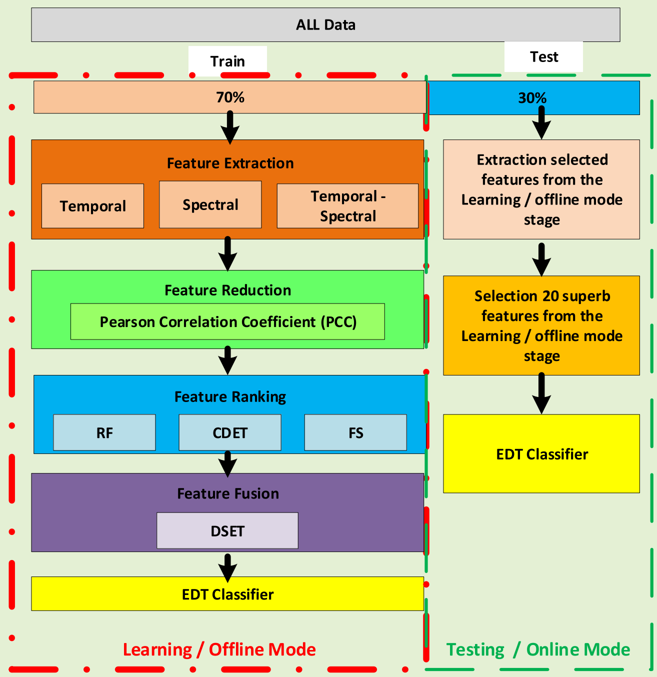
Multi-Feature Fusion Approach for Epileptic Seizure Detection From EEG Signals

Moein Radman, Milad Moradi, Ali Chaibakhsh, *Member, IEEE*, Mojtaba Kordestani, *Senior Member, IEEE*, and Mehrdad Saif, *Senior Member, IEEE*

***Abstract*—In this article, a new fusion scheme based on the Dempster–Shafer Evidence Theory (DSET) is introduced for Epileptic Seizure Detection (ESD) in brain disorders. Firstly, various features in temporal, spectral, and temporal-spectral domains are extracted from Electroencephalogram (EEG) signals. Afterward, a Correlation analysis via the Pearson Correlation Coefficient (PCC) is conducted on the extracted features to select and remove highly correlated features. It leads to the second feature set with about half numbers of the first feature set. Next, three separate filter-type feature selection techniques, including Relief-F (RF), Compensation Distance Evaluation Technique (CDET), and Fisher Score (FS), are conducted to this second feature set for ranking features. Following that, a feature fusion is engaged by the DSET through the individual feature ranking results to generate high qualified feature sets. Indeed, the DSET-based feature fusion is devoted to enhancing the feature selection confidence using the least superb ranked features. In the classification stage, an Ensemble Decision Tree (EDT) classifier, along with two common validation procedures, including hold out and 10-fold cross-validation, is appropriated to classify the selected features from the EEG signals as normal, pre-ictal (epileptic background), and ictal (epileptic seizure) classes. Finally, several test scenarios are investigated using experimental data of Bonn University to evaluate the proposed ESD performance. Moreover, a comparison with other research works on the same dataset and classes is accomplished. The obtained results indicate the effectiveness of the proposed feature fusion approach and superior accuracy compared to the traditional methods.**



***Index Terms*—Epileptic seizure detection (ESD), electroencephalogram (EEG) signals, feature selection, fusion approaches, ensemble-based classifier.**

I. Introduction

EPILEPSY is the most common neural disease, based on the World Health Organization (WHO) estimation. Each year, approximately 2.4 million new cases are identified, while 50 percent of the cases incur the disease when they are children or teenagers. Moreover, epilepsy disorders grow in older adults [1]. Generally, over 50 million people worldwide are affected by this neurological disorder [2]. The sick cases are consistently faced with unnatural sicknesses outbreak due to the brain's electric discharge during a certain period.

Epilepsy is often known as seizure repetitions during a time period. This chronic issue can happen from once a year to several times a day. Epileptic disorders and seizures are not identical. The epilepsy signs can be sudden attacks with no reason due to the central neural system interfering. This disease is as a result of a typical neural network procedure that suddenly transforms into a highly irritable network, mostly in the cerebrum region [3].

A. Motivation

The EEG directly records the brain activities through the electrodes attached to the scalp skin [4]. Most of the disorders

Manuscript received July 1, 2020; revised September 17, 2020; accepted September 18, 2020. Date of publication September 23, 2020; date of current version January 6, 2021. The work of Mehrdad Saif was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant 61873144. The associate editor coordinating the review of this article and approving it for publication was Dr. Ravibabu Mulaveesala. *(Corresponding author: Mojtaba Kordestani.)*

Moein Radman and Ali Chaibakhsh are with the Faculty of Mechanical Engineering, University of Guilan, Rasht 4199613776, Iran (e-mail: moeinradman@phd.guilan.ac.ir; chaibakhsh@guilan.ac.ir).

Milad Moradi, Mojtaba Kordestani, and Mehrdad Saif are with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON N9B 3P4, Canada (e-mail: moradih@uwindsor.ca; kordest@uwindsor.ca; msaif@uwindsor.ca).

Digital Object Identifier 10.1109/JSEN.2020.3026032

related to the brain malfunction, including epilepsy diseases, are examined using the EEG signals [5]. The EEG signals show the patterns before and during the seizure in the case of epilepsy, which is entirely different from the normal EEG signals. Some proposed ESD schemes divide the EEG signals as normal and abnormal classes (two classes) [6], [7], while others categorize EEG signals as normal, pre-ictal, and ictal conditions (three classes) [8]. Although various methods have been developed for the ESD [9], the EEG attack classification is a relatively difficult, yet challenging task due to overlap between classes.

B. Related Works

A typical procedure of EEG attack diagnosis often involves the following steps: a) Splitting the dataset into train and test sets with a specific portion, and applying EEG signal preprocessing techniques to exclude the noises and artifacts. b) Extracting signal's features by various time or/and frequency-domain methods. c) Selecting superb extracted features with the highest description. d) Implementing a classifier to identify the classes. e) Evaluating the attack diagnosis method using the test dataset [10].

Several works exist on publicly available benchmark databases [8], [11], [12] that classify epileptic seizures. Alam *et al.* [13] develop a seizure and epilepsy detection based on the Empirical Mode Decomposition (EMD). First, higher-order statical moments such as variance, kurtosis, and skewness are utilized to extract proper features. Following that, the features are given to an Artificial Neural Network (ANN) to identify seizures and epilepsy. Although this approach is faster than the time-frequency-based techniques, the classification accuracy is 80%, which is obtained only by three features.

Niknazar *et al.* [14] examine a system identification tool based on the time-delay method for ESD. Recurrence Quantification Analysis (RQA)-based features are utilized by Error-Correction Output Codes (ECOC) classifier. The RQA is adopted because there is no requirement to know about signal properties, such as its length, noise, etc. Therefore, the suggested ESD method does not need transformations or preexisting models. Besides, the ability to work with signals, including different morphology and spectrum, is another advantage. A hybrid ESD method is introduced in [15]. First, a Dual-Tree Complex Wavelet Transformation (DTCWT) is applied to reduce the data size and obtain features. Later, the features are presented to Complex-Valued Neural Networks (CVANN) for the classification task. Investigating the signals at various levels using wavelet transformations is a prominent part of this work.

