Expert Systems With Applications 232 (2023) 120585

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Patient-independent epileptic seizure detection by stable feature selection

Lina Abou-Abbas [a,b,*], Khadidja Henni [a,b], Imene Jemal [b,c], Amar Mitiche [c], Neila Mezghani [a,b]

[a] *Imaging and Orthopedics' Research Laboratory, The CHUM Research Center, Montreal, Canada*
[b] *Research Center LICEF, Teluq University, Montreal, Canada*
[c] *INRS- Centre Énergie, Matériaux et Télécommunications, Montréal, Canada*

ARTICLE INFO

Keywords:
Epileptic seizure detection
EEG signals
Linear and non-linear features
Features selection
Usupervised graph centrality
Stability measures
Random forest

ABSTRACT

The detection of epileptic seizures in EEG signals is a challenging task because it requires careful review of multi-channel EEG recordings over a lengthy time interval. In general, EEG-based seizure detection is strongly dependent on the ability to select descriptive features that are also stable in the sense that they are not sensitive to changes in the training data. This study proposes and investigates a patient-independent seizure detection model that uses stable EEG-based features obtained by comparing multiple feature selection methods. The schemes considered can be divided into five categories, often referred to as similarity, information theoretic, sparse learning, statistical, and graph centrality feature selection methods. The stability of the features these schemes produce was evaluated by random forest classification, using the intersection, frequency, correlation, and similarity measures. In experiments with Temple University Hospital's Seizure database of EEG records of 341 patients, unsupervised graph centrality feature selection was the most effective method, with a correct classification rate of 91.5%.

1. Introduction

Despite taking anti-seizure medication, about one-third of epileptic patients continue to experience unprovoked seizures. This persistent occurrence of uncontrolled seizures negatively affects the quality of life of patients and of caregivers as well. For a patient who suffers from recurrent epileptic seizures, the decision whether to administer an anti-seizure drug or undergo surgery is generally based on an accurate diagnosis of their seizures. A precise quantification, localization, and analysis of the seizures' pattern prior to, during, and after the onset of seizures could help clinicians to customize treatment objectively during epilepsy surgery or in hospital epilepsy monitoring units. Seizures are caused by abnormal electrical activity in one or more areas of the brain. Prolonged Electroencephalography (EEG) recording remains the most reliable and common way to detect abnormalities in brain functioning that might be causing seizures, as well as to determine their type and localization for treatment. Performing visual scanning of long-term EEG recordings lasting several days is time-consuming and prone to errors. To simplify a physician's work, attempts to develop automatic seizure systems based on the analysis of long-term EEG monitoring have been increasing in number and scope.

Several seizure detection algorithms have been proposed, using classical signal processing techniques (Abou-Abbas et al., 2022; Acharya et al., 2015; Alam & Bhuiyan, 2013; Fu et al., 2014; Sharma & Pachori, 2015). These typically involve three main stages: pre-processing, feature extraction and classification. The focus has been on identifying discriminant features to distinguish between seizure and seizure-free segments. Multi-channel EEG signals allow the extraction of a significant amount of data via linear and non-linear methods in time, frequency and time–frequency domains (Jemal et al., 2021b; Kalin et al., 2020; Shantha Selva Kumari & Prabin Jose, 2011; Tessy et al., 2017; Tzallas et al., 2009; Zhou et al., 2018). Time-domain analysis (TDA) involves extracting features such as energy, line length, and amplitude. The work in Tessy et al. (2017) developed a computationally efficient method of seizure detection using time-domain features and achieved 94.4% classification accuracy. Although TDA can be suitable for real-time applications and provide better spatial information, it lacks information on frequency content. A frequency-domain analysis (FDA) involves describing the signal in terms of its frequency by assuming its periodicity. Based on a comparison between FDA and TDA on two public databases, Zhou et al.. Zhou et al. (2018) concluded that FDA was more effective at detecting seizure segments than the TDA method. A main limitation of the FDA is the selection of an appropriate window size when the signal lacks stationarity. To address the non-linear and non-stationary nature of EEG signals, time–frequency-based techniques have been successfully used in a number of studies for the detection of epileptic seizures (Kalin et al., 2020; Shantha Selva Kumari

* Corresponding author at: Imaging and Orthopedics' Research Laboratory, The CHUM Research Center, Montreal, Canada.
E-mail addresses: lina.abou-abbas@teluq.ca (L. Abou-Abbas), Khadidja.henni@teluq.ca (K. Henni), Imene.Jemal@inrs.ca (I. Jemal), Amar.mitiche@inrs.ca (A. Mitiche), Neila.mezghani@teluq.ca (N. Mezghani).

https://doi.org/10.1016/j.eswa.2023.120585
Received 21 October 2022; Received in revised form 3 May 2023; Accepted 27 May 2023
Available online 15 June 2023

& Prabin Jose, 2011; Tzallas et al., 2009). Most of these methods use the short-time Fourier transform (STFT), the wavelet transform, and power spectral density (PSD). The STFT, which uses the fast Fourier transform (FFT) of successive segments of a signal, has been effectively applied to seizure detection (Kalin et al., 2020). Other studies have applied wavelet transforms rather than the FFT (Alickovic et al., 2018; Kumar & Kolekar, 2014). Empirical mode decomposition (EMD) has also proved beneficial to the analysis of nonlinear and non-stationary data (Orosco et al., 2009). Furthermore, a wide range of applications using power spectral density (PSD) to study the power distribution across a wide frequency range have been investigated (Greene et al., 2008; Tzallas et al., 2009). Recent studies suggest that epileptic seizures can be quantified by studying the changes in brain dynamics indicative of signal complexity; these have also been able to distinguish seizure and non-seizure segments. For this purpose, several entropies in the time, frequency and time–frequency domains have been used in epilepsy detection, including the Shannon entropy and approximation entropy, as well as spectral entropy and wavelet entropy. These outcomes reveal the existence of significant differences between seizure activity and non-seizure activity (Kannathal et al., 2005; Kumar et al., 2013; Li et al., 2014; Rajendra Acharya et al., 2012). In neonatal seizures, Temko et al. (2011) extracted 55 features grouped in the frequency domain (such as wavelet energy, peak frequency of the spectrum, normalized power in sub-bands), time-domain (including curve length, number of maxima and minima, Hjorth parameters, and zero crossings), and information theory (including Shannon entropy, spectral entropy, and Fisher information). The features were combined as a single feature vector and input to a support vector machine classifier. The results showed a good detection rate of 89% for neonatal seizures. Wavelet transforms and empirical mode decomposition were used in a study by Parvez and Paul (2014) to compute several features of EEG signals, such as the entropy and energy of high-frequency coefficients. In their study, Alickovic et al. (2018) analyzed six different statistical features and found that the wavelet packet-based feature vector (including the mean of coefficient's absolute values, the average power of coefficients, standard deviation, and the ratio of absolute mean values of adjacent sub-bands), resulted in the best overall accuracy. Along with feature extraction and selection, several substantial studies have been devoted to the statistical learning of classifiers, such as Support vector machines (SVM), k-nearest neighbors (k-NN), random forest (RF), decision tree (DT), and neural networks (NN), which have proven useful in seizure and type of seizure detections (Abou-Abbas et al., 2021a; Alickovic et al., 2018; Jemal et al., 2021a; Kumar & Kolekar, 2014; Song et al., 2012). Even though machine learning applications have enabled the development of automated EEG-based seizure detection approaches, further research is required to improve the sensitivity and reduce the false-positive rate of recurrent methods, as well as to confirm their accuracy and their clinical relevance by validating results across larger and more diverse cohorts.

Since both linear and nonlinear features extracted in time, frequency, and time–frequency domains have been effective to some degree in epileptic seizure research, picking one or some combined categories of these features will have its benefits and drawbacks.

Considering EEG data complexity and the large number of features in use in various studies, the selection of salient and stable features appears to be crucial because high-dimensional feature vectors complicate classification and, therefore, can negatively impact classification. Feature selection is driven by the fact that classifiers built using a smaller feature space are generally more robust and reproducible than those built using a large feature space (Khaire & Dhanalakshmi, 2019). Ensuring a smaller feature space is now considered a key step of almost all classification frameworks. By reducing the complexity of the feature space, the computational cost of the classifier is reduced, its performance generally improves, and its implementation in clinical practice can be justified. Moreover, controlling the quantity and quality of features in machine learning helps to reduce over-fitting, which leads to a better association between features and target classes, and an improved classification generalization ability. Research has been done on feature selection methods for epileptic seizure detection to assess which features are relevant for optimizing detection performance while minimizing computing cost by removing irrelevant features (Bou Assi et al., 2015; Direito et al., 2011; Senawi et al., 2017; Wei & Billings, 2006). These methods can be categorized as supervised or unsupervised, univariate, or multivariate, and further separated by whether they make use of a classifier during the selection process. Previous research focused on identifying relevant features for the distinction between seizure and seizure-free segments, but did not investigate how features contribute to detection or how stable they are when applied across different training samples or new datasets. The lack of evidence about feature stability may explain in part why these studies cannot be replicated and their results cannot be generalized, and reveals the lack of extensive analysis of EEG data aiming at retaining descriptive and stable features. Moreover, the majority of machine learning studies currently available in the literature are primarily patient-specific studies based on one classifier being trained for each subject, rather than a universal classifier trained for all subjects. This is partly due to the high inter-subject variability that leads patient-independent models to perform much below the level of patient-specific models. This inter-subject variability highlights the need to investigate techniques that effectively address the seizure detection task in a patient-independent manner. To that end, this study proposes a patient-independent model for the detection of epileptic seizures that provides accurate detection of seizure segments based on a stable subgroup of features. It has three key objectives: (1) to summarize feature representations and their interpretation related to epileptic seizure patterns with the use of multi-channel EEG data; (2) to resolve high dimensionality issues by considering seven supervised and unsupervised feature selection methods, including univariate and multivariate methods to determine a subgroup with a minimal set of informative attributes; and (3) to evaluate the stability of the selected subgroup by using various stability measures.

