Research Article

Ensemble Classifier for Epileptic Seizure Detection for Imperfect EEG Data

Khalid Abualsaud,[1,2] Massudi Mahmuddin,[2] Mohammad Saleh,[1] and Amr Mohamed[1]

[1]*Department of Computer Science & Engineering, College of Engineering, Qatar University, P.O. Box 2713, Doha, Qatar*
[2]*Computer Science Department, Graduate School of Computing, University Utara Malaysia (UUM), 06010 Sintok, Kedah, Malaysia*

Correspondence should be addressed to Khalid Abualsaud; k.abualsaud@qu.edu.qa

Brain status information is captured by physiological electroencephalogram (EEG) signals, which are extensively used to study different brain activities. This study investigates the use of a new ensemble classifier to detect an epileptic seizure from compressed and noisy EEG signals. This noise-aware signal combination (NSC) ensemble classifier combines four classification models based on their individual performance. The main objective of the proposed classifier is to enhance the classification accuracy in the presence of noisy and incomplete information while preserving a reasonable amount of complexity. The experimental results show the effectiveness of the NSC technique, which yields higher accuracies of 90% for noiseless data compared with 85%, 85.9%, and 89.5% in other experiments. The accuracy for the proposed method is 80% when SNR = 1 dB, 84% when SNR = 5 dB, and 88% when SNR = 10 dB, while the compression ratio (CR) is 85.35% for all of the datasets mentioned.

1. Introduction

Brain status information is captured by physiological electroencephalogram (EEG) signals, which are extensively used to study different brain activities. In particular, they provide important information pertaining to epileptic seizure disease, as reported previously [1–3]. Epilepsy is a neurological disorder involving disturbances to the nervous system that are induced by brain damage. It has been reported [4] that 1% of the population worldwide is affected by this disease. Visual inspection of EEG signals can be very difficult and time consuming due to the difficulty of maintaining a high level of concentration during a lengthy inspection; this difficulty increases operator errors [5, 6]. Therefore, artificial intelligence techniques are proposed to enhance the process of epileptic seizure detection.

Recently, ensemble methods for EEG signal classification have attracted growing attention from both academia and industry. Sun et al. [7] evaluated the performance of three popular ensemble methods, namely, bagging, boosting, and random subspace ensembles. They reported that the capability of the ensemble methods is subject to the type of base classifiers, particularly the settings and parameters used for each individual classifier. Dehuri et al. [8] presented the ensemble of radial basis function neural networks (RBFNs) method to identify epileptic seizures. This method was based on the bagging approach and used differential evolution- (DE-) RBFNs as the base classifier. He et al. [9] proposed a signal-strength-based combining (SSC) method to support decision making in EEG classification. The results show that the proposed SSC method is competitive with the existing classifiers. Wang et al. [5] proposed a bag-of-words model for biomedical EEG and ECG time series that are represented as a histogram of the code words. The results of the proposed model are insensitive to the used parameters and are also robust to noise.

Feature extraction techniques proposed in the literature can generally be categorized into time-domain- or frequency-domain-based according to the features used. These techniques were used in several research works [10, 11].

Time-domain features are easily computed, and their time complexity is usually manageable [10]. Vidaurre et al. [12] proposed a time-domain-parameter- (TDP-) based feature extraction method. It is a generalized form of the Hjorth parameter and can be computed efficiently. The TDP feature is then fed to a linear discriminant analysis (LDA) classifier that is utilized in a brain computer interface application. Mohamed et al. [13] proposed five time-domain features, namely, sum, average, standard deviation, zero crossing, and energy. Subsequently, they used a set of classifiers to detect epileptic seizures. The output of the classifiers was then combined, using the Dempster rule of combination, for a final system decision. A classification accuracy of 89.5% was achieved. Nigam and Graupe [14] proposed an automated neural network-based epileptic seizure detection model, called LAMSTAR. Two features, namely, the relative spike amplitude and the spike rhythmicity of the EEG signals, were calculated and utilized to train the neural network.

Frequency-domain features are usually obtained by transforming EEG signals into their basic frequency components [6]. The characteristics of these components primarily fall within four frequency bands [15]. One classification system uses a one-second time window to extract relevant features [16]. The fast Fourier transformation (FFT) is used to transform the data in the window into the frequency domain. To distinguish between several brain states, frequency components from 9 to 28 Hz were studied and presented to a modified version of Kohonen's learning vector quantization classifier. Wang et al. [17] proposed an EEG classification system for epileptic seizure detection. It consists of three main stages, namely, (1) the best basis-based wavelet packet entropy method, which is used to represent EEG signals by wavelet packet coefficients; (2) a k-NN classifier with the cross-validation method in the training stage of hierarchical knowledge base (HKB) construction; and (3) the top-ranked discriminating rules from the HKB used in the testing stage to compute the classification accuracy and rejection rate. They reported a classification accuracy of close to 100%; however, their experiments considered only healthy subjects which is class A and epileptic seizure active subjects which is class E data and never considered seizure-free intervals which are class C or class D. Trivially, neglecting such classes eliminated the main source of difficulty in this classification process. Moreover, the data of their experiments is only noiseless and used a single classifier, k-NN. Bajaj and Pachori [18] proposed a new method for classifying seizure and nonseizure states. The method used the empirical mode decomposition (EMD) technique based on bandwidth features. The features were used as an input to a least squares SVM classifier. Sharma et al. [19] also presented a classification method of two focal and nonfocal EEG signals. Data from five epilepsy patients who had longstanding drug resistance has been used to test the method. The only base classifier used was the least square support vector machine (LS-SVM). Average sample entropy and average variance of the intrinsic mode functions (IMFs) were obtained based on EMD of EEG signals. The results show that the proposed method gives a classification accuracy of 85%. The second-order difference plot method of IMF [20] has been used as a feature for epileptic seizure classification.

The computed area from the diagnostic signal demonstrates that the IMF detection is found to be a significant parameter for analysis of both healthy and unhealthy subjects [21]. The mean frequency feature of the IMFs has come up as a feature to identify variance between ictal and seizure-free EEG signals [22]. Wavelet and multiwavelet transformations have been included in analysis and classification of EEG time-frequency of the epileptic seizure [23]. However, these methods used noiseless data, while in this research both noiseless and noisy data were used. Compared with our methods, these datasets are only using the LS-SVM as a base classifier, while in this research 4 different classifiers were used. In another research work [24], the discrete wavelet transform (DWT) was used to transform EEG signals into their frequency components. For each wavelet subband, the maximum, minimum, mean, and standard deviation were then calculated and used as an input vector for a set of classifiers. The results revealed that the neural network classifier outperformed other classifiers with a 95% accuracy rate, while the k-NN classifier was more tolerant to imperfect data.

Other reported techniques utilize a mix of time- and frequency-domain features, such as in Valderrama et al. [25]. The first, second, third, and fourth statistical moments (i.e., mean, variance, skewness, and kurtosis, resp.) were extracted using the EEG amplitudes. Along with these time-domain features, energy and other frequency-domain features were extracted. A support vector machine (SVM) was then applied to the obtained features for seizure classification. Weng and Khorasani [26] proposed a method that uses the average EEG amplitude, average EEG duration, coefficient of variation, dominant frequency, and average power spectrum as features that are input to an adaptive structured neural network.

The classification techniques that are reported throughout the literature provide satisfactory performance data indicating that the EEG data are not contaminated by different factors. Although the raw EEG datasets (free of artifacts) were used, the lossy compression will introduce signal distortion, which will affect the reconstructed data. Therefore, wireless EEG data often are compressed before transmission, which means that some important information may get lost during the reconstruction process on the receiver side. Moreover, a wireless channel may augment the transmission problem by adding noise artifacts to the transmitted data. Therefore, a prospective classification technique should consider the uncertainty in the EEG data to guarantee the targeted performance.

In this paper, considering that the EEG signal is in nature bandwidth hungry, several works have considered in-network processing for either compressing EEG data [27] or transferring EEG features instead of delivering the raw uncompressed signal [28]. Another reason considering that the sensor is battery-operated, if the data is transmitted without compression, the battery power will be consumed faster. Therefore, we propose unified framework where the EEG data is compressed using compressive sensing (CS) and sent using two different types of channels. In the first, it was sent over a noiseless channel while the second was sent over the additive white Gaussian noise (AWGN) wireless channel in three

different cases where SNR = 1, 5, and 10 dB. On the other hand, the compressed data was reconstructed and statistical features were extracted. Finally, the data obtained was contaminated due to the reconstruction and the different values of noise. A distinct factor that distinguishes this research work is the proposal of a new framework and new noise-aware signal combination (NSC) method that improves the classification of the reconstructed and noisy EEG data. To address this scenario, a unified framework has been designed, which presents compressive sensing-based technique to send compressed EEG data over AWGN wireless channel, reconstruction, and feature extraction using time-frequency-domain analysis in preparation of data classification. Such framework makes this work more practical because it performs classification considering data imperfection due to compression and wireless channel transmission.

Thus, the main contributions of this paper are as follows: (1) a framework for EEG compression and classification using CS and AWGN channel transmission has been developed, (2) a new noise-aware signal combination (NSC) method that supports both types of biomedical reconstructed EEG data, noiseless and noisy, has been proposed, and (3) a series of comprehensive experiments are conducted to investigate the effectiveness and robustness of the NSC method for classifying EEG signals.