A data-driven method based on Multi-Layer Perceptron Neural Network (MLPNN) is developed in [16] for EEG classification. First, a Discrete Wavelet Transform (DWT) is adopted to decompose EEG signals into frequency subbands. Afterward, the K-means algorithm is employed to cluster the wavelet coefficients in each sub-band. Following that, the probability distributions based on the distribution of wavelet coefficients are computed. Finally, these distributions are applied as inputs to the MLPNN model. In this work, a clustering approach based on the k-means algorithm is allowed over the wavelet coefficients instead of using the basic statistics, which improves the performance.

Tiwari *et al.* [17] apply a Local Binary Pattern (LBP) method for epilepsy detection from EEG signals. Firstly, some filtered signals are detected by a pyramid of the Difference of Gaussian (DoG). Then, these signals are fed to a Support Vector Machine (SVM) classifier. The computational simplicity of LBP features, the ability to reach high detection accuracy with a smaller portion of the EEG signals, suitability for online epileptic detection with reduced computational burden are the clinical significance of this diagnosis method. In another work, Acharya *et al.* [18] apply several entropy-based extracted features to detect normal, pre-ictal, and ictal conditions. First, four different Entropy, including Approximate Entropy (ApEn), Sample Entropy (SampEn), Phase Entropy 1 (S1), and Phase Entropy 2 (S2), are investigated. Following that, the features are fed to seven different classifiers, which among them Fuzzy Sugeno Classifier (FSC), provides the best performance.

Tzallas *et al.* [19] develop an automatic ESD method to classify the epileptic attacks into three classes. Several extracted features from the time-frequency analysis are fed to Artificial Neural Network (ANN). Next, it is shown that the classifier reaches to the best performance using 40 features. Upadhyay *et al.* [20] apply integrated DWT-based features and Least Square-Support Vector Machine (LS-SVM) to classify the epilepsy attacks. First, utilizing the Maximum Energy to Permutation Entropy ratio, the best basis wavelet is chosen for feature extraction. Afterward, several feature ranking methods, including RF, FS, and Information Gain (IG), are applied for feature selection. Then, they are ranked and fed to the classifier for ESD. Simulation tests on the EEG signals indicate the high performance of the suggested ESD.

Hassan *et al.* [21] utilize Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) for epileptic seizure identification. They extract intrinsic mode functions through decomposing segments of EEG signals by the CEEMDAN. Afterward, Normal Inverse Gaussian (NIG) pdf parameters are utilized to model the extracted mode functions. Besides, ten classifiers are investigated in this work that Adaptive Boosting (AdaBoost) reaches the best classification performance. In another work, Li *et al.* [22] introduce a hybrid method for ESD by using Fuzzy Entropy (FuzzyEn)-based features which are obtained from EEG signals in a Fractional Fourier Transform - Wavelet Packet decomposition (FFT-WPT) domain. They also employ the Principal Component Analysis (PCA) to reduce dimensionality and produce uncorrelated variables. Finally, three different classifiers, including K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA) and SVM are applied for classification. In simulation, test results indicate that the SVM classifier produces the best performance.

An EEG classification method is developed in [23] using a combination of temporal and spectral features from EMD. Then, by concatenating the obtained temporal and spectral features of the EEG signals, the salient characteristics are extracted fed to an SVM classifier. Test results indicate an overall accuracy of 82 %. Hassan and Haque [24] introduce a

CEEMDAN for ESD and EEG signal analysis that resolves the mode mixing problem and gives better spectral separation of the modes. Then, various extracted statistical features by the CEEMDAN are fed to the ANN for ESD. Test results show an accuracy of 86.37 %. A Short-Time Fourier Transform (STFT) is utilized in [25] to extract features. Then, Four rule-based classifiers, including a decision tree algorithm, a random forest algorithm, an SVM-based decision tree algorithm, and an SVM-based random forest algorithm, are applied to detect seizures. Test results indicate that the random forest provides the best performance among the other investigated classifiers. Das *et al.* [26] developed an automated ESD for EEG classification. The ESD method utilizes statistical NIG parameters computed in the DTCWT domain to extract features. Then, an SVM is applied for the classification task. The proper overall performance and computational speed have been mentioned as two advantages of this work.

An integrated Variational Mode Decomposition (VMD) and Auto-Regression (AR)-based feature extraction is introduced in [27] for automated seizure detection. Then, the random forest classifier is implemented for the classification. Excellent robustness and adaptive adjustment of AR model parameters are two advantages of the suggested seizure detection. Hassan and Subasi [28] address the problem of automated ESD using single-channel EEG signals. To reach this aim, a CEEMDAN is employed to decompose segments of EEG signals. Then, the obtained train and test matrices are fed to Linear Programming Boost-ing (LPBoost) for classification. A scheme including Tunable-Q factor Wavelet Transform (TQWT), and bootstrap aggregating (Bagging) is proposed in [29] for the ESD problem. First, the EEG signal is decomposed into subbands by utilizing the TQWT, and then, various spectral features are extracted from the decomposed sub-bands. Finally, bagging is used to perform the classification task. The classifier reaches an accuracy of 98.40%.

C. The Proposed Failure Prognosis Method and Main Contributions

This article introduces a new fusion ESD scheme to isolate various conditions, including normal, pre-ictal, and ictal. For this aim, the various domains like temporal, spectral, and temporal-spectral features are firstly extracted from the EEG signals. Following that, a feature selection analysis based on correlation analysis is applied to reduce the feature number from eighty to forty-three. Following that, the extracted features that are now in the new reduced feature space are ranked using RF, CDET, and FS feature ranking techniques. Since then, the DSET is utilized to combine the features ranking methods and generate several new feature sets with the ranked high qualified properties. Finally, EDT classifier is employed to isolate and distinguish various Epileptic seizures.

The main contributions of this work are summarized as follows:

- Applying a combination of feature extraction methods from various domains, including temporal, spectral, and temporal-spectral domains, assists in producing rich feature sets. It helps to catch the epileptic seizure characteristic and therefore improves the detection rate and ESD's performance.
- Another novelty is to utilize the DSET to generate several new ranked feature sets by fusing the feature ranking methods. It enhances the classification performance compared to situations in which the classifier only applies each of the RF, CDET, and FS feature ranking methods individually.
- The proposed fusion ESD structure is appropriated well for online implementation. The suggested feature ranking based on DSET chooses the optimized features and consequently reaches a higher accuracy compared to [20] in a similar condition when it applies a holdout validation procedure. Therefore, this is suitable for online implementation.
- According to our best knowledge, this proposed ESD scheme has reached the best-achieved accuracy when data has been split into 70% and 30% for training and test, respectively.

D. The Paper Organization

The rest of this article is organized as follows. Section II illustrates materials and methods. Simulation studies are demonstrated in detail in Section III. Section IV provides discussion and details about the proposed EDT classifier using the integrated high-qualified features. Moreover, a comparison with other works in the literature is provided. Finally, a summary of the results is given in Section V.

