A review of feature extraction and performance evaluation in epileptic seizure detection using EEG

Poomipat Boonyakitanont [a], Apiwat Lek-uthai [a], Krisnachai Chomtho [b], Jitkomut Songsiri [a,*]

[a] Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand
[b] Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

ARTICLE INFO

ABSTRACT

Since the manual detection of electrographic seizures in continuous electroencephalogram (EEG) monitoring is very time-consuming and requires a trained expert, attempts to develop automatic seizure detection are diverse and ongoing. Machine learning approaches are intensely being applied to this problem due to their ability to classify seizure conditions from a large amount of data, and provide pre-screened results for neurologists. Several features, data transformations, and classifiers have been explored to analyze and classify seizures via EEG signals. In the literature, some jointly-applied features used in the classification may have shared similar contributions, making them redundant in the learning process. Therefore, this paper aims to comprehensively summarize feature descriptions and their interpretations in characterizing epileptic seizures using EEG signals, as well as to review classification performance metrics. To provide meaningful information of feature selection, we conducted an experiment to examine the quality of each feature independently. The Bayesian error and non-parametric probability distribution estimation were employed to determine the significance of the individual features. Moreover, a redundancy analysis using a correlation-based feature selection was applied. The results showed that the following features – variance, energy, nonlinear energy, and Shannon entropy computed on a raw EEG signal, as well as variance, energy, kurtosis, and line length calculated on wavelet coefficients – were able to significantly capture the seizures. When compared with a baseline method of classifying all epochs as normal, an improvement of 4.77–13.51% in the Bayesian error was obtained.

© 2019 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author.
  E-mail addresses: poomipatosk129@gmail.com (P. Boonyakitanont),
apiwat.l@chula.ac.th (A. Lek-uthai), kchomtho@yahoo.com (K. Chomtho),
jitkomut.s@chula.ac.th (J. Songsiri).

1. Introduction

An epileptic seizure, as defined by the International League Against Epilepsy [1], is a temporary event of symptoms due to synchronization of abnormally excessive activities of neurons in the brain. It has been estimated that approximately 65 million people around the world are affected by epilepsy [2]. Nevertheless, it is still a time-consuming process for neurologists to review continuous electroencephalograms (EEGs) to monitor epileptic patients. Therefore, several researchers have developed different techniques that help neurologists to identify an epilepsy occurrence [3–5]. The whole process of automated epileptic seizure analysis primarily consists of data acquisition, signal pre-processing, feature extraction, feature or channel selection, and classification. This paper focuses on a selection of features commonly used in the literature, including statistical parameters (mean, variance, skewness, and kurtosis), amplitude-related parameters (energy, nonlinear energy, line length, maximum and minimum values) and entropy-related measures. These features can be categorized based on their interpretation or the domain from which the features are calculated. While some studies have considered a particular group of features applicable to their proposed classification method [6–8], others have applied various groups of features extracted from the time, frequency, and time-frequency domains. For example, 55 features were used with a support vector machine (SVM) and post-processing for neonatal seizure detection, which provided a good detection rate of 89.2% with one false detection per hour [9]. Feature redundancy and relevance analysis were applied to 132 features to reduce the vector dimension [10]. Fast correlation based-filter proposed in [11], correlation-based selection (CFS) [12], and ReliefF established in [13], were utilized to select non-redundant features and the filtered features were classified by an artificial neural network (ANN). It turned out that 30-optimal selected features by ReliefF achieved the best result with 91% sensitivity and 95% specificity. These studies showed that there is the need to review feature selection as relevant features are directly related to the seizure classification performance.

There have been a number of review articles focusing on automated epileptic seizure detection (AESD). A brief investigation on the AESD methods with some mathematical descriptions was provided in [3]. The EEG analyses were categorized into time domain, frequency domain, time-frequency domain, and nonlinear methods. The authors then discussed the importance of analysis of surrogate data and methods of epilepsy classifications of two (normal and ictal) and three (normal, interictal, and ictal) classes, where the nonlinear features outperformed the others with an accuracy of more than 99%. However, implementation of statistical parameters was not addressed in this paper. Applications of using entropy for epilepsy analysis were reported in [14]. Many types of entropy have been utilized in the AESD, both jointly and independently. All entropies were mathematically stated in combination with their

advantages and disadvantages. The authors also showed differences of each feature in the normal and ictal groups using the *F*-test for checking the equality of variances. The three highest ranked features were Renyi's, sample, and spectral entropies. Moreover, the seizure detection techniques of the whole process, including feature extraction, feature selection, and classification methods, were summarized [15]. Feature extraction methods were divided into two spectral domains, *i.e.*, time-frequency analysis and temporal domain. However, only a brief demonstration of each method was depicted; there was no explanation in the mathematical formulae and the AESD implementation was not described in the paper. Characteristics of focal and non-focal seizure activities observed from EEG signals were also reviewed in [16]. Focal EEG signals and non-focal EEG signals were obtained from the epilepsy-affected and unaffected areas, respectively. Nonlinear features, such as Hjorth parameters, entropy, and fractal dimension, were used to characterize and compare the focal and non-focal EEG signals using the Bern-Barcelona EEG database [17]. Nevertheless, none of these features were expressed mathematically.

In addition, some studies have, in part, involved the AESD application. Studies on channel selection techniques for processing EEG signals and reducing feature dimensions in seizure detection and prediction are needed because considering every channel may cause overfitting problems [18]. The channel selection techniques of filtering, wrapper, embedded, hybrid, and human-based techniques were graphically explained via flowcharts and found to show a high accuracy and low computational cost function. Uses of deep learning, including ANNs, recurrent neural networks (RNNs), and convolutional neural networks (CNNs), in biological signal processing, including EEG, electromyography, and electrocardiography signals, were also reviewed in [19]. The author summarized that there were 24 studies relevant to EEG signal analysis; two of which were focused on AESD and its prediction. It was also reported that applying CNN on the AESD had an 86.9% accuracy. Furthermore, methods for non-stationary signal classification using time-frequency and time-scale representations, and their differences were discussed in depth in the seizure detection application [20]. Several features were extracted after time-frequency and time-scale transformations: spectrogram, Extended Modified B-Distribution, compact kernel distribution (CKD), Wigner-Ville distribution (WVD), scalogram, Affine WVD (AWVD), pseudo-smoothed AWVD, and CKD with feature selection, fed to a random forest (RF). The results showed that every time-frequency and time-scale transformation achieved an accuracy of more than 80%, while CKD with 50 optimal selected features yielded an accuracy of 86.41%.

A review mainly focusing on using the wavelet-based method for computer-aided seizure detection was described in [21]. The authors summarized approaches for seizure detection using different wavelets, *i.e.*, discrete wavelet transform (DWT) and continuous wavelet transform (CWT), with the use of nonlinear-based fea-

tures and chaos-based measurements. The authors showed that the Daubechies wavelet tap 4 was the most favorable wavelet for use in seizure detection. However, there was no explanation about the benefits of each wavelet, no mathematical expression of wavelet transforms, and no comparison of the results between using different wavelet-based techniques. In addition, a review of epileptic seizure prediction processes, including data collection, signal pre-processing, feature extraction, feature selection, classification, and performance evaluation, was reported in [22]. In the feature extraction, the authors categorized features into four groups of linear univariate, linear bivariate, nonlinear univariate, and nonlinear bivariate features. In this case, a nonlinear measure was defined as an attribute related to dynamic states of the brain. Moreover, univariate and bivariate features were extracted from individual and multiple channels, respectively. However, no mathematical description of all the processes was given.

From the past literature, we conclude that selecting distinguished features is required for applying machine learning (ML) approaches to AESD, but previous studies still lack some conclusive points. Firstly, many previous papers have applied a combination of features while providing neither a clear conclusion about which feature has the highest contribution nor an intuitive meaning of using such features [9,23,24]. Secondly, mathematical expressions of some complicated features, such as entropies, and Hurst exponent (HE), were not consistently stated and so may lead to an incorrect implementation [3,15]. Thirdly, most papers focused on a particular group of features that use the classification accuracy as the main indicator to conclude the effectiveness of those features [24–26]. However, we believe that this cannot be a fair indicator for a method comparison. To illustrate this, we consider using a default classifier that regards all signals as normal. This method already yields a high accuracy if the data set in consideration contains mostly normal signals, known as imbalanced data sets. Hence, in our opinion, a performance comparison of all features is required using the same experimental setting based on the Bayes classifier and should be done with a well-adjusted classification indicator. This paper, therefore, aims to explore all those issues by providing a review of the feature extraction process and investigating the contribution of each feature used in the seizure classification in EEG signals. We reported the accuracies and other related metrics (if available) of previous studies on the same EEG databases. Our conclusions related to feature significance are drawn from the experiments using commonly-used features and on a public database reported in the literature.

Literature search. In this review, we mainly gathered articles via Google Scholar to obtain publications from several databases, including ScienceDirect and IEEE Xplore. Studies that were published between January 2010 and December 2018 were primarily considered to reveal the current status of work. The first search aimed to find review articles using the following keywords: review, automated epileptic seizure detection, EEG signal, channel selection, feature selection, deep learning, and feature extraction. In the second search, we primarily used the keywords: feature, nonlinear feature, feature extraction, EEG signal, seizure detection, epilepsy detection, automated epileptic seizure detection, entropy, neural network, and fractal dimension for finding research articles focusing on the feature extraction. In addition, we also included more articles from ScienceDirect and Mendeley recommender systems and references of the included studies. Studies that focused on animals, devices, software, medical treatment, drugs, and surgery were excluded. Only relevant studies that used only EEG signals, applied non-typical features or combinations of widely used features, and had a well-described feature explanation were included. This led to a total of 55 papers, including nine review studies, to be thoroughly reviewed on the feature extraction. The other references were related to techniques used in feature extraction.

This manuscript is organized as follows. Section 2 describes the features extracted from each domain and their computational complexity. Features with detailed mathematical description are categorized according to the domains from which they were calculated: time, frequency, and time-frequency domains. In Section 3, we discuss evaluation metrics and summarize methods for the AESD based on the use of features. Section 4 explains the Bayes method for classification and CFS for assessing feature performance and redundancy, and the results are compared in Section 5.

2. Feature extraction

This section describes the details of features commonly used in the literature of EEG seizure detection, categorized by feature domains: time, frequency, and time-frequency domains. Time-domain features (TDFs) are those calculated on raw EEG signals or on pre-processed signals done in the time domain, such as empirical mode decomposition (EMD). Frequency-domain features (FDFs) are computed on discrete-Fourier transform of raw EEG signals. Time-frequency-domain features (TFDFs) are defined on transformed EEG signals that contain both time and frequency characteristics, such as short-time Fourier transform (STFT) spectrogram or DWT. In the last section, we briefly explain the computational complexity of each feature so that users can take this concern when performing a real-time implementation of AESD. To what follows, we denote $X = [x[0], x[1], \ldots, x[N-1]]$ a sequence of length N used for extracting a feature. For instance, X can be an epoch of a raw EEG segment, absolute values of a raw EEG segment, power spectral density (PSD), approximation or detail coefficients from any wavelet transform, or intrinsic mode functions (IMFs) from EMD.

2.1. Time-domain features (TDFs)

A TDF is calculated from a raw EEG signal or a decomposed signal performed on a time domain. In the literature, one example of signal transformation is EMD that decomposes a signal into IMFs [27] and is commonly applied to signals that apparently exhibit non-stationary properties. In this section, well-known TDFs are briefly described and those involving an entropy concept are explained with mathematical expressions.

1. Groups of statistical parameters have been frequently used to discriminate between ictal and normal patterns, because it is assumed that EEG statistical distributions during a seizure and normal periods are different. These parameters are *mean, variance, mode, median*, *skewness* (third moment describing data asymmetry), and *kurtosis* (fourth moment determining tailedness of the distribution). The *minimum* and *maximum* values are also used to quantify the range of data or the magnitude of signal baseline. Other statistical parameters include *coefficient of variation (CV)* defined as the ratio of the standard deviation (SD) to the sample mean that explains the dispersion of the data in relation to the population mean; *first* (Q_1) *and third quartiles* (Q_3) that quantify the data denseness; and *interquartile range (IQR)* that measures a deviation between the first and third quartiles.