The remainder of this paper is structured as follows. In Section 2, we present an EEG-based framework, including compressive sensing, the DCT method, and feature extraction, as well as the set of classifiers that have been used. Section 3 describes the proposed system model, which mainly includes an ensemble classification method, a description of the EEG datasets, an epileptic seizure detection system model, and a proposed noise-aware signal combination (NSC) method. The results and discussions of extensive experiments investigating the effectiveness and robustness of NSC for EEG signal classification are illustrated in Section 4, and the paper is concluded in Section 5.

2. Materials and Methods

Firstly, this section describes the framework of the implemented system and its architecture as well as the main components. Secondly, a description of the EEG datasets, which is being used to distinguish between healthy subjects and epilepsy subjects, is presented. Thirdly, the compressive sensing integrated with the discrete cosine transform and measurement matrix is being presented. Fourthly, feature extraction in described, and finally, a brief of classification methods are demonstrated.

2.1. Architecture of the Framework. The system model is composed of two main parts, the transmitter and the receiver, shown in Figure 1. The transmitter has 4096 samples raw electroencephalography (EEG) signals, represented by (x), and uses a CS technique to downsample the data based on a sparse measurement matrix. In this framework, we used DCT and the basis ψ for different quantities of M to obtain the compressed data $\hat{x}$ that will be transmitted over noiseless

and noisy channels (i.e., radio frequency (RF) or Bluetooth). Several sources of noise can alter the data, including wireless channel fading, path loss, and thermal noise at the receiver. In this paper, without loss of generality, we consider the thermal noise using the AWGN model at the receiving side as the most widely used model for representing thermal noise [29–32]. We control the noise level using the signal to noise ratio (SNR) to demonstrate data imperfection and to study the behavior of the different classification techniques in the presence of such noise.

The receiver, which receives the compressed signal of size M, reconstructs the EEG data using an inverse DCT (iDCT) and basis pursuit to obtain the reconstructed signal. The iDCT reconstruction algorithm is used for the DCT, or an optimization problem with certain constraints is solved for the CS [30, 33, 34]. For example, in the following, for a given compressed measurement y at the receiver, the signal x can be reconstructed by solving one of the following optimization problems:

$$\begin{aligned} \text{Minimum} \quad & \|x_0\|_2 \\ \text{Subject to} \quad & y_i = \langle \Phi_i \Psi x_{0i} \rangle . \end{aligned} \tag{1}$$

Using a trick of basis pursuit, find the vector x_0 with the lowest L_2 norm that satisfies the observations. For an N-dimensional EEG signal x,

$$x = \Psi \alpha, \tag{2}$$

where Ψ is a discrete cosine transform (DCT) basis, α is the wavelet, and both are domain coefficients. At the receiver side, once we detect α, iDCT will be utilized to reconstruct the original signal from α. Figure 1 shows the framework that has compressive sensing and data reconstruction as well as the classification processes for EEG-based epileptic seizure [24].

2.2. EEG Datasets Descriptions. The datasets used in this work originated from Andrzejak et al. [35], which are widely used for automatic epileptic seizure detection. It contains both normal and epileptic EEG datasets. The EEG datasets were collected from five patients. The datasets consist of five sets termed A, B, C, D, and E. Each set was composed of 100 single channel EEG segments of 23.6-second duration. For sets A and B, the patients were relaxed and awake with eyes open and eyes closed, respectively. Segments of sets A and B were taken from surface EEG recordings, which were performed using a standardized electrode placement scheme on five healthy subjects. The segments in set C were recorded from the hippocampal formation of the opposite from the epileptogenic zone. The segments in set D were recorded from within the epileptogenic zone. Sets C, D, and E originated from EEG archive of presurgical diagnosis. Sets C and D both contained only the activity measured during seizure-free intervals. Finally, only set E contained seizure activity. All EEG signals were recorded with the same 128-channel amplifier system (neglecting electrodes that have strong eye movement artifacts (A and B) or pathological activity (C, D, and E)). The data were constantly written at a sampling rate of 173.61 Hz to the disk of the data acquisition computer

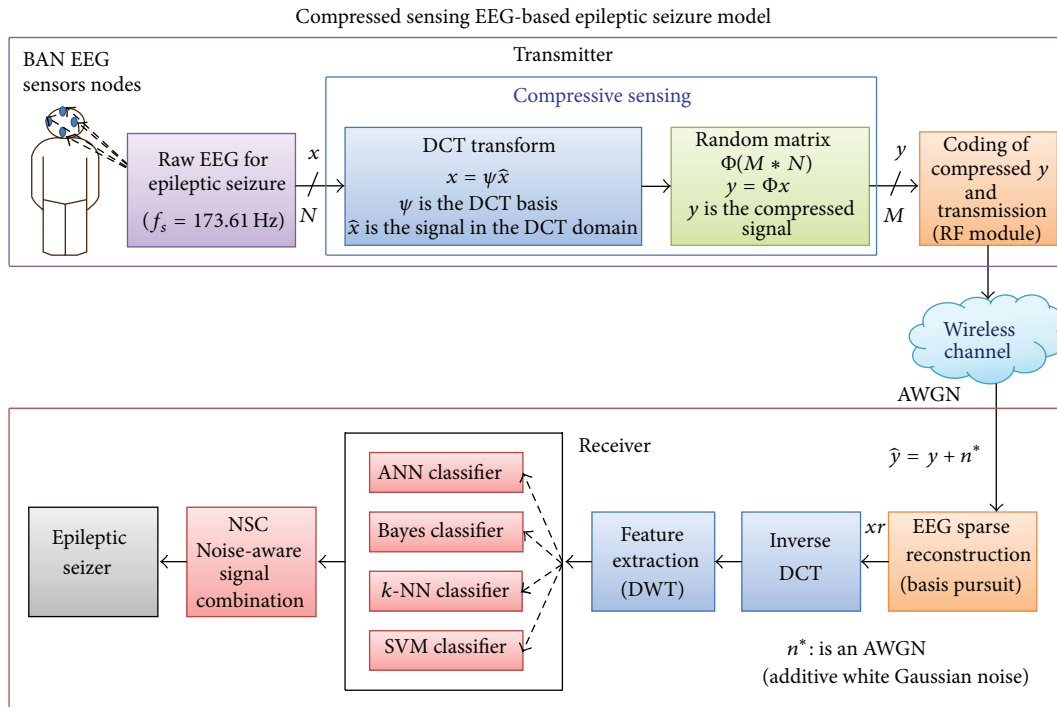Compressed sensing EEG-based epileptic seizure model

FIGURE 1: EEG-based epileptic seizure framework.

system. Kumar et al. [36] reported that when the performance of sets A and E was compared with set B and set E, it was concluded that set A and set E were more efficient [36]. In addition, set A and set B are similar in feature properties that are hard for the classifier to distinguish between both sets representing healthy patients. It is worth noting that, during performance evaluation, we have conducted many experiments using different groups of classes (i.e., one group was all 5 classes; another group was A, C, and E, etc.), and the best results were evident for the class groups of A, C, and E. Therefore, in this research paper, we opted to use set A to represent healthy subjects, set C to represent unhealthy with seizure-free interval subjects, and set E to represent the epileptic seizure active subjects. In this case, 300 EEG segments are used; each class consists of 100 segments. Figure 2 illustrates the ideal raw EEG signals of sets A, C, and E, respectively.

Typically, transmitters are mobile devices, which are equipped with battery sources; hence, the power consumption during data transmission is critical. Therefore, the compressive sensing (CS) and discrete cosine transform (DCT) methods have been utilized to reduce the amount of data before transmission because CS does not require much complexity for downsampling at the transmitter; this low complexity comes with the cost of higher complexity on the receiver side [29].

2.3. Compressive Sensing.
Compressive sensing (CS) technique [37] is used to reduce the size of the data that was sent from the transmitter to the receiver, and thus CS has been considered for efficient EEG acquisition and compression
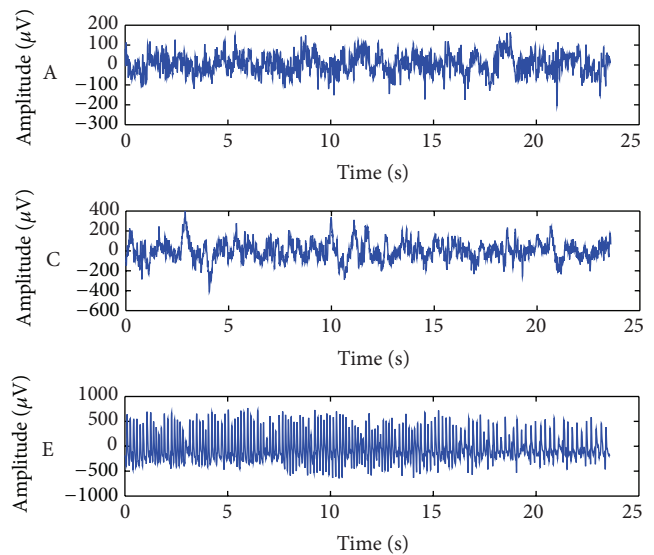
FIGURE 2: Example of three different classes of EEG signals taken from different subjects.

in several applications [31, 38, 39]. Signal acquisition is the critical part of most applications, where the acquisition time or the computational resources are limited, and the CS technique has the significant advantage of offloading the processing from the data acquisition step to the data reconstruction step. CS reduces the time consumed in processing at the transmitter, at the expense of higher complexity at the

receiver where more processing time and higher computational capacity are usually available. Previous research work [38, 39] focused on the sparse modeling of EEG signals and evaluating the efficiency of CS-based compression in terms of signal reconstruction errors and time required.

