Peer-reviewed version available at *Sensors* **2019**, *19*, 219; doi:10.3390/s19020219

Article

Automated recognition of epileptic EEG states using a combination of Symlet wavelet processing, a gradient boosting machine, and a grid search optimizer

Xiashuang Wang [1, 2], Guanghong Gong [2] and Ni Li [1, 2*]

[1] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China; wxs_sky@outlook.com
[2] Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China; ggh@buaa.edu.cn
* Correspondence: lini@buaa.edu.cn; Tel.: +86-135-8173-7778

Received: date; Accepted: date; Published: date

Abstract: Automatic recognition methods for non-stationary EEG data collected from EEG sensors play an essential role in neurological detection. The integrative approaches proposed in this study consists of Symlet wavelet processing, a gradient boosting machine, and a grid search optimizer for a three-level classification scheme for normal subjects, intermittent epilepsy, and continuous epilepsy. Fourth-order Symlet wavelets were adopted to decompose the EEG data into five time-frequency sub-bands, whose statistical features were computed and used as classification features. The grid search optimizer was used to automatically find the optimal parameters for training the classifier. The classification accuracy of the gradient boosting machine was compared with that of a support vector machine and a random forest classifier constructed according to previous descriptions. Multiple-index were used to evaluate the Symlet wavelet transform-gradient boosting machine-grid search optimizer classification scheme, which provided better classification accuracy and detection effectiveness than has recently reported in other work on three-level classification of EEG data.

Keywords: recognition of epilepsy EEG; Symlet wavelet; gradient boosting machine; grid search optimizer; multiple-index evaluation.

1. Introduction

Epilepsy is one of the most common neurological disorders, with one person in every hundred worldwide suffering from epilepsy [1]. Epileptic episodes are a clinical manifestation of paroxysmal abnormal ultra-synchronized electrical activity in the brain, which is iterative, sudden, and temporary. Automated detection of an epileptic episode and subsequent alerting can aid neurologists monitoring treatment in busy neurological wards, and could help to ensure patient safety [2]. However, the time frequencies of epileptic episodes are uncertain, and their clinical manifestations are not easy to detect. In the early stages of testing and monitoring patients for epilepsy, researchers attempted to use sensors [3] to collect biological data from the patient's surface, including electrocardiogram (ECG), electromyography (EMG) [4-5], motion data [6], and electrodermography (EDG). These data can be collected by wearable systems, including E-textiles [7], capacitive sensing [8], polymer materials such as carbon nanotube (CNT)-polydimethylsiloxane (PDMS) [9], Ag/AgCl electrodes [10], and micro-needle arrays [11]. Wearable sensor systems can non-invasively monitor biological signals from epileptic patients for long periods. However, such biological data has the drawback of insufficient spatial resolution.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at Sensors 2019, 19, 219; doi:10.3390/s19020219

Therefore, approaches to directly obtain epilepsy information from the brain have been researched, including positron emission tomography (PET), single photon emission computed tomography (SPECT), magnetic resonance imaging (MRI), and functional magnetic resonance imaging (fMRI) [12]. Currently, most research focuses on the use of video-electroencephalograms (EEGs)[13-14]. EEGs not only display temporal information, but also provide spatial information on electrical activity in the brain. The video-electroencephalogram technique has been considered a gold standard tool for the study of epilepsy. The real time information on an epileptic episode reflected by EEG cannot currently be replaced by any other physiological brain function monitoring method. As the physiological processes of seizure are typically non-stationary, dynamic, and nonlinear, the differentiation of rhythmic discharges from nonstationary processes brings great challenges to the analysis of EEG signals.

