, Shannon entropy, and spectral entropy. The permutation entropy is a complexity measure with a low computational cost for time series based on comparing neighboring values using the distribution of order patterns [16]. The sample entropy quantifies the regularity in the EEG signal regardless of its length [14]. The discrete wavelet transform is used to compute the wavelet entropies [5], whereas the wavelet packet decomposition is used to compute log energy entropy and Shannon entropy to measure the degree of uncertainty in the signal and to evaluate the dynamical order of the signal [4]. The spectral entropy is calculated using the normalized power spectral distribution of the EEG signal [14]. For the frequency domain analysis, we performed the Welch method, which takes an average of the periodograms obtained using fast Fourier transform (FFT). We calculated the absolute power density and relative power density within each of the five frequency sub-bands—delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (13–30 Hz), and gamma (30–70 Hz)—to increase the feature vector size to 19 per channel. The absolute power of a band can be calculated as the sum of

all of its power values in that frequency range, whereas the relative power (RP) index for each band was determined based on the absolute power in each frequency band, expressed as a percentage of the absolute power (AP) totaled across all frequency bands. Both absolute and relative PSD analysis are essential for achieving accurate brain analysis. Feature values were averaged using a scalar method before serving as input to the classifier. To investigate whether a subset of the extracted features had a greater significance compared to the entire set, we subdivided the features based on their category: frequency domain and information theory (see Table 2). A univariate feature selection method based on the one-way ANOVA test was also applied for comparison to investigate the effect of reducing the vector size. We evaluated the statistical difference of each feature between the two classes: seizure and seizure-free. In order to identify highly significant features, we sorted features based on their *p*-values, which correspond to the probability of the results being observed if the null hypothesis H0 is true. In this work, the null hypothesis H0 implies that there is no difference between the means of two groups (seizure and seizure-free records). In general, the lower the *p*-value, the higher the reliability of results. In medical applications, 0.05, 0.01, or 0.001 are often the recommended threshold values. The features' level of significance was implemented using Python (3.8.0), specifically the Sklearn library.

Table 2. Subset of features in correspondence to their category.

Type	Features	Category
Complexity measures	Sample entropy Logarithmic entropy Wavelet entropy Spectral entropy Shannon entropy Permutation entropy	Information theory
Oscillatory power	Absolute power relative power (delta, theta, alpha, beta, gamma)	Frequency domain

2.2.3. Training and Evaluation

Comparative supervised classification was carried out with four classifiers: random forest (RF), gradient-boosting decision tree (GBDT), support vector machine (SVM), and k-nearest neighbors (k-NN). The four classifiers were chosen based on findings from previous studies, which showed that they had superior classification compared to others [15,26,27,43,44]. The random forest algorithm, an ensemble learning method, involves multiple un-pruned decision tree classifiers. Every tree is an individual classifier, built at the training stage using randomly selected attributes from the original data and at each node to determine the best split. Classification is determined based on majority voting [45]. The gradient-boosting decision tree classifier runs an iterative algorithm to build multiple trees sequentially, where each tree learns and updates its model from the errors of all preceding ones [46]. The support vector machine, an extensively adopted supervised learning classifier, has been widely and successfully used in EEG binary classification problems and specifically in the medical field (automatic seizure detection, mental task classification, epileptic EEG classification, emotion recognition, etc.). The k-nearest neighbors algorithm assigns an unlabeled sample to the most frequently occurring class among the k-nearest labeled training samples. The Euclidean distance is used as the distance metric for k-NN [47].

These proposed ML techniques are based on the most widely used algorithms in previous studies and in various classification problems, as these algorithms have better predictive power and are designed to perform better than linear algorithms, especially for complex non-linearly separable EEG data. Additionally, these techniques have the great advantage of having fewer parameters and hyper-parameters, which makes optimization easier. The model was developed with a view toward being sufficiently transferable to enable the integration in the connected objects for real-time detection.

To avoid overfitting, for each experiment, the dataset was partitioned randomly into training and testing portions, using 10-fold cross-validation repeated 100 times so that one-fold served for testing and the remaining k-1 folds were used for training. The parameter settings of the classifiers in all experiments were as follows: for RF, the number of trees was 100, for GBDT it was 100 as well, for SVM the kernel type was linear, and for k-NN the number of nearest neighbors was 5.

To handle imbalanced data, which may occur due to uneven data representation of classes in the training dataset, with a ratio of seizure to seizure-free of approximately 1:4, we corrected the weights of minority and majority classes according to the distribution of the classes in the entire training set. Furthermore, the weights were inversely proportional to the frequency of classes in the dataset. In a final step, the accuracy and the area under the curve (AUC), an effective metric that combines sensitivity and specificity for classification performance evaluation, were computed for each subset of features and also for the entire set. An F-measure representing the harmonic mean of precision and recall was used as well to study performance. For clarity, results are given per 20 s window size input vector and with respect to the three set of features: complexity, power, and a combination of the two. We compared the results with and without the feature selection step.