An N-dimensional 4096-sample raw EEG signal x is considered to illustrate the CS compression and reconstruction. Assume that this signal is represented by a projection onto a different basis set Ψ:

$$x = \sum_{i=1}^{N} x_{0i} \Psi_i \quad \text{or} \quad x = \Psi x_0, \tag{3}$$

where x is the original signal, x_0 is the sparse of representation of x, and Ψ is an $N * N$ bases matrix.

The sparse vector x_{0i} can be calculated from the inner product of x and Ψ:

$$x_{0i} = \langle x, \Psi_i \rangle. \tag{4}$$

The basis (Ψ) can be a Gabor, Fourier, or discrete cosine transform (DCT) or a Mexican hat, linear spline, cubic spline, linear B-spline, or cubic B-spline function. In compressive sensing, Ψ is chosen such that x_0 is sparse. The vector x_0 is k-sparse if it has k nonzero entries and the remaining $(N - k)$ entries are all zeroes. In addition to the projection above, it is assumed that x can be related to another signal y:

$$y_{[M*1]} = \Phi_{[M*N]} \times x_{[N*1]}, \tag{5}$$

where Φ is a measurement matrix (also called sensing matrix) of dimensions $M * N$, and y is the compressive sensed version of x. Matrix y has dimensions $M * 1$, and data compression is achieved if $M < N$. It can be shown that this technique is possible if Φ and Ψ are incoherent. To satisfy this condition, Φ is chosen as a random matrix. The compression ratio (CR) is then defined as follows:

$$CR = \left(1 - \frac{M}{N}\right) * 100. \tag{6}$$

2.4. Discrete Cosine Transom (DCT) Method. The discrete cosine transform (DCT) is used as the basis to make the EEG signal sparse as part of the CS framework. It is a Fourier-related transform similar to the discrete Fourier transform (DFT); however, it only uses real numbers and has low computational complexity [24, 28]. Obtaining the signal $x(n)$ in the DCT domain will require a definition of the $(N + 1) * (N + 1)$ DCT transform matrix, whose elements are given by

$$[C]_{mn} = \sqrt{\frac{2}{N}} \left\{ k_m k_n \cos\left(\frac{mn\pi}{N}\right) \right\},$$

$$m, n = 0, 1, \dots, N,$$

$$k_i = \begin{cases} 1, & \text{for } i \neq 0 \text{ or } N, \\ \dfrac{1}{\sqrt{2}}, & \text{for } i = 0 \text{ or } N. \end{cases} \tag{7}$$

This matrix is unitary, and when it is applied to a data vector x of length $N + 1$, it produces a vector called X_c, where $X_c = [C] * x$, and its elements are given by

$$X_c(m) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N} k_m k_n \cos\left(\frac{mn\pi}{N}\right) x(n). \tag{8}$$

On the receiver side, the basis of the iDCT [28] is utilized in the CS decoder in order to obtain the reconstructed signal (x_r) as follows:

$$x_r(a) = \sum_{k=1}^{N} w(k) y(k) \cos\left[\frac{\pi(2a + 1)k}{2N}\right], \tag{9}$$

where N is the length of both time series and cosine transform signals, a is the time series index ($a = 1, 2, \dots, N$), k is the cosine transform index ($k = 1, 2, \dots, N$), and the window function $W(k)$ is defined as

$$w(k) = \begin{cases} \dfrac{1}{\sqrt{N}}, & k = 1, \\ \sqrt{\dfrac{2}{N}}, & 2 \leq k \leq N. \end{cases} \tag{10}$$

After obtaining the contaminated reconstructed signal (x_r), DWT is used as feature extraction and selection techniques.

2.5. Feature Extraction. EEG feature extraction plays a significant role in diagnosing most brain diseases. Obtaining useful and discriminant features depends largely on the feature extraction method used. Because EEG signals are time-varying and space-varying nonstationary signals, the discrete wavelet transform (DWT) method is widely used [17]. It captures both frequency and time location information [32, 40–42]. Using multiresolution wavelet analysis, DWT basically decomposes the EEG signals into different frequency bands.

EEG data are generally nonstationary signals, which are heavily dependent on the subject condition. The Daubechies 6 DWT was employed, where the data were sampled at a rate of 173.61 Hz. This means that the EEG data frequency is 86.81 Hz; thus, the filter length is long as well; the frequency wavelet subband is the same as the fundamental component of the EEG. Hence, decomposition level 7 was calculated based on the EEG frequency. In addition, considering our extensive experimental work on the reconstruction accuracy of different wavelet families, filter lengths, and decomposition levels [30], we used Daubechies 6 with 1–8 different decomposition levels in this research. We found that Daubechies 6 with decomposition level 7 is the optimum level in terms of classification accuracy and computational complexity of the EEG epileptic seizure category of data. Given the EEG signal $f(x)$, the wavelet series expansion is depicted [30] as follows:

$$f(x) = \sum_k c_{j0}(k) \varphi_{j0,k}(x)$$

$$+ \sum_{j=j_o}^{\infty} \sum_k d_j(k) \psi_{j,k}(x), \tag{11}$$

where $f(x) \in L^2(R)$, $L^2(R)$ is relative to the wavelet $\psi(x)$ and the scaling function $\varphi(x)$, and c_{j0} are the approximation coefficients.

In the first sum, the approximation coefficients c_{j0} can be represented as the outcome of the inner product process between the original signal $f(x)$ and the approximation function $\varphi_{j0,k}(x)$ as expressed by

$$c_{j_0}(k) = \left\langle f(x), \varphi_{jo,k}(x) \right\rangle. \tag{12}$$

In the second sum, a finer resolution is added to the approximation to provide increasing details. The function $d_j(k)$ represents the details coefficients and it can be obtained by the inner product between the original signal $f(x)$ and the wavelet function $\psi_{j,k}(x)$ calculated as

$$d_j(k) = \left\langle f(x), \psi_{j,k}(x) \right\rangle. \tag{13}$$

Generally, the classification accuracies improve when using a combination of time- and frequency-domain features rather than features solely based on either the frequency domain or the time domain [30]. Different implementation choices, including different wavelet families, filter lengths, and decomposition levels, have been utilized to extract features. Accordingly, the conventional statistical features (*maximum, minimum, mean, and standard deviation*) are extracted from each wavelet subband. The extraction rules for statistical features that have been implemented for the wavelet subband are as follows.

Maximum feature:

$$x_k \text{ such that } x_k > x_i, \quad \forall i \neq k, \ i = 1,\dots,n,$$
$$d_i(x) = \max_{i=1,\dots,k} \{d_i(x)\}. \tag{14}$$

Minimum feature:

$$x_k \text{ such that } x_k < x_i, \quad \forall i \neq k, \ i = 1,\dots,n,$$
$$d_i(x) = \min_{i=1,\dots,k} \{d_i(x)\}. \tag{15}$$

The mean can be calculated by the following:

$$\widehat{x} = \frac{1}{N}\sum_{i=1}^{n} x_i. \tag{16}$$

The standard deviation feature is given by the following:

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \widehat{x})^2. \tag{17}$$

The original EEG signal was analyzed for the wavelet subbands A7 and D7-D1. Eventually, four conventional statistical features are selected from each wavelet subband individually. As a consequence, 32 attributes are obtained from the whole subbands to be fed to the classifiers. So the features *maximum, minimum, mean, and standard deviation* contribute to the classification accuracy in this research. It has been found that these features are robust with the dynamic environment of the wireless channel [24, 28]. Meanwhile, these features have low computational complexity.

2.6. Classification Methods. EEG detection and classification play an essential role in timely diagnoses and analyze potentially fatal and chronic diseases proactively in clinical as well as various life settings [3, 43]. Liang et al. [44] proposed a systematic evaluation of EEGs by combining both complexity analysis and spectral analysis for epilepsy diagnosis and seizure detection. Approximately 60% of the features extracted from the dataset were used for training, while the remaining ones were used to test the performance of the classification procedure on randomly selected EEG signals [44].

In this research work, four different classifiers have been used, namely, ANN, naïve Bayes, k-NN, and SVM. Initially, the classifiers were developed to work individually to compare their performances. However, we developed a data fusion method for combining the output of all classifiers in order to reduce the effect of data imperfections while maximizing the classification accuracy. Each classifier belongs to a different family of classifiers and has been shown to be the best classifier in its family. However, it is expected that they may yield different classification results because they each use a different classification strategy [13, 17, 45–47]. The following provides a brief description of these classifiers.

2.6.1. Artificial Neural Network. An artificial neural network (ANN) is a mathematical model that is motivated by the structure and functional aspects of biological neural networks. To establish classification rules and perform statistical analysis, ANN is able to estimate the posterior probabilities [14, 47, 48]. The ANN has several parameters; in this paper, the ANN configuration uses training cycles = 500, learning rate = 0.3, and momentum decay = 0.2.