2. *Energy*, *average power*, and *root mean squared value (RMS)* are mutually relevant to amplitude measurements. The energy is a summation of a squared signal, the average power is the signal mean square, and the RMS is the square root of the average power.

3. *Line length*, sometimes called curve length, is the total vertical length of the signal defined as

$$L(X) = \sum_{i=1}^{N-1} |x[i] - x[i-1]|. \tag{1}$$

It was originally presented in [28] as an approximation of Katz's fractal dimension.

4 *Amplitude-integrated EEG* (*aEEG*) is a visual inspection that is widely used for seizure detection in hospitals [29]. The aEEG signal is obtained by calculating the difference between adjacent maximum and minimum values, *i.e.*, the peak-to-peak rectification. Consequently, the rectified signal is passed through the process of interpolation and semi-logarithmic compression proposed in [30]. The interpolated signal is linearly displayed in the range 0–10 μV and semi-logarithmically compressed in the range over 10 μV. This feature is applied based upon the assumption that during seizure events the amplitude of signals are shifted up from its baseline during normal events.

5 *Nonlinear energy* (*NE*), firstly established in [31], extends the concept of energy (quadratic measure) to including indefinite terms of shifted and lagged sequences, defined as

$$NE(X) = \sum_{i=1}^{N-2} \left(x^2[i] - x[i+1]x[i-1] \right). \tag{2}$$

In [31], if a signal has a simple harmonic motion with the amplitude A and the oscillation frequency ω, it can be derived that the NE is proportional to $A^2\omega^2$ when the sampling frequency is high. Hence, high values of NE can indicate both shifted values in a high frequency of oscillation and amplitude.

6 *Shannon entropy* (*ShEn*) [32] reflects the uncertainty in random process or quantities. It is defined as

$$ShEn(X) = -\sum_i p_i \log p_i, \tag{3}$$

where p_i is the probability of an occurrence of each of value in X.

7 *Approximate entropy* (*ApEn*) [33] is a measure of the regularity and fluctuation in a time series derived by comparing the similarity patterns of template vectors. The *template vector* of size m is defined as a windowed signal: $u[i] = \begin{bmatrix} x[i] & x[i+1] & \cdots & x[i+m-1] \end{bmatrix}^T$, and we first consider the self-similarity of the template vector $u[i]$ with a tolerance r, defined by

$$C_i^m(r) = \frac{1}{N-m+1} \sum_{j=0}^{N-m} \Theta \left(r - ||u[i] - u[j]||_\infty \right),$$

where $\Theta(x)$ is the Heaviside step function, *i.e.*, $\Theta(x)$ is one when $x \geq 0$, and zero otherwise. When X is mostly self-similar, then $u[i]$ and $u[j]$ sequences are very close and thus C_i is high. The ApEn aggregates the self-similarity indices over all shifted possibilities of template vectors given template length and tolerance. The ApEn is defined as

$$ApEn(X, m, r) = \frac{1}{N-m+1} \sum_{i=0}^{N-m} \log C_i^m(r) - \frac{1}{N-m} \sum_{i=0}^{N-m-1} \log C_i^{m+1}(r). \tag{4}$$

8 *Sample entropy* (*SampEn*) [34] is based upon a concept similar to the ApEn, where the SampEn compares the total number of template vectors of size m and $m+1$. The SampEn differs from the ApEn in that the self-similarity of *all pairs* of template vectors $u[i]$ and $u[j]$ with a tolerance r is calculated by

$$\phi^m(r) = \sum_{j=0, j \neq i}^{N-m} \sum_{i=0}^{N-m} \Theta \left(r - ||u[i] - u[j]||_\infty \right).$$

If the signals are self-similar, $\phi^m(r)$ is high. The SampEn is defined by

$$SampEn(X, m, r) = \log \phi^m(r) - \log \phi^{m+1}(r). \tag{5}$$

9 *Permutation entropy* (*PE*) [35] is a measure of the local complexity in a signal. With the template vector $u[i]$ and a permutation pattern π_k of order m consisting of $m!$ patterns, the permutation pattern probability for all $k = 1, 2, \ldots, m!$ is defined as the probability that a template vector has the same pattern as the permutation pattern:

$$p(\pi_k) = \frac{1}{N-m+1} \sum_{i=0}^{N-m} f(u[i], \pi_k), \tag{6}$$

where $f(u[i], \pi_k) = 1$ when $u[i]$ and π_k are the same pattern, and zero otherwise. In this case, the pattern is defined by the order of $u[i]$ corresponding to its element. Then, the PE is defined as

$$PE(X) = -\sum_{k=1}^{m!} p(\pi_k) \log p(\pi_k). \tag{7}$$

10 *Weighted-permutation entropy* (*WPE*) [36] is a measure of the process complexity extended from PE by putting the variance of the template vector as a weight (w_i) in computing the probability of each permutation pattern, as calculated by

$$p(\pi_k) = \left(\sum_{i=0}^{N-m} f(u[i], \pi_k) w_i \right) / \left(\sum_{i=0}^{N-m} w_i \right), \tag{8}$$

where $f(u[i], \pi_k) = 1$ when $u[i]$ and π_k are the same pattern, and zero otherwise. Finally, the WPE is calculated by

$$WPE(X) = -\sum_{k=1}^{m!} p(\pi_k) \log p(\pi_k). \tag{9}$$

11 *Fuzzy entropy* (*FuzzEn*) is another measure of the process uncertainty, where its calculation is based on a zero-mean sequence [37]. Consider a windowed signal with mean removed: $u[i] = \begin{bmatrix} x[i] & x[i+1] & \cdots & x[i+m-1] \end{bmatrix}^T - \bar{x}[i]\mathbf{1}$, where $\bar{x}[i] = (1/m)\sum_{j=0}^{m-1} x[i+j]$, and $\mathbf{1}$ is a column vector of ones. Like other entropy measures, the sequence u is used to compute a self-similarity index, ϕ, but for FuzzEn, a Gaussian function is used as defined by

$$\phi^{(m)}(r) = \frac{1}{(N-m)(N-m-1)} \sum_{j=0, j \neq i}^{N-m} \sum_{i=0}^{N-m} e^{d_{ij}^2/2r^2},$$

where $d_{ij} = ||u[i] - u[j]||_\infty$. If X is self-similar, then d_{ij} is small and relative to the Gaussian curve with variance r^2, and so we can determine how high $\phi^{(m)}(r)$ is. The FuzzEn is then defined to reflect the change of ϕ as the window length m is increased to $m+1$,

$$FuzzEn(X, m, r) = \log \phi^m(r) - \log \phi^{m+1}(r). \tag{10}$$

12 *Distribution entropy* (*DistEn*) measures the uncertainty of a process by estimating the probability distribution with a histogram [38]. With the zero-mean vector defined previously in FuzzEn, the distance between two template vectors (windowed signals) is $d_{ij} = ||u[i] - u[j]||_\infty$ and is used to construct a histogram with B

bins. The DistEn is then defined to be the entropy of the random process d_{ij}:

$$\text{DistEn}(X, m, B) = -\frac{1}{\log B}\sum_{i=1}^{B} p_i \log p_i, \tag{11}$$

where p_i is the relative frequency of each bin. A high DistEn value reflects unpredictable natures of the original sequence X.

13 *Singular Value Decomposition (SVD) entropy (SVDEn)*, first established in a brain-computer interface application, was proposed to measure the temporal and spatial complexity of a process based on entropy [39]. Any rectangular matrix A has an SVD of $A = U\Sigma V^T$, where $\Sigma = \mathbf{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m)$ containing the singular values of A. The number of non-zero singular values of A determines its rank. In feature extraction, the normalized singular value is given by $\tilde{\sigma}_j = \sigma_j/\sum_i^m \sigma_i$ and the SVDEn is defined as the entropy form of these normalized singular values:

$$\text{SVDEn}(X) = -\sum_j \tilde{\sigma}_j \log \tilde{\sigma}_j. \tag{12}$$

For the temporal measurement, a delay method of time τ is used to construct each row of the matrix A. For the spatial SVDEn, each row of the matrix A is a sequence of the EEG signal from each channel. It is claimed that when the EEG signals are noisy, potentially all singular values are non-zero (A is full rank). More importantly, it is observed that singular values associated with noisy signals are significantly smaller than the deterministic (noiseless) signal. Therefore, a low SVDEn value spatially and temporally indicates patterns of epileptic seizure in EEG signals because the singular values from the EEG signals from a seizure are higher than those of the normal signals.

14 *Hurst exponent (HE)* [40] indicates a degree of time series tendency. We call a *partial time series* of X as a chopped signal having length m, such that $m < N$, e.g. $m = N/2$. A *cumulative deviate series*, or so-called *partial cumulative time series*, is defined as

$$z[t] = \sum_{j=0}^{t}(x[j] - \bar{x}), \tag{13}$$

where $\bar{x} = (1/m)\sum_{i=0}^{m-1} x[i]$ is the average of the partial time series. With these auxiliary signals, we consider $R(m)$ as a cumulative range: $R(m) = \max_t z[t] - \min_t z[t]$, and $S(m)$ as the SD of the partial time series. Then, the HE is defined as the slope of the straight line that fits $\log\left(R(m)/S(m)\right)$ as a function of $\log m$ by the least-square method.

$$\log\left(R(m)/S(m)\right) = \text{HE} \cdot \log m.$$

With a fixed m, we observe how the signal is fluctuated from its mean via z and the range of this fluctuation is explained in R, relative to the SD (normalized by S). Thus, the HE tends to capture how this range grows as m increases and approximates it as a linear growth in a log-scale.

15 *Fractal dimension* is a mathematical index for measuring signal complexity [41]. The Higuchi dimension (D_H) was proposed to calculate the fractal dimension for a time series in the time domain [42]. A partial length, denoted as $L_m(\tau)$, is calculated from a partial time series with a time interval τ and an initial time $m = 0, 1, 2, \ldots, \tau$ as

$$L_m(\tau) = \frac{1}{\tau}\left(\sum_{i=1}^{\lfloor\frac{N-m}{\tau}\rfloor} |x[m+i\tau] - x[m+(i-1)\tau]|\right)\frac{N-1}{\lfloor\frac{N-m}{\tau}\rfloor\tau}. \tag{14}$$

The slope of the straight-line fitting $\log L(\tau)$ as a function of $\log \tau$ is equal to $-D_H(X)$, where $L(\tau)$ is the average of $L_m(\tau)$ over τ. Moreover, there is another approach, the Minkowski dimension or box-counting dimension (D_B), that measures the complexity of images and signals by counting square boxes that cover the object [41]. Here, $D_B(X)$ is defined as

$$D_B(X) = \lim_{\varepsilon\to 0}\frac{\log N(\varepsilon)}{\log(1/\varepsilon)}, \tag{15}$$

where $N(\varepsilon)$ is the number of boxes of size ε of each side required to cover the signal.

16 *Local Binary Pattern (LBP)*, *Local Neighbor Descriptive Pattern (LNDP)*, and *Local Gradient Pattern (LGP)* are encoding techniques that transform a partial time series into encoded patterns [5,43]. For $i = 0, 1, \ldots, N - m - 1$ when m is a positive even integer, the LBP of the partial time series is defined as the difference between the partial time series and its center:

$$\text{LBP}_i(X) = \sum_{j=0}^{m/2-1} \Theta\left(x[i+j] - x[i+m/2]\right) 2^j$$
$$+ \sum_{j=m/2}^{m-1} \Theta\left(x[i+j+1] - x[i+m/2]\right) 2^j, \tag{16}$$

where $\Theta(x)$ is the Heaviside step function. The LNDP captures the relationship of signals from the neighbor sequences and is defined by

$$\text{LNDP}_i(X) = \sum_{j=0}^{m-1} \Theta(x[i+j] - x[i+j+1])2^j. \tag{17}$$

The LGP represents the changes in a signal as measured by the difference between signal gradients and the average gradient, and is defined by

$$\text{LGP}_i(X) = \sum_{j=0}^{m/2-1} \Theta\left(|x[i+j] - x[i+m/2]|\right.$$
$$\left. - \frac{1}{m}\sum_{k=0}^{m} |x[i+k] - x[i]|\right) 2^j$$
$$+ \sum_{j=m/2}^{m-1} \Theta\left(|x[i+j+1] - x[i+m/2]|\right.$$
$$\left. - \frac{1}{m}\sum_{k=0}^{m} |x[i+k] - x[i]|\right) 2^j. \tag{18}$$

These three parameters are pattern indicators and their histograms can form into features. The main advantage of these parameters is their invariant properties. The LBP is globally invariant, while the LNDP and LGP are not affected by local variation in the signal. From the algorithms, it was observed that histograms from normal and seizure EEG signals are different; the histograms of LBP, LNDP, and LGP from the abnormal signals have a high frequency at some specific values.