The remainder of this paper is organized as follows: Section 2 gives an overview of the dataset used in this study 2. Section 3 presents a description of the framework, as well as details and background for each step of the framework. The results and a comparison between the different features selection techniques are presented in Section 4. Section 5 contains a discussion of the results obtained and compares them to those obtained by others. Finally, Section 6 contains a conclusion and some suggestions for future work.

2. Dataset

The data used in this study were obtained from the Temple University Hospital EEG Seizure Corpus (TUSZ) v1.5.1. This dataset was recently made openly available for researchers to address the lack of EEG datasets in EEG seizure research. The data is from a total of 341 patients, of which 188 are female. It consists of 886 sessions broken up into 7634 files, with each file lasting between one second and an hour. A total of 1760 of these files contain EEG seizure segments, comprising a total of 40.41 h, which represents approximately 6% of the overall data. The EEG signals were recorded in a real-time clinical environment using the 10/20 international standard system with over 19 channels and a minimum sampling rate of 250 Hz. The start and end of each seizure were manually annotated by trained researchers using an open-source annotation tool. All portions of artifacts and eye blinks were purged from the recordings. Metadata information relevant to the EEG interpretation including signal conditions, type and location of the session, and patient demographics including gender, age, as well as medication history are also provided. Further details on the data used in this study are described in Table 1.

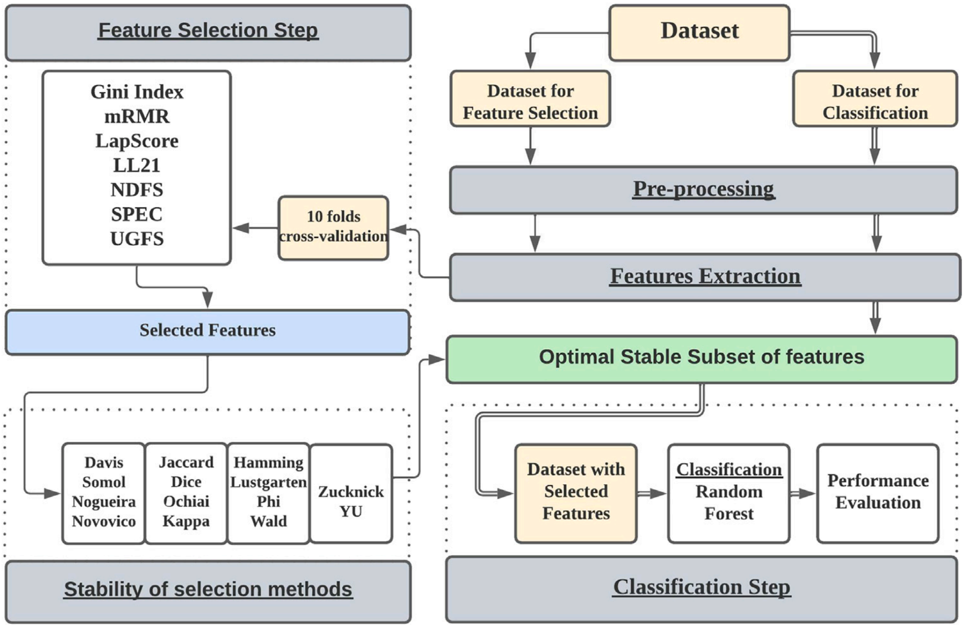

Fig. 1. The framework of the proposed methodology. It includes five main steps: Pre-processing, feature extraction, feature selection, stability assessment, and classification, followed by a performance evaluation.

Table 1
Overview of the subset of the Temple University Hospital EEG Seizure Corpus (TUSZ) used in our seizure detection experiments.

Nb of patients (Female)	341 (188 F)
Nb of patients with seizure (Female)	133 (72 F)
Total Nb of sessions	886
Total Nb of files	7634
Nb of Seizure files	1780
Nb of Seizure-free files	5854
Total duration in h.	655.36

3. Methods

The framework for building the classification model consists of two stages: a training stage and a testing stage. The training stage consists of five steps: data preprocessing, feature extraction, feature selection, stability assessment, and classification. Fig. 1 gives a block diagram of the approach.

3.1. Pre-processing

The EEG data were digitized at a sampling rate of 256 Hz, and then filtered with cut-off frequencies of (0.5:75) Hz followed by a 60 Hz notch filter to smooth out noise and power line interference. To provide a better description of the signal, the retained data was uniformly segmented into 4-, 5-, 10-, 15-, and 20-second intervals and then analyzed separately.

3.2. Features extraction

Given the inherent irregularity of EEG signals, it is insufficient to rely solely on linear features for the analysis of EEG activity. In light of this, a set of 44 univariate features were selected based on a combination of linear and nonlinear signal analysis techniques. The choice of these features was influenced by the commonly used features in EEG analysis and epilepsy seizure detection. Recent research findings have emphasized the significance of these features, which have been classified them into three main groups: time-domain, frequency-domain, time–frequency domain Abou-Abbas et al. (2021b), Alam and Bhuiyan (2013), Bandarabadi et al. (2015), Bosl et al. (2011), Hjorth (1970), Kumar et al. (2013), Orosco et al. (2009), Park et al. (2011), Rajendra Acharya et al. (2012), Ramadhani et al. (2019), Rezek and Roberts (1998), Şen et al. (2014). The list of all the extracted features used in this study is displayed in Table 2

Table 2
List of all the 44 features extracted.

Number	Features
1–5	Average value of Power Spectral Density (Delta, Theta, Alpha, Beta, Gamma)
6–10	Absolute value of Power Spectral Density (Delta, Theta, Alpha, Beta, Gamma)
11–15	Relative value of Power Spectral Density (Delta, Theta, Alpha, Beta, Gamma)
16–18	Skewness, Variance and Kurtosis
19–24	Features of Empirical Mode Decomposition (Energy, Spectral Entropy, Mean, Standard deviation, Moment, Skewness)
25–29	Sample Entropy, Permutation Entropy (4 levels)
30–32	Hjorth (Mobility, Activity, Complexity)
33	Spectral Entropy of PSD
34	Features of Discrete Wavelet Transform- Shannon Entropy
35–36	Features of Wavelet Packet Decomposition (Log Energy Entropy and Shannon Entropy)
37	Successive Decomposition Index
38–39	Mean Energy and its cumulative sum
40–44	Features of wavelet Decomposition (Percentage of energy-5 levels)

3.2.1. Time-domain features

A number of studies have shown the advantages of the time domain features (Hjorth, 1970; Ramadhani et al., 2019; Şen et al., 2014). This study uses nine time-domain features calculated directly from the EEG signals: three statistical features (skewness, variance, and Kurtosis); the three Hjorth parameters (activity, mobility, and complexity); the successive decomposition index; and two entropy-based features.

1. Statistical features:

   (a) *Skewness of the raw signal*:

   $$Skew = \frac{\sum_{i=1}^{N} \left(\frac{x_i - \overline{x}}{\sigma}\right)^3}{N-1}$$

   (b) *Variance of the raw signal*:

   $$v = \sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2$$

   (c) *Kurtosis of the raw signal:*

   $$\mathrm{Kurt} = \frac{\sum_{i=1}^{N} \left(\frac{x_i - \overline{x}}{\sigma}\right)^3}{N-1} - 3$$

2. The Hjorth parameters: time domain parameters

   (a) *Mobility* indicates the mean frequency of the power spectrum:

   $$H_{\mathrm{mob}} = \sqrt{\frac{\mathrm{var}\ (y'(t))}{\mathrm{var}\ (y(t))}}$$

   (b) *Activity* represents the mean power of the signal:

   $$H_{\mathrm{act}} = var\,(y(t))$$

   (c) *Complexity* represents the change in frequency:

   $$H_{\mathrm{com}} = \frac{\mathrm{mobility}\ \left(y'(t)\right)}{\mathrm{mobility}\ (y(t))}$$

3. The successive decomposition index (SDI) is derived by the motivation of the discrete wavelet transform. The SDI is a novel matrix determinant that has been used successfully in previous studies (Raghu et al., 2019) . The square matrix is formed using four coefficients calculated by an iterative process (Raghu et al., 2019).

   $$SDI = \log_{10} \frac{n}{L}(X^{+}X^{++} - X^{-}X^{--})$$

   with EEG in time series $x = \{x_1, x_2, x_3, \dots, x_n\}$, X^+: average of $|x|$, $X^+ = \frac{1}{n}\sum_{i=1}^{n} |x_i|$
   X^-: difference average of x, $X^{++} = \frac{X^+ + X^-}{2}$ *and* $X^{--} = \frac{X^+ - X^-}{2}$

4. Entropy features:

   (a) The sample Entropy is defined as:

   $$sampEn = \lim_{N\to\infty} \left\{ -\ln \left[\frac{A^m(r)}{B^m(r)}\right] \right\}$$

   For a given N data points of a time series $x(n) = x(1), x(2), \dots, x(N)$. $B^m(r)$ is the probability that two sequences will match for m points. $A^m(r)$ is the probability that two sequences will match for $m+1$ points.

   (b) Permutation entropy is defined as:

   $$E_{perm} = \sum_{j=1}^{D!} p_j \log_2 p_j$$

   where p_j are the relative frequencies of the signal and D is the relative order.