2.6.2. Naïve Bayes. The naïve Bayes (NB) classifier is a statistical classifier. It is a simple probabilistic classifier based on the application of Bayes' theorem. The NB method involves an assumption that makes the calculation of the NB classifiers more efficient than the exponential complexity. Simply, it works by considering that the presence of certain features of a class is irrelevant to any other features. The NB classifier considers each feature independently to calculate the feature properties that contribute to the probability of a certain class to be the outcome of the classification. It then uses Laplace correction to prevent high encounters of zero probabilities as the default configuration [13, 24, 46].

2.6.3. k-Nearest Neighbor (k-NN). The k-nearest neighbor (k-NN) algorithm compares a given test sample with training samples that are alike, where k parameter is a small positive and odd integer value. This algorithm combines two steps. First, find the k training samples that are closest to the invisible sample. Second, take the commonly occurring classification for these k samples and find the average of the values of its k-nearest neighbors in the regression. It can be defined by a distance metric called the normalized Euclidean distance, as indicated in the following equation, given two

points $Y_1 = (y_{11}, y_{12}, \ldots, y_{1n})$ and $Y_2 = (y_{21}, y_{22}, \ldots, y_{2n})$ [6, 24, 45]:

$$\text{dist}(Y_1, Y_2) = \sqrt{\sum_{i=1}^{n}(y_{1i} - y_{2i})^2}. \qquad (18)$$

In this research, the k-NN configuration uses value of $k = 10$, and *mixed measures* were selected as the measure type, which makes the *mixed Euclidean distance* the only available option.

2.6.4. Support Vector Machine. The support vector machine (SVM) learner is a strong classifier based on statistical learning theory. SVM constructs an ideal hyperplane in order to separate the data into two different classes to minimize the risks. SVM takes a set of input data and predicts, for each given input, which of the two possible classes involves the input. SVM is an integrated and powerful method for both classification and regression as well as distribution estimation. SVM operator supports types *C-SVC* and *nu-SVC* for classification tasks; *epsilon-SVR* and *nu-SVR* types for regression tasks. Finally, the *one-class* type is used for distribution estimation [13, 24, 46, 49]. In this research, SVM configuration is consist of both *nu-SVC* and *radial basis function* kernel were used for SVM configurations consist of both classification technique.

3. Ensemble Detection and Classification

Ensemble methods are introduced first, followed by the proposed ensemble system model and, finally, the ensemble method in this section.

3.1. Ensemble Classification Methods. Several combination techniques have been introduced in the literature, and each offers certain advantages and suffers from certain limitations. However, given several classifiers, the combination (fusion) method must address two critical issues: the dependency among the potentially combined classifiers and the consistency of the information contained in each classifier.

For the first issue, the classifiers must be independent because we consider each classifier to be a source of information. This means that each classifier simply works on the input feature set independently, while the classification is based on combining the outcomes of all classifiers simultaneously.

For the second issue, the classifiers may have conflicting decisions because different classifiers are expected to consider different viewpoints of the current system state. To address this anticipated conflict, an effective mechanism that is capable of quantifying the assurance in the decision of each classifier is desirable.

One of these well-known combination techniques is the majority voting. The majority voting (MV) rule technique collects the votes of all classifiers and investigates the class name that is mostly reported by the classifiers. It then chooses that class as a final decision [50]. However, MV is based on the idea that the classifiers participating in the voting process have the same weight. It completely ignores the inconsistency

that may arise among the classifiers. This, of course, can cause less capable classifiers to override more capable classifiers. Thus, the performance of the classification system can be deteriorated. Because the classifier models proposed in this work are expected to have different discriminant weights, the MV technique is not suitable as a combination method.

In contrast, in probability-based voting schemes, the combination method should assign a probability value (p) that reflects the confidence of a classifier in its viewpoint. One of these schemes can be based on an accumulated experience. For instance, a given classifier is correct in identifying a certain hypothesis 75% percent of the time, while another classifier can correctly identify a different hypothesis 30% of the time. These values can actually be interpreted as probability assignments.

If the classifiers happen to provide these different and conflicting hypotheses as an explanation of the current system state, then the classifiers should not be treated equally at the classification stage. Clearly, the first classifier is more confident in its decision than the first one. This valuable information should be incorporated into the fusion (combination) process.

For instance, we may assign a weight (p) of 0.75 to the first classifier while assigning only 0.30 as a weight to the second classifier.

Let T be the set of classifiers:

$$T = \{C_1, C_2, \ldots, C_n\}, \qquad (19)$$

and let C be the set of classes:

$$\{p_1, p_2, \ldots, p_n\}. \qquad (20)$$

Then, let $d_{i,j}$ be the decision of the classifier i and have the following definition:

$$d_{i,j} \in \{0, 1\}, \qquad (21)$$

where $i = 1, \ldots, T$ and $j = 1, \ldots, C$.

Let p_i represent the weight of the classifier i. Then, the probability-based voting decision is calculated as

$$\sum_{i=1}^{|T|} p_i d_{i,j} = \max_{j=1}^{C} \sum_{r=1}^{|T|} p_r d_{r,j}. \qquad (22)$$

Considering the weight of each classifier, (22) counts the votes from the participating classifiers.

3.2. Proposed Ensemble System Model. The proposed model consists of three stages for detecting electroencephalogram seizures, namely, statistical feature extraction, classifier prediction, and proposed noise-aware signal combination (NSC) method. The extraction of statistical features was discussed in Section 2.5. For classifier prediction, four classifiers are utilized in this model, namely, ANN, Bayes, k-NN, and SVM. These classification methods are trained using the most popular data mining tools that are an industry standard and widely used tools for research. The training process is conducted on similar data adhering to various combinations

Features: $\mathrm{DWT}(o_k) = f_{k,0}, f_{k,1}, \ldots, f_{k,l}$,
where $k = 0, \ldots, s-1$

| ANN algorithm | Bayes algorithm | k-NN algorithm | SVM algorithm |

$\mathrm{PL}_{i,j}, \mathrm{PC}_{i,j}, \mathrm{PR}_{i,j}, \mathrm{RE}_{i,j}, \mathrm{AC}_j$

NSC:
Noise-aware signal combination method
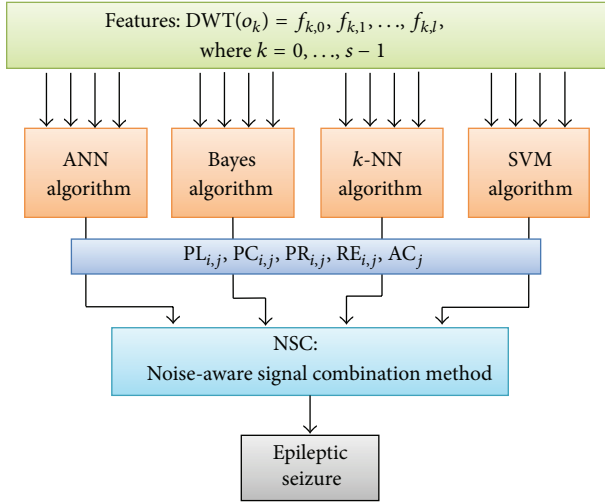
Epileptic seizure

FIGURE 3: Proposed ensemble system.

of SNRs and downsampling rates. After exhaustive iterated experiments, the trained models are saved, and their averaged performances in different scenarios are reported to the NSC. The NSC is our proposed ensemble method using combinations of probability estimates. Eventually, the ultimate classification accuracy is obtained for the epileptic seizure detection. The proposed system model is shown in Figure 3.

There are s tabular observations $O = \{o_0, o_1, \ldots, o_{s-1}\}$, where each o_i is a t-tuple of readings $R_i = (r_0, r_1, \ldots, r_{t-1})$. These observations fall into (s/m) different categories of classes $= \{c_0, c_1, \ldots, c_{m-1}\}$.

The DWT is applied to the set of observations O to obtain an l-tuple of features $R_i = (f_0, f_1, \ldots, f_{l-1})$ for each $o_i \in O$. In other words, $\mathrm{DWT} : O \rightarrow F$ such that $\mathrm{DWT}(o_k) = (f_{k,0}, f_{k,1}, \ldots, f_{k,l})$, where $f_{k,j}$ is an l-tuple extracted feature for the observation o_k obtained by DWT.

Hence, $\mathrm{DWT}(O) = \{(f_{i,0}, f_{i,1}, \ldots, f_{i,l} \mid i = 0, 1, \ldots, s-1)\}$ is the training and testing tabular l-tuple format representing the input data for the classification model in this research work.

3.3. The Ensemble Method. Several classifiers (n) built on various hypotheses $H = \{h_0, h_1, \ldots, h_{n-1}\}$ are fed with input data. The data are $\mathrm{DWT}(O)$ in a tabular l-tuple format, as discussed above. Each classifier k built on hypothesis h_k is trained on the data to predict the label representing the class c_j that best describes a given set of features $(f_{i,0}, f_{i,1}, \ldots, f_{i,l})$ corresponding to the observation o_i.

At the end of the training of each classifier, a set of performance measurements of interest is recorded. Table 1 shows some of these performance measurements. The trained model will then be saved for application to various categories of testing data. This process is replicated and repeated to yield an output that can be averaged to describe the model behavior for long run times.

The proposed ensemble classification method is fed with the output of the m trained classifiers. In a sense, the training data are bundled first into two parts and are used to train

TABLE 1: Classifier's performance measurements.