In general, the automated detection of EEG signals includes the two core methods of feature extraction and classification. The features extracted can be divided into four categories: statistical features, fractal dimension features, entropy features, and time-frequency domain features. Several studies have used combined time and frequency features for the automatic recognition of non-stationary EEG at the onset of epilepsy. Gotman was a pioneer in the exploration of automatic seizure detection technology based on EEG, and to capture transient behavior during long-term EEG monitoring he decomposed the EEG into half waves, which recorded the typical peak value as a morphological feature [15]. Discrete wavelet transforms (DWT) have the ability to capture the time-frequency information in epileptic EEG [16-18], and many researchers have used Daubechies wavelets to analyze epileptic EEG, as they considered the db4 wavelet to be similar to the spike wave of EEG. A classic example is the Welch spectral analysis method introduced into the feature analysis of epileptic seizure detection. Tzallas et al. used different time-frequency domain methods to extract the power spectrum density of computerized EEG during epileptic seizures [19]. Independent components analysis (ICA) [20] and linear discriminant analysis have also been reported for EEG signal extraction, while a multiscale radial basis function algorithm recently showed promising results in the decoding of EEG of epileptic seizures [21]. After consideration of the above-mentioned literature, Polat et al. proposed a hybrid model for seizure detection using a fast Fourier transform (FFT) for feature extraction [22]. However, the FFT has several disadvantages: first, the FFT cannot do a good job of solving the non-stationary EEG problem using a fixed window function; second, it is very time-consuming. Therefore, a short-time Fourier transform (STFT) was used to extract the frequency information from the raw EEG recording. The original signal was truncated into smaller sections and windowed, and the discrete Fourier transform was applied to the signals. The STFT algorithm performed time-frequency analysis of non-stationary EEG signals by adjusting different time windows to avoid the disadvantages of the FFT [23-24]. Boashash et al. extracted statistical and image features according to their time-frequency distribution to handle multichannel EEG from neonates [25]. A sensitivity of the criterion is that it is taken into consideration in the feature selection, resulting in a reduction in computational cost and improvement in detection performance [26]. Flexible wavelet transforms and the fractal dimension of the time-frequency method have also been used for the detection of seizure segments in long-term EEG [27-29]. From the above literature, we believe that the wavelet transform is the most commonly used method for extracting EEG features, although this extraction method ignores the overall statistical information. Therefore, we aimed to find an EEG analysis method combining time-frequency information and statistical information.

For the automatic detection of EEG by machine learning, most studies adopt a supervised learning paradigm. Regardless of the categories of the input EEGs, the EEGs used to train classifiers are labeled according to prior knowledge. He et al.'s neural network (NN) classification technique used machine learning applied to the field of brain science [30]. Boser et al. [31], Kai Fu et al. [32], and Ying Gue et al. [14] all used support vector machines (SVM) to identify the EEG signals of epilepsy patients, and obtained a relatively good recognition performance. Brabanter et al. proposed a least squares support vector machine (LS-SVM) for the classification of two-levels of seizure and non-seizure

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at Sensors 2019, 19, 219; doi:10.3390/s19020219

EEG signals from the small seizure dataset of Bonn University. They obtained 98.0%–99.5% accuracy using a radial basis function (RBF) kernel, and 99.5%–100% accuracy using a Morlet kernel [33]. Sun et al. used an Ada-Boost classifier to achieve good accuracy for spike detection of epilepsy seizures [34]. However, the choice of a suitable strategy for machine learning is a difficult one, numerous classification strategies have been developed for seizure detection, including random forests (RF), K-nearest neighbors (KNN) [35], principle component analysis (PCA), Bayesian neural networks [36], and empirical mode decomposition (EMD) [37]. The classification results indicate that these pattern recognition systems can achieve high levels of classification accuracy, from 93% to 99.66%. Nevertheless, these accuracy scores were the results of two-level EEG classification, and the above mentioned schemes were too inconvenient and time consuming for practical clinical applications. Recently, Wang et al. explored a three-level classification problem, analyzing continuous ictal epilepsy patients, intermittent epilepsy patients, and healthy subjects. Using an SVM recognition system, they achieved an accuracy of 93.9% for the Bonn datasets [38]. A more effective classification scheme needs to be developed to solve the multi-level classification problem presented in this work.

After completing the feature extraction and classification procedures, it is also essential to perform a reasonable assessment to verify their accuracy. In the use of machine learning for the assessment of EEG, only pursuit the high classification accuracy of recognition system cannot satisfy comprehensive assessing of the classification effect of the classifiers. However, some other verification indicators revealing the causes of error in classification are also important in epilepsy detection. Recognition systems achieving high levels in these verification indicators could help fill gaps in the analysis of seizure monitoring devices, and could reduce the rate of missed detections in clinical situations [39]. In 2014, the Mayo Clinic and the University of Pennsylvania hosted a competition to find robust seizure detection and prediction systems. Participants used SVM and random forest (RF) machine learning techniques to recognize canine and human cortical electroencephalogram (ECoGs) datasets, and achieved high sensitivity and low false-positive rates [40-41].