II. Materials and Methods

A. Dataset

The dataset studied in this work is obtained from Bonn University in Germany [8], which is accessible in the Epileptology department of the university. This dataset consists of five subsets of A, B, C, D, and E; each of them carries 100 pieces of single-channel EEG signal with a length of 23.6 seconds. The A and B sets comprise EEG pieces recorded from the scalp skin of five healthy volunteers based on the $10-20$ standard electrode placement procedure. Volunteers in set A are awake with their eyes open, while B's volunteers are awake with their eyes closed. Sets C, D, and E are collected from the patient's EEG archive before the surgery. Based on table 1, normal signals are gathered from 5 healthy subjects, attack signals, and signals between the attacks. Set D contains EEG recorded from inside the brain, and C includes recorded EEG from the brain's hippocampus region. Set C and D restrain brain activity between two epileptic attacks. Nevertheless, Set E holds activity during the epileptic attack. All the EEG signals are recorded by a 128-channel amplifier system using an average standard reference after 12-bit analog to digital transformation. The sampling frequency of the datasets is 173.63 Hz. Datasets are split into A and B as a normal condition, C and D as pre-ictal, and finally E as ictal conditions.

B. Feature Extraction

Processing the raw EEG information is very challenging due to the encoded information in EEG signals. Therefore, some

TABLE I
FEATURE EXTRACTION METHODS

Domains	Features
Temporal	Root mean square Variance Mean absolute value Kurtosis Skewness
Spectral	Power Spectral Ratio Skewness BP Variance Median power frequency
Wavelet	Skewness Kurtosis Variance Mean FDMandelbrot [32] SampEn [33]

mathematical transformations are often practiced to the EEG signals to extract their properties from a particular perspective. The output, or in other words, the obtained coefficients from these transforms are called features [30].

The standard feature extraction methods for the EEG signals are divided into three general categories: 1) temporal, 2) spectral, and 3) spatial features [31]. The spatial features are often employed in the Brain-Computer Interface (BCI) [4]. To extract features, firstly, the Butterworth filter is adopted to decompose the brain signal into five frequency sub-bands named delta, theta, alpha, beta, and gamma. It is noteworthy that each band contains useful information. Later, the features are extracted for each frequency band individually. Besides, several various features in different domains like temporal, spectral, and temporal-spectral are extracted to attain the highest distinct characteristics of the EEG signals. The extracted features are summarized in Table I.

First, several EEG signal characteristics in the temporal domain, such as root mean square, variance, skewness, kurtosis, and mean absolute value, are exercised to obtain features for five frequency bands. Afterward, the signals are transformed into the spectral domain by the Fourier Transform (FT). Then, particular spectral features, according to Table I, are extracted. The spectral features manifest the signal's change rate in a specific frequency bound. Therefore, they confer the intensity and quantity of the signal in each frequency bound.

Temporal-spectral domain features are obtained by Wavelet Transform (WT). Commonly, wavelet coefficients are employed to represent the signal in both time and spectral domains. In the first step of discrete WT, approximate coefficients are obtained by passing a signal from a low-pass filter. Then, detailed coefficients are acquired by passing from a high-pass filter in the second step [34]. In the next level, decomposition is recursively executed to the low-pass approximate coefficients to reach the desired decomposed level. In this work, resampling is applied to acquire proper WT's coefficients. WT's coefficients, including $A4$, $D4$, $D3$, $D2$, and $D1$ are related to Delta, Theta, Alpha, Beta, and Gamma, respectively. Following that, the statistical and entropy features are extracted from these coefficients.

C. Feature Reduction, Ranking, Selection

Choosing effective features is recognized as an essential step for designing an appropriate ESD system. In general, some appropriate features are often defined and extracted according to the chosen dominant variables for ESD. It is noted that selecting weak or irrelative features may cause incorrect judgments of the decision-maker system. Feature selection removes irrelevant or redundant features to obtain a minimum set of features (or classes) [35].

Feature selection techniques can be categorized into two major groups, including filter-type and wrapper-type methods. In filter-type methods, variable selection is accomplished by setting an appropriate ranking criterion to score and ordering the variables [36]. The optimal feature subset is obtained in the wrapper-type method by evaluating each feature subset's performance, which generally achieves a higher performance compared to filter-type methods. Regarding the feature selection operating mechanism, it is clear that the filter-type methods are faster than the wrapper-type methods. Despite the lower accuracy of filter-type methods, they are often preferred due to their computational efficiency, especially for high-dimensional feature spaces [37], [38].

In this section, four different filter-type approaches, including the Pearson Correlation Coefficient (PCC), RF, CDET, and FS, are employed for feature selection purposes. In the following, each method is briefly described.

1) Pearson Correlation Coefficient (PCC): Correlation indicates a relationship among measured data values or mathematical variables. The PCC is defined by a static gain in a range of -1 to $+1$ for vectors X and Y. The gain close to one intimates that two variables are highly correlated. In contrast, the gain near zero implies a wealy relationship among the variables. The PCC is calculated as follows [39]:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

where $X = \{x_1, \ldots, x_i, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_i, \ldots, y_n\}$ represent the two vectors with a length of n. Furthermore, $\bar{x}$ and $\bar{y}$ denotes the mean values for the vectors, respectively.

2) Relief-F (RF): One of the common filter-type methods that works based on ranking features is RF [40]. The RF identifies conditional dependency among attributes, and it is classified as a supervised technique. The priority goal in this ranking approach is to determine the features' weight individually using target data. The rating mechanism is based on feature maximum distance. It is worth mentioning that handling the multi-class problem and proper robustness in noisy datasets are some of its merits. In contrast, the requirement for selecting a threshold is its drawback [41]. In this method, the feature quality $W(A)$ for all attributes A based on a specified instance I_0 is computed as follows:

$$W(A) = \sum_{i=1}^{a}[\sum_{B \neq class(A)}[\frac{P(B) \times diff(A, I_0, I_M)}{i}] - \frac{diff(A, I_0, I_M)}{i}] \quad (2)$$

where I_H indicates the nearest hit instance from the same class, I_M denotes the nearest miss instance from a different class. Moreover, a presents the nearest neighbor from the various classes. Besides, $P(B)$ implies each class's prior probability, which is predicted from the training set. Additionally, the function $diff(A, I0, I)$ shows the difference between the values of A for both instances of I_0 and I [41].

3) Compensation Distance Evaluation Technique (CDET): The CDET is a well-known feature ranking method, which was introduced by Lei *et al.* (2008) [42]. In this filter type approach, features are ordered using their CDET scores in a range of [0, 1]. Following that, the ranked features with lower significance are neglected. A feature set of A condition is assumed as follows:

$$\{T_{jk}^{k}, q = 1, 2, \ldots, Q_j; j = 1, 2, \ldots, J; k = 1, 2, \ldots, K\} \quad (3)$$

where T_{ij}^{k} represents kth eigenvalue from qth sample under the jth condition. Furthermore, K denotes the feature number for each condition. Besides, Q_j indicates the sample number for jth condition. It is noted that the ranking mechanism is accomplished based on the average distance of the same condition samples, DS_k, divided to the average distance between different condition samples, DD_k, which are formulated in the following [42]:

$$DS_k = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{Q_j \times (Q_j - 1)}\sum_{l,q=1}^{Q_j}|T_{ij}^{k} - T_{lj}^{k}| \quad (4)$$

where l and q take values among $1, 2, \ldots, Q_j$ and $l \neq q$.

$$DD_k = \frac{1}{J \times (J-1)}\sum_{j,e}^{J}|T_{e}^{k} - \frac{1}{Q_j}\sum_{i-1}^{Q_j}T_{ij}^{k}| \quad (5)$$

where j and e hold values among $1, 2, \ldots, J$ and $j \neq e$.