3. Results

3.1. Univariate Data Analysis

Each feature was subjected to a univariate data analysis to identify its level of significance with the studied classes, i.e., normal or seizure class, and to investigate its discrimination capability in characterizing seizure records. Tables 3 and 4 show the normalized values of the extracted measures (mean ± standard deviation) for seizure and seizure-free segments. Significant *p*-values (obtained using the ANOVA test) indicate a high discrimination capability in characterizing seizure segments. The values displayed in Tables 3 and 4 correspond to the highest *p*-value per feature. It is evident from Table 3 that seizure-free records had lower complexity values compared to epileptic seizure records. Results in Table 4 show that the relative power in the alpha band had the lowest *p*-value. The results also show that oscillatory power change features had higher levels of importance than complexity measures (Tables 3 and 4), which indicates that oscillatory power change in all channels can contribute to better classification. In addition, Shannon entropy and relative power in the alpha band had lower *p*-values, which indicates that they had higher levels of importance than other features of this category.

Table 3. Range of values of the complexity measures for the two classes of seizure and seizure-free. The lowest *p*-value of 4.130×10^{-98} for multi-permutation entropy at level 2 indicates the great discrimination capability of this feature in characterizing seizure records.

Features	Seizure-Free (Mean ± Std)	Seizure (Mean ± Std)	<i>p</i> -Value
Shannon entropy	0.242 ± 0.125	0.1556 ± 0.095	6.159×10^{-15}
Sample entropy	0.279 ± 0.132	0.199 ± 0.118	1.760×10^{-96}
Spectral entropy	0.288 ± 0.143	0.203 ± 0.123	2.193×10^{-95}
Log energy entropy	0.320 ± 0.152	0.217 ± 0.139	7.184×10^{-11}
Wavelet entropy	0.262 ± 0.125	0.181 ± 0.112	6.657×10^{-11}
Multi-permutation 1	0.267 ± 0.122	0.196 ± 0.117	1.212×10^{-80}
Multi-permutation 2	0.296 ± 0.145	0.207 ± 0.129	4.130×10^{-98}
Multi-permutation 3	0.275 ± 0.130	0.194 ± 0.125	3.050×10^{-90}
Multi-permutation 4	0.291 ± 0.141	0.196 ± 0.123	3.467×10^{-12}

Table 4. Range of values of the spectral power measures for the two classes of seizure and seizure-free. The lowest p -value of 7.173×10^{-145} for the relative power at the alpha band indicates the great discrimination capability of this feature in characterizing seizure records.

Features	Band	Seizure-Free (Mean \pm Std in mV)	Seizure (Mean \pm Std in mV)	p -Value
Absolute power	Delta	0.89 ± 0.02	0.87 ± 0.03	5.694×10^{-124}
	Theta	0.89 ± 0.02	0.87 ± 0.03	1.156×10^{-129}
	Alpha	0.89 ± 0.02	0.87 ± 0.03	7.125×10^{-133}
	Beta	0.90 ± 0.02	0.87 ± 0.04	9.067×10^{-133}
	Gamma	0.90 ± 0.02	0.87 ± 0.04	3.423×10^{-133}
Relative power	Delta	0.88 ± 0.02	0.86 ± 0.03	8.007×10^{-132}
	Theta	0.88 ± 0.02	0.53 ± 0.04	9.594×10^{-128}
	Alpha	0.25 ± 0.16	0.15 ± 0.11	7.173×10^{-145}
	Beta	0.26 ± 0.15	0.16 ± 0.11	1.367×10^{-129}
	Gamma	0.27 ± 0.14	0.19 ± 0.11	1.344×10^{-93}

3.2. Performance of Group of Features Extracted from All Channels without Feature Selection

We evaluated classification performance corresponding to the three sets of features: complexity measures (9 features per channel), oscillatory power (10 features per channel), and a combination of the two (19 features per channel). Table 5 displays the performance of the three sets of features. Classification of seizure versus seizure-free reached an accuracy and F-measure of 90.68% and 91.05%, respectively, using the complexity measures, and 90.95% and 91.33%, respectively, using relative and absolute power in the alpha, beta, gamma, theta, and delta bands. A high AUC value of 95%, an F-measure of 91.41%, and an accuracy of 91.07% were obtained using the RF classifier when the entire set of features was given as classification input. RF and GBDT outperformed SVM and k-NN in all experiments. There was only a slight difference in performance between RF and GBDT.

Table 5. Classification performance for the complexity measures, oscillatory power, and a combination of the two without the univariate feature selection.