Measure	Description
$\mathrm{PL}_{i,j}$	Predicted label of o_i using hypothesis h_j
$\mathrm{PC}_{i,j}$	Confidence value predicting c_i using hypothesis h_j
$\mathrm{PR}_{i,j}$	Precision value of c_i using hypothesis h_j
$\mathrm{RE}_{i,j}$	Class recall value of c_i using hypothesis h_j
AC_j	Accuracy value of for applying hypothesis h_j

Preliminaries
  Let $O = \{o_0, o_1, \ldots, o_{s-1}\}$ be the set of observations
  Let $C = \{c_0, c_1, \ldots, c_{m-1}\}$ be the set of class labels
  Let $H = \{h_0, h_1, \ldots, h_{n-1}\}$ be the set of hypotheses
  Let h_n be the hypothesis of the combined classifier
  Let $\mathrm{PL}_{i,j}, \mathrm{PC}_{i,j}, \mathrm{PR}_{i,j}, \mathrm{RE}_{i,j}, \mathrm{AC}_j$ be the j predicted class,
  class confidence, class accuracy, class recall,
  and accuracy of h_i
PROCESS
  $\forall h_i \mid i = 0, \ldots, n-1$
  $$\overline{\mathrm{PR}}_{i,j} = \frac{\mathrm{PR}_{i,j}}{\sum_{j=0}^{n-1} \mathrm{PR}_{i,j}}, \quad \overline{\mathrm{RE}}_{i,j} = \frac{\mathrm{RE}_{i,j}}{\sum_{j=0}^{n-1} \mathrm{RE}_{i,j}},$$
  $$\overline{\mathrm{AC}}_j = \frac{\mathrm{AC}_j}{\sum_{j=0}^{n-1} \mathrm{AC}_j}$$
  if $\exists\, O$ /** *on receiving a periodical batch*
  of observations */
    $\forall o_i \mid i = 0, \ldots, s$
    $W = 0$
    $\forall h_j \mid j = 0, \ldots, m-1$
    $k = \mathrm{PL}_{i,j} \mid k \in C$
    $w(h_j) = \sigma(\overline{\mathrm{PR}}_{k,j} + \overline{\mathrm{RE}}_{k,j}) + (1 - \sigma)\overline{\mathrm{AC}}_j$
    $W \mathrel{+}= w(h_j)$
    $\forall h_i \mid i = 0, \ldots, n-1$
    $P(h_i) = \dfrac{w(h_i)}{W}$
    $\mathrm{PL}_{i,n} = \max(P(h_i)\vert_{i=0}^{n-1})$
    Record prediction
  Calculate h_n *Performance measurements of interest*

ALGORITHM 1: Noise-aware signal combination algorithm.

the m classifiers on the patterns within each bundle. Finally, the classification decision of a testing sample is obtained from an ensemble of the decisions from the corresponding m classifiers at each layer using the noise-aware signal combination method. A subset of the performance measures of each classifier together with the predicted class label $c \in C$ for an observation $o \in O$ provided by each classifier with hypothesis $h \in H$ are the input to the hypothesis used by this combined classifier.

The confusion matrix for each hypothesis h_k based on the reported performance results of the trained hypothesis h_k is calculated using the algorithm shown in Algorithm 1. An entry MRP (i, j) in the matrix of reported performance results for hypothesis h_k represents the frequency of predicting class j as being class i. Therefore, MRP (i, i) represents the frequency of correct predictions being in class i, while $\sum_{j \neq i}^{m} \mathrm{MRP}\,(i, j)$ is

TABLE 2: Confusion matrix for noiseless data.

(a)

	True class label			$\overline{\text{PR}}$
	c_0	c_1	c_2	
Predicted class label				
c_0	0.88	0.12	0.02	0.273
c_1	0.12	0.85	0.02	0.252
c_2	0.00	0.03	0.96	0.245
$\overline{\text{RE}}$	0.247	0.288	0.251	
$h_0 = \text{ANN}$		$\overline{\text{AC}_0}$	0.259	

(b)

	True class label			$\overline{\text{PR}}$
	c_0	c_1	c_2	
Predicted class label				
c_0	0.90	0.23	0.00	0.252
c_1	0.10	0.75	0.01	0.256
c_2	0.00	0.02	0.99	0.248
$\overline{\text{RE}}$	0.253	0.254	0.258	
$h_1 = \text{NB}$		$\overline{\text{AC}_0}$	0.254	

(c)

	True class label			$\overline{\text{PR}}$
	c_0	c_1	c_2	
Predicted class label				
c_0	0.96	0.37	0.01	0.226
c_1	0.04	0.63	0.09	0.244
c_2	0.00	0.00	0.90	0.253
$\overline{\text{RE}}$	0.270	0.214	0.235	
$h_2 = k\text{-NN}$		$\overline{\text{AC}_0}$	0.239	

(d)

	True class label			$\overline{\text{PR}}$
	c_0	c_1	c_2	
Predicted class label				
c_0	0.97	0.19	0.07	0.249
c_1	0.03	0.81	0.12	0.248
c_2	0.00	0.00	0.81	0.253
$\overline{\text{RE}}$	0.230	0.244	0.256	
$h_3 = \text{SVM}$		$\overline{\text{AC}_0}$	0.248	

TABLE 3: Confusion matrix for noisy data (SNR = 1 dB).

(a)

	True class label			$\overline{\text{PR}}$
	c_0	c_1	c_2	
Predicted class label				
c_0	0.88	0.38	0.04	0.249
c_1	0.12	0.61	0.02	0.267
c_2	0.00	0.01	0.94	0.249
$\overline{\text{RE}}$	0.257	0.251	0.255	
$h_0 = \text{ANN}$		$\overline{\text{AC}_0}$	0.257	

(b)

	True class label			$\overline{\text{PR}}$
	c_0	c_1	c_2	
Predicted class label				
c_0	0.82	0.40	0.00	0.248
c_1	0.18	0.59	0.00	0.252
c_2	0.00	0.01	1.00	0.249
$\overline{\text{RE}}$	0.239	0.243	0.271	
$h_1 = \text{NB}$		$\overline{\text{AC}_0}$	0.217	

(c)

	True class label			$\overline{\text{PR}}$
	c_0	c_1	c_2	
Predicted class label				
c_0	0.92	0.46	0.00	0.246
c_1	0.08	0.53	0.07	0.256
c_2	0.00	0.01	0.93	0.249
$\overline{\text{RE}}$	0.268	0.218	0.252	
$h_2 = k\text{-NN}$		$\overline{\text{AC}_0}$	0.264	

(d)

	True class label			$\overline{\text{PR}}$
	c_0	c_1	c_2	
Predicted class label				
c_0	0.81	0.30	0.05	0.257
c_1	0.19	0.70	0.13	0.225
c_2	0.00	0.00	0.82	0.252
$\overline{\text{RE}}$	0.236	0.288	0.222	
$h_3 = \text{SVM}$		$\overline{\text{AC}_0}$	0.262	

the frequency of wrong predictions of other classes that are in class i.

Hence, PR_i, the precision of class i, is $(\text{MRP}(i, i)/\sum_{j\neq i}^m \text{MRP}(i, j))\%$, and RE_i, the recall of class j, is $(\text{MRP}(i, i)/\sum_{i\neq j}^m \text{MRP}(i, j))\%$. Finally, AC_i, the accuracy using hypothesis h_k, is the averaged precision of the classes and is given by $(\sum_i^m \text{PR}_i/m) \times 100\%$.

In the confusion matrix, an entry $\text{CM}(i, j)$ is the weighted entry $\text{MRP}(i, j)$ on class i recall. That is, $\text{CM}(i, j) = \text{MRP}(i, j)/\sum_i^m \text{MRP}(i, j)$, where $\sum_i^m \text{CM}(i, j) = 1$. The $\overline{\text{PR}}_i$ in the confusion matrix is the weighted PR_i across the set of hypotheses H given by $\text{PR}_i/\sum_j^n \text{PR}_i$, where $\sum_i^n \overline{\text{PR}}_i = 1$. $\overline{\text{RE}}_i$

and $\overline{\text{AC}}_i$ are also weighted across H and are calculated in the same manner.

The prediction of the combined classifier is calculated following the hypothesis with the highest probability calculated as

$$P\left(h_j\right) = \frac{\sigma\left(\overline{\text{PR}}_{k,j} + \overline{\text{RE}}_{k,j}\right) + (1 - \sigma)\,\overline{\text{AC}}_j}{\sum_{i=0}^{n-1} \sigma\left(\overline{\text{PR}}_{k,i} + \overline{\text{RE}}_{k,i}\right) + (1 - \sigma)\,\overline{\text{AC}}_i}, \qquad (23)$$

where k is the label of the predicted class and $\sum_{j=0}^{n-1} P(h_j) = 1$.

Tables 2–5 show the confusion matrices for the four classifiers, namely, ANN, naïve Bayes, k-NN, and SVM. These matrices represent the finalized weighted performance of the

TABLE 4: Confusion matrix for noisy data (SNR = 5 dB).