17 Three *Hjorth parameters* (*activity, mobility*, and *complexity*) were established to characterize the spectral properties of EEG signals in the time domain [44]. The activity is the variance of a signal and is high during seizure events in terms of a high amplitude variation from its mean. The mobility is a measure of a quantity related to the SD of the signal PSD. The mobility is expressed as the ratio of the SD of a signal derivative to the SD of the signal.

Complexity represents the difference between the signal and a pure sine wave. The complexity is defined by the ratio of the mobility of a signal derivative to the mobility of the signal.

18 Two *bandwidths* – *amplitude modulation bandwidth* (B_{AM}) and *frequency modulation bandwidth* (B_{FM}) – calculated from the analytic signal of IMF are measures of a frequency spread in terms of changes in amplitude and average frequency [45]. Let $z(t)$ be the analytic signal of IMF $y(t)$ defined as

$$z(t) = y(t) + jy_H(t) = A(t)e^{j\phi(t)}, \tag{19}$$

where $y_H(t) = y(t) * 1/(\pi t)$ is the Hilbert transform of $c(t)$, $A(t)$ denotes the amplitude of the analytic signal, and $\phi(t)$ represents the instantaneous phase [46]. Calculations of B_{AM} and B_{FM} are as follows:

$$B_{AM} = \frac{1}{E}\int\left(\frac{dA(t)}{dt}\right)^2 dt, \quad B_{FM} = \frac{1}{E}\int\left(\frac{d\phi(t)}{dt} - \bar{\omega}\right)^2 A^2(t)dt, \tag{20}$$

where E is the analytic signal energy, and $\bar{\omega} = \frac{1}{E}\int\frac{d\phi(t)}{dt}A^2(t)dt$ is the center frequency.

19 *Detrended fluctuation analysis* (*DFA*) is a method of examining the self-correlation of a signal, which is similar to the HE. Its calculation is conducted by the following steps [47]. A *partial cumulative time series*, $z[t]$, as defined in (13) is used to determine a trend of the signal. Subsequently, $z[t]$ is then segmented into a signal of length n. A local trend of the segment is estimated by a least-square straight line $\hat{z}[t]$. The mean-squared residual $F[n]$ is then obtained from

$$F[n] = \sqrt{\frac{1}{N}\sum_{t=0}^{N-1}\left(z[t] - \hat{z}[t]\right)^2}. \tag{21}$$

DFA(X) is defined as the slope of the straight-line fitting $\log F[n]$ as a function of $\log n$ in a regression setting. The result indicates the fluctuation of the signal when its local trend is removed.

20 *Number of zero-crossings* is an indirect measurement of the frequency characteristics of a signal. If this number is large, it means the signal contains high frequency components and more uncertainty.

21 *Number of local extrema* is the total number of local maxima and minima in a signal. It is similar to the number of zero-crossings that indirectly represents the frequency measurement of the signal.

2.2. Frequency-domain features (FDFs)

Frequency domain analysis is also crucial, since a frequency representation of an EEG signal provides some useful information about patterns in the signal. The PSD and the normalized PSD (by the total power) are mostly used to extract features that represent the power partition at each frequency. This section describes features that are extracted from the PSD.

1 The *energy* is extracted from some specified frequency ranges, usually corresponding to normal EEG activities, to determine the EEG rhythmicity in each frequency range.

2 *Intensity weighted mean frequency* (*IWMF*), also known as the mean frequency, gives the mean of the frequency distribution using the normalized PSD, and is defined as

$$IWMF(X) = \sum_k x[k]f[k], \tag{22}$$

where $x[k]$ is the normalized PSD of an EEG epoch at the frequency $f[k]$.

3 *Intensity weighted bandwidth* (*IWBW*), or the SD frequency, is a measure of the PSD width in terms of the SD, and is defined as

$$IWBW(X) = \sqrt{\sum_k x[k](f[k] - IMWF(X))^2}, \tag{23}$$

where $x[k]$ is the normalized PSD. According to the seizure patterns in typical EEG signals, the PSD is sharper during seizure activities. Therefore, the IWBW is smaller during those activities.

4 *Spectral edge frequency* (*SEF*), denoted as SEFα, is the frequency at which the total power of the normalized PSD, $x[k]$, is equal to α percent. We define SEFα as the frequency $f[k_{SEF}]$ such that $\sum_{k=0}^{k_{SEF}} x[k] = 0.01\alpha$. By this definition, the *median frequency* is a related feature, defined as the frequency at which the total power is equally separated and is, therefore, equal to the SEF50.

5 *Spectral entropy* (SE) is a measure of the random process uncertainty from the frequency distribution. A low SE value means the frequency distribution is intense in some frequency bands. Its calculation is similar to that for the ShEn but replaces the probability distribution with the normalized PSD as follows:

$$SE(X) = -\sum_k x[k]\log x[k]. \tag{24}$$

6 *Peak frequency*, also called dominant frequency, is the frequency at which the PSD of the highest average power in its full-width-half-max (FWHM) band has the highest magnitude. Since the peak frequencies of normal and seizure are located differently, it can be used to differentiate the two events.

7 *Bandwidth* of the dominant frequency is defined as the FWHM band corresponding to the peak frequency. Regarding the rhythmicity of an EEG signal, it can be implied that, during seizure activities, there are frequency ranges that are more intense. Hence, the bandwidth is supposedly small during the seizure activities.

8 *Power ratio* of the current and background epochs at the same frequency range is determined to compare the powers. The power of a seizure period is relatively higher than that of the background.

2.3. Time-frequency-domain features (TFDFs)

Since an seizure occurrence can be captured by both TDFs and FDFs, many transformation and decomposition techniques that give information in terms of time and frequency have been widely considered, *e.g.*, STFT spectrogram and DWT analyses. In this section, all the features are computed from the results of decomposition techniques. For example, when the DWT was used, features were calculated from the decomposition coefficients of some specific levels.

1 Statistical parameters are also common in the time-frequency domain. Like TDFs, these attributes are computed to distinguish statistical characteristics of normal and seizure EEG signals in each specified frequency sub-band. Features applied in previous studies included the *mean*, *absolute mean*, *variance*, *SD*, *skewness*, *kurtosis*, *absolute median*, and *minimum and maximum values*.

2 The *energy*, *average power*, and *RMS* are used to observe the amplitudes of signals corresponding to specific frequency sub-bands.

3 The *line length* is used to estimate the fractal dimension of the DWT decomposition coefficients.

4 The *ShEn, ApEn, PE*, and *WPE*, entropy-based indicators, determine the uncertainties and complexities of decomposed signals.

5 *Weighted multi-scale Rényi permutation entropy* (*WMRPE*) is an extension of multi-scale Rényi permutation entropy measuring the time-series complexity [48]. Considering a

coarse-grained signal of $x[n]$: $v[i] = (1/l)\sum_{j=(i-1)l+1}^{il} x[j]$, a multi-scale template vector of length m is denoted as $u[i] = \begin{bmatrix} v[i] & v[i+\tau] & \cdots & v[i+(m-1)\tau] \end{bmatrix}^T$, where τ is a time delay. The probability $p(\pi_k)$ of each permutation pattern π_k is then defined as in (8) and the WMRPE is finally calculated as

$$\text{WMRPE}(X) = \frac{1}{1-q} \log \left(\sum_{k=1}^{m!} p(\pi_k) \right)^q, \tag{25}$$

when $q \neq 1$. If $q = 1$, the entropy eventually changes to *weighted multi-scale permutation entropy* (*WMPE*) as computed in (9).

6 Additionally, *HE* and D_H are applied to demonstrate randomness in a time series corresponding to the DWT coefficients of each sub-band.

7 *Autoregressive* (*AR*) and *autoregressive-moving-average* (*ARMA*) *model parameters* of a time series model are fitted to DWT coefficient sequences. It is assumed that the time-series models explaining the dynamics of ictal and normal EEG patterns are different. Time-series model parameters are then estimated for each sub-band where stationarity assumption holds. However, it is claimed that ARMA parameters were significant features only in some sub-bands [49].

2.4. Complexity of feature extraction

Suppose a raw EEG time series is split into epochs; each of which contains N time samples that are extracted into a feature. Based on mathematical expressions of each feature, we describe its computational complexity as a function of N samples. Firstly, simple features that require only $\mathcal{O}(N)$ are standard statistical parameters, except those which need a prior sorting process, including mode, median, Q_1, Q_3, and IQR, which require $\mathcal{O}(N \log N)$, energy, NE, RMS, line length, Hjorth parameters, number of zero-crossing, and number of local extrema. The HE, DFA, and fractal dimension features involve calculating a partial time series and solving a regression with one variable, so the first task is the more dominant. The complexity of these features is still linear in N, $\mathcal{O}(N)$. Similarly, the main calculation of LBP, LNDP, and LGP are from computing partial time series of length m. Since the computational complexity of computing each partial time series is linear in m and there exists $N - m$ partial time series where $m \ll N$, then the total complexity for calculation of these features is $\mathcal{O}(N)$, or $\mathcal{O}(Nm)$ (to be more precise).

Complexities of entropy-related features differ in the methods of approximating the probability distribution (a bubble sort algorithm requires $\mathcal{O}(N^2)$, while a quicksort takes $\mathcal{O}(N \log N)$ [50]). For exmaple, the ShEn needs $\mathcal{O}(N \log N)$. Also, the ApEn, FuzzyEn, DistEn, and SampEn normally require complexities of $\mathcal{O}(N^2)$ because there are approximately $N(N-1)/2$ similarity comparisons between two template vectors [34,51]. Additionally, some complexities of entropy calculation can be reduced to $\mathcal{O}(N \log N)$ using a sorting algorithm before computing the self-similarity [52,53]. Moreover, the PE and WPE require comparison of $m!$ permutation patterns and a template vector of length m. So there exist $(m!)(N - m + 1)$ comparisons, making the computational complexity $\mathcal{O}((m!)N)$. On the other hand, the SVDEn needs SVD computation at the beginning followed by the entropy calculation. The complexity of calculating SVD is $\mathcal{O}(m^2N + m^3)$, where m is the number of the singular values and $m < N$ and the entropy calculation requires $\mathcal{O}(m)$. Therefore, the complexity of SVDEn is dominated by the SVD computation of $\mathcal{O}(m^2N + m^3)$.

Computing FDFs firstly requires the Fourier transform that needs $\mathcal{O}(N \log N)$ for the fast-Fourier transform (FFT). Subsequent processes in calculating energy, IWMF, IWBW, SEF are just linear in N, so the complexity of the FDFs is $\mathcal{O}(N \log N)$. Computational complexities of TFDFs can be dominated by two processes: signal decomposition and feature extraction, depending on which step has a higher complexity. For example, the DWT has the complexity of $\mathcal{O}(N)$, while the STFT has the complexity of $\mathcal{O}(N \log m)$, where m is a size of the moving window. If the energy feature is calculated from STFT, then the overall complexity is $\mathcal{O}(N \log m)$. On the other hand, the total complexity is $\mathcal{O}(N \log N)$ when the ApEn is calculated from DWT coefficients.

3. Epileptic seizure detection

3.1. Performance metrics

The method to assess AESD is critical for comparing detection algorithms. As a binary classification, seizure detection results from a classifier are comprised of true positive (TP), false positive (FP), false negative (FN), and true negative (TN). These quantities can be formed into a true positive rate (TPR) and a false positive rate (FPR), or as sensitivity (Sen), specificity (Spec) and accuracy (Acc) [54]. The accuracy is the overall classification performance, where the sensitivity and specificity are regarded as how well the algorithm detects seizure and normal outcomes, respectively. Currently, there are two sets of evaluation metrics in the AESD: epoch-based and event-based metrics [55].