4) Fisher Score (FS): Another standard feature ranking procedure is FS [20]. The main idea is to select a feature subset in a dataset such that data points in the same classes hold the smallest distances, whereas data points in different classes possess the largest distances [43]. The FS is expressed as follows:

$$FS = \frac{(\mu_{xn} - \mu_{yn})^2}{\sigma_{xn}^2 + \sigma_{yn}^2} \quad (6)$$

where for N features with class labels $l = (l_1, l_2, \ldots, l_n)$, μ_{xn} and μ_{yn} denote mean value in x and y classes, respectively. Moreover, σ_{xn} and σ_{yn} indicate standard deviation in x and y classes, respectively.

D. Feature Fusion

Feature fusion is an intermediate-level fusion that is utilized to combine features to achieve an optimal feature set for decision-making or classification purposes [44]–[46].

1) Dempster-Shafer Evidence Theory (DSET): The evidence theory was developed by Dempster and Shafer (1976) [47]. This algorithm has a great ability to deal with uncertainty. The DSET is acknowledged as the generalization of Bayesian theory to handle uncertainty in datasets, and enhance confidence in data mining approaches such as classification [48]. The DSET is formulated by the Basic Probability Assignment (BPA or m) as follows:

$$m : \Gamma(X) \longrightarrow [0 \quad 1]$$
$$m(\oslash) = 0, \sum_{A \in P} m(A) = 1 \quad (7)$$

where $\Gamma(X)$ represents a power set of X. Moreover, $A \in \Gamma(X)$ which indicates that A is a subset in the power set $\Gamma(X)$. Furthermore, $\oslash$ denotes a null set. Additionally, $m(A)$ holds a value in a range of $[0 \quad 1]$ for the subset A. It is noted that the DSET for several BPAs can be formulated using the orthogonal property of the subsets and the belief rule as follows [49]:

$$m = m_1 \oplus m_2 \oplus \ldots \oplus m_P \Rightarrow$$
$$m(A) = \frac{1}{1-K}\sum_{\forall A_i \Rightarrow \cap_{i=1}^{p} A_i = A}\prod_{j=1}^{r} m_j(A_i)$$
$$m(\oslash) = 0, \quad K = \sum_{\forall A_i \Rightarrow \cap_{i=1}^{p} A_i = \oslash}\prod_{j=1}^{r} m_j(A_i) \quad (8)$$

where m_i denotes a belief value for subset A_i. Moreover, p and r represents the number of classes and classifiers, respectively. In addition, K measures the conflict degree, which shows a probability mass, concerning the conflicts among the evidence sources.

E. Ensemble Decision Tree (EDT) Classifier

The Decision Tree (DT) classifier is a famous algorithm in machine learning methods due to its several advantages such as flexibility, easy understanding, and easy debugging [50]. It has a flowchart like a tree structure that helps to build a decision rule generation model. The DT is classified as a non-parametric supervised learning model that works based on the divide and conquer process. The mechanism is based on the recursive partitioning of input features into many different subspaces to classify unlabeled data [51].

In machine learning, the Ensemble algorithm is known as a technique that leads to increasing the accuracy of classifiers by combining several of them [52]. The ensemble methods consist of bagging, boosting [52], and stacking methods [53]. An important characteristic of ensemble methods is their frequent improvement in predictive performance [54]. In the boosting methods, which are not easy to overfit, at the first, equal initial uniform weight allots to all instances. Then, at a new learning stage, each instance that has correctly classified obtains less weight in comparison with instances that wrongly classified until the system can concentrate on misclassified instances. It leads to classifying them correctly during the next learning step. Finally, the classifiers' results are combined to find the final prediction through a majority voting [55], [56]. Figure 1 shows the boosted EDT classifier structure.

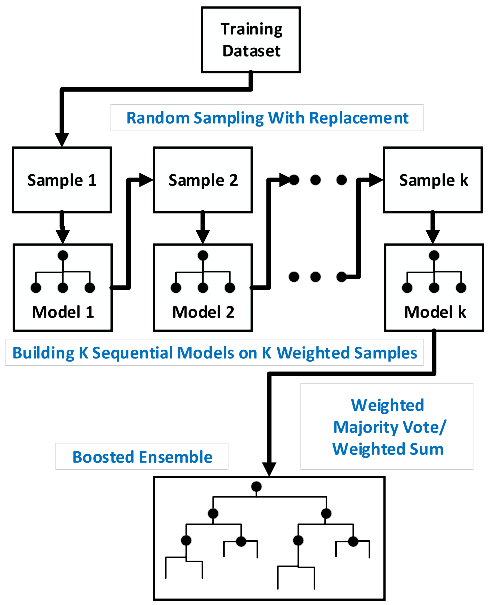


Fig. 1. The boosted EDT classifier structure.

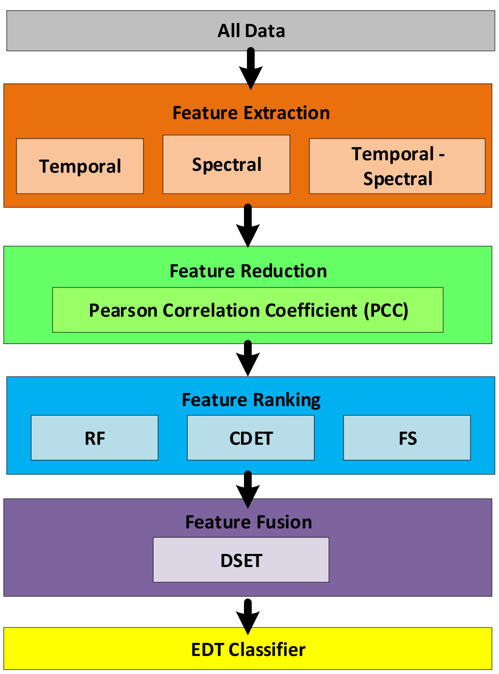


Fig. 2. Flowchart of the Proposed ESD method.

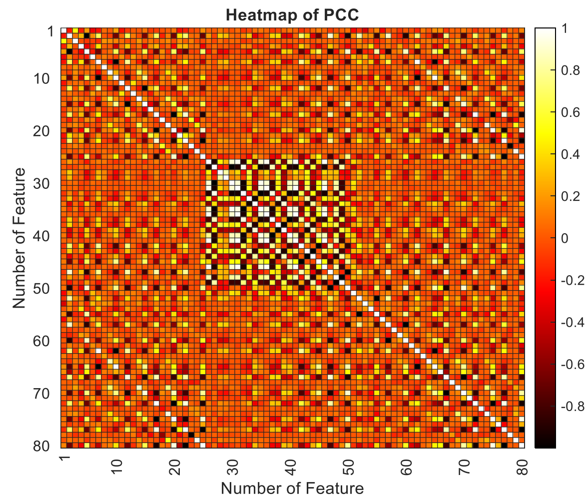


Fig. 3. The heatmap of PCC analysis.