3.2.2. Frequency-domain features

The frequency-domain features used in this study are based on the power spectral density on each segment of each channel in the EEG data. The average of the power spectral density (PSD) was calculated using the Welch method over all electrodes in the five frequency bands: delta (0.4–4 Hz); theta(4–8 Hz); alpha(8–13 Hz); beta(13–30 Hz) and gamma(30–75 Hz). Additionally, the Spectral Entropy was determined using the normalized power spectral distribution of the EEG signal (P_f), it is calculated as:

$$SpEn = -\sum_{k=1}^{N-1} P_k \log_2 P_k$$

3.2.3. Time–frequency domain features

1. EMD-based features
   The Empirical Mode Decomposition (EMD) technique has gained widespread use in the analysis of EEG signals due to its ability to address the non-stationarity inherent in such signals. Abou-Abbas et al. (2021b), Alam and Bhuiyan (2013), Oweis and Abdulhay (2011), Wang et al. (2012), Zeng et al. (2019).
   The EMD approach divides the signal into multiple frequency bands, preserving the temporal variation of frequency. The sifting algorithm is then applied to decompose the signal into a finite number of intrinsic modes. After several evaluations, it was determined that only the IMF1, IMF2, and IMF3 levels are relevant for the classification task, as IMF4 and higher levels provide minimal benefit. A total of 6 features describing the changes and statistical distribution of the IMF were used.

   (a) Energy

   $$F_{IMF_i} = -\sum_{n=0}^{N-1} |IMF_i[n]|^2$$

   (b) Spectral Entropy

   $$SpEn_{IMF_i} = -\sum_{k=1}^{N-1} IMF_k \log_2 IMF_k$$

   (c) Mean

   $$\mu_k = \frac{1}{N} \sum_{n=0}^{N-1} IMF_k[n]$$

   (d) Standard Deviation

   $$\sigma = \sqrt{sum_N^{i=1} \frac{(IMF_k[i] - \overline{IMF_k})^2}{N-1}}$$

   (e) Skewness

   $$SkewIMF = \frac{\sum_{i=1}^{N} \left(\frac{IMF_k[i] - \overline{IMF_k}}{\sigma}\right)^3}{N-1}$$

2. Wavelet-based features
   In the present paper, the signal was analyzed through 5 levels of decomposition using 5 Daubechies wavelet functions. Tables 3 and 4 show the five-level decomposition of EEG signals using wavelet and wavelet packet transforms, respectively.
   Entropy-based wavelet features have gained prominence in recent times due to their capability to quantify the complexity and irregularity of EEG signals across multiple time scales, as well as to determine the level of uncertainty in the signal and evaluate its dynamic order. These features have demonstrated success in the detection of seizures in several prior studies. Bosl et al. (2011), Kumar et al. (2013), Rajendra Acharya et al. (2012), Rezek and Roberts (1998).

   (a) Wavelet Shannon Entropy based on discrete wavelet transform

   $$WavEnt = -\sum_{j=1}^{D!} p_j \log_2 p_j$$

   Where p_j is the normalized wavelet energy.

Table 3
Five levels of decomposition of EEG signal using a wavelet transform.

Decomposition level	Frequency range (Hz)	EEG band
D1	64–75	Higher gamma
D2	32–64	Lower gamma
D3	16–32	Beta
D4	8–16	Alpha
D5	4–8	Theta
A5	0–4	Delta

Table 4
Five levels of decomposition of EEG signal using wavelet packet transform.

Decomposition level	Frequency range (Hz)
D1	64–75
A1	0–60
D2	32–64
A2	0–30
D3	16–32
A3	0–16
D4	8–16
A4	0–8
D5	4–8
A5	0–4

(b) Log Energy Entropy based on wavelet packet transform

$$LogEnt = -\sum_{i=0}^{N-1} (\log_2 (p_i(x)))^2$$

(c) Shannon Entropy based on wavelet packet transform

$$ShEn = -\sum_{i=0}^{N-1} (p_i(x))^2 (\log_2 (p_i(x)))^2$$

3.3. Features selection

In view of the large number of features, it is important to consider the feature selection step as a key component of the framework, a step that removes redundant and irrelevant features, increases classification performance, and helps to ensure feature stability.

Our study compared seven feature selection (FS) algorithms, four of them filter methods that do not require a learning process and feedback from predictors: the Gini-index method, the minimum redundancy maximum relevance (mRMR), the spectral feature selection (SPEC) and Laplacian score (LapScore), and three from embedded methods that are tightly coupled with an embedded clustering algorithm and that depend on the classification result: LL21, non-negative discriminative feature selection (NDFS) and unsupervised graph-based feature selection (UGFS). The filter methods differ according to the ranking concept. Gini-Index uses statistical information; LapScore and SPEC methods use similarity and mRMR uses information theory. The embedded methods are further categorized based on the clustering technique: sparse learning (NDFS and LL21) or graph centrality (UGFS). An FS method would be considered univariate if it evaluates each feature individually, regardless of its relevance, and determines the relationship between each feature and the class. Meanwhile, if it considers a subset of features as well as their interactions, then it would be considered multivariate. Fig. 2 shows the hierarchical structure of the selection methods used in this study.

Feature selection based on the Gini index was originally developed by Breiman et al. (2017) to sort features based on their impurity. It is a multivariate, supervised filter-based method. The lower the value of the Gini index, the better the feature.

The mRMR, a multivariate feature selection approach, was originally developed by Ding and Peng (2005). It is a supervised filter-based

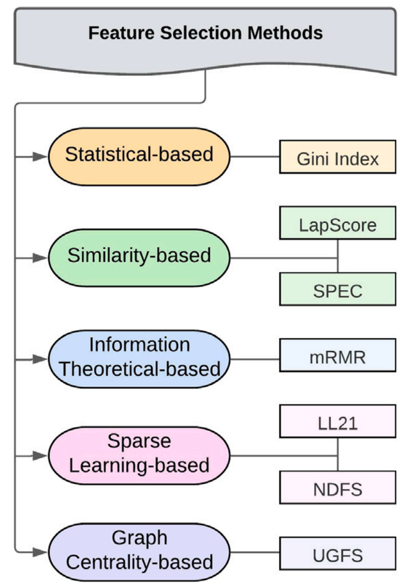

Fig. 2. The hierarchical structure of the feature selections methods used in this research grouped in categories based on the procedures employed for selection.

method that combines similarity and importance criteria to select features based on mutual information. For similarity, the mRMR considers the minimum redundancy between features, and for importance, the maximum relevance between individual features and the target class. Each feature is given an index, where 0 is the most relevant feature and so forth. The mRMR is a powerful method for feature selection and previous studies indicate that it leads to significant improvement in classification accuracy (Memar & Faradji, 2018; Peng et al., 2005). The main advantages of this approach are its speed and scalability.

The Laplacian score (LS) method proposed by He et al. (2005) can be used for both supervised and unsupervised feature selection. It is a univariate filter-based feature selection approach based on finding spectral similarity between features. It estimates a feature's importance based on their Laplacian score to rank its relevance and evaluate its locality-preserving ability: this is a measure of similarity between one feature and the nearby instances on the graph. The lower the Laplacian score, the more significant the feature.

The SPEC algorithm proposed by Zhao and Liu (2007) is a univariate filter-based method for supervised and unsupervised feature selection that uses spectral graph theory to find structural information. The highest score is given to relevant features that offer better separability. After building a similarity set, constructing the graph representation, and evaluating the features of the graph using its spectrum, the algorithm ranks each feature according to its relevance.

Li et al. introduced an unsupervised non-negative Discriminative Feature Selection (NDFS) (Li et al., 2012). This method computes the weight matrix to score features. It is comprised of two steps, The first investigates the cluster structure of the random data by means of spectral analysis of the non-negative matrix factorization, and the second selects the features over the whole feature space by using sparsity regularization to preserve the estimation cluster structure and to provide the scores to the most discriminative features. Shi et al. (2014).

An unsupervised graph-based feature selection (UGFS) method was developed by Henni et al. (2018, 2020). UGFS is an unsupervised feature selection algorithm that operates by iteratively selecting and removing the most irrelevant features from a dataset. The algorithm uses subspace preference clusters to define relationships between features and PageRank to rank the importance of each feature. The subspace preference clusters are sets of data points that share similar dense

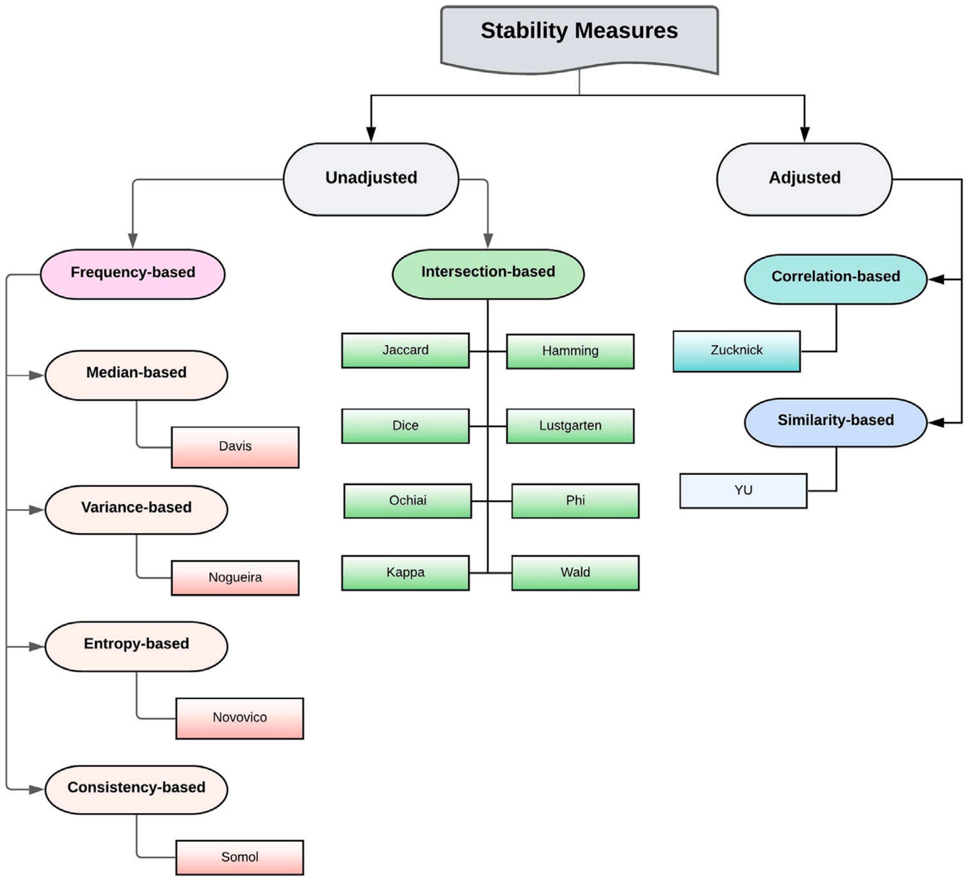

Fig. 3. Hierarchical structure of the stability measures.