(a)

Predicted class label	c_0	c_1	c_2	$\overline{PR}$
c_0	0.94	0.31	0.02	0.260
c_1	0.06	0.68	0.03	0.267
c_2	0.00	0.01	0.95	0.249
$\overline{RE}$	0.320	0.192	0.529	
h_0 = ANN		$\overline{AC_0}$	0.258	

(b)

Predicted class label	c_0	c_1	c_2	$\overline{PR}$
c_0	0.94	0.36	0.00	0.254
c_1	0.06	0.62	0.00	0.275
c_2	0.00	0.02	1.00	0.247
$\overline{RE}$	0.252	0.172	0.391	
h_1 = NB		$\overline{AC_0}$	0.258	

(c)

Predicted class label	c_0	c_1	c_2	$\overline{PR}$
c_0	0.91	0.50	0.00	0.227
c_1	0.09	0.50	0.011	0.216
c_2	0.00	0.00	0.89	0.252
$\overline{RE}$	0.313	0.252	0.471	
h_2 = k-NN		$\overline{AC_0}$	0.231	

(d)

Predicted class label	c_0	c_1	c_2	$\overline{PR}$
c_0	0.89	0.22	0.10	0.259
c_1	0.11	0.78	0.08	0.243
c_2	0.00	0.00	0.82	0.252
$\overline{RE}$	0.306	0.185	0.575	
h_3 = SVM		$\overline{AC_0}$	0.252	

TABLE 5: Confusion matrix for noisy data (SNR = 10 dB).

(a)

Predicted class label	c_0	c_1	c_2	$\overline{PR}$
c_0	1.00	0.52	0.00	0.263
c_1	0.00	0.48	0.12	0.252
c_2	0.00	0.00	0.88	0.248
$\overline{RE}$	0.235	0.275	0.264	
h_0 = ANN		$\overline{AC_0}$	0.257	

(b)

Predicted class label	c_0	c_1	c_2	$\overline{PR}$
c_0	0.86	0.38	0.08	0.252
c_1	0.14	0.62	0.36	0.260
c_2	0.00	0.00	0.56	0.246
$\overline{RE}$	0.246	0.240	0.272	
h_1 = NB		$\overline{AC_0}$	0.254	

(c)

Predicted class label	c_0	c_1	c_2	$\overline{PR}$
c_0	1.00	0.22	0.08	0.232
c_1	0.00	0.78	0.12	0.234
c_2	0.00	0.00	0.80	0.253
$\overline{RE}$	0.262	0.206	0.239	
h_2 = k-NN		$\overline{AC_0}$	0.237	

(d)

Predicted class label	c_0	c_1	c_2	$\overline{PR}$
c_0	1.00	0.30	0.02	0.253
c_1	0.00	0.66	0.00	0.253
c_2	0.00	0.04	0.98	0.253
$\overline{RE}$	0.257	0.279	0.225	
h_3 = SVM		$\overline{AC_0}$	0.252	

trained classifiers based on noiseless data and three different levels of data noise, SNR = 1 dB, 5 dB, and 10 dB, for EEG-based epileptic seizures at M = 600 downsampling. Also, these tables show that c_0, c_1, and c_2 are representing class A, class C, and class E, respectively.

For example, Table 2 represents noiseless EEG data, classes c_0, c_1, and c_2 in vertical line are representing the predicted class label; on the other hand, in the horizontal line, we show the true class label. The normalized precision $\overline{PR}$ of class A in the first row of the four matrices is 0.273, 0.252, 0.226, and 0.249 for ANN, NB, k-NN, and SVM, respectively.

The normalized class recall $\overline{RE}$ of class A in the first four matrices is 0.247, 0.253, 0.270, and 0.230 for ANN, NB, k-NN, and SVM, respectively. Furthermore, the normalized overall accuracy $\overline{AC}$ is 0.259, 0.254, 0.239, and 0.248 for the same set of classifiers, respectively.

At the end of each experiment, the algorithm calculates the performance of each classifier, based on the recorded test results. The next section reports the obtained results and provides illustrations and discussions relevant to the performance of NSC compared with that of the other individual classifiers.

TABLE 6: Performance of the classifiers with CR = 85.35% for SNR = 1, 5, and 10 dB and noiseless EEG data.

	SNR	PR			RE			AC	
		c_0	c_1	c_2	c_0	c_1	c_2	AVG	STD
ANN	1	0.66	0.73	0.99	0.80	0.61	0.94	0.78	0.17
	5	0.74	0.88	0.99	0.84	0.68	0.95	0.82	0.14
	10	0.80	0.85	0.98	0.88	0.75	0.96	0.86	0.11
	Noiseless	0.86	0.82	0.92	0.84	0.82	0.96	0.87	0.08
NB	1	0.67	0.75	0.99	0.80	0.59	0.96	0.78	0.19
	5	0.71	0.79	0.98	0.88	0.62	0.96	0.82	0.18
	10	0.76	0.83	0.97	0.90	0.69	0.96	0.85	0.14
	Noiseless	0.80	0.87	0.98	0.90	0.75	0.99	0.86	0.12
k-NN	1	0.64	0.73	0.99	0.92	0.53	0.93	0.77	0.23
	5	0.65	0.71	1.00	0.91	0.50	0.89	0.78	0.21
	10	0.71	0.80	1.00	0.98	0.59	0.87	0.81	0.20
	Noiseless	0.72	0.83	1.00	0.96	0.63	0.90	0.83	0.18
SVM	1	0.70	0.69	1.00	0.81	0.70	0.82	0.78	0.07
	5	0.73	0.78	1.00	0.87	0.78	0.82	0.82	0.05
	10	0.76	0.84	1.00	0.94	0.80	0.82	0.85	0.08
	Noiseless	0.79	0.84	1.00	0.97	0.81	0.81	0.86	0.09
NSC	1	0.65	0.86	0.98	0.92	0.48	1.00	**0.80**	0.28
	5	0.70	0.93	0.98	0.96	0.56	1.00	**0.84**	0.24
	10	0.77	1.00	0.94	1.00	0.64	1.00	**0.88**	0.21
	Noiseless	0.83	0.95	0.94	0.98	0.74	0.98	**0.90**	0.14

4. Results and Discussion

This research work addresses EEG-based epileptic seizure data classification considering noiseless and noisy data with different values of SNR. For each point on the graphs, we have conducted 10 experiments and calculated the average accuracy and its standard deviation accordingly. The standard deviation describes the distribution range, describing how much difference occurs between successful computations, which correspond to the data imperfection. In this case, the standard deviation (SD) is important to show the difference between successive measurements to make sure that the classifiers are not affected by data imperfection. Table 6 shows the calculated performance measures of the studied classifiers with EEG-epileptic seizure data compressed with CR = 85.35% for noiseless and added noise of SNR = 1, 5, and 10 dB. The class precision (PR), class recall (RE) and the classification average (AVG) accuracy (AC), and standard deviation (STD) for each classifier for different SNR and noiseless channel conditions are also shown in Table 6.

The results for each of the individual classifiers ANN, NB, k-NN (with $k = 10$), and SVM in each SNR case together with the results of NSC are plotted to illustrate the differences in their performances. Figures 4–8 show the performance for noiseless and SNRs of 1, 5, and 10 dB, respectively. The corresponding accuracies in Table 6 are emphasized in Figures 4–7 with the line drawn at CR = 85.35%. The constraint on the desired accuracy in the case of noiseless data is to achieve 90%. The CR of 84.35% was the cutting edge of achieving this desired goal. Therefore, the performance of the classifiers at this CR value is the most
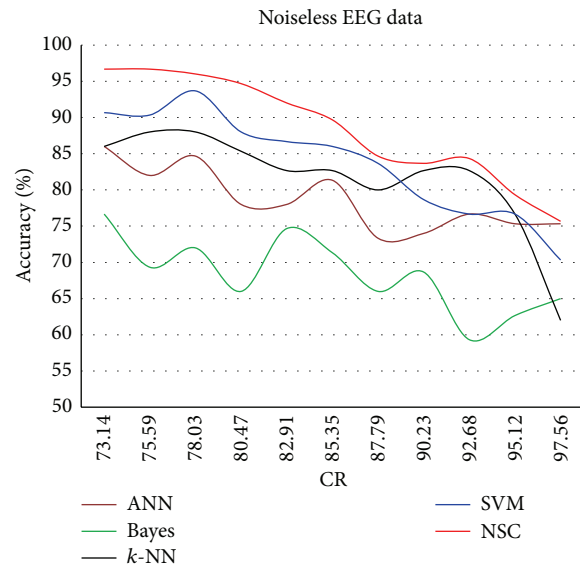

FIGURE 4: Classification accuracy against CR for noiseless data.

important to us. The overall accuracy results of all of the experiments show that this constraint is met at CR = 85.35%, while a high accuracy of 80% was maintained with very noisy data at SNR = 1 dB.