Epoch-based metrics are indicators used to evaluate the detection performance when each epoch segmented from a long EEG signal is supposed to be a separate data sample. The epoch-based metrics are calculated from the numbers of TP, FP, FN, and TN evaluated on all samples. For example, accuracy, sensitivity, and specificity have been reported in many studies [3,7,10]. Nevertheless, these epoch-based metrics can lead to wrong diagnosis when seizure events are incorrectly recognized. For instance, an algorithm wrongly classifying one totally short seizure event as normal still achieves a high epoch-based performance since there are many other epochs that are correctly detected.

Event-based metrics, on the other hand, are used to evaluate a classifier based on seizure events in EEG signals. Two common metrics, good detection rate (GDR) and FPR per hour (FPR/h), are calculated based on the intersection of detection results and annotations [56–58]. Here, GDR is defined as the percentage of detected seizure events that have an overlap with the annotations and FPR/h is the proportion of events declared as a seizure without any intersection with the annotations in one hour. A higher GDR indicates a higher number of correctly detected seizure events, while a small FPR/h refers to having a lower number of wrongly recognized seizure events. However, care is required with these high event-based metrics to avoid being misled into a conclusion of a correct detection when a duration is considered. For example, declaring an occurrence of seizure at the last second of an actual seizure event is still counted as good detection even though the detection system nearly misses the whole seizure event.

3.2. Selected features used in epileptic seizure detection

From the lists of common features described in Section 2, this section summarizes how those features were applied to AESD in the literature. Performances of the detection algorithms generally depend on both feature selection and the classifier, so we report the classification method, features and categories by domains, and the performance results in Tables 1 and 2 for studies using the CHB-MIT Scalp EEG database [59] and in Tables 3 and 4 for those using the Bonn University database [60]. The description of the CHB-MIT Scalp EEG database is explained in Section 5. For the Bonn University database, there are five sets of data, each of which contains 100 single-channel EEG samples. Only one data set consisted of seizure activities recorded intracranially from epileptic patients,

Table 1
Summary of automated epileptic seizure detection using the *CHB-MIT* Scalp EEG database when *single-domain* features were used.

Domain	Features	Method	Performance	Ref.
Time	Raw signal	ANN	Acc = 100%	[26]
	aEEG	Thresholding	Sen = 88.50%, FPR/h = 0.18	[58]
	Line length, NE, variance, average power, max	RBF SVM	Acc = 95.17%, Sen = 66.35%, Spec = 96.91%	[61]
	Absolute mean values, average power, SD, ratio of absolute mean values, skewness, kurtosis	MSPCA + EMD + RF	Acc = 96.90%	[23]
		MSPCA + EMD + SVM	Acc = 97.50%	[23]
		MSPCA + EMD + ANN	Acc = 96.90%	[23]
		MSPCA + EMD + k-NN	Acc = 94.90%	[23]
	D_B	RVM	Sen = 97.00%, FPR/h = 0.24	[56]
Frequency	Energy	RBF SVM	Sen = 96.00%, FPR/h = 0.08	[57]*
Time-frequency	Spectrogram	STFT + SSDA	Acc = 93.82%	[62]
	Mean, ratio of variance, SD, skewness, kurtosis, mean frequency, peak frequency	DWT + ELM	Acc = 94.83%	[24]
	Log of variance	DWT + thresholding	Acc = 93.24%, Sen = 83.34%, Spec = 93.53%	[63]
		DWT + RBF SVM	Acc = 96.87%, Sen = 72.99%, Spec = 98.13%	[61]
	Absolute mean, average power, SD, ratio of absolute mean, skewness, kurtosis	MSPCA[1] + DWT + RF	Acc = 100%	[23]
		MSPCA + DWT + SVM	Acc = 100%	[23]
		MSPCA + DWT + ANN	Acc = 100%	[23]
		MSPCA + DWT + k-NN	Acc = 100%	[23]
		MSPCA + WPD[2] + RF	Acc = 100%	[23]
		MSPCA + WPD + SVM	Acc = 100%	[23]
		MSPCA + WPD + ANN	Acc = 100%	[23]
		MSPCA + WPD + k-NN	Acc = 100%	[23]
	Energy	HWPT[3] + RVM	Sen = 97.00%, FPR/h = 0.25	[56]
	Energy, D_B	HWPT + RVM	Sen = 97.00%, FPR/h = 0.10	[56]

[1] Multi-scale principal component analysis, [2] wavelet packet decomposition, [3] harmonic wavelet packet transform. * Use all data records.

Table 2
Summary of automated epileptic seizure detection using the *CHB-MIT* Scalp EEG database when *multi-domain* features were used.

Time	Frequency	Time-frequency	Method	Performance	Ref.
Variance, RMS, skewness, kurtosis, SampEn	Peak frequency, median frequency		LDA	Sen = 70.00%, Spec = 83.00%	[64]
			QDA	Sen = 65.00%, Spec = 92.00%	[64]
			Polynomial classifier	Sen = 70.00%, Spec = 83.00%	[64]
			Logistic regression	Sen = 79.00%, Spec = 86.00%	[64]
			k-NN	Sen = 84.00%, Spec = 85.00%	[64]
			DT	Sen = 78.00%, Spec = 80.00%	[64]
			Parzen classifier	Sen = 61.00%, Spec = 86.00%	[64]
			SVM	Sen = 79.00%, Spec = 86.00%	[64]

whereas the others were regarded as non-epileptic EEG signals. More data description is available in [60]. We have separated these summaries according to the databases so that a fair comparison of performance results can be concluded. However, even though these studies tested the methods on the same database, the selection of data records in each paper can be varied or this information was not available. In this case, the interpretation of performance figures should be performed with care, and we have put a footnote to such publications. The methods are described as a series of transformations or pre-processing steps (if performed) followed by classifiers.

3.2.1. Time-domain features (TDFs)

Previous studies using TDFs in both databases are depicted in Tables 1 and 3. The most obvious approach was to use the raw EEG signals as inputs but a deep learning process to extract the patterns. Using the raw signal with ANN [26] and one-dimensional CNN following the *z*-score normalization process [65] provided promising results. Similarly, proposed studies of CNN-based classification using raw signals and IMFs from EMD achieved nearly perfect classification [66]. However, data used in [26] were only chosen from female patients, where the records contained epileptic activities in the frontal lobe and the activities were only simple and complex partial epilepsy. A thresholding technique using aEEG was used in [58], where the threshold could be simultaneously updated. The proposed method effectively detected a high-amplitude seizure,

but it also responded to high-amplitude artifacts. Thus, it is unsuitable for detecting a seizure when the EEG signals are contaminated with artifacts. Another disadvantage of the detection algorithm was that it required the EEG signal to be normal at the beginning and ictal patterns must be adequately long. Hence, only 85 seizure events were left for evaluation. When a single feature was extracted to detect epilepsy, amplitude or uncertainty-related features were commonly used with a ML technique. Results of using LBP with k-nearest neighbor (k-NN) showed that histograms between normal and epileptic patterns were different [5]. Similarly, LNDP and LGP were individually applied to k-NN, SVM, decision tree (DT), and ANN to distinguish their patterns in epileptic EEG signals [43]. As a result, both LNDP and LGP achieved capabilities of analyzing seizures for any classifiers while k-NN had the fastest computation. In addition, PE was employed to distinguish seizures from EEG signals where it was concluded that PE from a normal EEG signal was higher than that from a seizure EEG signal [4]. Furthermore, WPE, which was calculated from sub-segments of one-long epoch and then concatenated into a feature vector, was used with three classifiers: linear SVM, radial basis function kernel SVM (RBF SVM), and ANN to find the performance of each classifier [67]. Combining WPE and RBF SVM outperformed the other two classifiers. Unlike the ML approach, the results of using a thresholding method with ApEn as a single feature showed that the seizure period obtained lower ApEn than the normal period [68]. An error energy obtained from a linear predictive filter was used in a thresholding method in

Table 3
Summary of automated epileptic seizure detection using the *Bonn University* database when *single-domain* features were used.

Domain	Features	Method	Performance	Ref.
Time	Raw signal	CNN	Acc = 88.70%, Sen = 95.00%, Spec = 90.00%	[65]
			Acc = 98.80%	[66]
	IMFs	EMD + CNN	Acc = 98.80%	[66]
	Linear prediction error energy	Thresholding	Sen = 92.00%, Spec = 94.50%	[69]
	LBP	k-NN	Acc = 99.33%	[5]
	LNDP	ANN	Acc = 98.72%, Sen = 98.30%, Spec = 98.82%	[43]
	1D-LGP	ANN	Acc = 98.65%, Sen = 98.44%, Spec = 98.70%	[43]
	FuzzyEn, DistEn	QDA	Acc = 92.80%, Sen = 90.67%, Spec = 92.80%	[70]
	ApEn	Thresholding	Acc = 73.00%	[68]
	WPE	Linear SVM	Acc = 91.63%	[67]
		RBF SVM	Acc = 93.38%	[67]
		ANN	Acc = 91.86%	[67]
	B_{AM}, B_{FM}	EMD + Wavelet SVM	Acc = 99.75%, Sen = 100%, Spec = 99.69%	[45]
Frequency	PSD	DT	Acc = 98.72%, Sen = 99.40%, Spec = 99.31%	[73]
	Coefficients	DFT + CNN	Acc = 99.40%	[66]
	Peak amplitude, peak frequency	PSD (ARMA) + GMM	Sen = 70.00%, Spec = 53.33%	[74]
		PSD (ARMA) + ANN	Sen = 91.67%, Spec = 83.33%	[74]
		PSD (ARMA) + RBF SVM	Sen = 96.67%, Spec = 86.67%	[74]
		PSD (YW) + GMM	Sen = 78.33%, Spec = 83.33%	[74]
		PSD (YW) + ANN	Sen = 98.33%, Spec = 90.00%	[74]
		PSD (YW) + RBF SVM	Sen = 96.67%, Spec = 86.67%	[74]
		PSD (Burg) + GMM	Sen = 75.00%, Spec = 93.33%	[74]
		PSD (Burg) + ANN	Sen = 98.33%, Spec = 96.67%	[74]
		PSD (Burg) + RBF SVM	Sen = 98.33%, Spec = 96.67%	[74]
Time-frequency	log of Fourier spectrum at scale 4 and 5	DT-CWT[4] + k-NN	Acc = 100%	[75]
	Coefficients	DWT + CNN	Acc = 99.20%	[66]
	Mean, variance, skewness, kurtosis, entropy	EWT[5] + RBF SVM	Acc = 100%	[25]
	Absolute coefficients, absolute mean, absolute median, absolute SD, absolute max, absolute min	RDSTFT[6] + ANN	Acc = 99.80%, Sen = 99.90%, Spec = 99.60%	[76]
	Absolute mean, average power, SD	DWT + PCA + RBF SVM	Acc = 98.75%, Sen = 99.00%, Spec = 98.50%	[77]
		DWT + LDA + RBF SVM	Acc = 100%, Sen = 100%, Spec = 100%	[77]
		DWT + ICA + RBF SVM	Acc = 99.50%, Sen = 99.00%, Spec = 100%	[77]
	Absolute max, absolute min, absolute mean, SD	WPD + k-NN	Acc = 99.45%	[78]
	Line length	DWT + ANN	Acc = 97.77%	[7]
	Fractional energy	SPWVD[7] + ANN	Acc = 99.92%	[79]
	ApEn	MWT[8] (GHM) + ANN	Acc = 96.69%, Sen = 98.62%, Spec = 89.91%	[80]
		MWT (CL) + ANN	Acc = 95.15%, Sen = 96.57%, Spec = 89.21%	[80]
		MWT (SA4) + ANN	Acc = 98.27%, Sen = 99.00%, Spec = 95.50%	[80]
		DWT + Thresholding	Acc = 96.00%	[68]
	WPE	DWT + Linear SVM	Acc = 86.50%	[67]
		DWT + RBF SVM	Acc = 88.25%	[67]
		DWT + ANN	Acc = 86.63%	[67]
	WMRPE	FBS[9] + DT	Acc = 98.60%	[81]
	HE, ARMA parameters	DCT + RBF SVM	Acc = 97.79%, Sen = 97.97%, Spec = 97.60%	[49]
	HE, fractal dimension, PE	DT-CWT + RBF SVM	Acc = 98.87%, Sen = 98.20%, Spec = 100%	[82]
		DT-CWT + k-NN	Acc = 97.80%, Sen = 97.20%, Spec = 98.60%	[82]
		DT-CWT + DT	Acc = 90.33%, Sen = 90.00%, Spec = 94.00%	[82]
		DT-CWT + RF	Acc = 98.13%, Sen = 98.20%, Spec = 98.40%	[82]

[4] Dual-tree complex wavelet transform, [5] empirical wavelet transform, [6] rational discrete STFT, [7] smoothed pseudo Wigner-Ville distribution, [8] multiwavelet transform, [9] Fourier-Bessel series. Note that all publications used every set of the data but none of them explained which segments were selected.