regions, which are associated with a set of features that have low variance in the neighborhood of the points. These features are considered to be relevant and are preserved in the clusters, while the irrelevant features are filtered out. The PageRank algorithm computes a normalized and propagated value for each feature in the dataset based on the importance of all connected features. This importance is based on the expected sum of the importance of all connected features, with the direction of the edges being defined by the correlation between the features. The resulting PageRank scores are used to rank the importance of each feature, with the least important features being removed in each iteration of the algorithm. The iterative process of UGFS continues until a stopping criterion is met, such as reaching a predefined number of selected features or reaching a certain level of classification accuracy. UGFS has been shown to be effective in reducing the dimensionality of high-dimensional datasets while preserving the classification accuracy of the original dataset.

3.4. Stability

The stability of a feature selection approach is determined by the ability to generate similar features regardless of perturbations in the training data. A stable feature selection approach will select a reproducible set of relevant features. This set could be, for example, a dataset split into different folds for training and validation. Disregarding the feature selection algorithm's stability issue can lead to erroneous conclusions and the inability to replicate and generalize the results. Therefore, the best approach is to combine the feature selection approach with stability analysis to realize a high-quality, consistent, and reliable subset of features.

Stability measures exist in many forms to assess feature selection stability, but it remains unclear which measure is the best to use. Therefore, we looked at the stability of feature selection methods using 14 stability measures split into two categories, either unadjusted or adjusted. The unadjusted measures are grouped according to the criteria they are based on: intersection-based stability measures that consider the cardinalities of all pairwise intersections, and frequency-based stability measures that focus on the frequency of selectivity of all features using the following criteria: median, variance, entropy and consistency. The adjusted measures are grouped based on the similarity and correlation between features. Fig. 3 shows a hierarchical structure of the stability measures used in our study and Tables 5, 6 and 7 show the mathematical foundation of these measures as well as their maximum and minimum values. A detailed review of these measures can be found in Bommert (2020). with $X_1, X_2 \dots X_n$ features extracted and $V_1, V_2 \dots V_m$ the set of chosen features for the m subsets h_j the absolute frequency with which feature X_j is chosen.

3.5. Classification and performance evaluation

Several classifiers have been presented for the detection of epileptic seizures using EEG signals (Bhattacharyya et al., 2018; Faust et al., 2010; Mahjoub et al., 2020; Orhan et al., 2011; Sharma et al., 2017). In this work, the detection of epileptic seizures was performed using the random forest classifier, as it has shown good performance in several previous works (Bhattacharyya et al., 2018; Mahjoub et al., 2020; Orhan et al., 2011).

A Random Forest algorithm (Breiman, 2001) is an ensemble learning method in which each tree of the ensemble is an individual predictor created using a randomly selected subset of data. This algorithm is based on the concepts of random subspaces and bagging.

We evaluated our proposed method based on accuracy, F1 score, precision, sensitivity, and the area under the ROC curve. To avoid

Table 5
Intersection-based stability measures.

Stability measures	Formula	Min value	Max value
Dice	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{2\|V_i \cap V_j\|}{\|V_i\|+\|V_j\|}$	0	1
Hamming	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\|V_i \cap V_j\|+\|V_i^c \cap V_j^c\|}{p}$	0	1
Jaccard	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\|V_i \cap V_j\|}{\|V_i^c \cup V_j^c\|}$	0	1
Kappa	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\|V_i \cap V_j\|-\frac{\|V_i\|.\|V_j\|}{p}}{\frac{\|V_i\|+\|V_j\|}{2}-\frac{\|V_i\|.\|V_j\|}{p}}$	−1	1
Lustgarten	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\|V_i \cap V_j\|-\frac{\|V_i\|.\|V_j\|}{p}}{min\{\|V_i\|,\|V_j\|\}-max\{0,\|V_i\|+\|V_j\|-p\}}$	−1	1
Ochiai	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\|V_i \cap V_j\|}{\sqrt{\|V_i\|.\|V_j\|}}$	0	1
Phi	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\|V_i \cap V_j\|-\frac{\|V_i\|.\|V_j\|}{p}}{\sqrt{\|V_i\|(1-\frac{\|V_i\|}{p}).\|V_j\|(1-\frac{\|V_j\|}{p})}}$	−1	1
Wald	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\|V_i \cap V_j\|-\frac{\|V_i\|.\|V_j\|}{p}}{min\{\|V_i\|,\|V_j\|\}-\frac{\|V_i\|.\|V_j\|}{p}}$	1-p	1

Table 6
Frequency-based stability measures.

Stability measures	Formula	Min value	Max value
Davis	$\max\left\{0,\frac{1}{\|V\|}\sum_{j=1}^{p}\frac{h_j}{m}-\frac{penalty}{p}.median\{\|V_1\|,\|V_2\|,\ldots\|V_m\|\}\right\}$	0	1
Nogueira	$1-\frac{\frac{1}{p}\sum_{j=1}^{p}\frac{m}{m-1}\frac{h_j}{m}\left(1-\frac{h_j}{m}\right)}{\frac{q}{mp}\left(1-\frac{q}{mp}\right)}$	−1	1
Novovicova	$\frac{1}{q\log_2(m)}\sum_{j:X_{j\in V}}h_j log_2(h_j)$	0	1
Somol	$\frac{\left(\sum_{j=1}^{p}\frac{h_j}{q}\frac{h_{j1}}{m-1}\right)-c_{min}}{c_{max}-c_{min}}$	0	1

Table 7
Adjusted stability measures.

Stability measures	Formula	Min value	Max value
Yu	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{I(V_i,V_j)-E(I(V_i,V_j))}{\frac{\|V_i\|+\|V_j\|}{2}-E(I(V_i,V_j))}$	NA	1
Zucknick	$\frac{2}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\frac{\|V_i^c \cap V_j^c\|+C(V_i,V_j)+C(V_j,V_i)}{\|V_i^c \cup V_j^c\|}$	0	1

overfitting, a 10-fold cross validation procedure was used, in which the classifier was trained using 9 folds and tested on the remaining fold. This procedure was repeated ten times and the average of the performance measures across all experiments was considered. Additionally, to ensure that no patient information was used in both the training and testing phases, we employed the GroupKFold technique. This method ensures that samples from the same patient are always kept in the same fold, thereby avoiding any contamination of the test set with training information.

4. Experimental results

In this work, we explored a set of linear and non-linear features and different feature selection techniques for epileptic seizure detection. The goal of feature selection is to help in the development of a better classifier by highlighting important features, while simultaneously lowering computational overload. We evaluated the performance of our methodology using EEGs from 341 patients in Temple University Hospital's EEG seizure database. Our approach extracted 44 quantitative EEG features from the time, frequency, and time–frequency domains, including linear and nonlinear features, combined them across 19 channels to reveal the spatial and temporal patterns of seizures and to allow them to be used as a high-dimensional input vector for a total of 856 features per each window size.

The entire dataset was split into two sets, one containing 33%, reserved for the selection procedure, and the other set containing 67%, reserved for the modeling and classification steps. To ensure that the samples from each patient are kept together, the data was grouped by patient ID and split into training and testing sets using a stratified split by ID. The training set was then balanced using SMOTE to reduce any bias towards the majority class and improve the performance of the classifier. Feature selection (FS) methods were then applied to the set reserved for selection to score each feature by an index or weight. LapScore, NDFS, SPEC, and UGFS were used as unsupervised methods, while Gini Index, mRMR, and LL21 were used as supervised methods. To create a model that was capable of distinguishing seizure segments from non-seizure segments, we used a random forest classifier (RF) following the feature selection step. Classification experiments were

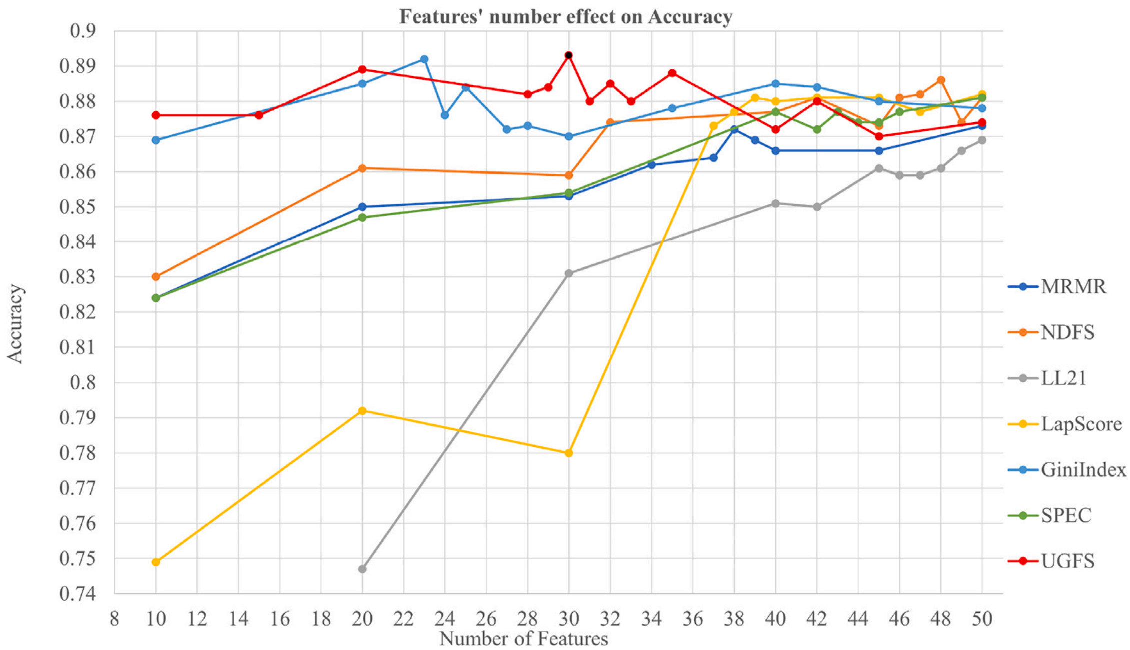

Fig. 4. Seizure detection classification performance using features from seven different feature selection methods. The impact of features' number on the performance. A high value of performance is obtained using UGFS method for a total number of features of 30.