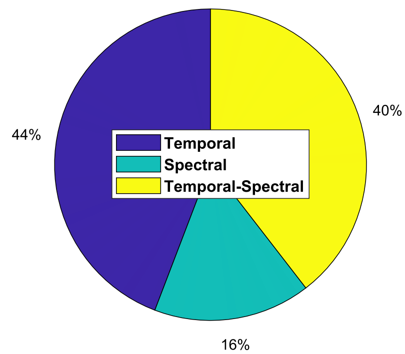


Fig. 4. The percentage of the features in the shrunk feature set.

III. Simulation Studies and Design Implementations

In this section, simulation tests and design implementations of the proposed fusion ESD method are demonstrated. Figure 2 shows the proposed ESD block diagram. It is noted from Figure 2 that the proposed ESD method includes feature extraction, feature reduction, feature ranking, feature fusion, and EDT classification block. In the following, the design implementation of each block is briefly illustrated.

A. Feature Extraction

As we mentioned, the Butterworth filter is applied to EEG signals to obtain five frequency sub-bands. Then, sixteen features, correspond to Table I, are extracted from each frequency sub-band. Hence, eighty features become available for all frequency sub-bands. Theses features include temporal, spectral, and temporal-spectral (wavelet) domains, and therefore, they can represent all properties of the EEG signals.

B. Feature Reduction

Generally, highly correlated features create redundant input data, which negatively affects classifier performance. Therefore, the PCC analysis applies to these eighty features to identify similar ones. Figure 3 shows the heatmap of PCC analysis for eighty features. It is noted from Eq. (1) that PCC's values more than $+0.9$ and less than -0.9 refer to highly correlated features. These values in Figure 3 correspond to the points with white and black colors, respectively. At this stage, redundant features are eliminated to shrink the feature space to a slimmer space with only forty-three members. Figure 4 shows the percentages of the selected forty-three features related to the various domains in which features are extracted from them.

It is noted from Figure 4 that the temporal participates 44%, which is the highest percentage. Notably, the temporal-spectral domain features are in the second rank.

C. Feature Ranking

At the feature ranking block, all the forty-three remaining features are fed to one of three feature ranking methods until

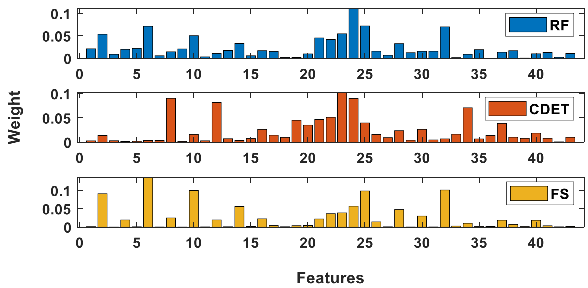


Fig. 5. Associated weights to features from feature ranking methods.

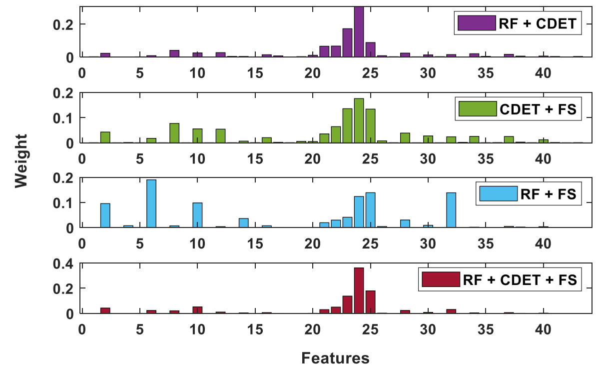


Fig. 6. Associated weights to features from feature fusion methods.

the feature's weight and ranks are assigned based on the corresponding ranking methods' criteria. Indeed, through this occurrence, it is tried to sort remaining features in the reduced extracted feature sets based on their weights and eligibility for better separation between classes. Each filter-based ranking method allocates a weight to each feature that the weight is obtained based on their inner operation.

It is noted that the weights can take values in different ranges. For instance, the RF weights are between -1 and $+1$, while the CDET weights are between 0 and $+1$. Therefore, It is necessary to rescale the weights for the feature ranking methods in a range of $[0 \quad 1]$. The rescaling mechanism is straightforward. For each proposed feature ranking method, the feature weight is divided into the sum of all features' weights. At the end of this procedure, there will be three feature sets in which the summation of all features' weights in each method is equal to 1. Figure 5 shows the weights of the feature rankings methods.

It is indicated from Figure 5 that all three feature ranking methods associate considerable weights to Features 20-25. These features, which belong to spectral, are more likely to be correlated with Epileptic seizures.

D. Feature Fusion

Feature fusion combines the feature ranking methods using the DSET, and obtains the integrated features. The main goal is to utilize the DSET to obtain high-qualified features with less amount of uncertainty using individual feature ranking methods. There are three ranking methods available. Therefore, four combinations can be made, which are shown in Figure 6 Similarly, It is noted from Figure 6 that Features 20-25 gain more weights. Thus, these features must influence the EDT classifier.

E. EDT Classifier

In this section, the design implementation of the EDT classifier is illustrated. As we mentioned before, the datasets include the normal, pre-ictal, and ictal classes. First, the datasets are divided into two parts with 70% for the training, and 30% for the test. Following that, Two common validation procedures, including hold-out and 10-fold cross-validation, are utilized to evaluate classifier performance. Boosting approach is applied to design the classifier in both validation techniques. Moreover, 487 and 426 trees are utilized in the hold-out and 10-fold approaches, respectively. All design characteristics and parameters are summarized in TableII.

TABLE II
THE MOST IMPORTANT PARAMETERS OF PROPOSED ESD

Parameter	Description or value when data selection is based on hold-out	Description or value when data selection is based on 10-fold
Number of the superb Features	20	20
Number of Train Data	350 Samples	450
Number of Test Data	150 Samples	50
Number Of Tree in the EDT classifier	487	426
Ensemble Type	Boosting	Boosting
Boosting algorithm Method	AdaBoost.M2	AdaBoost.M2
Method of combining decision trees in the EDT classifier	Weighted Sum	Weighted Sum
Minimum Leaf Size	2	3
Learn Rate	0.9820	0.9840
Number of Learning Cycles	292	14

To determine the best configuration of the proposed EDT classifier, the ranked feature sets, built in the previous sections, are fed into the classifier inputs, and their accuracies are monitored. There exist three feature sets (Figure 5) from the feature ranking methods, and also four integrated feature sets (Figure 6) from the DSET method. These seven feature spaces can be utilized to feed to the EDT classifier. Then, the performance of the EDT classifier is presented and compared to select the best feature sets that can maximize the classification accuracy.

In this presentation, the weighing factors allocated to each feature indicates the influence and participation of this feature in the final classifier decision. Therefore, the features are sorted with respect to their importance in each feature set to feed the EDT classifier. Afterward, for each feature set, the first superb feature is fed to the classifier, and the classification accuracy is plotted. We continue to add the features from the feature set to maximize the classifier accuracy. Figures 7 and 8 illustrate the classification accuracy concerning the numbers of the features added to the EDT classifier in training and test phases, respectively.