Following-on from the above-mentioned literature and analysis, this paper adopts a wavelet transform to analyze the time-frequency information of epilepsy and avoid the shortcomings of Fourier, STFT, and Welch spectral analysis. The Symlet wavelet is used to decompose the EEG signals into $\gamma,\beta,\alpha,\theta,\delta$ sub-bands. Then, the statistical information of the five sub-bands is extracted to generate the features for feeding into the feature recognition sensor module. Most previous studies have applied recognition algorithms to the Bonn epilepsy dataset, which is classified into seizure EEG epochs and non-seizure EEG epochs for two-level classification. Such two-level classification schemes are not ideal for practical applications, because in reality there are multiple degrees of epileptic seizure. To obtain an efficient three-level classification scheme, we propose a gradient boosting machine-grid search optimization (GBM-GSO) to classify the Bonn EEG dataset into three categories representing normal subjects, intermittent epilepsy, and continuous epilepsy. We also implement two other state-of-the-art machine learning classifiers, an SVM and an RF, and compared them with the GBM classifiers. This comparison demonstrated that the GBM classifier was the most effective for identifying epileptic state EEG. This recognition scheme not only ran faster than the SVM and RF, but also effectively avoided the misdiagnoses or missed diagnoses caused by manual tuning of parameters. Our auxiliary medical diagnostic system can directly recognize three classifications from epilepsy EEG signals: continuous ictal epilepsy patients, intermittent epilepsy patients, and healthy subjects.

We believe that our machine learning approach has the following innovations and advantages for training on and classification of epilepsy EEG.

a) It not only allows visualization of the core time-frequency information of EEG through wavelet transforms, but also extracts the statistical information by key statistical techniques. The statistical information of time-frequency features are as the latter recognition features, and these features reflect the overall characteristics of the data. Simultaneously, a principle component analysis algorithm used to reduce the dimensionality of the data. Thus, the new method reduces the hardware calculation under the premise of ensuring the accuracy of the classifier.

b) The proposed GBM recognition system is highly parallelized to improve operational efficiency. Another advantage is that it can process large-scale data. However, the recognition system generates many parameters in the course of the training process, and it can be difficult to determine the optimal parameters by manual tuning. This paper proposes a grid search optimizer to optimize these parameters and determine the best recognition system filtering parameters repeatedly by variable step size way. To prevent over-fitting in the GBM training process, we adopt a 10-fold cross-validation (CV) strategy, which ensures that the optimized system is more robust.

c) The SW-GBM-GSO integrative techniques can perform three classifications: healthy subjects, intermittent epilepsy patients, and continuous ictal epilepsy patients. We use multiple indicators to evaluate and verify the diagnostic system, with these being not only limited to classification accuracy, but also including other indicators such as accuracy (ACC), a confusion matrix (CM), a precision recall curve (PRC), the receiver operating characteristic curve (ROC), and the area under curve (AUC). Multiple indicators can make a more thorough and clearer analysis of the error rate resulting from misclassification. This strategy is pivotal in medical screening.

The automatic integrative epileptic seizure EEG technology described in this paper comprises five major modules, as illustrated in Fig. 1. In the first step, signals are collected from EEG sensors in the monitoring module. In the third step, time-frequency and statistical features are extracted by the feature extraction and selection module. PCA is applied to reduce the feature dimensionality, which is beneficial in respect to the computer operating time. In the fourth step, pre-classified testing data are fed into the feature classification module. Then, an optimization procedure is performed to search the hyper-parameters and optimize the recognition system. Finally, we use multiple indices to evaluate the scheme, and a verification module to detect the performance in the three classifications of seizure status.

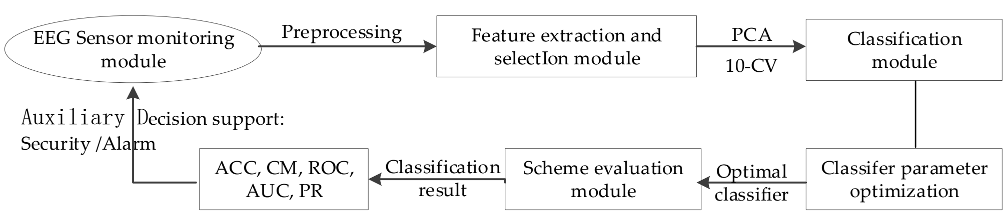


Figure 1. Auxiliary medical diagnostic system for Epilepsy EEG

The remainder of the paper is organized as follows. In Section 2, we apply the time-frequency and statistical methods to real EEG data after first preprocessing it. This study adopts the PCA method to reduce the dimensionality of the EEG features. In Section 3, we build the novel automatic GBM recognition system using 10-fold CV. In Section 4, we apply the automatic detection method to real EEG data to classify the three categories of seizure, light-seizure, and non-seizure, and verify the effectiveness of the machine learning system. The experimental results are analyzed using accuracy, CM, PRC, ROC, and AUC generated from the sensitivity and specificity. Finally, the contributions and future contemplated work are summarized in Section 5.