Table 8
Seizure detection classification mean performance metrics in terms of F1-score, precision, and sensitivity of the seven feature selection methods obtained from 10-fold cross-validation as well as the standard deviation in brackets. The performance of the baseline classifier without any feature selection in the last row as well as the optimal number of features obtained for each approach in the last column.

Feature selection	F1-score	Precision	Sensitivity	Nb of features
Gini Index	90.13 (0.016)	90.62(0.025)	90.12 (0.029)	23
mRMR	87.18 (0.011)	87.26(0.024)	87.18(0.011)	38
Lap Score	88.8(0.022)	88.98(0.039)	88.79(0.027)	39
LL21	87.1(0.022)	87.31(0.039)	87.11(0.027)	45
NDFS	88.56(0.019)	88.74(0.038)	88.56(0.029)	42
SPEC	88.89(0.042)	89.13(0.068)	88.89(0.050)	42
UGFS	93.20(0.054)	91.90(0.072)	94.50(0.070)	30
Baseline Without FS	75.62(0.076)	74.26(0.103)	78.43(0.094)	856

repeated many times by varying the feature numbers while respecting the scores given to them by each of the FS methods. To tune the hyperparameters of the Random Forest model, we performed a grid search over a range of values for the number of estimators (100, 200, 300, 400 and 500). The optimal number of trees was found to be 100. It was selected based on the highest accuracy score on the validation set, which was obtained through 3-fold cross-validation. The classification performance in terms of accuracy resulting from different numbers of features is shown in Fig. 4. As indicated in the figure, all models perform better as the number of features increases until 20 features, at which point their performance begins to remain constant. This is particularly true for UGFS, mRMR, Gini Index, NDFS, and SPEC methods. The peak performances were achieved using UGFS with 30 features and Gini Index with 23 features. As the number of features increases to 50, none of the models improve their performance. This figure shows that UGFS methods provide much better results than other selection methods.

Performance levels in terms of F1-score, precision and sensitivity are indicated in Table 8 for the models created using the seven feature selection methods, as well as for the model generated using the full set of features without any feature selection step. The optimal number of features varies with each approach. Table 8 offers details about the optimal number of features. With UGFS, a satisfying F1-score of 93.20%, precision of 91.90%, and sensitivity of 94.50% were achieved. The Gini Index, with 23 features, also produced a satisfactory F1 score of 90.13%, precision of 90.62%, and sensitivity of 90.12%. For UGFS, the number of features decreased from 856 to 30 while the F1 score increased from 75.62 to 93.20%, the sensitivity improved from 78.43 to 90.12%, and the precision grew from 74.26 to 91.90%. These results were obtained with 10-fold cross validation repeated ten times. The performance measures are the average over the 10 cross validation folds. It is important to note that all the feature selection algorithms gave a satisfactory performance value of 86% or higher, a significant improvement over not using a feature selection algorithm.

Experiments were performed to investigate the stability of selected features. Ten-fold cross-validation was used to split each feature selection dataset into ten different folds and then apply a feature selection step to each fold separately. In this step, the feature selection algorithm produces a feature preference for each fold in the training dataset. A total of 10 feature vectors were obtained for each FS method, and fourteen stability measures were then computed. The results in Table 9 show that the NDFS, SPEC and UGFS methods are more stable than the other four. The features' ranking remains the same with different training sets, at least for the 30 first features with the UGFS and the first 42 features for the NDFS and SPEC methods. A low stability level was obtained using LapScore.

Fig. 5 shows the seven subsets of channels from which the output vectors of feature selection methods were extracted. The results indicate that features selected using mRMR, NDFS, SPEC and UGFS were extracted mainly from channels F8 (in green) and T3 (in blue), located in the frontal and temporal regions, respectively.

Table 10 shows the feature selection results according to the indices (check Table 2 for feature indices) and the channels. The relative power

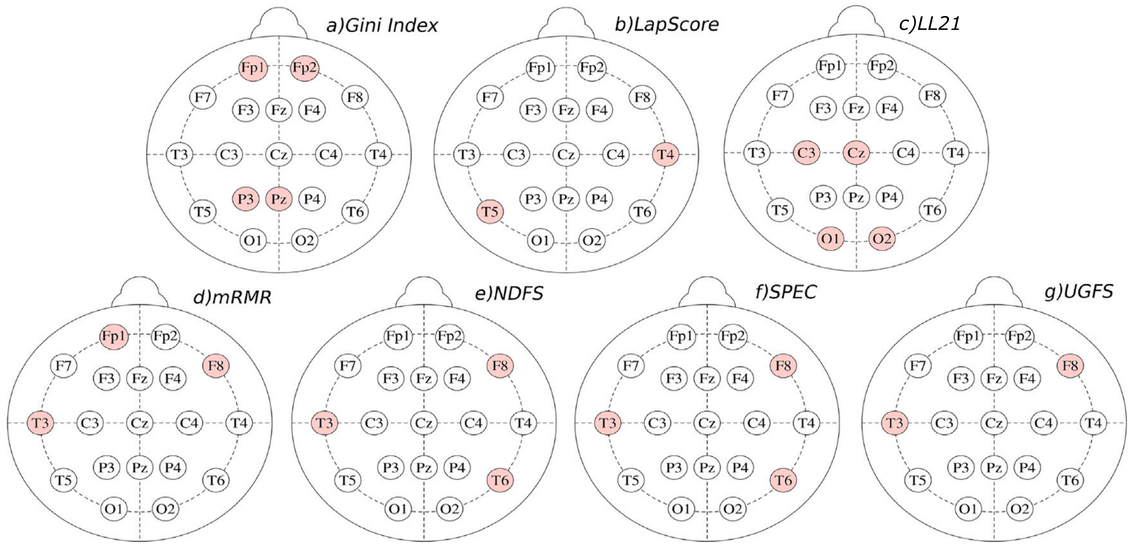

Fig. 5. Seven subsets of the EEG channels from which the output vectors of FS methods was extracted.

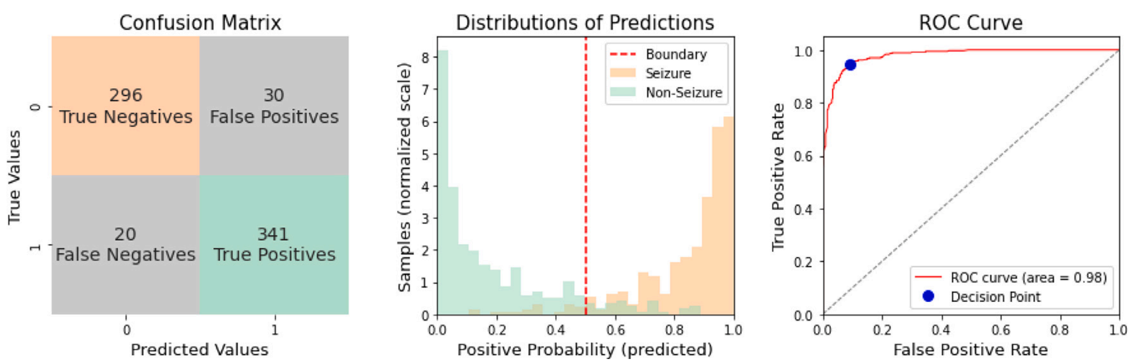

Fig. 6. From left to right: Confusion matrix, Prediction distribution, and the ROC curve of the best model created using Random Forest, the UGFS method and a 20-second window. In the confusion matrix, the blocks from top to bottom and from left to right indicate the number of true negatives, false negatives, false positives and true positives, respectively. Zero (0) represents the non-seizure class and 1 the seizure class. For the prediction score distribution, any score lower than the boundary will be predicted as 0, and any score above the boundary will be predicted as 1. For the ROC curve, with an area of 0.98, it can be said that the test accurately distinguishes between seizure and non-seizure segments. Precision: 0.919; Recall: 0.945; F1 Score: 0.932; Accuracy: 0.927; Sensitivity: 0.945; and Specificity: 0.908.

Table 9
Stability results for the different stability measures computed for each of the seven feature selection methods.