It is noted from Figures 7 and 8 that the EDT classifier has reached to the best performance with the first 20 superb features. For the integrated RF and FS method, the classifier accuracies are 100% and 99.33% for the training and test phases, respectively. Among the three basic feature selection methods, including RF, CDET, and FS, none of them could reach this performance. This fact shows that the obtained

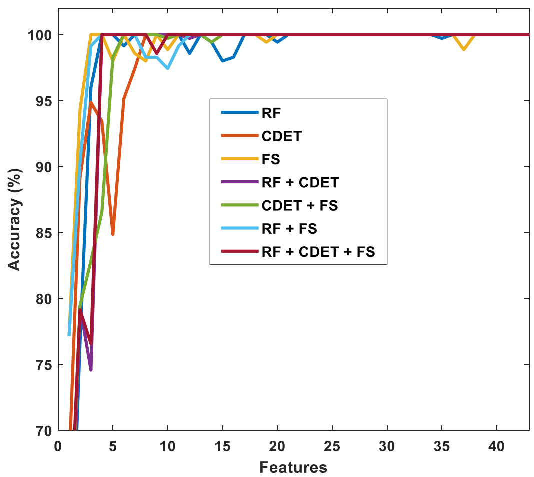


Fig. 7. The classification accuracy with respect to the numbers of the features added in training phase for the seven feature sets.

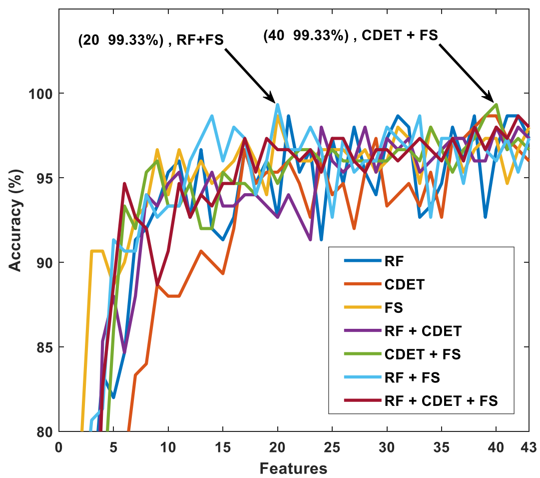


Fig. 8. The classification accuracy with respect to the numbers of the features added in test phase for the seven feature sets.

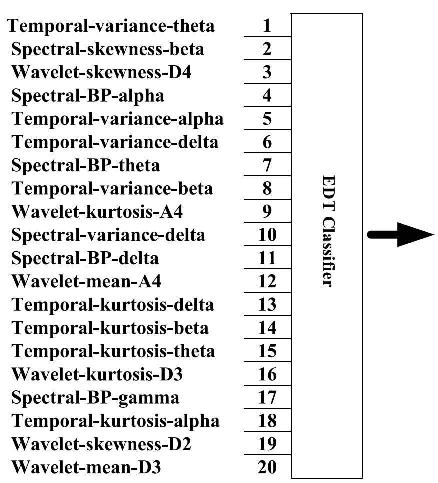


Fig. 9. The input-output structure of the proposed EDT classifier using the first superb twenty features.

feature set by utilizing the feature fusion approach leads to better performance of the classifier for the ESD task. Therefore, the proposed EDT classifier is developed using the integrated RF and FS feature set. Figures 9 depicts the input-output structure of the proposed EDT classifier using the superb twenty features.

Fig. 10. The confusion matrix of the proposed EDT classifier for the training phase.

Fig. 11. The confusion matrix of the proposed EDT classifier for the test phase.

IV. Classification Results and Discussions

In this section, test result discussions are provided using the proposed EDT classifier using the integrated RF and FS feature set. Figures 10 and 11 show the confusion matrix of the proposed EDT classifier in the training and test phases, respectively.

It is noted from Figures 10 that the proposed EDT classifier identifies all samples correctly in the training phase. Regarding Figures 11, only one sample from pre-ictal is not detected and mis-classified into health class.

Furthermore, sensitivity, specificity, and accuracy are defined based on the number of True Positive (TP), False Positive (FP), True Negative(TN), and False Negative (FN) as follows:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{10}$$

TABLE III
EDT CLASSIFIER PERFORMANCE IN TERMS OF ACCURACY, SENSITIVITY, SPECIFICITY, COHEN'S KAPPA COEFFICIENT, AND AUC

Data selection approach	Accuracy (%)	Sensitivity (%)			Specificity (%)			Cohen's Kappa co-efficient	AUC
		normal	pre-ictal	ictal	normal	Pre-ictal	ictal		
10-fold	100	100	100	100	100	100	100	1.0	1.0
Hold-out (70% - 30%) Test phase	99.33	100	98.33	100	98.88	100	100	0.9896	0.9988

TABLE IV
A COMPARISON WITH THE OTHER METHODS IN THE LITERATURE BASED ON THE SAME USED DATASET AND CLASSES (NORMAL, PRE-ICTAL, AND ICTAL), IN TERM OF ACCURACY

Reference	Feature extraction	Classification	Number of used features	Overall accuracy	Test accuracy when data is split	Data selection
[13]	EMD, Higher order moments	ANN	3	80%	-	Hold-out 60% - 5% -35% (train – validation - test)
[23]	EMD, Temporal and spectral	SVM	-	82%	-	10-fold
[24]	CEEMDAN, Statistical	ANN	26	86.37%	-	Hold-out 60% - 5% -35% (train – validation - test)
[16]	DWT	ANN	56	95.60%	-	Hold-out (50% - 50%)
[25]	STFT, Energy	Random Forest	5	96%	-	10-fold
[26]	DTCWT, NIG	SVM	12	96.28%	-	Hold-out (50% - 50%)
[27]	VMD, AR	Random Forest	-	97.352%	-	10-fold
[21]	CEEMDAN, NIG	AdaBoos	-	97.60%	-	10-fold
[28]	CEEMDAN, Spectral	LPBoost	-	97.60%	-	10-fold
[19]	Time-Frequency (TF) analysis	ANN	40	97.72%	-	Hold-out (50% - 50%)
[18]	Entropy	FSC	4	98.10%	-	3-fold
[15]	DTCWT	CVANN	-	98.28%	-	10-fold
				-	97.79%	Hold-out (60% - 40%)
[29]	TQWT, Spectral	Bagging	-	98.40%	-	Hold-out (50% - 50%)
[14]	RQA	ECOC	-	98.67%	-	Hold-out (70% - 30%)
[17]	Histogram	SVM	-	98.80%	-	10-fold
[22]	FuzzyEn, FrFT-WPT	SVM	21	98.84%	-	10-fold
[20]	DWT	LS-SVM	20	100%	-	10-fold
			20	-	98.80%	Hold-out (70% - 30%)
The proposed method	**Temporal, Spectral, Temporal-Spectral**	**EDT**	**20**	**100%**	**-**	**10-fold**
				-	**99.33%**	**Hold-out (70% - 30%)**

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (11)$$

In addition, Cohen's kappa metric, which is known as a more robust measure than a regular percent agreement metric due to considering coincidentally occurring agreements [21], is used as another performance evaluation criterion. Indeed, this criterion evaluates the agreement between human expert and classifier output. It is worth mentioning that AUC, which is the area under Receiver Operating Characteristic (ROC), is another classifier performance criterion that ranges from zero to one. In the AUC criterion, the value which is closer to one indicates a better classifier performance. Table III collects the accuracy, sensitivity, specificity, Cohen's kappa coefficient, and the Area Under Curve (AUC) for the proposed ESD method.