Table 4
Summary of automated epileptic seizure detection using the *Bonn University* database when *multi-domain* features were used.

Time	Frequency	Time-frequency	Method	Performance	Ref.
Raw signal, IMFs	Coefficients	Coefficients	EMD + FFT + DWT + CNN	Acc = 99.50%	[66]
Mean, energy, SD, max value		Mean, energy, SD, max value	DWT + HHT + NNE	Acc = 98.78%	[83]
			DWT + HHT + ANN	Acc = 88.00%	[83]
			DWT + HHT + RNN	Acc = 91.33%	[83]
			DWT + HHT + SVM	Acc = 94.67%	[83]
			DWT + HHT + k-NN	Acc = 95.33%	[83]
			DWT + HHT + LDA	Acc = 92.67%	[83]
Max, min, mean, SD, kurtosis, skewness, Q1, Q3, IQR, median, mode, mobility, complexity, SampEn, HE, DFA		Max, min, mean, SD, ShEn	DWT + RF	Acc = 97.40%, Sen = 97.40%, Spec = 97.50%	[84]
D_B		Energy	HWPT + RVM	Acc = 99.80%, Sen = 100%, Spec = 99.00%	[56]

Note that all publications used every set of the data but none of them explained which segments were selected.

[69], where the epileptic group had a higher error energy than the normal group.

On the other hand, making use of several TDFs applied to ML techniques was found in [23,70,45] (Tables 1 and 3). In [23], statistical features, including absolute mean, SD, skewness, kurtosis, ratio of absolute mean, and average power of each IMF obtained from EMD, were computed following an artifact removal using a multiscale principal component analysis (MSPCA), and then passed to

many classifiers, including SVM, ANN and k-NN. Applying FuzzEn and DistEn in [70] to a multi-length segmentation approach and a quadratic discriminant analysis (QDA) showed that this combination could identify seizures. However, their performance was inferior to that in [43] which used only single features with ANN or k-NN. Furthermore, B_{AM} and B_{FM} computed from each IMF were analyzed independently and fed to SVMs with many kernel functions including RBF and wavelet kernel functions [45]. The authors reported that using Morlet wavelet with those features could effectively separate seizure epochs from normal epochs with accuracy, sensitivity, and specificity of 99.75%, 100%, and 99.69%, respectively. Alternatively, among pioneer works in epilepsy detection (not listed in the tables as the data were from different sources), there were studies considering a group of features applied to a criterion-based classifier as opposed to ML approaches [6,71,72]. The average amplitude relative to the background, average duration, and CV of durations of half-wave transformation were applied to detect epilepsy [6]. A half-wave established in [71] was proposed to determine the amplitude and the duration of the wave. The results showed that the CV should be small, indicating a measure of rhythmicity of the EEG signal, the average duration was limited in some specific range, and the average amplitude was high. Similarly, after the half-wave transformation, duration, amplitude, and sharpness (defined by the slope of the half-wave) were extracted and then compared to its background activity [72], which revealed that these features were in a specific range during epileptic seizures. However, the specific range was not provided in [72]. Energies from three IMFs performed by a thresholding method from the epilepsy period were higher compared to the normal [8].

3.2.2. Frequency-domain features (FDFs)

From Tables 1 and 3, commonly-used features in the frequency domain were PSD, peak amplitude, peak frequency, and energy. From the advantages of deep learning, coefficients computed from DFT and a CNN model were used to detect seizures [66]. The classification accuracy of this approach was 99.40% accuracy. In [73], the combination of PSD based on the Welch method and DT as a classifier achieved a high performance, both in terms of sensitivity and specificity. The first four local extrema and their frequencies obtained from the PSD of the signal computed by several PSD estimation methods (ARMA, Yule-Walker equation, and Burg) were used to identify epileptiform activities [74]. These features computed from any estimated PSD performed well with ANN and RBF SVM but the performance dropped when using these features with a Gaussian mixture model (GMM). By means of filter bank analysis, the energy obtained from each frequency range of three consecutive epochs was concatenated into a feature vector for temporal and spectral information, and then applied to RBF SVM for classification [57]. As a result, the energies from the seizure group tended to be higher than that of the normal group.

Furthermore, some features in the frequency domain were used jointly to detect epilepsy. Both criteria-based and ML methods were used in the AESD. The peak frequency, bandwidth of the peak frequency, and power ratio of the current and background epochs in the same frequency range were applied to the AESD in newborns [85]. These features with appropriate thresholds could identify a large part of seizures but still incorrectly classified some seizure events, such as slow or spike waves.

3.2.3. Time-frequency-domain features (TFDFs)

Common transformed signals still containing information of both time and frequency domains were analyzed by STFT or DWT. Firstly, we describe the literature that regards the transformed signals as raw features. By using a deep learning approach, the STFT spectrogram of a raw EEG signal was considered as a feature with a modified stacked sparse denoising autoencoder (mSSDA) as a classifier [62]. This combination was then compared to the other classifiers, where the result showed that the mSSDA could successfully distinguish epilepsy from normality. Similarly, the logarithms of the Fourier spectrum of dual-tree complex transform (DT-CWT) coefficients at scales 4 and 5 were performed with k-NN for detecting epilepsy [75]. This approach performed well with an accuracy of 100% and could be implemented in real time because of its low computational complexity, consuming 14.4 ms for processing.

The other group of previous studies further extracted features from STFT or DWT sequences. Amplitude and uncertainty measurements were commonly used individually, whereas statistical parameters were jointly employed in general. Secondly, we describe studies that applied a single feature on transformed signals. A logarithm of variance of each DWT sub-band coefficients from a chosen channel were used independently with a threshold [63] to detect patient-specific epileptic seizure. The use of this feature obtained outstanding average performances of 93.24% accuracy, 83.34% sensitivity, and 93.53% specificity from the best classification result from each chosen patient. The author also applied combinations of those features to SVM and thoroughly compared the results with TDFs [61]. The TDFs were line length, NE, variance, power, and maximum value of raw EEG signals. As a result, the best average detection performances from each patient using SVM with wavelet-based features were average accuracy of 96.87%, sensitivity of 72.99%, and specificity of 98.13%. On the other hand, the best average outcomes using TDFs were 95.17% accuracy, 66.35% sensitivity, and 96.91% specificity, respectively. The energy calculated from each transformation was commonly used in seizure detection. The energy from each DWT sub-band coefficients with RBF SVM was used to present the spectrum in each sub-band [86]. Similarly, fractional energies, energies in specific frequency bands and time windows of a Smoothed pseudo WVD (SPWVD) were applied to principal component analysis (PCA) and ANN, resulting in a high average accuracy of 95.00% [79].

Furthermore, energies from Cohen's class transformation, including SPWVD, were employed to analyze EEG signals [87]. As a result, the energy computed from any transformation was successfully able to distinguish epileptic EEG from normal EEG. The line length calculated from DWT coefficients with ANN could determine seizures properly [7]. For the uncertainty-related features, the ApEn, calculated from DWT coefficients, was also applied to a criteria-based detection system [68] and the ApEn computed from multiwavelet transform (MWT) was used with ANN [80] in the AESD. Three famous multiwavelets, namely Gernoimo-Hardin-Massopust (GHM), Chui-Lian (CL), and SA4, were exploited. Both approaches achieved an accuracy of 95-98%. In addition, as a result in [68], the ApEn had small values in a seizure group and using DWT could improve and provide some useful information for the AESD. When WPE was extracted from the DWT coefficients in [67], the work tested the methods of ANN, linear SVM, and RBF SVM, and the results indicated that RBF SVM could surpass the two other classifiers, but with lower accuracies than those of [80] evaluated on the same data set. Moreover, WMRPE computed from Fourier-Bessel series (FBS) coefficients corresponding to five normal EEG rhythms was used to analyze epileptic seizures using three classifiers, namely, RF, SVM, and DT [81]. The authors reported that using the WMRPE with the DT achieved the highest accuracy of 98.60%.

Thirdly, we discuss TFDFs that were applied jointly. In this case, statistical parameters and amplitude-related measures were commonly used. The absolute mean, SD, average power of DWT coefficients of each level, and ratio of absolute mean values of the adjacent sub-bands were combined and applied to the dimensionality reduction techniques of PCA, independent component analysis (ICA), and linear discriminant analysis (LDA) to extract useful features and reject meaningless features [77]. The dimension-reduced feature vector was classified by RBF SVM and the results showed

that this feature-based approach could firmly detect seizure with an accuracy of 98-100%. The mean, SD, skewness, kurtosis, and ratio of variance of the adjacent sub-bands computed from the DWT coefficients of every decomposition level, and mean frequency and peak frequency calculated from the PSD of each decomposition level were cooperated with an extreme learning machine (ELM) to classify an epileptic EEG signal [24]. Due to the efficiency of computation in the ELM [88], this combination could perform the detection accurately and quickly in a short time. Nevertheless, a total of 20 recordings from three patients was used to evaluate the classification performance. The absolute mean, SD, skewness, kurtosis, ratio of absolute mean, and average power of each decomposition level were applied with DWT and wavelet packet decomposition (WPD) in the AESD [23]. A noise reduction of the multi-channel EEG signals was done by the MSPCA and when using this with the RF, SVM, ANN, and k-NN classifiers, it achieved an accuracy of 100%. However, the authors only used 2,000 8-second EEG segments, 1,000 segments for each class, as samples in the experiments. Moreover, the mean, variance, skewness, kurtosis, and ShEn of each sub-band in the empirical wavelet transform (EWT) were extracted and the use of these features could potentially detect focal epilepsy automatically by RBF SVM with a reported accuracy of 100% [25]. The absolute mean, absolute median, absolute maximum, absolute minimum, and absolute values of coefficients from Rational discrete STFT (RDSTFT) were investigated with diverse classifiers [76]. The results indicated that the ANN was the most optimal classifier, and the absolute median and some absolute coefficients were the most dominant features.

In addition to statistical parameters, combinations of uncertainty and similarity measurements were also established to indicate a periodic pattern of seizure. The use of TFDFs including energy and D_B from improved eigenvalue decomposition of Hankel matrix and Hilbert transform potentially distinguished segments of seizure and seizure-free [89]. The PE, HE, and D_H extracted from some DT-CWT sub-band coefficients were used as the features for RBF SVM, k-NN, DT and RF analyses [82]. These features from each decomposition level could separate the seizure activities from the normal group. The HE estimated from discrete-time fractional Brownian motion process and discrete-time fractional Gaussian motion process and ARMA parameters from multi-rate filter bank were computed and applied to the RBF SVM to detect seizure [49]. As a result, the HE of each sub-band was different according to the stationarity of each sub-band, and the HE and the ARMA parameters computed from some sub-band coefficients were significant.

3.2.4. Multi-domain features

In terms of the complementary information that cannot be achieved in one domain, some studies have used features from different domains combined into a feature vector. Table 4 shows that features from the time domain and time-frequency domain were commonly used together since the time-frequency domain also provides spectral information. A combination of the mean, SD, energy, and maximum value from EEG signal and from the envelope spectrum of some DWT sub-bands were also formed as features and applied to a neural network ensemble (NNE) [83]. The authors reported that features extracted from the envelope spectrum of the decomposition levels of the seizure group were normally higher than in the normal group [83]. The mean, SD, maximum and minimum values of EEG signal combined with the energy ratio of each detail and approximation coefficients from the DWT were employed with an Ant Colony classifier [90]. As a result, these features were typically higher in epileptic seizure periods, which agreed with previous studies.