Classifiers	RF			GBDT			SVM			K-NN		
	AUC%	F%	ACC%	AUC%	F%	ACC%	AUC%	F%	ACC%	AUC%	F%	ACC%
Complexity measures	95 ± 01	91.05	90.68	93 ± 01	89.86	89.48	79 ± 03	71.44	73.57	85 ± 02	88.92	87.90
Oscillatory power	95 ± 01	91.33	90.95	94 ± 01	90.68	90.40	88 ± 02	79.29	80.65	89 ± 02	89.67	88.73
Complexity and power	95 ± 01	91.41	91.07	94 ± 01	90.95	90.67	86 ± 02	77.24	78.72	88 ± 02	90.09	89.16

3.3. Performance of Group of Features Extracted from All Channels with Feature Selection

For this study, we executed three sets of experiments to investigate the effect of the recommended thresholds. Consistent with previous findings [48,49], we found that reducing the features by the criteria that their p -values be less than 0.001 led to improved performance. All features with a p -value higher than 0.001 were discarded from the feature vector as non-descriptive. Classification performance was evaluated using AUC, F-measure, and accuracy. Table 6 displays the performance of the three sets of features. RF was able to achieve an accuracy of 90.30% and an F-measure of 90.76%, with a decrease of 1%, compared to the experiments done without the feature selection step. There was no significant change in the performance of the three classifiers RF, GBDT, and k-NN between the experiments done with and without univariate feature selection. SVM accuracy increased from 78.72% to 81.92%.

Table 6. Classification performance for the complexity measures, oscillatory power, and a combination of the two with the univariate feature selection.

Classifiers	RF			GBDT			SVM			K-NN		
	AUC%	F%	ACC%	AUC%	F%	ACC%	AUC%	F%	ACC%	AUC%	F%	ACC%
Complexity measures	87 ± 02	88.65	86.00	83 ± 03	87.13	84.52	81 ± 03	73.09	75.07	84 ± 02	88.80	87.47
Oscillatory power	94 ± 01	90.07	89.60	93 ± 01	89.56	89.21	91 ± 01	81.73	82.96	88 ± 02	89.50	88.43
Complexity and power	94 ± 01	90.76	90.30	94 ± 01	89.76	89.43	90 ± 01	80.63	81.92	88 ± 02	90.04	89.07

3.4. Performance of Classifiers with Features Extracted from One Channel at a Time

Table 7 displays the results of RF- and GBDT-based classification using the combined complexity and power features for the 19 EEG channels. A total of 19 features was considered per run. The best AUC, F-measure, and accuracy of 94%, 89.96%, and 89.51% respectively, were obtained with the RF classifier in the Pz channel.

Table 7. Classification performance per channel using complexity measures and oscillatory power.

Channel	RF			GBDT		
	AUC%	F%	ACC%	AUC%	F%	ACC%
FP1	91 ± 01	87.27	86.41	88 ± 01	85.63	84.85
FP2	94 ± 01	89.78	89.38	92 ± 01	88.09	87.70
F3	79 ± 03	88.15	83.86	73 ± 03	87.50	82.78
F4	80 ± 02	87.82	83.39	73 ± 02	87.44	81.89
F7	90 ± 02	88.47	87.02	88 ± 02	87.07	85.55
F8	78 ± 03	87.86	81.70	73 ± 03	86.90	81.00
Fz	80 ± 02	88.05	83.78	73 ± 02	87.76	81.25
C3	92 ± 01	88.66	88.08	90 ± 01	87.15	86.76
C4	88 ± 01	86.43	85.55	87 ± 01	85.05	84.82
Cz	88 ± 01	85.96	85.11	87 ± 01	85.10	84.61
T3	92 ± 01	89.00	88.36	89 ± 01	86.34	85.66
T4	86 ± 02	86.64	84.14	85 ± 02	84.77	83.28
T5	91 ± 01	87.60	86.94	89 ± 01	86.25	85.71
T6	92 ± 01	88.20	87.52	89 ± 01	86.06	85.41
P3	90 ± 01	87.75	86.91	87 ± 01	85.43	84.69
P4	90 ± 01	88.22	87.00	87 ± 01	86.66	85.00
Pz	94 ± 01	89.96	89.51	92 ± 01	88.16	87.72
O1	80 ± 02	88.19	84.33	73 ± 02	87.83	81.63
O2	85 ± 02	88.13	85.25	80 ± 03	86.32	83.27

3.5. ROC Analysis of RF and GBDT Classifiers

The performance of this study's proposed approach was measured using the receiver operating curve (ROC) analysis and AUC metric. A higher AUC designates better performance. Figure 3 displays a comparison between ROC curves when the entire set of features is used. The AUC for each fold is shown in the caption of each figure, and the mean of the AUC was computed. Figure 3 illustrates the ranking performance of 100 iterations of 10-fold cross-validation. The ROC curve summarizes the results and shows that performance was good for all folds, giving an average AUC equal to 95% for RF and 94% for GBDT. The

RF classifier yielded better performance for each of the 10 folds compared to GBDT. It is remarkable for both classifiers that the fluctuations of the AUC were smaller among the different folds, indicating the stability of the feature set.