The results in Figure 4 show a trend in which the classification accuracy increases almost linearly with the decrease in CR. The decay in the accuracy seems to be reasonable in all regions, and NSC has the best accuracy, which starts to
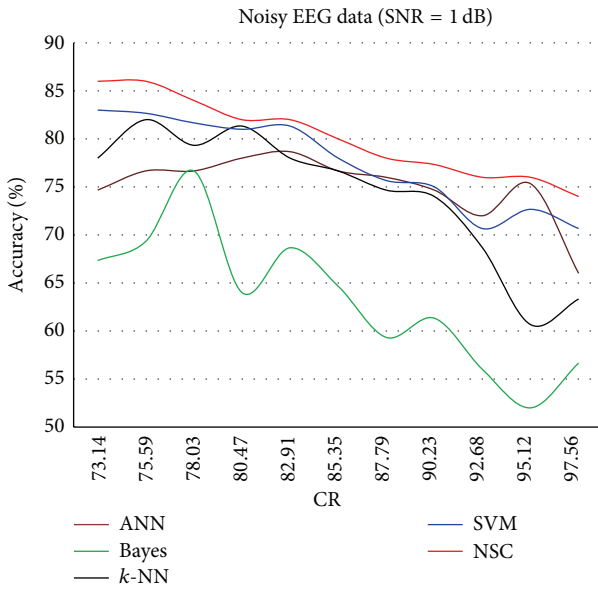
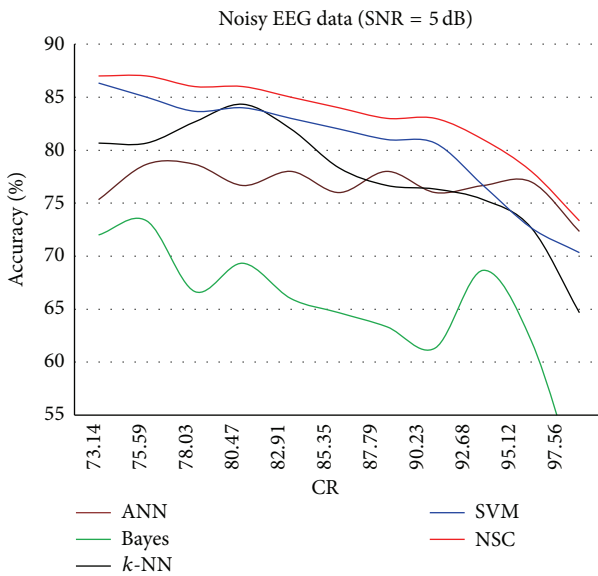Figure 5: Classification accuracy against CR for SNR = 1 dB.

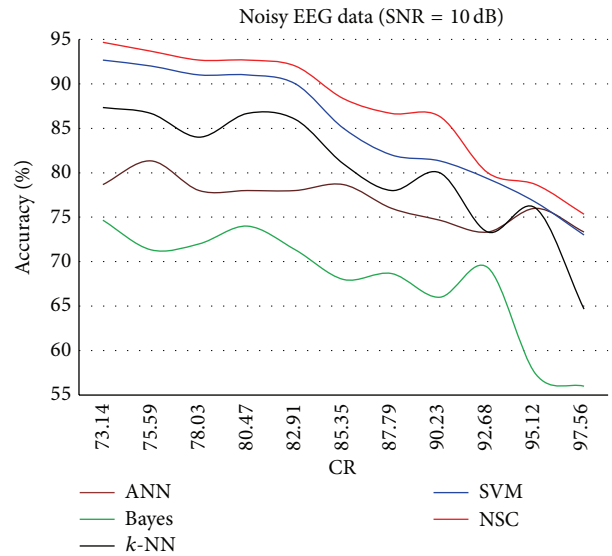Figure 7: Classification accuracy against CR for SNR = 10 dB.
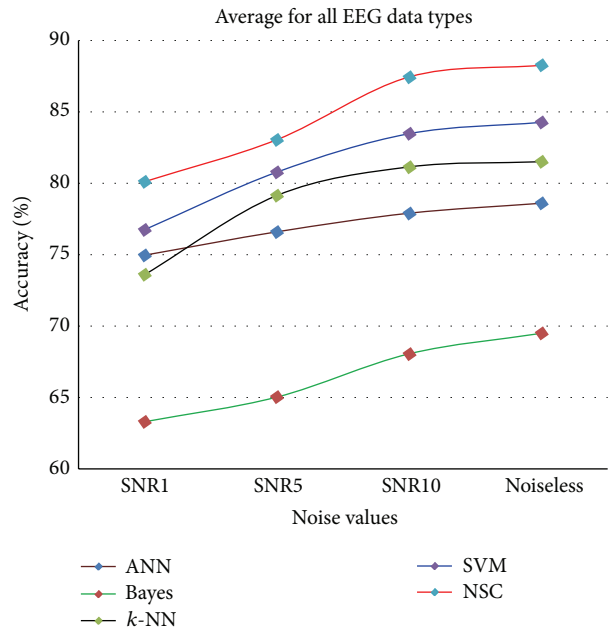

Figure 6: Classification accuracy against CR for SNR = 5 dB.

Figure 8: Average accuracy of classifiers for all CR values with SNR = 1, 5, and 10 dB and noiseless EEG data.

decay exponentially similar to the accuracy of all of the other individual classifiers.

Figure 5 shows the lower accuracies for all classifiers because of the injected quantity of AWGN (SNR = 1 dB), which is the highest noise injected in all experiments. In this case, the NSC continued to perform consistently better than the rest of the classifiers. In addition, the Bayes classifier continues to exhibit the poorest performance. The exact reported results at CR = 85.35% can be observed in Table 6.

As expected, Figure 6 shows that increasing the CR results in decreased overall accuracy for all classifiers.

Finally, Figure 7 shows a slightly different behavior for all classifiers. The classification accuracy of 90% starts to decay

after CR = 82.91%. The effect of the AWGN is much less when SNR = 10 dB, which is close to the EEG data. To the best of our knowledge, no reported work has been found that employ similar evaluation approach of EEG-based epileptic seizure in which AWGN considers different SNR values. Moreover, new interesting results could be realized that the thermal noise using AWGN clearly affects the classification accuracy.

Overall and regardless of the compression ratio value, Figure 8 shows the results for the average classification accuracy; the NSC accuracy is constantly better than the

Table 7: Comparisons between previous works and the proposed method.

Authors	Noiseless data	Imperfect data	Classifiers	Accuracy
Sharma et al., 2014 [19]	Two different classes	**N/A**	LS-SVM	85%
Sadati et al., 2006 [15]	A, D, and E	**N/A**	SVM, FBNN, ANFIS, and proposed ANFN	85.9%
Mohamed et al., 2013 [13]	A, B, C, D, and E	**N/A**	NB, MLP, k-NN, LDA, and SVM	89.5%
Liang et al., 2010 [44]	A, D, and E	**N/A**	PCA and GA	80%–90%
Tzallas et al., 2009 [11]	A, B, C, D, and E	**N/A**	ANN	89%
Proposed method	**A, C, and E**	**A, C, and E**	**Ensemble NSC**	**90%**

accuracy of any individual classifier. This statement is valid for both noiseless and noisy EEG-epileptic seizure data.

Compared with previous works, the proposed NSC classification accuracy of noiseless EEG data has achieved 90%, which is 5% higher than the accuracy done in Sharma [19], 4.1% higher than the work done in Sadati et al. [15] (85.9% accuracy) especially for sets A, D, and E, and 0.5% higher than that reported in Mohamed et al. [13] (89.5% accuracy). In addition, Liang et al. achieved classification accuracy between 80% and 90% [44]. Tzallas et al. [11] achieved 89% only for noiseless dataset using one classifier. All of those approaches considered the same EEG dataset. In contrast to these methods, the proposed method achieved the desired and improved classification accuracy with noisy data using different SNR values: 80% for SNR = 1 dB, 84% for SNR = 5 dB, and 88% for SNR = 10 dB. These results were obtained at a CR of 85.35%. Moreover, the proposed method provides several significant benefits such as simplicity and the improvement of the overall classification accuracy. Table 7 shows the comparisons between the proposed NSC and others reported in the literature.

5. Conclusion

In this paper, an EEG noise-aware signal combination method for EEG-based epileptic seizure detection applications is proposed and investigated. Compression paradigms with low complexity are achieved by utilizing the iDCT method for data reconstruction. Features are extracted from the reconstructed data using DWT. The proposed noise-aware signal combination (NSC) method together with the classifiers ANN, naïve Bayes, k-NN, and SVM is tested with different categories of EEG-based epileptic seizure data. Noise is introduced to the data at different levels: SNRs of 1, 5, and 10 dB. The proposed NSC combination method constantly performs better than any of the above four classifiers. The experimental results show that the proposed NSC technique is effective with noisy data of 80% for SNR = 1 dB, 84% for SNR = 5 dB, and 88% for SNR = 10 dB while being effective with 90% accuracy for noiseless data. These results were obtained at CR = 85.35%.

Conflict of Interests

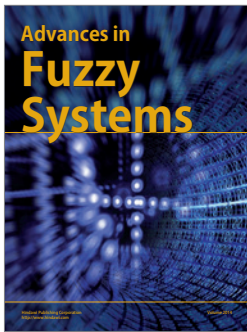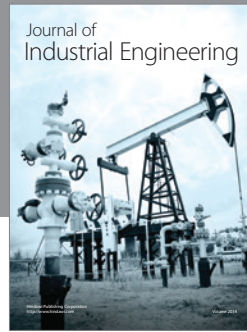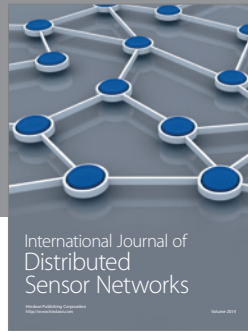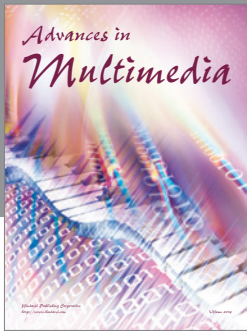The authors declare that they have no conflict of interests regarding the publication of this paper.