The mean, SD, skewness, kurtosis, maximum and minimum values, median, mode, Q_1, Q_3, IQR, mobility, complexity, HE, DFA, ShEn, and SampEn were computed from an EEG signal, and the

mean, SD, maximum and minimum values were obtained from each sub-band of DWT [84]. Subsequently, a novel feature selection, the improved correlation-based feature selection, was proposed to select significant features based on their correlation and then the selected features were classified by the RF. The results showed that the feature dimension was extremely reduced and the remaining features could improve the detection system compared to features selected by the CFS. The variance, skewness, kurtosis, RMS, and SampEn calculated from each sub-band and complete bandwidth based on the second order Butterworth filter, and peak frequency and median frequency computed from the PSD of the bands were fed to several feature selection techniques and classifiers to find the best combination [64]; see performance in Table 2. From the best results reported in the article, the top five uncorrelated features via LDA a with backward search, a feature selection method, in each region were mostly the RMSs of some frequency range. Moreover, compared to several classifiers, k-NN performed the best with this approach. However, the authors selected only records containing seizure activities to perform the experiments. In [56], the D_B from raw EEG signals and energies from each Harmonic wavelet packet transform (HWPT) sub-bands used with a relevance vector machine (RVM) could be successfully applied in real-time settings.

On the other hand, some literature also included features from three domains to obtain information from all different aspects. A deep learning approach, namely, CNN, was exploited to extract features and detect seizures [66]. Combining several transformations, this method could successfully differentiate ictal and normal epochs with an accuracy of 99.50%. Twenty-one features from different domains, which were RMS, NE, line length, number of zero-crossings, number of local extrema, three Hjorth parameters, ShEn, ApEn, SVDEn, and error from validating an AR model on other segment obtained from an EEG signal, IWMF, IWBW, SEF90, the total power, peak frequency and bandwidth of the peak frequency calculated from PSD, and energy extracted from a specific DWT sub-band that corresponded to 1.25–2.5 Hz, were tested with LDA to determine their abilities on neonatal seizure detection in the EEG signal [91]. It showed that all the features combined achieved the best outcome, although the RMS, number of local extrema, and line length were three most significant features.

Based on our review, the first conclusive point is that a raw EEG signal or coefficients of any transformation were typically combined with a deep learning technique. Most studies were in favor of a ML approach rather than a thresholding technique, since the threshold of each feature can vary with subjects and baselines of EEG signals in a normal period. Secondly, no statistical parameter was used individually in an AESD. In fact, statistical parameters were always applied with other features that were related to amplitude or uncertainty measurement and with a ML technique. This implies that statistical parameters alone do not attain sufficient information to distinguish the seizures from the normality. On the other hand, features, such as quantities of amplitude and similarity, were commonly used separately from other features. In particular, the energy was the most widely used feature that contributed to the meaning of amplitude. Moreover, attributes of amplitude and uncertainty were also typically considered when being employed with statistical parameters. This indicates that the amplitude-related features and similarity-uncertainty measurements have the capability to distinguish the seizures from the normality in an EEG signal.

4. Methods for feature evaluation

Our paper aims to examine the significance of a single feature to the classification performance through the Bayesian method.

However, multiple features are commonly applied in practice, so a redundancy analysis is also needed. Therefore, this section describes the brief concepts about the Bayes classifier and the correlation-based feature selection.

4.1. Bayes error rate

Common feature selection and dimensionality reduction methods in AESD [10,79,92] do not provide information about how a single feature independently improves the classification performances. Hence, to determine the significance of each feature on the AESD, we employed the Bayes classifier. If we define x as a feature, $C(x)$ as a class in which feature x is classified, and C_i denotes a class i labeled from the data, then a misclassification error is generally defined as

$$\text{err} = \int \sum_{C_i \neq C(x)} P(C_i|x)p(x)dx, \tag{26}$$

where $P(C_i|x)$ is the posterior probability of x in class i, and $p(x)$ is the probability density function of x. Intuitively, we can interpret that the error is the total joint probability that the feature is incorrectly classified. The Bayes optimal classifier gives the minimum error (the *Bayes error rate*) by choosing the class of which the posterior probability is the highest [93]. As a result, the Bayes error rate (err_b) is obtained from

$$\text{err}_b = \int \sum_{C_i \neq C_{\max}} P(C_i|x)p(x)dx, \tag{27}$$

where $C_{\max}$ is the class of which the posterior probability is maximum.

Practically, the distribution of a likelihood function is unknown, so a non-parametric distribution estimation was exploited in our experiment and the posterior probability was obtained by Bayes' rule:

$$P(C_i|x) = \frac{p(x|C_i)P(C_i)}{p(x)}, \tag{28}$$

where $p(x|C_i)$ is the likelihood function, $P(C_i)$ is the prior distribution, and $p(x)$ is the evidence. For all real x values, the likelihood function can be estimated using the non-parametric kernel smooth function

$$p(x|C_i) = \frac{1}{N_i h} \sum_{j=1}^{N_i} K\left(\frac{x - x_j}{h}\right), \tag{}$$

where N_i is the sample size of class C_i, x_j is a sample in the class, $K(x)$ is the Gaussian kernel and h is a bandwidth [94,95]. Regarding a classification problem, the prior $P(C_i)$ is assumed to be binomial estimated by the size of class C_i divided by the total number of samples. Finally, $p(x)$ is obtained by the total probability and it completes the calculation of $P(C_i|x)$ in (28).

Our problem was a two-class classification (normal/seizure) problem, so the Bayesian error was reduced to

$$\text{err}_b = \int \min_{i=1,2} P(C_i|x)p(x)dx, \tag{29}$$

where C_1 and C_2 stand for the seizure and normal classes. For the non-parametric kernel function, $h \approx 1.06\hat{\sigma}N_i^{-\frac{1}{5}}$ was chosen to be the optimal bandwidth, where $\hat{\sigma}$ is the sample SD of the sample.

We have observed that using conventional performance metrics, such as accuracy, cannot give the significance of each individual feature. Hence, to evaluate the performance of individual features, we proposed to use an improvement rate (**rate**) from a standard condition (err_0) as follows:

$$\text{err}_0 = \int_{-\infty}^{\infty} P(C_2|x)p(x)dx = P(C_2), \quad \textbf{rate} = \frac{\text{err}_0 - \text{err}_b}{\text{err}_0} \times 100\%. \tag{30}$$

4.2. Correlation-based feature selection (CFS)

The CFS introduced in [96] is a feature selection method based on the hypothesis that *a good subset of features is highly correlated with the class, but uncorrelated with others.* For a feature subset F containing k features, an index called heuristic merit is exploited to measure feature-feature and feature-class correlations and is defined by

$$\text{Merit}_F = \frac{k\bar{r}_{fc}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}, \tag{31}$$

where $\bar{r}_{fc}$ is the mean value of the correlation of feature and class, and $\bar{r}_{ff}$ is the average of the feature-feature correlation. To find the subset F with the highest merit score, we applied the CFS algorithm provided in [97], where the correlations are estimated by conditional entropy. The algorithm initially assigns the subset F to be empty. New feature subsets are constructed by adding another feature that has not been previously selected to F. All new subsets are then evaluated and the subset having the best merit score is used as the subset F in the next iteration. This process is repeated until F has m features. The final subset F contains features ranked in the descending order by the merit score.

5. Experimental results

The experiment was performed on the public CHB-MIT Scalp EEG database [59] that contains 24 EEG recordings from 23 patients: five males aged 3-22 years, 17 females aged 1.5-19 years, and one anonymous subject. All EEG data were collected at the Children's Hospital Boston and the international 10-20 system was used to locate electrode placements. All signals were recorded with a sampling frequency of 256 Hz digitized at 16-bit resolution and stored in an EDF file [98]. This full database is available downloaded at PhysioNet (https://physionet.org/physiobank/database/chbmit/).

We randomly chose two records from each case, subject to the inclusion condition that every record must contain at least one seizure activity, so the records for assessing the improvement rate of the Bayes error rate of each feature are shown in Table 5. The channels were sequentially listed as follows: FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, and P8-O2.

An EEG epoch was defined by segmenting a raw EEG signal in every channel with a 4-second width. The next consecutive epoch was segmented from the raw EEG signal by moving the window for 1 s. These choices were selected from inspecting the processing step using the commercial Persyst [99] software. After the process of segmenting the raw EEG signals, a feature was then computed from each epoch and each channel independently. Only commonly used features in the literature were selected in this experiment (Table 6). For the ApEn and SampEn, the template length, m, and the tolerance, r, were set to $m = 2$ and $r = 0.2$SD, where SD was the sample SD of the segment. All TDFs were calculated from a raw EEG signal, whereas features from the frequency domain were extracted from PSD and TFDFs were computed from DWT coefficients with the Daubechies wavelet tap 4 for five levels.

The EEG channel selection is still an open research question and using multi-channel EEG signals may be redundant. Moreover, some commercial software also analyzes the seizure activity over the left and right sides of the brain. For these reasons, we have not

Table 5
List of all records used to evaluate features.

Records							
chb01_04	chb01_16	chb02_16+	chb02_19	chb03_03	chb03_35	chb04_08	chb04_28
chb05_06	chb05_13	chb06_01	chb06_04	chb07_13	chb07_19	chb08_02	chb08_05
chb09_06	chb09_08	chb10_38	chb10_89	chb11_92	chb11_99	chb12_33	chb12_38
chb13_19	chb13_55	chb14_04	chb14_18	chb15_06	chb15_15	chb16_17	chb16_18
chb17a_04	chb17b_63	chb18_32	chb18_35	chb19_29	chb19_30	chb20_14	chb20_16
chb21_21	chb21_22	chb22_21	chb22_22	chb23_06	chb23_09	chb24_04	chb24_11

Table 6
List of features for the Bayesian error rate evaluation and redundancy analysis.

Domains	Features
Time	Mean, variance, CV, skewness, kurtosis, max, min, energy, NE, line length, ShEn, ApEn, SampEn, number of zero-crossing, number of local extrema, mobility, complexity
Frequency	IWMF, IWBW, SE, peak frequency, peak amplitude
Time-frequency	Mean, absolute mean, variance, skewness, kurtosis, max, min, energy, line length

Table 7
Bayes error (err_b) and improvement rate of time-domain and frequency-domain features.

Feature	Left side		Right side	
	err_b	rate	err_b	rate
(a) **Time-domain** features				
Mean	0.0174	0.00	0.0174	0.05
Variance	0.0160	**8.20**	0.0166	**4.78**
CV	0.0174	0.00	0.0174	0.00
Skewness	0.0174	0.00	0.0174	0.00
Kurtosis	0.0174	0.08	0.0174	0.17
Max	0.0174	0.00	0.0174	0.00
Min	0.0174	0.00	0.0174	0.00
Energy	0.0160	**8.18**	0.0166	**4.77**
NE	0.0157	**10.07**	0.0160	**8.36**
Line length	0.0166	1.92	0.0167	1.16
ShEn	0.0174	0.33	0.0165	**5.64**
ApEn	0.0174	0.00	0.0174	0.00
SampEn	0.0174	0.00	0.0174	0.00
Local extrema	0.0174	0.00	0.0174	0.00
Zero-crossing	0.0174	0.09	0.0174	0.09
Mobility	0.0174	0.00	0.0174	0.00
Complexity	0.0174	0.00	0.0174	0.00
(b) **Frequency-domain** features				
IWMF	0.0174	0.00	0.0174	0.00
IWBW	0.0174	0.00	0.0174	0.00
SE	0.0174	0.00	0.0174	0.00
Peak amplitude	0.0174	0.00	0.0174	0.00
Peak frequency	0.0174	0.00	0.0174	0.00

explored the channel selection topic but rather used a spatial averaging of features over the left and right sides of the brain, denoted as x_{left} and x_{right}, respectively. We then estimated the posterior probability distributions and computed the Bayes error rates of using the left and right feature representatives.