Moreover, a comparison with other methods in the literature is illustrated in Table IV. According to Table IV, our proposed EDT classifier has the best performance in comparison with all the other methods in the literature. Particularly, the proposed EDT classifier reaches a better accuracy than [20] when the holdout technique with a 70% to 30% ratio is chosen. Moreover, the proposed EDT classifier achieves the accuracy of 100% for 10-fold cross-validation technique.

V. CONCLUSION

This article introduced a multi-level fusion strategy for the ESD in brain disorders. For this aim, the Butterworth filter is adopted to decompose the brain signal into five frequency sub-bands. Then, eighty features were extracted from these frequency sub-bands in temporal, spectral, and temporal-spectral domains. Following that, the PCC-based feature selection methods were utilized to remove highly correlated redundant features and obtain a shrunk feature set with only forty-three features. Afterward, three feature ranking methods, including the RF, CDET, and RF, were applied to rescale and sort the selected features. Later, the DSET was adapted to integrate the feature ranking methods and obtain seven high-quality feature sets. Finally, the seven high-quality features were fed into the EDT classifier to detect various types of Epileptic seizures. The EDT classifier using integrated RF and FS methods reached the highest accuracy compared to the EDT classifier with the other feature ranking methods and the ESD methods in the literature. This is due to eliminating redundant features and choosing a few high-qualified features in the proposed ESD structure.

The main contribution was to employ various temporal, spectral, and temporal-spectral domains to extract rich features. It helped to catch the seizure characteristics and consequently enhanced the detection rate. Another novelty was to utilize the DSET to produce high-quality feature sets. It improved the classification accuracy.

The main advantage of the proposed ESD is its optimized feature selection. Therefore, a few features are required in the classification task, which is proper for online implementation. However, the utilized data in this approach is artifact-free. This means that before applying the proposed ESD scheme to the EEG signals, all types of noise and artifacts must be removed using a de-noising method.

References

[1] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain*, vol. 130, no. 2, pp. 314–333, Feb. 2007.

[2] E. Alickovic, J. Kevric, and 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*, vol. 39, pp. 94–102, Jan. 2018.

[3] S. Chakrabarti, A. Swetapadma, and P. K. Pattnaik, "A review on epileptic seizure detection and prediction using soft computing techniques," in *Smart Techniques for a Smarter Planet*. Cham, Switzerland: Springer, 2019, pp. 37–51.

[4] M. Radman, A. Chaibakhsh, N. Nariman-zadeh, and H. He, "Generalized sequential forward selection method for channel selection in EEG signals for classification of left or right hand movement in BCI," in *Proc. 9th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Oct. 2019, pp. 137–142.

[5] J. A. de la O Serna, M. R. A. Paternina, A. Zamora-Mendez, R. K. Tripathy, and R. B. Pachori, "EEG-rhythm specific Taylor–Fourier filter bank implemented with O-splines for the detection of epilepsy using EEG signals," *IEEE Sensors J.*, vol. 20, no. 12, pp. 6542–6551, Jun. 2020.

[6] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8659–8666, Dec. 2010.

[7] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1084–1093, May 2007.

[8] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, Nov. 2001, Art. no. 061907.

[9] O. Hanosh, R. Ansari, K. Younis, and A. E. Cetin, "Real-time epileptic seizure detection during sleep using passive infrared sensors," *IEEE Sensors J.*, vol. 19, no. 15, pp. 6467–6476, Aug. 2019.

[10] P. Boonyakitanont, A. Lek-uthai, K. Chomtho, and J. Songsiri, "A review of feature extraction and performance evaluation in epileptic seizure detection using EEG," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101702.

[11] J. Kevric and A. Subasi, "The effect of multiscale PCA de-noising in epileptic seizure detection," *J. Med. Syst.*, vol. 38, no. 10, p. 131, Oct. 2014.

[12] A. Subasi, J. Kevric, and M. Abdullah Canbaz, "Epileptic seizure detection using hybrid machine learning methods," *Neural Comput. Appl.*, vol. 31, no. 1, pp. 317–325, Jan. 2019.

[13] S. M. S. Alam and M. I. H. Bhuiyan, "Detection of seizure and epilepsy using higher order statistics in the EMD domain," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 2, pp. 312–318, Mar. 2013.

[14] M. Niknazar, S. R. Mousavi, B. Vosoughi Vahdat, and M. Sayyah, "A new framework based on recurrence quantification analysis for epileptic seizure detection," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 3, pp. 572–578, May 2013.

[15] M. Peker, B. Sen, and D. Delen, "A novel method for automated diagnosis of epilepsy using complex-valued classifiers," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 108–118, Jan. 2016.

[16] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the k-means clustering and a multilayer perceptron neural network model," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13475–13481, 2011.

[17] A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi, "Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 4, pp. 888–896, Jul. 2017.

[18] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, and J. S. Suri, "Automated diagnosis of epileptic EEG using entropies," *Biomed. Signal Process. Control*, vol. 7, no. 4, pp. 401–408, Jul. 2012.

[19] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Comput. Intell. Neurosci.*, vol. 2007, pp. 1–13, Aug. 2007.

[20] R. Upadhyay, P. K. Padhy, and P. K. Kankar, "A comparative study of feature ranking techniques for epileptic seizure detection using wavelet transform," *Comput. Electr. Eng.*, vol. 53, pp. 163–176, Jul. 2016.

[21] A. R. Hassan, A. Subasi, and Y. Zhang, "Epilepsy seizure detection using complete ensemble empirical mode decomposition with adaptive noise," *Knowl.-Based Syst.*, vol. 191, Mar. 2020, Art. no. 105333.

[22] M. Li, W. Chen, and T. Zhang, "FuzzyEn-based features in FrFT-WPT domain for epileptic seizure detection," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 9335–9348, Dec. 2019.

[23] F. Riaz, A. Hassan, S. Rehman, I. K. Niazi, and K. Dremstrup, "EMD-based temporal and spectral features for the classification of EEG signals using supervised learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 28–35, Jan. 2016.

[24] A. R. Hassan and M. A. Haque, "Epilepsy and seizure detection using statistical features in the complete ensemble empirical mode decomposition domain," in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2015, pp. 1–6.

[25] G. Wang, Z. Deng, and K.-S. Choi, "Detection of epilepsy with electroencephalogram using rule-based classifiers," *Neurocomputing*, vol. 228, pp. 283–290, Mar. 2017.

[26] A. B. Das, M. I. H. Bhuiyan, and S. M. S. Alam, "Classification of EEG signals using normal inverse Gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection," *Signal, Image Video Process.*, vol. 10, no. 2, pp. 259–266, Feb. 2016.

[27] T. Zhang, W. Chen, and M. Li, "AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier," *Biomed. Signal Process. Control*, vol. 31, pp. 550–559, Jan. 2017.