Stability measure	GiniIndex	LapScore	LL21	mRMR	NDFS	SPEC	UGFS
Davis	0.77	0.32	0.82	0.95	1	1	1
Dice	0.90	0.36	0.92	0.99	1	1	1
Hamming	0.99	0.94	0.99	0.99	1	1	1
Jaccard	0.82	0.27	0.86	0.98	1	1	1
Kappa	0.89	0.33	0.92	0.99	1	1	1
Lustgarten	0.87	0.31	0.87	0.95	1	1	1
Nogueira	0.89	0.33	0.92	0.99	1	1	1
Novovicova	0.94	0.57	0.95	0.99	1	1	1
Ochiai	0.90	0.36	0.93	0.99	1	1	1
Phi	0.89	0.33	0.92	0.99	1	1	1
Somol	0.90	0.36	0.93	0.99	1	1	1
Wald	0.89	0.33	0.92	0.99	1	1	1
Yu	0.89	0.30	0.94	0.99	1	1	1
Zucknick	0.82	0.28	0.86	0.98	1	1	1

of the Delta, Theta, Alpha, Beta and Gamma bands are among the features selected specifically for the three most stable methods. The absolute power of the gamma band, sample entropy, and Shannon entropy using wavelet decomposition, as well as the log energy entropy and Shannon entropy derived from wavelet packet decomposition and finally the permutation entropy at levels 1, 3 and 4 are among feature vectors selected by the three most stable methods.

An investigation of the efficacy of five different window sizes was also conducted. The region that conserves signal stationarity is between 0 and 20 s, and so we limited our experiments to this interval. The training and testing processes were repeated for each window size using the feature vector from the most stable feature selection approach: UGFS. As shown in Table 11, performance in terms of accuracy reaches its highest value with a 20-second window size of 91.55%. From 4 to 20 s, accuracy increased by 2.22% and sensitivity by 1.61%. Overall, the accuracy only varied by small amounts. Fig. 6 displays the confusion matrix, the score distribution of the binary classification model and the ROC curve of the best model created using the Random Forest classifier, the UGFS method and a 20-second window. The area under the ROC curve of 0.98 can be interpreted as indicating that the test accurately distinguishes a seizure from a non-seizure.

5. Discussion

Diagnosis methods for seizures, both machine and deep learning-based, have limitations in their approach as they are primarily focused on patient-specific methods. They train and test EEG samples from the same or similar patient groups, which results in the models only being able to account for factors within a patient, but not differences between patients. This makes accurate seizure diagnosis even more challenging in a clinical setting where the testing patient has not been seen during

Table 10
The set of extracted features by indices for each FS method. The channel from where the features were extracted is also illustrated.

Feature selection method	Nb of features	Channel	Selected features according to their indices
GiniIndex	23	FP1	[40, 42, 16, 17, 2, 32, 21]
		FP2	[9, 13, 8, 15]
		Pz	[13, 12, 25, 11, 7, 15, 8]
		P3	[44, 3, 30, 22, 4]
LapScore	38	T4	[24, 17, 18, 2, 3, 4, 38, 39, 40, 41, 43, 44]
		T5	[31, 30, 32, 19, 20, 22, 23, 24, 17, 34, 16, 1, 2, 3, 4, 5, 38, 39, 40, 18, 35, 29, 27]
LL21	45	C3	[3, 1, 2, 26, 27, 26, 19, 16, 18, 30, 20, 21, 22]
		Cz	[2, 3, 4, 5, 29]
		O1	[44, 43, 42, 41, 24, 4, 18, 17, 32]
		O2	[36, 35, 29, 14, 2, 5, 6, 3, 33, 12, 15, 6, 9, 28, 25, 26, 13, 27]
mRMR	38	F8	[38, 40, 43]
		FP1	[6]
		T3	[35, 22, 36, 20, 1, 32, 17, 27, 28, 16, 18, 34, 31, 23, 29, 30, 20, 21, 11, 10, 2, 7, 19, 33, 8, 9, 32, 25, 13, 6, 14, 15]
NDFS	42	F8	[35, 24, 23, 22, 21, 20, 19, 32, 30, 31, 34, 36, 29, 8, 28, 26, 25, 33, 25, 15, 14, 13, 12, 11, 10, 17, 16, 18, 1, 44, 43]
		T3	[15, 14, 13, 12, 11, 10, 9, 8, 7, 6]
		T6	[37]
SPEC	42	F8	[35, 24, 23, 22, 21, 20, 19, 32, 30, 31, 34, 36, 29, 8, 28, 26, 33, 25, 15, 14, 13, 12, 11, 10, 17, 16, 18, 1, 44, 43]
		T3	[15, 14, 13, 12, 11, 10, 9, 8, 7, 6]
		T6	[37]
UGFS	30	F8	[6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 25, 33, 26, 27, 28, 29, 35, 36, 34]
		T3	[4, 5, 30, 32, 38, 39, 40, 41, 42, 43, 44]

Table 11
Performance results in terms of accuracy, F1 score, Precision, Sensitivity and ROCAUC for features selected using UGFS methods and 5 window sizes.

Window size	Accuracy	F1 score	Precision	Sensitivity	ROCAUC
4s	0.8933 (0.0088)	0.8932 (0.0088)	0.8707 (0.0145)	0.9251 (0.0068)	0.8929 (0.8929)
5s	0.8859 (0.0140)	0.8857 (0.0140)	0.8574 (0.0187)	0.9265 (0.0174)	0.8858 (0.8858)
10s	0.8994 (0.0099)	0.8993 (0.0100)	0.8781 (0.0158)	0.9263 (0.0098)	0.8996 (0.8996)
15s	0.9057 (0.0083)	0.9056 (0.0083)	0.8835 (0.0145)	0.9361 (0.0080)	0.9056 (0.9056)
20s	0.9155 (0.0114)	0.9155 (0.0114)	0.8959 (0.0176)	0.9412 (0.0116)	0.9156 (0.9156)

training. The field of seizure detection using EEG data has seen limited progress in developing patient-independent methods, which can effectively diagnose seizures regardless of the patient being seen during training. In response to this challenge, our study contributes to this field by proposing a unique approach that considers stability in feature selection, which results in a higher accuracy and sensitivity in seizure detection compared to the state-of-the-art methods. To the best of our knowledge stability has never been considered as a factor in seizure detection methods before. The subset of characteristics is termed stable when several iterations of a selection process from the training data with different random seeds result in only minor variations in the selection outcomes. Few studies have explored the patient-independent scenario (Orosco et al., 2016; Zhang et al., 2020; Zhao et al., 2022) In a patient-independent study by Adversarial Representation Learning for Robust Patient-Independent Epileptic Seizure Detection (Zhang et al., 2020), the best accuracy achieved using Random Forest was 61.9%, SVM 64.3%, KNN 69.9%, and deep learning 80.5% using a limited number of 14 patients from the same EEG seizure database. In another study (Orosco et al., 2016) authors proposed a set of features for 18 patients from the CHB-MIT database and reported an average sensitivity of 87.5%. In another study conducted in 2022 (Zhao et al., 2022) an accuracy of 76.36%, a specificity of 76.32% and sensitivity of 77.42%.

Our method, which incorporates an unsupervised graph-based feature selection technique and a Random Forest classifier, achieved a significantly higher accuracy of 91.55% and sensitivity of 94.12%. This result showcases the competitiveness of our approach in comparison to the previous studies, especially when considering the larger publicly available EEG seizure database used. The use of a larger and more diverse database increases the generalizability of our findings, as seizure detection algorithms that perform well on smaller datasets may not yield the same results on larger datasets. It is worth mentioning that the detection rate of 89% reported in Temko et al. (2011) study is also noteworthy, but our approach has demonstrated a higher performance in terms of accuracy and sensitivity.

The results also showed that our approach exhibits a high degree of stability based on frequency, intersection, correlation, and similarity. Our study sought to examine the stability of various feature selection methods, a critical aspect in ensuring that the optimal performance and minimum number of features selected are not just random occurrences.

The experimental work showed a significant increase in the performance of all FS methods compared to a model without any FS method. These results are not surprising, given that prior research showed the performance accuracy reached 100%. Additionally, other studies such as (Wang et al., 2021) that utilized smaller databases such as BONN, Neurology and Sleep Centre (NSC) and CHB-MIT reported similar results with performance accuracy reaching 100%. It is worth noting that the results of these studies are based on limited patient data.

The novelty of our research is its study of the stability of such feature selection methods to guarantee that a minimal number of features and performance obtained are not a result of a random selection and can remain constant despite many changes in training and testing datasets. By comparing stability profiles of the different feature selection algorithms, we found NDFS, SPEC and UGFS have same identical maximal stability scores of 1. We can conclude that the group of unsupervised methods (NDFS, SPEC and UGFS) based on similar principles of selection such as clustering and graphic representation techniques could share very similar high stability characteristics. Even though LapScore belongs to the group of unsupervised methods, its lower stability profile could be explained by the fact that LapScore is based on similarity techniques and not clustering or graph techniques.

Comparing the FS methods with respect to their stability measures in Table 9 and their performance measures in Table 8, we can conclude that even though the Gini Index gives the optimal reduced feature set with a satisfactory performance, for stability we prefer UGFS, SPEC and NDFS. Among those three, UGFS is preferable due to its higher performance and minimal number of selected features compared to SPEC and NDFS. By clustering the selected features by channel to determine the contribution of channels to the seizure detection, the results were surprising, leading us to conclude that most of the selected features from the most stable FS methods (mRMR, NDFS, SPEC and UGFS) were extracted largely from channels F8 and T3. A literature review supports our findings, with several studies highlighting the importance of stability in feature selection methods (Khaire & Dhanalakshmi, 2019; Nogueira et al., 2017). This is because stability ensures that the optimal performance and minimum number of features selected are not just a result of random chance, but are consistent and reliable across different training and testing datasets. Furthermore, the selection of features from specific channels such as F8 and T3 can improve efficiency and reduce the risk of overfitting, making it a valuable consideration in future research.