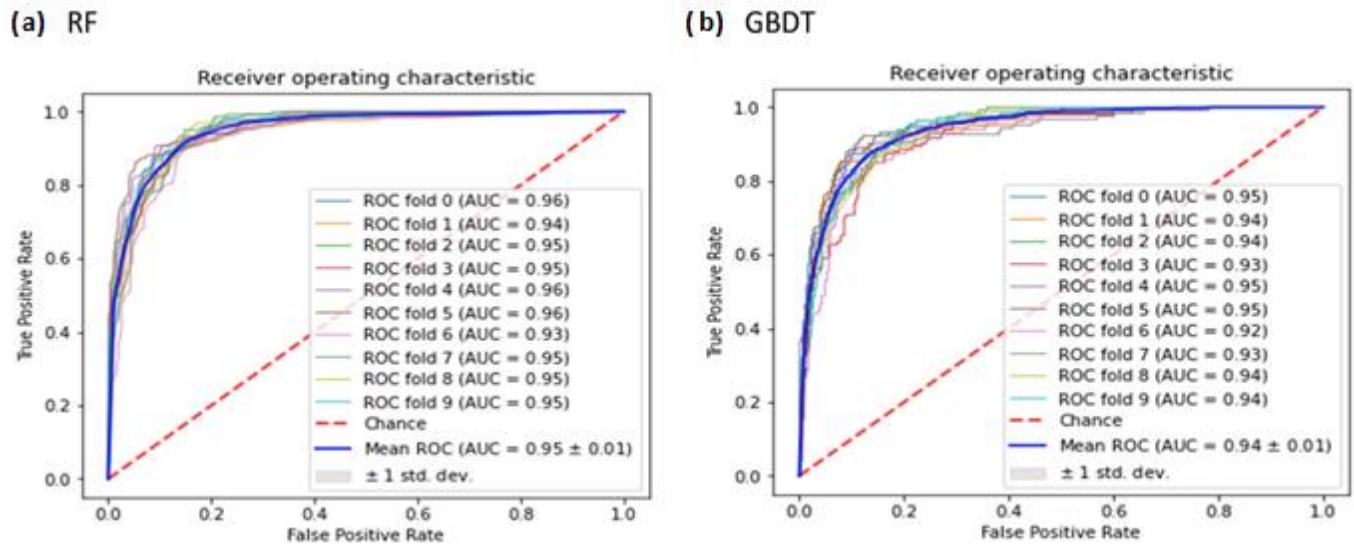


Figure 3. Comparison between ROC curves showing the best results achieved when the entire set of features is used. (a) The left ROC curve corresponds to the RF classifier, and (b) the right curve corresponds to the GBDT classifier. Each curve denotes one-fold of the 10-fold cross-validation, and the area under the curve (AUC) is displayed for each fold in the figure legend. The mean AUC is computed for the 10-fold.

3.6. Comparison of ROC Curves and Accuracies of Classifiers

A comparison between the ROC curves of the four classifiers is displayed in Figure 4. All four classifiers achieved a good AUC. RF produced the highest AUC of 96% and differed slightly from GBDT at 95%. RF outperformed k-NN and SVM in all experiments. The 10-fold cross-validation was performed and the mean and standard deviation of the cross-validation scores over 100 iterations were computed for each classifier. A comparison between accuracies achieved by the classifiers is shown in Figure 5. It is remarkable that the performance of RF classification was superior in both ROC and accuracy. An average accuracy of 90.90% was reached using RF, followed by 89.87%, 89.16%, and 78.72% with GBDT, k-NN, and SVM, respectively.

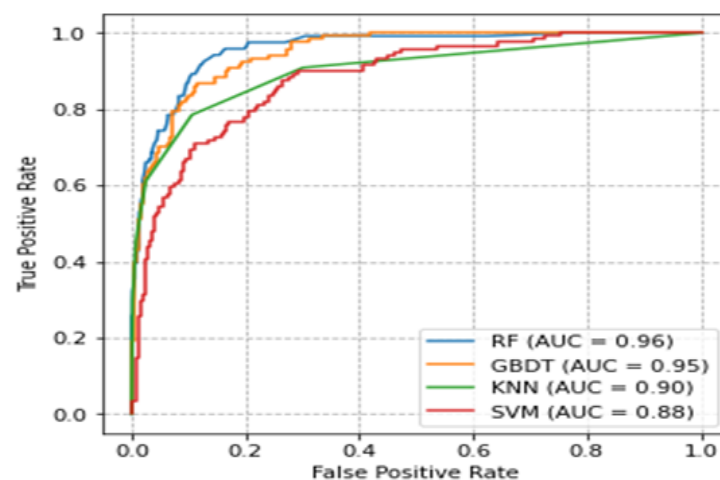


Figure 4. A comparison between ROC curves of the four classifiers: RF, GBDT, KNN and SVM. The AUC of each classifier is displayed.