2. Feature Extraction and Selection Module

There are many approaches for extracting the features of EEG signals. First, we discuss the time-frequency feature methods for EEG. The EEG scalp signals at time t can be defined as a vector:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at Sensors 2019, 19, 219; doi:10.3390/s19020219

$$\mathrm{s(t)}=\left(s(\mathrm{t}_1),s(\mathrm{t}_2),\cdots,s(\mathrm{t}_n)\right)=\begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{pn} \end{pmatrix} \quad (1)$$

where s(t) represents the EEG signals. Fourier transforms, multitaper spectral analysis, and STFTs can all be used to describe the time-frequency features of the signals. To a certain degree, the most commonly used Fourier transform reflects the frequency characteristics of the entire signal. Using FFTs, it is possible to smooth and slow signals that change over time. However, EEG signals are nonlinear and non-stationary, and their frequency changes rapidly with time; fast changing frequencies are effectively "averaged" by the FFT, and it can only give the overall effect of the signals, it cannot reflect the frequency variation characteristics of the signals themselves. The STFT will move a fixed length window function over the signal during signal processing. Under the assumption that the windowed signals represent stationary signals in different finite time widths, the power spectrum at different moments can be calculated. The STFT considers non-stationary EEG signals as stationary signals and superimposes a series of short signals.

Discrete wavelet analysis has seen rapid development in recent years [17, 38, 42]. It is an analysis method that combines both the time and frequency domains. The wavelet transform decomposes the signal into different frequency bands, and studies the nature and characteristics of the signal according to the "wavelet family" $\zeta_{a,b}(t)$ in these different frequency bands. The wavelet transform generally performs better than the FFT and STFT without prior knowledge. It can be used to analyze different frequency components of burst and non-stationary signals using variable windows.

$$\zeta_{a,b}(t)=\frac{1}{\sqrt{|a|}}\zeta\left(\frac{t-b}{a}\right) \quad (2)$$

where $a \neq 0, a, b \in R, a, b$ are different ratios and conversion parameters and t is time. The continuous wavelet transform for EEG s(t) is defined as the correlation between a and b:

$$\mathrm{WT}_s(\mathrm{a,b})=\frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty}\mathrm{s}(t)\zeta^*\left(\frac{t-b}{a}\right)\mathrm{d}t \quad (3)$$

Following work based on the above analysis applied to practice signals, we selected the DWT to construct a wavelet basis function. The wavelet basis function and its decomposition impact on feature extraction effort of signal. There are a variety of continuous wavelet bases, including Daubechies, Symlets, Haar, Morlet, Mexican Hat, and Meyer types. Each wavelet transform is suitable for different task applications. We compared the characteristics of different wavelets according to their weight orthogonality, tight support, support length, and symmetry, as shown in Table 1. In this work, we considered the Symlet wavelet (SW) to be appropriate for the nonlinear EEG analysis. The Symlet wavelet is an improvement on the Daubechies wavelet, making up for the shortcoming of approximate asymmetry present with the Daubechies wavelet. Secondly, the support range and the vanishing moment of the Symlet wavelet are 2N-1 and N, respectively. The Symlet wavelet basis has better regularity than Daubechies, and this can reduce the phase distortion in the analysis and reconstruction of signals.

Table 1. Comparison of the properties of wavelet transforms

Wavelet	Daubechies	Symlets	Haar	Morlet	Mexican Hat	Meyer
Orthogonality	Yes	Yes	Yes	No	No	Yes
Tight support	Yes	Yes	Yes	No	No	No
Support length	2N-1	2N-1	1	Finite	Finite	Finite
Symmetry	No	Yes	Yes	No	Yes	Yes

As the energy loss of the epileptic EEG signal in the first two scales of the Symlet wavelet transform is less than 30%, we used the fourth-order Symlet wavelet to transform the scalp EEG signal on the third scale to obtain the EEG spectrum. After applying a band-pass filter to preprocess the EEG to the bands between 0 and 50 Hz, we divided the frequency information into five bands of

EEG component $\gamma(25\sim 50\,\mathrm{Hz}),\beta(12\sim 25\,\mathrm{Hz}),\alpha(6\sim 12\,\mathrm{Hz}),\theta(3\sim 6\,\mathrm{Hz}),\delta(0\sim 3\,\mathrm{Hz})$. The approximation coefficient a_4 and detail coefficients d_1, d_2, d_3 and d_4 are generated in the decomposition structure process, as shown in Fig. 2.