6. Conclusion

The main purpose of a seizure detection system is to improve long-term patient care by allowing remote monitoring, rapid intervention, and timely adjustment of therapy or treatment, such as anti-seizure drugs and electrical stimulation that could decrease the evolution of seizures. For the development of epileptic seizure detection systems, integrating advanced EEG signal processing and machine learning techniques can be considered an effective approach. In summary, in this study we proposed a patient-independent method for epileptic seizure detection based on a selected set of stable features. A common practice in many research studies is to select fewer variables prior to classification, as researchers seek to obtain the smallest set of variables that will likely produce satisfactory results. The stability of the selected vector of features was evaluated using several stability measures to assure the features' consistency when training data is changed. In spite of the fact that all the feature selection methods were effective in promoting the classification performance, our results showed that the seizure detection algorithm used with an unsupervised graph-based feature selection technique and a Random Forest classifier showed a higher accuracy, 91.55%, and sensitivity, 94.12%, than all the other feature selection methods evaluated. Even though performances as high as 100% have been reported widely in previous studies, ours is unique, as we used the largest publicly available EEG seizure database (from Temple University), and because we considered the stability of the features to determine their general applicability as compared to existing methods.

However, there are limitations to our method that should be acknowledged. One of the major limitations is the computational time, as UGFS method involves mapping features on an affinity graph and calculating feature importance scores, which can be computationally intensive and time-consuming, especially when dealing with large datasets. However, we believe that the trade-off in computational time is worth it as the method has been shown to be effective in selecting relevant features that is can be computationally intensive, particularly when dealing with large datasets. Another limitation is the requirement for threshold setting. The method requires setting a threshold to determine the cut-off point for selecting the most important features. This threshold can impact the results of the feature selection process, and it is important to carefully consider the appropriate value for each individual dataset and application. To address these limitations in future studies, one approach could be to employ more efficient algorithms for feature selection, such as those based on meta-heuristics or swarm intelligence. Additionally, the sensitivity of the algorithm to parameter settings can be reduced by employing more advanced machine learning techniques, such as Bayesian optimization or reinforcement learning, to optimize the parameters automatically. This would allow for more accurate and robust results, and ultimately lead to improved patient care through the development of advanced seizure detection systems. In addition to evaluating the proposed method for detecting seizures, we plan to expand its applicability by incorporating it into the classification of both local and generalized seizures. This will provide a more comprehensive understanding of the effectiveness of the proposed method in a wider range of seizure types. Furthermore, we aim to explore the potential of deep learning techniques for seizure detection using the same set of stable features identified in this study. By comparing the performance of the deep learning approach with that of the raw EEG data, we hope to gain insights into the most effective methods for accurately detecting seizures and improving patient care. Overall, our future work aims to build upon the results of this study and further advance the field of epileptic seizure detection.

CRediT authorship contribution statement

Lina Abou-Abbas: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Khadidja Henni:** Methodology, Software, Validation, Formal analysis, Writing – original draft. **Imene Jemal:** Methodology, Investigation, Writing – original draft. **Amar Mitiche:** Writing – original draft, Supervision. **Neila Mezghani:** Writing – original draft, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Lina Abou-Abbas reports financial support was provided by Quebec Research Fund Nature and Technology. Neila Mezghani reports financial support was provided by Canada Research Chair on Biomedical Data Mining.

Data availability

Public database.

Acknowledgments

We want to offer our special thanks to Youssef Ouakrim for his technical contribution in the coding stage. Research reported in this publication was supported by the Quebec Research Fund Nature and Technology, Canada (L.A.A-298991) and the Canada Research Chair on Biomedical Data Mining (N.M. 950-231214).

References

Abou-Abbas, L., Jemal, I., Henni, K., Mitiche, A., & Mezghani, N. (2021). Focal and generalized seizures distinction by rebalancing class data and random forest classification. In *Bioengineering and biomedical signal and image processing: first international conference, BIOMESIP 2021, Meloneras, Gran Canaria, Spain, July 19-21, 2021, proceedings 1* (pp. 63–70). Springer, http://dx.doi.org/10.1007/978-3-030-88163-4_6.

Abou-Abbas, L., Jemal, I., Henni, K., Ouakrim, Y., Mitiche, A., & Mezghani, N. (2022). EEG oscillatory power and complexity for epileptic seizure detection. *Applied Sciences*, *12*(9), 4181. http://dx.doi.org/10.3390/app12094181.

Abou-Abbas, L., Noordt, S., Desjardins, J., Cichonski, M., & Elsabbagh, M. (2021). Use of empirical mode decomposition in ERP analysis to classify familial risk and diagnostic outcomes for autism spectrum disorder. *Brain Sciences*, *11*(4), 409. http://dx.doi.org/10.3390/brainsci11040409.

Acharya, U., Fujita, H., Sudarshan, V., Bhat, S., & Koh, J. (2015). Application of entropies for automated diagnosis of epilepsy using EEG signals: A review. *Knowledge-Based Systems*, *88*, 85–96. http://dx.doi.org/10.1016/j.knosys.2015.08.004.

Alam, S., & Bhuiyan, M. (2013). Detection of seizure and epilepsy using higher order statistics in the EMD domain. *IEEE Journal of Biomedical and Health Informatics*, *17*(2), 312–318. http://dx.doi.org/10.1109/JBHI.2012.2237409.

Alickovic, E., Kevric, J., & Subasi, A. (2018). Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomedical Signal Processing and Control*, *39*, 94–102. http://dx.doi.org/10.1016/j.bspc.2017.07.022.

Bandarabadi, M., Teixeira, C., Rasekhi, J., & Dourado, A. (2015). Epileptic seizure prediction using relative spectral power features. *Clinical Neurophysiology*, *126*(2), 237–248. http://dx.doi.org/10.1016/J.CLINPH.2014.05.022.

Bhattacharyya, A., Sharma, M., Pachori, R., Sircar, P., & Acharya, U. (2018). A novel approach for automated detection of focal EEG signals using empirical wavelet transform. *Neural Computing and Applications*, *29*(8), 47–57. http://dx.doi.org/10.1007/s00521-016-2646-4.

Bommert, A. (2020). Integration of feature selection stability in model fitting.

Bosl, W., Tierney, A., Tager-Flusberg, H., & Nelson, C. (2011). EEG complexity as a biomarker for autism spectrum disorder risk. *BMC Medicine*, *9*(1), 1–16. http://dx.doi.org/10.1186/1741-7015-9-18.

Bou Assi, E., Sawan, M., Nguyen, D., & Rihana, S. (2015). A hybrid mRMR-genetic based selection method for the prediction of epileptic seizures. In *IEEE biomedical circuits and systems conference: engineering for healthy minds and able bodies, BioCAS 2015 - proceedings*. http://dx.doi.org/10.1109/BIOCAS.2015.7348367.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. http://dx.doi.org/10.1002/widm.8.

Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, *3*(2), 185–205. http://dx.doi.org/10.1142/S0219720005001004.

Direito, B., Duarte, J., Teixeira, C., Schelter, B., Le Van Quyen, M., Schulze-Bonhage, A., Sales, F., & Dourado, A. (2011). Feature selection in high dimensional EEG features spaces for epileptic seizure prediction. *IFAC Proceedings*, *44*(1), 6206–6211. http://dx.doi.org/10.3182/20110828-6-IT-1002.03331.

Faust, O., Acharya, U. R., Min, L. C., & Sputh, B. H. (2010). Automatic identification of epileptic and background EEG signals using frequency domain parameters. *International Journal of Neural Systems*, *20*(02), 159–176. http://dx.doi.org/10.1142/S0129065710002334.

Fu, K., Qu, J., Chai, Y., & Dong, Y. (2014). Classification of seizure based on the time-frequency image of eeg signals using HHT and SVM. *Biomedical Signal Processing and Control*, *13*, 15–22. http://dx.doi.org/10.1016/j.bspc.2014.03.007.

Greene, B., Faul, S., Marnane, W., Lightbody, G., Korotchikova, I., & Boylan, G. (2008). A comparison of quantitative EEG features for neonatal seizure detection. *Clinical Neurophysiology*, *119*(6), 1248–1261. http://dx.doi.org/10.1016/J.CLINPH.2008.02.001.

He, X., Cai, D., & Niyogi, P. (2005). Laplacian score for feature selection. *Advances in Neural Information Processing Systems*, *18*.

Henni, K., Mezghani, N., & Gouin-Vallerand, C. (2018). Unsupervised graph-based feature selection via subspace and pagerank centrality. *Expert Systems with Applications*, *114*, 46–53. http://dx.doi.org/10.1016/J.ESWA.2018.07.029.

Henni, K., Mezghani, N., & Mitiche, A. (2020). Cluster density properties define a graph for effective pattern feature selection. *IEEE Access*, *8*, 62841–62854. http://dx.doi.org/10.1109/ACCESS.2020.2981265.

Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalography and Clinical Neurophysiology*, *29*(3), 306–310. http://dx.doi.org/10.1016/0013-4694(70)90143-4.

Jemal, I., Mitiche, A., Abou-Abbas, L., Henni, K., & Mezghani, N. (2021). An effective deep neural network architecture for cross-subject epileptic seizure detection in EEG data. In *Proceedings of CECNet 2021* (pp. 54–62). IOS Press, http://dx.doi.org/10.3233/FAIA210389.

Jemal, I., Mitiche, A., & Mezghani, N. (2021). A study of EEG feature complexity in epileptic seizure prediction. *Applied Sciences*, *11*(4), 1579. http://dx.doi.org/10.3390/app11041579.

Kalin, F., Akinci, T., Türkpence, D., Seker, S., & Korkmaz, U. (2020). Detection of epileptic seizure using STFT and statistical analysis. In *Advances in neural signal processing. IntechOpen*. http://dx.doi.org/10.5772/intechopen.89026.