5.1. Feature significance

From the chosen records, there were 263,424 samples in the normal group and 4,677 epochs belonging to the seizure group, and so $err_0 = 4,677/(263,424 + 4,677) = 0.0174$. The features with an improvement rate higher than 4.5% were considered as significant. Table 7 shows the Bayes error rate and improvement rate of each TDF and FDF. The results from this data set showed that most features achieved almost the same Bayes errors that were close to err_0, except for the *variance, energy, NE,* and *ShEn* that obtained high improvement rates of 4.77% to 10.07%, where NE reached the

highest improvement rate. Figure 1 displays the improvement rates of all TFDFs in each wavelet decomposition level. Note that D1, D2, D3, D4, D5, and A5 represent sub-bands from which the features are extracted. Overall, the results showed that the *variance* and *energy* of the wavelet coefficients in all decomposition levels yielded relatively high improvement rates, compared to other features computed on the coefficients of other levels. Specifically, *energy* from level D1 of the left half brain accomplished the highest improvement rate of 13.51%. *Line length* from levels D5 and A5, and *kurtosis* from D1 of the left hemisphere also achieved significant improvement rates. We found that the most significant features were related to amplitudes and variations of the signals, such as variance, energy, and NE. Additionally, features that can capture changes in amplitude, frequency, and rhythmicity gain some improvement, since there is continuous evolution of amplitude, frequency, and rhythms during seizure activities compared to the background [100]. On the other hand, FDFs do not help improve the performance from the baseline because, in this data set, there are artifacts causing the probability of the seizure occurrence to be less than that of the other class.
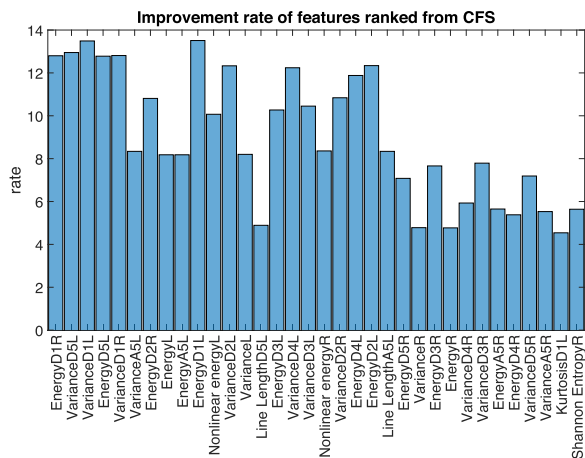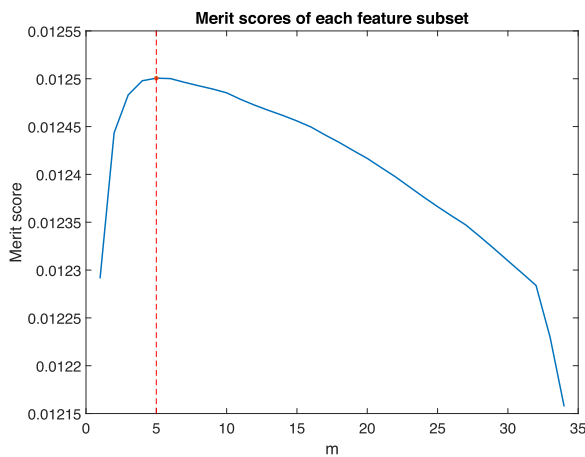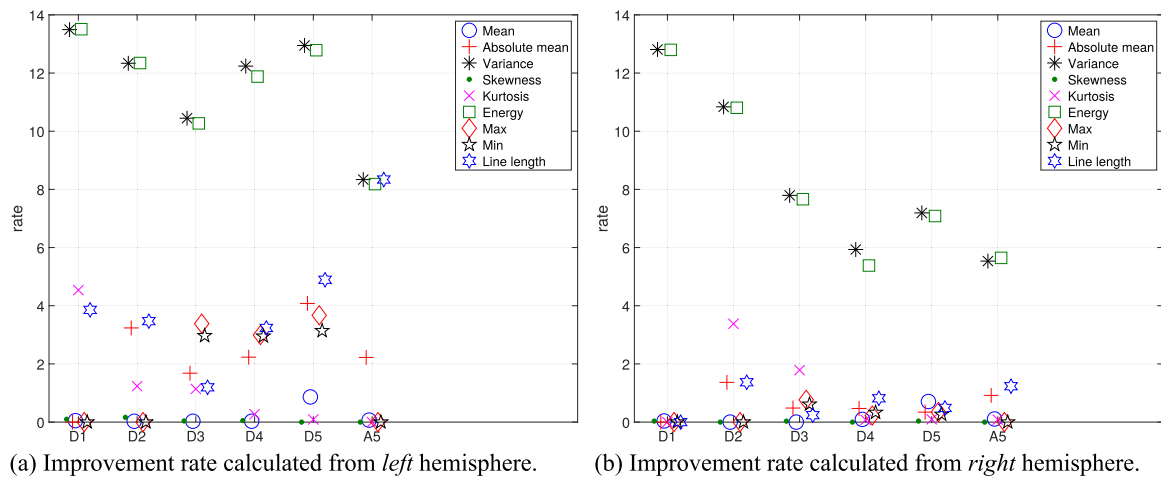
5.2. Feature redundancy analysis

The significant features achieving the improvement rate higher than 4.5% (in total, 34 features) were then applied to the CFS algorithm to examine their redundancy. Figure 2 shows that the merit score initially increases as m (the number of features in the subset) increases until the feature subset contains five features, with a maximum score of 1.25×10^{-2}. As m increases beyond five, the merit score decreases.

In addition, the features and their improvement rates in the final feature subset were ranked in the descending order by the CFS as shown in Figure 3, where L and R stand for features computed from the left side and right side, respectively. The features in the optimal subset were *EnergyD1R, VarianceD5L, VarianceD1L, EnergyD5L*, and *VarianceD1R*. Accordingly, these five features also achieved relatively high improvement rates among the significant features. As shown in Figure 2, the merit scores of the optimal feature subset of each size were not that different. This indicated that the results of feature significance from our experiments also agree with the outputs of the CFS. The results in [10] showed that the classification accuracy was not increased when other features were added to the optimal feature subset. Moreover, these five features require just only $\mathcal{O}(N)$ for the total computation. Hence, we suggest that these five features should be at least used as features in the AESD because of their high improvement rates and low computational complexity.

6. Conclusions

This paper aimed to review details of features in epileptic seizure detection using EEG signals. We provided mathematical descriptions and computations of features in detail to reduce inconsistencies of some complicated feature definitions that have appeared in the literature. We also summarized feature usages

(a) Improvement rate calculated from *left* hemisphere.

(b) Improvement rate calculated from *right* hemisphere.

Fig. 1. Improvement rate based on the Bayesian method of time-frequency domain features calculated from **DWT** using Daubechies 4 wavelet.

Fig. 2. Merit scores of feature subsets. The subset size achieving the highest merit score is 5.

Fig. 3. The features ranked by CFS and their improvement rates. All features in the optimal subset also obtained high improvement rates.

and intuitive meanings for detecting seizures according to their algorithms. Based on our review, deep learning techniques were used in combination with raw EEG signals or coefficients of a transformation, while shallow neural networks always required features as inputs. A single statistical parameter (as a feature) was never used but combinations of them were applied. Such common

combinations, including energy and entropy, were applied with a classification from ML approaches, while other types of features, such as entropy and fractal dimension, had sometimes been used independently. It was evident that the feature most widely used to quantify a relation with amplitude was energy.

Furthermore, we conducted two experiments to verify the significance of each widely used feature from the time, frequency, and time-frequency domains, and to analyze the redundancy of those features. Based on our experiment using the standard Bayes classifier, we concluded that variance, energy, NE, and ShEn extracted from raw EEG signals were useful for distinguishing between seizure and normal EEG signals, and could improve the detection performance from the baseline condition. However, other TDFs and FDFs gave insignificant outcomes for the detection of seizures. Furthermore, features extracted from the DWT coefficients from some specific levels also provided intrinsically useful information about seizure that could not be achieved from raw EEG signals. Moreover, based on redundancy analysis, the top significant features selected from CFS were the energy and variance from the DWT coefficients of some decomposition levels. Therefore, the energy and variance calculated from the DWT coefficients were typically recommended to be used in both online and offline AESD due to their significance and low computational complexity.

Acknowledgement

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. Cross, C. Elger, J.J. Engel, L. Forsgren, J. French, M. Glynn, et al., ILAE official report: a practical clinical definition of epilepsy, Epilepsia 55 (4) (2014) 475–482.
[2] D. Thurman, E. Beghi, C. Begley, A. Berg, J. Buchhalter, D. Ding, D. Hesdorffer, W. Hauser, L. Kazis, R. Kobau, et al., Standards for epidemiologic studies and surveillance of epilepsy, Epilepsia 52 (7) (2011) 2–26.
[3] U. Acharya, S. Sree, G. Swapna, R. Martis, J. Suri, Automated EEG analysis of epilepsy: a review, Knowl-Based Syst. 45 (2013) 147–165.

[4] J. Li, J. Yan, X. Liu, G. Ouyang, Using permutation entropy to measure the changes in EEG signals during absence seizures, Entropy 16 (6) (2014) 3049–3061.

[5] T. Kumar, V. Kanhangad, R. Pachori, Classification of seizure and seizure-free EEG signals using local binary patterns, Biomed. Signal Process. Control 15 (2015) 33–40.

[6] J. Gotman, Automatic recognition of epileptic seizures in the EEG, Electroencephalogr. Clin. Neurophysiol. 54 (5) (1982) 530–540.

[7] L. Guo, D. Rivero, J. Dorado, J. Rabunal, A. Pazos, Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks, J. Neurosci. Methods 191 (1) (2010) 101–109.

[8] L. Orosco, E. Laciar, A. Correa, A. Torres, J. Graffigna, An epileptic seizures detection algorithm based on the empirical mode decomposition of EEG, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE (2009) 2651–2654.

[9] A. Temko, E. Thomas, W. Marnane, G. Lightbody, G. Boylan, EEG-based neonatal seizure detection with support vector machines, Clin. Neurophysiol. 122 (3) (2011) 464–473.

[10] A. Aarabi, F. Wallois, R. Grebe, Automated neonatal seizure detection: a multistage classification system through feature selection based on relevance and redundancy analysis, Clin. Neurophysiol. 117 (2) (2006) 328–340.

[11] L. Yu, H. Liu, Efficient feature selection via analysis of relevance and redundancy, J. Mach. Learn. Res. 5 (Oct) (2004) 1205–1224.

[12] H. Hall, Correlation-based Feature Selection for Machine Learning, Ph.D. Thesis, University of Waikato Hamilton, New Zealand, 1999.

[13] I. Kononenko, E. Šimec, M. Robnik-Šikonja, Overcoming the myopia of inductive learning algorithms with RELIEFF, Appl. Intell. 7 (1) (1997) 39–55.

[14] U. Acharya, H. Fujita, V. Sudarshan, S. Bhat, J. Koh, Application of entropies for automated diagnosis of epilepsy using EEG signals: A review, Knowl-Based Syst. 88 (2015) 85–96.

[15] M. Ahmad, M. Saeed, S. Saleem, A. Kamboh, Seizure detection using EEG: a survey of different techniques, Proceedings of the 2016 International Conference on Emerging Technologies, IEEE (2016) 1–6.

[16] U.R. Acharya, Y. Hagiwara, S.N. Deshpande, S. Suren, J.E.W. Koh, S.L. Oh, N. Arunkumar, E.J. Ciaccio, C.M. Lim, Characterization of focal EEG signals: a review, Future Gener. Comput. Syst. 91 (2018) 290–299.

[17] R.G. Andrzejak, K. Schindler, C. Rummel, Nonrandomness nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients, Phys. Rev. E 86 (4) (2012), 046206.

[18] T. Alotaiby, F. El-Samie, S. Alshebeili, I. Ahmad, A review of channel selection algorithms for EEG signal processing, EURASIP J. Adv. Signal Process. 2015 (1) (2015) 66–86.

[19] O. Faust, Y. Hagiwara, T. Hong, O. Lih, U. Acharya, Deep learning for healthcare applications based on physiological signals: a review, Comput. Methods Programs Biomed. 161 (2018) 1–13.

[20] B. Boashash, S. Ouelha, Designing high-resolution time-frequency and time-scale distributions for the analysis and classification of non-stationary signals: a tutorial review with a comparison of features performance, Digit. Signal Process. 77 (2018) 120–152.

[21] O. Faust, U.R. Acharya, H. Adeli, A. Adeli, Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis, Seizure 26 (2015) 56–64.

[22] E. Assi, D. Nguyen, S. Rihana, M. Sawan, Towards accurate prediction of epileptic seizures: a review, Biomed. Signal Process. Control 34 (2017) 144–157.