[28] A. R. Hassan and A. Subasi, "Automatic identification of epileptic seizures from EEG signals using linear programming boosting," *Comput. Methods Programs Biomed.*, vol. 136, pp. 65–77, Nov. 2016.

[29] A. R. Hassan, S. Siuly, and Y. Zhang, "Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating," *Comput. Methods Programs Biomed.*, vol. 137, pp. 247–259, Dec. 2016.

[30] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Hoboken, NJ, USA: Wiley, 2013.

[31] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *J. Neural Eng.*, vol. 4, no. 2, p. R32, 2007.

[32] B. B. Mandelbrot, D. E. Passoja, and A. J. Paullay, "Fractal character of fracture surfaces of metals," *Nature*, vol. 308, no. 5961, pp. 721–722, Apr. 1984.

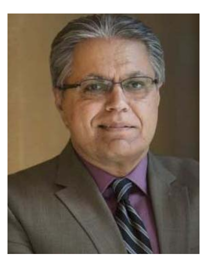
[33] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.
[34] D. Zhang, "Wavelet transform," in *Fundamentals Image Data Mining*. Cham, Switzerland: Springer, 2019, pp. 35–44.
[35] H. Liu and H. Motoda, *Computational Methods of Feature Selection*. Boca Raton, FL, USA: CRC Press, 2007.
[36] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.
[37] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.
[38] T. Zhang, T. Zhu, P. Xiong, H. Huo, Z. Tari, and W. Zhou, "Correlated differential privacy: Feature selection in machine learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 2115–2124, Mar. 2020.
[39] M. Kordestani, A. Chaibakhsh, and M. Saif, "SMS—A security management system for steam turbines using a multisensor array," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3813–3824, Sep. 2020.
[40] V. Ablavsky and M. R. Stevens, "Automatic feature selection with applications to script identification of degraded documents," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, Aug. 2003, pp. 750–754.
[41] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.
[42] Y. Lei, Z. He, Y. Zi, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique," *Mech. Syst. Signal Process.*, vol. 22, no. 2, pp. 419–435, Feb. 2008.
[43] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," 2012, *arXiv:1202.3725*. [Online]. Available: http://arxiv.org/abs/1202.3725
[44] F. Castanedo, "A review of data fusion techniques," *Sci. World J.*, vol. 2013, pp. 1–19, 2013.
[45] M. Moradi, A. Chaibakhsh, and M. Mohammadi, "Multi-sensor feature fusion and grey wolf optimizer-based support vector machine for transient fault detection in a once-through power plant," in *Proc. 27th Iranian Conf. Electr. Eng. (ICEE)*, Apr. 2019, pp. 2067–2073.
[46] M. Rezamand, M. Kordestani, R. Carriveau, D. S.-K. Ting, and M. Saif, "An integrated feature-based failure prognosis method for wind turbine bearings," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 3, pp. 1468–1478, Jun. 2020.
[47] G. Shafer, *A Mathematical Theory of Evidence*, vol. 42. Princeton, NJ, USA: Princeton Univ. Press, 1976.
[48] M. Beynon, D. Cosker, and D. Marshall, "An expert system for multi-criteria decision making using dempster shafer theory," *Expert Syst. Appl.*, vol. 20, no. 4, pp. 357–367, May 2001.
[49] K. Sentz and S. Ferson, *Combination of Evidence in Dempster-Shafer Theory*, vol. 4015. Albuquerque, NM, USA: Sandia National Laboratories, 2002.
[50] X. Liu, Z. Liu, G. Wang, Z. Cai, and H. Zhang, "Ensemble transfer learning algorithm," *IEEE Access*, vol. 6, pp. 2389–2396, 2018.
[51] S. Huda, J. Yearwood, H. F. Jelinek, M. M. Hassan, G. Fortino, and M. Buckland, "A hybrid feature selection with ensemble classification for imbalanced healthcare data: A case study for brain tumor diagnosis," *IEEE Access*, vol. 4, pp. 9145–9154, 2016.
[52] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, Boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.
[53] J. Yoon, W. R. Zame, and M. van der Schaar, "ToPs: Ensemble learning with trees of predictors," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2141–2152, Apr. 2018.
[54] A. Verikas, Z. Kalsyte, M. Bacauskiene, and A. Gelzinis, "Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: A survey," *Soft Comput.*, vol. 14, no. 9, pp. 995–1010, Jul. 2010.
[55] L. N. Eeti and K. M. Buddhiraju, "Comparison of AdaBoost.M2 and perspective based model ensemble in multispectral image classification," in *Proc. IEEE Annu. India Conf. (INDICON)*, Dec. 2016, pp. 1–5.
[56] F. Li, L. Xu, P. Siva, A. Wong, and D. A. Clausi, "Hyperspectral image classification with limited labeled training samples using enhanced ensemble learning and conditional random fields," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2427–2438, Jun. 2015.

Moein Radman received the bachelor's degree from the University of Guilan, Rasht, Iran, in 2012, and the master's degree in mechanical engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2015, respectively. He is currently pursuing the Ph.D. degree with the University of Guilan. During the Ph.D. studies, he is also a member of ISACLAB. His current research interests include the brain–computer interface, signal processing, data mining, and machine learning, with applications to biomedical and health monitoring.

Milad Moradi received the B.Sc. and M.Sc. degrees from the University of Guilan, Rasht, Iran, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the University of Windsor. The selection as a brilliant talent student during his B.Sc. course, winning a prestigious prize from Iran's National Elites Foundation, and working as a Research Assistant with ISACLAB, University of Guilan from 2016 to 2020 are his honors and research experiences. His research interests include machine learning, fault diagnosis, and application of these areas in biomedical engineering, and other complex engineering systems areas.

Ali Chaibakhsh (Member, IEEE) received the B.Sc. degree from the University of Guilan, Rasht, Iran, in 2002, and the M.Sc. degree and the Ph.D. degree in mechanical engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2004 and 2009, respectively. He is an Associate Professor of Mechanical Engineering with industrial and academic research backgrounds in process control and instrumentation. His main research interests include intelligent systems design and their applications in industrial process systems. His special expertise areas are within control systems, modeling and simulation, and fault diagnosis.

Mojtaba Kordestani (Senior Member, IEEE) received the bachelor's degree in electronic engineering from Malek Ashtar University in 2002, and the master's degree in control engineering from Tehran Azad University, Iran, in 2008, and the Ph.D. degree in electrical engineering from the University of Windsor, Canada, in 2018. He is a Postdoctoral Fellow with the University of Windsor. His research interests include control, estimation, fault diagnostics, and failure prognosis. He has published about 40 refereed journal articles and conference papers in these areas.



Mehrdad Saif (Senior Member, IEEE) received the B.S., M.S., and D.Eng. degrees in electrical engineering from Cleveland State University, Cleveland, OH, USA, in 1982, 1984, and 1987, respectively. He has been the Dean of the Faculty of Engineering, University of Windsor, since July 2011. Dr. Saif's research interests include systems and control, estimation and observer theory, model-based fault diagnostics, condition monitoring, diagnostics and prognostic, and application of these areas to automotive, power, autonomous systems, and other complex engineering systems. He has published over 350 refereed journal article and conference papers plus an edited book in these areas.