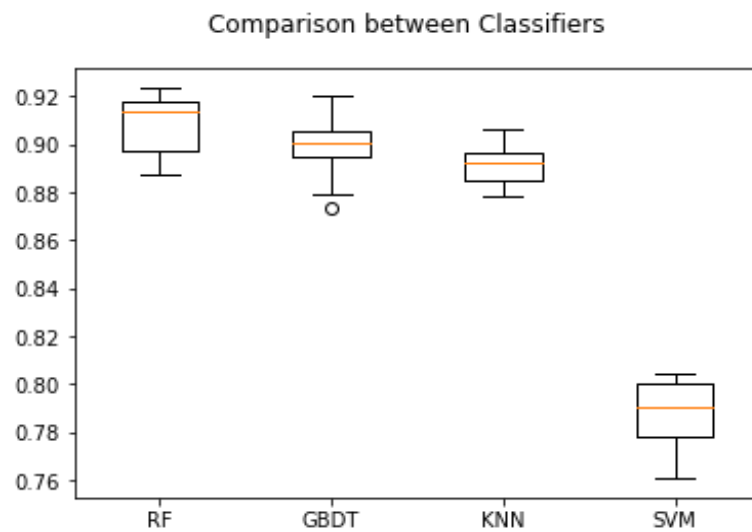


Figure 5. Box plots corresponding to the accuracy of the classification of the two classes of seizure and seizure-free. The orange line in the boxes marks the median. The final accuracy score is determined by averaging the 10 ROC curves and calculating the AUC of the mean ROC.

4. Discussion

This study addressed the problem of seizure vs. seizure-free EEG record classification. The classifier design and validation used data from 341 subjects in the TUSZ database, a larger amount of data than previously used.

Several studies have highlighted the advantage of complexity measures as EEG signal description in seizure detection [2,4,5,9–11,13,15,19,20], and others showed the importance, as well, of signal power in different frequency bands [3,7,8,19]. We inquired into the performance of these features with the larger TUSZ database to gain some understanding of the effect of the experimentation database size on seizure detection potency. We followed this with the study of a new method that exploits both types of features. We investigated different feature-grouping options, followed by machine-learning classification of selected features to identify seizure segments.

Two feature subsets of prevailing features, as well as their combination, were investigated to evaluate their seizure detection potency. In contrast to earlier studies that focused on patient-specific seizure detection, this study addressed patient-independent detection, using data from 341 subjects. Signal complexity and power both yielded good classification, with signal power providing slightly better results. When used jointly, signal complexity and power gave superior performance: 95% AUC with RF classification, 94% with GBDT, 86% with SVM, and 88% with k-NN. RF outperformed other classifiers systematically in all experiments. This may be explained by the fact that decisions made by the RF classifier correspond to unpruned and diverse trees that lead to high resolution in feature space. Random operations in the training and voting procedures of RF also contribute to better classification by addressing the issue of overfitting [50].

This study showed the classification significance of the features at the channel level. Unlike most others, which limited features to one channel (e.g., BONN dataset) or six channels (e.g., Freiburg dataset), the TUSZ database used in validating the method in this study has more than 19 channels corresponding to the 10–20 standard system. This allowed us to investigate the performance of the method using only one channel at each run. Results showed that a maximum distinction between seizure and seizure-free records was obtained from the Pz channel, with a corresponding AUC of 94%. In addition, the method results show that the parietal and frontal-parietal regions were effective in extracting features that can discriminate seizure from seizure-free records, wherein the performance for each channel obtained was above 90%. Additionally, this study addressed the effect of adding a feature selection step prior to classification. The analysis removed the less informative

features using the p -value test: If the p -value was greater than 0.001, the feature extracted from a specified channel was discarded from the group considered. Only a slight decrease in performance occurred for RF and GBDT with feature selection, keeping in mind that the method is computationally more efficient when running on a smaller set of features.

The results are consistent with findings in a previous study [51] that compared random forest variable selection methods for classification. Indeed, the latter concluded that there was practically no difference in prediction error rates between all methods. However, it was obvious that SVM was more sensitive to the number of features, with a decrease in accuracy rate as the number of features increased.

It may be important to note that the good performance of all classifiers is an indirect indication that the feature set is stable in addition to being representative of EEG data for the problem of seizure detection. An analysis of the significance level of a set of entropy properties determined that seizure records have invariably lower signal complexity than seizure-free records. This finding agrees with previous studies that concluded that the EEG signal during seizure is less complex than when seizure free, and therefore, a reduction in signal information content and complexity can be inferred [11,12,14,15]. Additionally, an analysis of the significance level of relative and absolute signal power in frequency bands, typically, delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (13–30 Hz), and gamma (30–70 Hz), indicated that these power features can contribute to higher classification, explaining in part their successful use in [3,7,8,19]. A higher level of importance for oscillatory power-based feature vectors in comparison to complexity measures was apparent and could justify their outperformance in classification.