References

[1] C. P. Shen, C. M. Chan, F. S. Lin et al., "Epileptic seizure detection for multichannel EEG signals with support vector machines," in *Proceedings of the 11th IEEE International Conference on Bioinformatics and Bioengineering (BIBE' 11)*, pp. 39–43, October 2011.

[2] I. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 113–121, 2005.

[3] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

[4] L. D. Iasemidis, D.-S. Shiau, W. Chaovalitwongse et al., "Adaptive epileptic seizure prediction system," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 616–627, 2003.

[5] J. Wang, P. Liu, M. F. H. She, S. Nahavandi, and A. Kouzani, "Bag-of-words representation for biomedical time series classification," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 634–644, 2013.

[6] D. Rivero, E. Fernandez-Blanco, J. Dorado, and A. Pazos, "A new signal classification technique by means of Genetic Algorithms and kNN," in *Proceedings of the IEEE Congress of Evolutionary Computation (CEC '11)*, pp. 581–586, New Orleans, La, USA, June 2011.

[7] S. Sun, C. Zhang, and D. Zhang, "An experimental evaluation of ensemble methods for EEG signal classification," *Pattern Recognition Letters*, vol. 28, no. 15, pp. 2157–2163, 2007.

[8] S. Dehuri, A. Kumar Jagadev, and S. Cho, "Epileptic seizure identification from electroencephalography signal using DE-RBFNs ensemble," *Journal of Procedia Computer Science*, vol. 23, pp. 84–95, 2013.

[9] H. He and Y. Cao, "SSC: a classifier combination method based on signal strength," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1100–1117, 2012.

[10] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Computational Intelligence and Neuroscience*, vol. 2007, Article ID 80510, 13 pages, 2007.

[11] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in electroencephalograms using time-frequency analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.

[12] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for EEG-based brain computer interfaces," *Neural Networks*, vol. 22, no. 9, pp. 1313–1319, 2009.

[13] A. Mohamed and K. B. Shaban, "Evidence theory-based approach for epileptic seizure detection using EEG signals," in *Proceedings of the 12th IEEE International Conference on Data Mining Workshops (ICDMW '13)*, pp. 79–85, Brussels, Belgium, 2013.

[14] V. P. Nigam and D. Graupe, "A neural-network-based detection of epilepsy," *Journal of Neurological Research*, vol. 26, no. 1, pp. 55–60, 2004.

[15] N. Sadati, H. R. Mohseni, and A. Maghsoudi, "Epileptic seizure detection using neural fuzzy networks," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 596–600, Vancouver, Canada, July 2006.

[16] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 7, no. 4, pp. 413–419, 1999.

[17] D. Wang, D. Miao, and C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection," *Expert Systems with Applications*, vol. 38, no. 11, pp. 14314–14320, 2011.

[18] V. Bajaj and R. B. Pachori, "Classification of seizure and nonseizure EEG signals using empirical mode decomposition," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1135–1142, 2012.

[19] R. Sharma, R. B. Pachori, and S. Gautam, "Empirical mode decomposition based classification of focal and non-focal EEG signals," in *Proceedings of the International Conference on Medical Biometrics (ICMB '14)*, pp. 135–140, Shenzhen, China, May-June 2014.

[20] R. B. Pachori and S. Patidar, "Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 494–502, 2014.

[21] R. B. Pachori and V. Bajaj, "Analysis of normal and epileptic seizure EEG signals using empirical mode decomposition," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. 373–381, 2011.

[22] R. Pachori, "Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition," *Research Letters in Signal Processing*, vol. 2008, Article ID 293056, 5 pages, 2008.

[23] V. Joshi, R. B. Pachori, and A. Vijesh, "Classification of ictal and seizure-free EEG signals using fractional linear prediction," *Biomedical Signal Processing and Control*, vol. 9, no. 1, pp. 1–5, 2014.

[24] K. Abualsaud, M. Mahmuddin, R. Hussein, and A. Mohamed, "Performance evaluation for compression-accuracy trade-off using compressive sensing for EEG-based epileptic seizure detection in wireless tele-monitoring," in *Proceedings of the 9th International Wireless Communications and Mobile Computing Conference (IWCMC '13)*, pp. 231–236, Sardinia, Italy, July 2013.

[25] M. Valderrama, C. Alvarado, S. Nikolopoulos et al., "Identifying an increased risk of epileptic seizures using a multi-feature EEG-ECG classification," *Biomedical Signal Processing and Control*, vol. 7, no. 3, pp. 237–244, 2012.

[26] W. Weng and K. Khorasani, "An adaptive structure neural networks with application to EEG automatic seizure detection," *Neural Networks*, vol. 9, no. 7, pp. 1223–1240, 1996.

[27] A. Awad, A. Mohamed, A. A. El-Sherif, and O. A. Nasr, "Interference-aware energy-efficient cross-layer design for healthcare monitoring applications," *Computer Networks*, vol. 74, pp. 64–77, 2014.

[28] R. Hussein, A. Mohamed, and M. Alghoniemy, "Adaptive compression and optimization for real-time energy-efficient wireless EEG monitoring systems," in *Proceedings of the 6th Biomedical Engineering International Conference (BMEiCON '13)*, 2013.

[29] S. Fauvel and R. K. Ward, "An energy efficient compressed sensing framework for the compression of electroencephalogram signals," *Sensors*, vol. 14, no. 1, pp. 1474–1496, 2014.

[30] S. Mallat, *A Wavelet Tour of Signal Processing. Textbook*, Wavelet Analysis & Its Applications, Academic Press, 1999.

[31] A. Abdulghani, A. Casson, and E. Rodriguez-Villegas, "Quantifying the feasibility of compressive sensing in portable electroencephalography systems," in *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience: Proceedings of the 5th International Conference, FAC 2009 Held as Part of HCI International 2009 San Diego, CA, USA, July 19–24, 2009*, vol. 5638 of *Lecture Notes in Computer Science*, pp. 319–328, Springer, Berlin, Germany, 2009.

[32] J. Proakis and D. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*, Pearson Prentice Hall, 4th edition, 2007.

[33] P. J. Durka and K. J. Blinowska, "A unified time-frequency parametrization of EEGs," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 5, pp. 47–53, 2001.

[34] S. Aviyente, E. M. Bernat, S. M. Malone, and W. G. Iacono, "Analysis of event related potentials using PCA and matching pursuit on the time-frequency plane," in *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMS '06)*, vol. 3, pp. 2454–2457, New York, NY, USA, September 2006.

[35] 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," *Physical Review E*, vol. 64, no. 6, 2001.

[36] S. Pravin Kumar, N. Sriraam, P. G. Benakop, and B. C. Jinaga, "Entropies based detection of epileptic seizures with artificial neural network classifiers," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3284–3291, 2010.

[37] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2456–2466, 2011.

[38] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proceedings of the IEEE/SP 14th WorkShoP on Statistical Signal Processing (SSP '07)*, pp. 181–184, August 2007.

[39] S. Şenay, L. F. Chaparro, M. Sun, and R. J. Sclabassi, "Compressive sensing and random filtering of EEG signals using Slepian basis," in *Proceedings of the 16th European Signal Processing Conference (EUSIPCO '08)*, Lausanne, Switzerland, August 2008.

[40] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal of NeuroEngineering and Rehabilitation*, vol. 7, article 21, 2010.

[41] S. Sun and C. Zhang, "Adaptive feature extraction for EEG signal classification," *Medical and Biological Engineering and Computing*, vol. 44, no. 10, pp. 931–935, 2006.

[42] P. J. Durka, H. Klekowicz, K. J. Blinowska, W. Szelenberger, and S. Niemcewicz, "A simple system for detection of EEG artifacts in polysomnographic recordings," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 4, pp. 526–528, 2003.

[43] K. A. Al-Saud, M. Mahmuddin, and A. Mohamed, "Wireless body area sensor networks signal processing and communication framework: survey on sensing, communication technologies, delivery and feedback," *Journal of Computer Science*, vol. 8, no. 1, pp. 121–132, 2012.

[44] S.-F. Liang, H.-C. Wang, and W.-L. Chang, "Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection," *Eurasip Journal on Advances in Signal Processing*, vol. 2010, Article ID 853434, 15 pages, 2010.

[45] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, New York, NY, USA, 2nd edition, 2006.

[46] Y. Yang and X. Liu, "A re-examination of text categorization methods," in *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval*, pp. 42–49, Berkeley, Calif, USA, 1999.

[47] A. Subasi and E. Erçelebi, "Classification of EEG signals using neural network and logistic regression," *Computer Methods and Programs in Biomedicine*, vol. 78, no. 2, pp. 87–99, 2005.

[48] D. Rivero, V. Aguiar-Pulido, E. Fernández-Blanco, and M. Gestal, "Using genetic algorithms for automatic recurrent ANN development: an application to EEG signal classification," *International Journal of Data Mining, Modelling and Management*, vol. 5, no. 2, pp. 182–191, 2013.

[49] D. Rivero, L. Guo, J. Seoane, and J. Dorado, "Using genetic algorithms and k-nearest neighbour for automatic frequency band selection for signal classification," *IET Signal Processing*, vol. 6, no. 3, pp. 186–194, 2012.

[50] A. Ahangi, M. Karamnejad, N. Mohammadi, R. Ebrahimpour, and N. Bagheri, "Multiple classifier system for EEG signal classification with application to BCI," *Journal of Neural Computing and Applications*, vol. 23, no. 5, pp. 1319–1327, 2013.