Kannathal, N., Choo, M., Acharya, U., & Sadasivan, P. (2005). Entropies for detection of epilepsy in EEG. *Computer Methods and Programs in Biomedicine*, *80*(3), 187–194. http://dx.doi.org/10.1016/j.cmpb.2005.06.012.

Khaire, U., & Dhanalakshmi, R. (2019). Stability of feature selection algorithm: A review. *Journal of King Saud University - Computer and Information Sciences*, http://dx.doi.org/10.1016/J.JKSUCI.2019.06.012.

Kumar, Y., Dewal, M., & Anand, R. (2013). Wavelet entropy based EEG analysis for seizure detection. In *2013 IEEE international conference on signal processing*. http://dx.doi.org/10.1109/ISPCC.2013.6663415.

Kumar, A., & Kolekar, M. (2014). Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals. In *International conference on medical imaging, M-health and emerging communication systems* (pp. 412–416). http://dx.doi.org/10.1109/MedCom.2014.7006043.

Li, J., Yan, J., Liu, X., & Ouyang, G. (2014). Using permutation entropy to measure the changes in EEG signals during absence seizures. *Entropy*, *16*, 3049–3061. http://dx.doi.org/10.3390/e16063049.

Li, Z., Yang, Y., Liu, J., Zhou, X., & Lu, H. (2012). Unsupervised feature selection using nonnegative spectral analysis. *26*, In *Proceedings of the AAAI conference on artificial intelligence*. (1), http://dx.doi.org/10.1609/aaai.v26i1.8289.

Mahjoub, C., Jeannès, R. L. B., Lajnef, T., & Kachouri, A. (2020). Epileptic seizure detection on EEG signals using machine learning techniques and advanced preprocessing methods. *Biomedical Engineering/Biomedizinische Technik*, *65*(1), 33–50. http://dx.doi.org/10.1515/bmt-2019-0001.

Memar, P., & Faradji, F. (2018). A novel multi-class EEG-based sleep stage classification system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(1), 84–95. http://dx.doi.org/10.1109/TNSRE.2017.2776149.

Nogueira, S., Sechidis, K., & Brown, G. (2017). On the stability of feature selection algorithms. *Journal of Machine Learning Research*, *18*(1), 6345–6398. http://dx.doi.org/10.5555/3122009.3242031.

Orhan, U., Hekim, M., & Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications*, *38*(10), 13475–13481. http://dx.doi.org/10.1016/J.ESWA.2011.04.149.

Orosco, L., Correa, A. G., Diez, P., & Laciar, E. (2016). Patient non-specific algorithm for seizures detection in scalp EEG. *Computers in Biology and Medicine*, *71*, 128–134. http://dx.doi.org/10.1016/j.compbiomed.2016.02.016.

Orosco, L., Laciar, E., Correa, A., Torres, A., & Graffigna, J. (2009). An epileptic seizures detection algorithm based on the empirical mode decomposition of EEG. In *Proceedings of the 31st annual international conference of the IEEE engineering in medicine and biology society: engineering the future of biomedicine* (pp. 2651–2654). EMBC: http://dx.doi.org/10.1109/IEMBS.2009.5332861.

Oweis, R., & Abdulhay, E. (2011). Seizure classification in EEG signals utilizing Hilbert-Huang transform. *BioMedical Engineering Online*, *10*(1), 1–15. http://dx.doi.org/10.1186/1475-925X-10-38.

Park, Y., Luo, L., Parhi, K., & Netoff, T. (2011). Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia*, *52*(10), 1761–1770. http://dx.doi.org/10.1111/J.1528-1167.2011.03138.X.

Parvez, M., & Paul, M. (2014). Epileptic seizure detection by analyzing EEG signals using different transformation techniques. *Neurocomputing*, *145*, 190–200. http://dx.doi.org/10.1016/j.neucom.2014.05.044.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238. http://dx.doi.org/10.1109/TPAMI.2005.159.

Raghu, S., Sriraam, N., Vasudeva Rao, S., Hegde, A., & Kubben, P. (2019). Automated detection of epileptic seizures using successive decomposition index and support vector machine classifier in long-term EEG. *Neural Computing and Applications*, *32*(13), 8965–8984. http://dx.doi.org/10.1007/S00521-019-04389-1.

Rajendra Acharya, U., Molinari, F., Sree, S., Chattopadhyay, S., Ng, K.-H., & Suri, J. (2012). Automated diagnosis of epileptic EEG using entropies. *Biomedical Signal Processing and Control*, *7*, 401–408. http://dx.doi.org/10.1016/j.bspc.2011.07.007.

Ramadhani, I., Saputro, D., Maryati, N., Solihati, S., Wijayanto, I., Hadiyoso, S., & Patmasari, R. (2019). Seizure type classification on EEG signal using support vector machine recent citations fractal based feature extraction method for epileptic seizure detection in long-term EEG recording seizure type classification on EEG signal using support vector machine. http://dx.doi.org/10.1088/1742-6596/1201/1/012065, 12065.

Rezek, I., & Roberts, S. (1998). Stochastic complexity measures for physiological signal analysis. *IEEE Transactions on Biomedical Engineering*, *45*(9), 1186–1191. http://dx.doi.org/10.1109/10.709563.

Şen, B., Peker, M., Çavuşoğlu, A., & Çelebi, F. V. (2014). A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *Journal of Medical Systems*, *38*(3), 1–21. http://dx.doi.org/10.1007/s10916-014-0018-0.

Senawi, A., Wei, H.-L., & Billings, S. A. (2017). A new maximum relevance-minimum multicollinearity (MRmMC) method for feature selection and ranking. *Pattern Recognition*, *67*, 47–61. http://dx.doi.org/10.1016/j.patcog.2017.01.026.

Shantha Selva Kumari, R., & Prabin Jose, J. (2011). Seizure detection in EEG using time frequency analysis and SVM. In *International conference on emerging trends in electrical and computer technology, ICETECT* (pp. 626–630). http://dx.doi.org/10.1109/ICETECT.2011.5760193.

Sharma, M., Dhere, A., Pachori, R., & Acharya, U. (2017). An automatic detection of focal EEG signals using new class of time–frequency localized orthogonal wavelet filter banks. *Knowledge-Based Systems*, *118*, 217–227. http://dx.doi.org/10.1016/j.knosys.2016.11.024.

Sharma, R., & Pachori, R. (2015). Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Systems with Applications*, http://dx.doi.org/10.1016/j.eswa.2014.08.030.

Shi, L., Du, L., & Shen, Y. (2014). Robust spectral learning for unsupervised feature selection. *ICDM, 2015-January(January*, In *Proceedings - IEEE international conference on data mining* (pp. 977–982). http://dx.doi.org/10.1109/ICDM.2014.58.

Song, Y., Crowcroft, J., & Zhang, J. (2012). Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. *Journal of Neuroscience Methods*, *210*(2), 132–146. http://dx.doi.org/10.1016/j.jneumeth.2012.07.003.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., & Boylan, G. (2011). EEG-based neonatal seizure detection with support vector machines. *Clinical Neurophysiology*, *122*(3), 464–473. http://dx.doi.org/10.1016/j.clinph.2010.06.034.

Tessy, E., Muhammed Shanir, P., & Manafuddin, S. (2017). Time domain analysis of epileptic EEG for seizure detection. In *International conference on next generation intelligent systems, ICNGIS 2016*. http://dx.doi.org/10.1109/ICNGIS.2016.7854034.

Tzallas, A., Tsipouras, M., & Fotiadis, D. (2009). Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Transactions on Information Technology in Biomedicine*, *13*(5), 703–710. http://dx.doi.org/10.1109/TITB.2009.2017939.

Wang, Q., Wei, H.-L., Wang, L., & Xu, S. (2021). A novel time-varying modeling and signal processing approach for epileptic seizure detection and classification. *Neural Computing and Applications*, *33*, 5525–5541. http://dx.doi.org/10.1007/s00521-020-05330-7.

Wang, T., Zhang, M., Yu, Q., & Zhang, H. (2012). Comparing the applications of EMD and EEMD on time-frequency analysis of seismic signal. *Journal of Applied Geophysics*, http://dx.doi.org/10.1016/j.jappgeo.2012.05.002.

Wei, H.-L., & Billings, S. A. (2006). Feature subset selection and ranking for data dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(1), 162–166. http://dx.doi.org/10.1109/TPAMI.2007.250607.

Zeng, W., Li, M., Yuan, C., Wang, Q., Liu, F., & Wang, Y. (2019). Classification of focal and non focal EEG signals using empirical mode decomposition (EMD), phase space reconstruction (PSR) and neural networks. *Artificial Intelligence Review*, *52*(1), 625–647. http://dx.doi.org/10.1007/s10462-019-09698-4.

Zhang, X., Yao, L., Dong, M., Liu, Z., Zhang, Y., & Li, Y. (2020). Adversarial representation learning for robust patient-independent epileptic seizure detection. *IEEE Journal of Biomedical and Health Informatics*, *24*(10), 2852–2859. http://dx.doi.org/10.1109/JBHI.2020.2971610.

Zhao, Z., & Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. *Vol. 227*, In *ACM international conference proceeding series* (pp. 1151–1157). http://dx.doi.org/10.1145/1273496.1273641.

Zhao, Y., Zhang, G., Zhang, Y., Xiao, T., Wang, Z., Xu, F., & Zheng, Y. (2022). Multi-view cross-subject seizure detection with information bottleneck attribution. *Journal of Neural Engineering*, *19*(4), Article 046011. http://dx.doi.org/10.1088/1741-2552/ac7d0d.

Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., Guo, H., & Xiang, J. (2018). Epileptic seizure detection based on EEG signals and CNN. *Frontiers in Neuroinformatics*, *12*, 95. http://dx.doi.org/10.3389/fninf.2018.00095.