It is noteworthy to mention that one of the limitations of this study is related to the medical frequency sub-bands that were not considered in this work. This is the case of beta frequency Beta-1, 13–15 Hz; Beta-2, 15–18 Hz; Beta-3, 18–25 Hz; and Hi-Beta, 25–30 Hz. Recent reports [52,53] indicate that additional frequency band analysis is beneficial for epilepsy detection and shows the impact of the frequency sub-bands to the epileptic EEG classification accuracy, and the obtained results revealed several frequency sub-band combinations that achieved high classification accuracy, including the medical frequency sub-bands.

5. Conclusions

This study investigated the use of various ML methods for seizure vs. seizure-free EEG record classification. The classification performance of three sets of extracted features—complexity, oscillatory power, and a combination of the two—was evaluated. Our results showed that abnormalities in a seizure EEG record could be distinguished from a seizure-free EEG record by using its complexity measures and oscillatory power, collected as one feature vector fitted to a random forest classifier. This study's proposed approach could provide a prominent contribution to the development of a fully automated seizure detection system. In a future work, we will expand our study to include pre-ictal and ictal. As there are eight types of seizures present in the TUSZ database, we plan to classify seizures by type, incorporating gender as a predictor, since there are certain differences associated with epilepsy patterns between genders. In addition, we will include a cross-database evaluation to validate the effectiveness of our proposed method and to confirm its generalization ability in seizure classification.

Author Contributions: Methodology, L.A.-A.; software, L.A.-A. and Y.O.; validation, L.A.-A., Y.O. and I.J.; formal analysis, L.A.-A.; writing—original draft preparation, L.A.-A.; writing—review and editing, L.A.-A., I.J., K.H., A.M. and N.M.; supervision, N.M. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fonds de recherche du Quebec—Nature et technologies (L.A.-A.) and the Canada Research Chair on Biomedical Data Mining (950-231214).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used for this study is available as open source from the TUH EEG seizure corpus (TUSZ) v1.5.1. This version of data was released in 20 March 2020. It can be downloaded using the following link: https://www.isip.piconepress.com/projects/tuh_eeg/html/downloads.shtml. Informed consent was obtained from all subjects involved in the study at Temple Hospital. The data has been properly anonymized before being released from Temple Hospital.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alam, S.M.S.; Bhuiyan, M.I.H. Detection of Seizure and Epilepsy Using Higher Order Statistics in the EMD Domain. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 312–318. [[CrossRef](#)] [[PubMed](#)]
2. Martis, R.J.; Acharya, U.R.; Tan, J.H.; Petznick, A.; Yanti, R.; Chua, C.K.; Ng, E.Y.; Tong, L. Application of Empirical Mode Decomposition (Emd) for Automated Detection of Epilepsy using Eeg Signals. *Int. J. Neural Syst.* **2012**, *22*, 1250027. [[CrossRef](#)]
3. Birjandtalab, J.; Heydarzadeh, M.; Nourani, M. Automated EEG-Based Epileptic Seizure Detection Using Deep Neural Networks. In Proceedings of the 2017 IEEE International Conference on Healthcare Informatics (ICHI 2017), Park City, UT, USA, 23–26 August 2017; pp. 552–555.
4. Harlalka, V.; Puntambekar, V.P.; Raviteja, K.; Mahalakshmi, P. Detection of Epileptic Seizure Using Wavelet Analysis based Shannon Entropy, Logarithmic Energy Entropy and Support Vector Machine. *Int. J. Eng. Technol.* **2018**, *7*, 935–939. [[CrossRef](#)]