[23] E. Alickovic, J. Kevric, A. Subasi, Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction, Biomed. Signal Process. Control 39 (2018) 94–102.

[24] S. Ammar, M. Senouci, Seizure detection with single-channel EEG using extreme learning machine, Proceedings of the 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering, IEEE (2016) 776–779.

[25] S. Anand, S. Jaiswal, P. Ghosh, Automatic focal epileptic seizure detection in EEG signals, Proceedings of the 2017 IEEE International WIE Conference on Electrical and Computer Engineering, IEEE (2017) 103–107.

[26] S. Chakraborti, A. Choudhary, A. Singh, R. Kumar, A. Swetapadma, A machine learning based method to detect epilepsy, Int. J. Inf. Technol. (2018) 1–7.

[27] N. Huang, Hilbert-Huang Transform and its Applications, vol. 16, World Sci, 2014.

[28] R. Esteller, J. Echauz, T. Tcheng, B. Litt, B. Pless, Line length: an efficient feature for seizure onset detection, Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 2, IEEE (2001) 1707–1710.

[29] L. Hellström-Westas, L. de Vries, I. Rosén, An Atlas of Amplitude-integrated EEGs in the Newborn, CRC Press, 2008.

[30] D. Maynard, P. Prior, D. Scott, Device for continuous monitoring of cerebral activity in resuscitated patients, BMJ 4 (5682) (1969) 545–546.

[31] J. Kaiser, On a simple algorithm to calculate the energy of a signal, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, IEEE (1990) 381–384.

[32] C. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.

[33] S. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. 88 (6) (1991) 2297–2301.

[34] J. Richman, J. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (6) (2000) H2039–H2049.

[35] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88 (17) (2002), 174102.

[36] B. Fadlallah, B. Chen, A. Keil, J. Príncipe, Weighted-permutation entropy: a complexity measure for time series incorporating amplitude information, Phys. Rev. E 87 (2) (2013), 022911.

[37] W. Chen, Z. Wang, H. Xie, W. Yu, Characterization of surface EMG signal based on fuzzy entropy, IEEE Trans. Neural. Syst. Rehabil. Eng. 15 (2) (2007) 266–272.

[38] P. Li, C. Liu, K. Li, D. Zheng, C. Liu, Y. Hou, Assessing the complexity of short-term heartbeat interval series by distribution entropy, Med. Biol. Eng. Comput. 53 (1) (2015) 77–87.

[39] S. Roberts, W. Penny, I. Rezek, Temporal and spatial complexity measures for electroencephalogram based brain-computer interfacing, Med. Biol. Eng. Comput. 37 (1) (1999) 93–98.

[40] H. Hurst, Long-term storage capacity of reservoirs, Trans. Am. Soc. Civ. Eng. 116 (1951) 770–799.

[41] K. Falconer, Fractal Geometry: Mathematical Foundations and Applications, John Wiley & Sons, 2004.

[42] T. Higuchi, Approach to an irregular time series on the basis of the fractal theory, Phys. D Nonlinear Phenom. 31 (2) (1988) 277–283.

[43] A. Jaiswal, H. Banka, Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals, Biomed. Signal Process. Control 34 (2017) 81–92.

[44] B. Hjorth, EEG analysis based on time domain properties, Electroencephalography Clin. Neurophysiol. 29 (3) (1970) 306–310.

[45] V. Bajaj, R.B. Pachori, Classification of seizure and nonseizure EEG signals using empirical mode decomposition, IEEE Trans. Inf. Technol. Biomed. 16 (6) (2011) 1135–1142.

[46] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454 (1971) (1998) 903–995.

[47] R. Bryce, K. Sprague, Revisiting detrended fluctuation analysis, Sci. Rep. 2 (2012) 315–320.

[48] S. Chen, P. Shang, Y. Wu, Weighted multiscale Rényi permutation entropy of nonlinear time series, Physica A: Stat. Mech. Appl. 496 (2018) 548–570.

[49] A. Gupta, P. Singh, M. Karlekar, A novel signal modeling approach for classification of seizure and seizure-free EEG signals, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (5) (2018) 925–935.

[50] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, MIT Press, 2009.

[51] G. Manis, Fast computation of approximate entropy, Comput. Methods Programs Biomed. 91 (1) (2008) 48–54.

[52] G. Fele-Žorž, A faster algorithm for calculating the sample entropy, Proceedings of the Middle-European Conference on Applied Theoretical Computer Science, Slovenia (2013).

[53] G. Manis, M. Aktaruzzaman, R. Sassi, Low computational cost for sample entropy, Entropy 20 (1) (2018) 61–75.

[54] D. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, J. Mach. Learn. Techonol. 2 (2011) 37–63.

[55] A. Temko, E. Thomas, W. Marnane, G. Lightbody, G. Boylan, Performance assessment for EEG-based neonatal seizure detectors, Clin. Neurophysiol. 122 (3) (2011) 474–482.

[56] L. Vidyaratne, K. Iftekharuddin, Real-time epileptic seizure detection using EEG, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (11) (2017) 2146–2156.

[57] A. Shoeb, J. Guttag, Application of machine learning to epileptic seizure detection, Proceedings of the 27th International Conference on Machine Learning (2010) 975–982.

[58] C. Satirasethawong, A. Lek-Uthai, K. Chomtho, Amplitude-integrated EEG processing and its performance for automatic seizure detection, Proceedings of the 2015 IEEE International Conference on Signal and Image Processing Applications, IEEE (2015) 551–556.

[59] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, H. Stanley, PhysioBank, PhysioToolkit, and PhysioNet, Circulation 101 (23) (2000) e215–e220.

[60] R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E 64 (6) (2001), 061907.

[61] S. Janjarasjitt, Epileptic seizure classifications of single-channel scalp EEG data using wavelet-based features and SVM, Med. Biol. Eng. Comput. 55 (10) (2017) 1743–1761.

[62] Y. Yuan, G. Xun, K. Jia, A. Zhang, A multi-view deep learning method for epileptic seizure detection using short-time fourier transform, Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, ACM (2017) 213–222.

[63] S. Janjarasjitt, Performance of epileptic single-channel scalp EEG classifications using single wavelet-based features, Australas. Phys. Eng. Sci. Med. 40 (1) (2017) 57–67.

[64] P. Fergus, A. Hussain, D. Hignett, D. Al-Jumeily, K. Abdel-Aziz, H. Hamdan, A machine learning system for automated whole-brain seizure detection, Appl. Comput. Inf. 12 (1) (2016) 70–89.

[65] U. Acharya, S. Oh, Y. Hagiwara, J. Tan, H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, Comput. Biol. Med. 100 (1) (2018) 270–278.

[66] R. San-Segundo, M. Gil-Martín, L.F. D'Haro-Enríquez, J.M. Pardo, Classification of epileptic EEG recordings using signal transforms and convolutional neural networks, Comput. Biol. Med. 109 (2019) 148–158.

[67] N. Tawfik, S. Youssef, M. Kholief, A hybrid automated detection of epileptic seizures in EEG records, Comput. Electr. Eng. 53 (2016) 177–190.

[68] H. Ocak, Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy, Expert Syst. Appl. 36 (2) (2009) 2027–2036.

[69] S. Altunay, Z. Telatar, O. Erogul, Epileptic EEG detection using the linear prediction error energy, Expert Syst. Appl. 37 (8) (2010) 5661–5665.

[70] P. Li, C. Karmakar, J. Yearwood, S. Venkatesh, M. Palaniswami, C. Liu, Detection of epileptic seizure based on entropy analysis of short-term EEG, PLoS One 13 (3) (2018), e0193691.

[71] J. Gotman, P. Gloor, Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG, Electroencephalography Clin. Neurophysiol. 41 (5) (1976) 513–529.

[72] A. Dingle, R. Jones, G. Carroll, W. Fright, A multistage system to detect epileptiform activity in the EEG, IEEE Trans. Biomed. Eng. 40 (12) (1993) 1260–1268.

[73] K. Polat, S. Güneş, Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform, Appl. Math. Comput. 187 (2) (2007) 1017–1026.

[74] O. Faust, U. Acharya, L. Min, B. Sputh, Automatic identification of epileptic and background EEG signals using frequency domain parameters, Int. J. Neural Syst. 20 (02) (2010) 159–176.

[75] G. Chen, Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features, Expert Syst. Appl. 41 (5) (2014) 2391–2394.

[76] K. Samiee, P. Kovacs, M. Gabbouj, Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform, IEEE Trans. Biomed. Eng. 62 (2) (2015) 541–552.

[77] A. Subasi, M. Gursoy, EEG signal classification using PCA, ICA, LDA and support vector machines, Expert Syst. Appl. 37 (12) (2010) 8659–8666.

[78] D. Wang, D. Miao, C. Xie, Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection, Expert Syst. Appl. 38 (11) (2011) 14314–14320.

[79] A. Tzallas, M. Tsipouras, D. Fotiadis, Automatic seizure detection based on time-frequency analysis and artificial neural networks, Comput. Intell. Neurosci. 2007 (2007).

[80] L. Guo, D. Rivero, A. Pazos, Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks, J. Neurosci. Methods 193 (1) (2010) 156–163.

[81] V. Gupta, R.B. Pachori, Epileptic seizure identification using entropy of FBSE based EEG rhythms, Biomed. Signal Process. Control 53 (2019) 101569.

[82] M. Li, W. Chen, T. Zhang, Automatic epileptic EEG detection using DT-CWT-based non-linear features, Biomed. Signal Process. Control 34 (2017) 114–125.

[83] M. Li, W. Chen, T. Zhang, Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble, Biomed. Signal Process. Control 31 (2017) 357–365.

[84] M. Mursalin, Y. Zhang, Y. Chen, N. Chawla, Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier, Neurocomputing 241 (2017) 204–214.

[85] J. Gotman, D. Flanagan, J. Zhang, B. Rosenblatt, Automatic seizure detection in the newborn: methods and initial evaluation, Electroencephalography Clin. Neurophysiol. 103 (3) (1997) 356–362.

[86] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. Treves, J. Guttag, Patient-specific seizure onset detection, Epilepsy Behav. 5 (4) (2004) 483–498.

[87] A. Tzallas, M. Tsipouras, D. Fotiadis, Epileptic seizure detection in EEGs using time-frequency analysis, IEEE Trans. Inf. Technol. Biomed. 13 (5) (2009) 703–710.

[88] G. Huang, Q. Zhu, C. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1-3) (2006) 489–501.

[89] R.R. Sharma, R.B. Pachori, Time-frequency representation using IEVDHM-HT with application to classification of epileptic EEG signals, IET Science, Meas. Technol. 12 (1) (2017) 72–82.

[90] O. Salem, A. Naseem, A. Mehaoua, Epileptic seizure detection from EEG signal using discrete wavelet transform and ant colony classifier, Proceedings of the 2014 IEEE International Conference on Communications, IEEE (2014) 3529–3534.

[91] B. Greene, S. Faul, W. Marnane, G. Lightbody, I. Korotchikova, G. Boylan, A comparison of quantitative EEG features for neonatal seizure detection, Clin. Neurophysiol. 119 (6) (2008) 1248–1261.

[92] M. Bandarabadi, C. Teixeira, J. Rasekhi, A. Dourado, Epileptic seizure prediction using relative spectral power features, Clin. Neurophysiol. 126 (2) (2015) 237–248.

[93] L. Devroye, L. Györfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, vol. 31, Springer Science & Business Media, 2013.

[94] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd edition, Academic Press Inc., 1990.

[95] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Stat. 33 (3) (1962) 1065–1076.

[96] M. Hall, L. Smith, Feature subset selection: a correlation based filter approach, Proceedings of the 4th International Conference on Neural Information Processing and Intelligent Information Systems, Springer (1997) 855–858.

[97] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, H. Liu, Advancing feature selection research, ASU Feature. Sel. Repository (2010) 1–28.

[98] A. Shoeb, Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment, Ph.D. Thesis, Massachusetts Institute of Technology, 2009.

[99] Seizure Detection, https://www.persyst.com/technology/seizure-detection/. (Accessed 25 April 2019).

[100] F. Pauri, F. Pierelli, G. Chatrian, W. Erdly, Long-term EEG-video-audio monitoring: computer detection of focal EEG seizure patterns, Electroencephalography Clin. Neurophysiol. 82 (1) (1992) 1–9.