5. Kumar, Y.; Dewal, M.L.; Anand, R.S. Wavelet entropy based EEG analysis for seizure detection. In Proceedings of the 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC 2013), Solan, India, 26–28 September 2013.
6. Sharma, R.; Pachori, R.B. Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Syst. Appl.* **2015**, *42*, 1106–1117. [[CrossRef](#)]
7. Tzallas, A.T.; Tsipouras, M.G.; Fotiadis, D.I. Epileptic Seizure Detection in EEGs Using Time–Frequency Analysis. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 703–710. [[CrossRef](#)]
8. Tzallas, A.T.; Tsipouras, M.G.; Fotiadis, D.I. Automatic Seizure Detection Based on Time-Frequency Analysis and Artificial Neural Networks. *Comput. Intell. Neurosci.* **2007**, *2007*, 080510. [[CrossRef](#)] [[PubMed](#)]
9. Raghu, S.; Sraam, N.; Kumar, G.P. Classification of epileptic seizures using wavelet packet log energy and norm entropies with recurrent Elman neural network classifier. *Cogn. Neurodynamics* **2017**, *11*, 51–66. [[CrossRef](#)] [[PubMed](#)]
10. Jemal, I.; Mitiche, A.; Mezghani, N. A Study of EEG Feature Complexity in Epileptic Seizure Prediction. *Appl. Sci.* **2021**, *11*, 1579. [[CrossRef](#)]
11. Kannathal, N.; Choo, M.L.; Acharya, U.R.; Sadasivan, P. Entropies for detection of epilepsy in EEG. *Comput. Methods Programs Biomed.* **2005**, *80*, 187–194. [[CrossRef](#)]
12. Acharya, U.R.; Fujita, H.; Sudarshan, V.K.; Bhat, S.; Koh, J.E. Application of entropies for automated diagnosis of epilepsy using EEG signals: A review. *Knowl. Based Syst.* **2015**, *88*, 85–96. [[CrossRef](#)]
13. Orosco, L.; Laciari, E.; Correa, A.G.; Torres, A.; Graffigna, J.P. An epileptic seizures detection algorithm based on the empirical mode decomposition of EEG. In Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009, Minneapolis, MN, USA, 3–6 September 2009; pp. 2651–2654.
14. Acharya, U.R.; Molinari, F.; Sree, S.V.; Chattopadhyay, S.; Ng, K.-H.; Suri, J.S. Automated diagnosis of epileptic EEG using entropies. *Biomed. Signal Process. Control.* **2012**, *7*, 401–408. [[CrossRef](#)]
15. Sharma, R.; Pachori, R.B.; Acharya, U.R. Application of Entropy Measures on Intrinsic Mode Functions for the Automated Identification of Focal Electroencephalogram Signals. *Entropy* **2015**, *17*, 669–691. [[CrossRef](#)]
16. Li, J.; Yan, J.; Liu, X.; Ouyang, G. Using Permutation Entropy to Measure the Changes in EEG Signals during Absence Seizures. *Entropy* **2014**, *16*, 3049–3061. [[CrossRef](#)]
17. Wendling, F.; Bartolomei, F.; Bellanger, J.-J.; Bourien, J.; Chauvel, P. Epileptic fast intracerebral EEG activity: Evidence for spatial decorrelation at seizure onset. *Brain* **2003**, *126*, 1449–1459. [[CrossRef](#)]
18. Shantha Selva Kumari, R.; Prabin Jose, J. Seizure detection in EEG using time frequency analysis and SVM. In Proceedings of the 2011 International Conference on Emerging Trends in Electrical and Computer Technology, ICETECT 2011, Nagercoil, India, 23–24 March 2011; pp. 626–630.
19. Ridouh, A.; Boutana, D.; Bourenane, S. EEG Signals Classification Based on Time Frequency Analysis. *J. Circuits Syst. Comput.* **2017**, *26*, 1750198. [[CrossRef](#)]
20. Myers, M.H.; Jolly, E.; Li, Y.; de Jongh Curry, A.; Parfenova, H. Power Spectral Density Analysis of Electrooculogram Recordings during Cerebral Hypothermia in Neonatal Seizures. *Ann. Neurosci.* **2017**, *24*, 12–19. [[CrossRef](#)] [[PubMed](#)]
21. Zhang, Z.; Parhi, K.K. Seizure detection using regression tree based feature selection and polynomial SVM classification. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 6578–6581.
22. Bandarabadi, M.; Teixeira, C.; Rasekhi, J.; Dourado, A. Epileptic seizure prediction using relative spectral power features. *Clin. Neurophysiol.* **2015**, *126*, 237–248. [[CrossRef](#)]

23. Chen, J.; Zhou, X.; Jin, L.; Lu, Q.; Sun, H.; Liu, Q.; Huang, Y. Can Spectral Power Be Used as a Candidate Seizure Marker of the Periodic Discharges Pattern? *Front. Neurol.* **2021**, *12*, 642669. [[CrossRef](#)]
24. Subha, D.P.; Joseph, P.K.; Acharya, R.; Lim, C.M. EEG Signal Analysis: A Survey. *J. Med. Syst.* **2010**, *34*, 195–212. [[CrossRef](#)]
25. Kotsiantis, S.B. Supervised machine learning: A review of classification techniques. *Informatica* **2007**, *31*, 249–268.
26. Fu, K.; Qu, J.; Chai, Y.; Dong, Y. Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM. *Biomed. Signal Process. Control* **2014**, *13*, 15–22. [[CrossRef](#)]
27. Temko, A.; Thomas, E.; Marnane, W.; Lightbody, G.; Boylan, G. EEG-based neonatal seizure detection with Support Vector Machines. *Clin. Neurophysiol.* **2011**, *122*, 464–473. [[CrossRef](#)]
28. Gómez, C.; Arbeláez, P.; Navarrete, M.; Alvarado-Rojas, C.; Le Van Quyen, M.; Valderrama, M. Automatic seizure detection based on imaged-EEG signals through fully convolutional networks. *Sci. Rep.* **2020**, *10*, 21833. [[CrossRef](#)] [[PubMed](#)]
29. Aarabi, A.; Wallois, F.; Grebe, R. Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis. *Clin. Neurophysiol.* **2006**, *117*, 328–340. [[CrossRef](#)]
30. Shoeb, A.H. Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. Ph.D. Thesis, Mass NSL of Technology, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
31. Khan, Y.U.; Rafiuddin, N.; Farooq, O. Automated seizure detection in scalp EEG using multiple wavelet scales. In Proceedings of the 2012 IEEE International Conference on Signal Processing, Computing and Control (ISPCC 2012), Solan, India, 15–17 March 2012.
32. Yuan, Q.; Zhou, W.; Liu, Y.; Wang, J. Epileptic seizure detection with linear and nonlinear features. *Epilepsy Behav.* **2012**, *24*, 415–421. [[CrossRef](#)] [[PubMed](#)]
33. Williamson, J.R.; Bliss, D.W.; Browne, D.W.; Narayanan, J.T. Seizure prediction using EEG spatiotemporal correlation structure. *Epilepsy Behav.* **2012**, *25*, 230–238. [[CrossRef](#)]
34. Patnaik, L.; Manyam, O.K. Epileptic EEG detection using neural networks and post-classification. *Comput. Methods Programs Biomed.* **2008**, *91*, 100–109. [[CrossRef](#)]
35. González Otárula, K.A.; Mikhaeil-Demo, Y.; Bachman, E.M.; Balaguera, P.; Schuele, S. Automated seizure detection accuracy for ambulatory EEG recordings. *Neurology* **2019**, *92*, e1540–e1546. [[CrossRef](#)]
36. Marín, M.R.; Martínez, I.V.; Bermúdez, G.R.; Porfiri, M. Integrating old and new complexity measures toward automated seizure detection from long-term video EEG recordings. *iScience* **2021**, *24*, 101997. [[CrossRef](#)]
37. Shum, J.; Friedman, D. Commercially available seizure detection devices: A systematic review. *J. Neurol. Sci.* **2021**, *428*, 117611. [[CrossRef](#)]
38. Koren, J.; Hafner, S.; Feigl, M.; Baumgartner, C. Systematic analysis and comparison of commercial seizure-detection software. *Epilepsia* **2021**, *62*, 426–438. [[CrossRef](#)]
39. Bruno, E.; Viana, P.F.; Sperling, M.R.; Richardson, M.P. Seizure detection at home: Do devices on the market match the needs of people living with epilepsy and their caregivers? *Epilepsia* **2020**, *61*, S11–S24. [[CrossRef](#)] [[PubMed](#)]
40. Obeid, I.; Picone, J. The Temple University Hospital EEG Data Corpus. *Front. Neurosci.* **2016**, *10*, 196. [[CrossRef](#)] [[PubMed](#)]
41. Shah, V.; von Weltin, E.; Lopez, S.; McHugh, J.R.; Veloso, L.; Golmohammadi, M.; Obeid, I.; Picone, J. The Temple University Hospital Seizure Detection Corpus. *Front. Neuroinformatics* **2018**, *12*, 83. [[CrossRef](#)]
42. Ochal, D.; Rahman, S.; Ferrell, S.; Elseify, T.; Obeid, I.; Picone, J. The Temple University Hospital EEG Corpus: Annotation Guidelines. *Inst. Signal Inf. Process. Rep.* **2020**, *1*, 1–28.
43. Kursu, M.B. Robustness of Random Forest-based gene selection methods. *BMC Bioinform.* **2014**, *15*, 8. [[CrossRef](#)]
44. Detti, P.; De Lara, G.Z.M.; Bruni, R.; Pranzo, M.; Sarnari, F.; Vatti, G. A Patient-Specific Approach for Short-Term Epileptic Seizures Prediction through the Analysis of EEG Synchronization. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 1494–1504. [[CrossRef](#)]
45. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
46. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
47. Pereira, F.; Mitchell, T.; Botvinick, M. Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage* **2009**, *45*, S199–S209. [[CrossRef](#)]
48. Sharma, R.; Pachori, R.B.; Sircar, P. Seizures classification based on higher order statistics and deep neural network. *Biomed. Signal Process. Control* **2020**, *59*, 101921. [[CrossRef](#)]
49. Tibdewal, M.N.; Dey, H.R.; Mahadevappa, M.; Ray, A.; Malokar, M. Multiple entropies performance measure for detection and localization of multi-channel epileptic EEG. *Biomed. Signal Process. Control* **2017**, *38*, 158–167. [[CrossRef](#)]
50. Biau, G.; Scornet, E. A random forest guided tour. *TEST* **2016**, *25*, 197–227. [[CrossRef](#)]
51. Speiser, J.L.; Miller, M.E.; Tooze, J.; Ip, E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **2019**, *134*, 93–101. [[CrossRef](#)]
52. Tsiouras, M.G. Spectral information of EEG signals with respect to epilepsy classification. *EURASIP J. Adv. Signal Process.* **2019**, *2019*, 10. [[CrossRef](#)]
53. Pérez-Elvira, R.; Ultra-Cucarella, J.; Carrobbles, J.; Teodoru, M.; Bacila, C.; Neamtu, B. Individual Alpha Peak Frequency, an Important Biomarker for Live Z-Score Training Neurofeedback in Adolescents with Learning Disabilities. *Brain Sci.* **2021**, *11*, 167. [[CrossRef](#)]