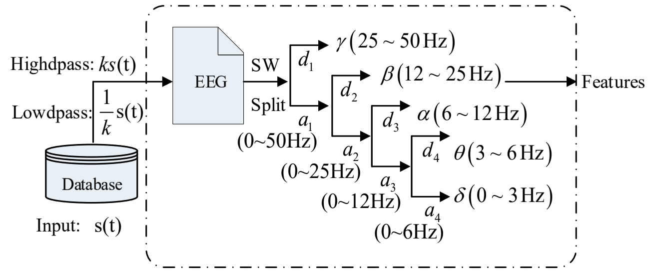


Figure 2. Wavelet transform decomposition process for EEG

The mean and standard deviations show the density of the value center and possible EEG signal values, which are defined as the formulas. In practice, it is necessary to extract statistical information from the time-frequency features of EEG signals, which can be understood as continuous random variables [43].

$$\omega = \int_{-\infty}^{+\infty} s(t)\, P(s(t))\, ds(t) \tag{4}$$

$$\delta^2 = \int_{-\infty}^{+\infty} (s(t)-\omega)^2\, P(s(t))\, d\,s(t) \tag{5}$$

It is also necessary to use PCA in the process of feature extraction to achieve low dimension features after using the Symlet wavelet. PCA ensures the information is as relevant as possible, with a new feature subspace being constructed from the information derived from the existing features. This procedure reduces the load on the recognition system and increases computational efficiency.

3. Classification Module

3.1. Analysis for the machine learning classification

There are many machine learning pattern classifiers that could be used to classify EEG data, and it is difficult to choose the most suitable one for the analysis of multilevel epilepsy EEG data. In the following section we discuss the most widely used SVM and RF classifiers and the gradient boosting machine classifier proposed in this paper.

3.1.1. Support vector machine

Among the pattern recognition systems available, SVMs are the most popular machine learning algorithms for classifying EEG data, because they offer a good classification performance and excellent generalization ability [44-48]. The essence of the SVM is to search for a separating hyperplane $\omega^T + b$ maximizing the boundary distance between two types of data feature vectors. SVMs are not very suitable for three or multi-classification problems, which is why they have been mainly used to separate EEG data into two types, rather than three or more categories. The pseudo code for the SVM is shown in Table 2.

Table 2. Support vector machine classifier in pseudo-code

ALGORITHM 1: Support vector machine (SVM)
Data: n observed data features {T-F features, statistical features $\overline{s(t_i)}$} classifier: SVM.
1. The unit step function $f_{w,b}$ acts on the classified super-plane $g(\omega^T + b)$, where $g(\omega^T + b) = 1$ if $\omega^T + b \geq 0$ and -1 otherwise.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 219; doi:10.3390/s19020219

2. Calculate the distance l of the EEG data to the separated hyperplane and find ω,b st. $\min(l)$.
3. Then, maximize the minimum distance.

$\arg\max_{\omega,b}\{\min_{n}(\text{label}(\omega^{T}+b))\cdot\|\omega\|^{-1}\}$,st. $\text{label}(\omega^{T}+b)=1$, $\max(\|\omega\|^{-1})$

Result: label, minimum distance l

In the above optimization problem, we find the optimal value given the constraints $\text{label}(\omega^{T}+b)\geq 0$. Lagrange multipliers can be used for this type of optimization problem, and the optimization objective function can be written as:

$$\max\left[\sum_{i=1}^{m}\alpha-\frac{1}{2}\sum_{i,j=1}^{m}label^{(i)}\cdot label^{(j)}\cdot\alpha_{i}\cdot\alpha_{j}\left\langle x^{(\mathrm{i})},x^{(\mathrm{j})}\right\rangle\right] \tag{6}$$

The constraint is $\alpha>0$ and $\sum_{i=1}^{m}\alpha_{i}label^{(i)}=0$. A Gaussian kernel achieved consistently better performance than a linear kernel.

3.1.2. Random Forests

The RF [49] is an effective integrated machine learning classifier combining many decision trees, and is an extended variant of Bagging. First, a bootstrap sample Z^{*} was randomly selected from the training set in a returning way. Taking the randomly selected data in the above steps as the training data, T_{b} decision trees were established. Second, a subset of M features is randomly selected from the feature set of each node of the decision tree. The RF tree is grown to enhance the binding EEG data by recursively repeating the above steps for each terminal node of the decision tree until the decision tree can accurately identify the training data set while achieving the minimum node size. For the process of training the recognition system, this study used a classification and regression tree algorithm to split the nodes, and the Gini value of the Gini index was used as the basis of the splitting node. The sample training set Z^{*} contains different characteristics, and the Gini index of this training set is:

$$GINI(\mathrm{k})=1-\sum_{i=1}^{k}p_{i} \tag{7}$$

where p_{i} is the probability of a category $\mathbf{i}$ feature. The number of features corresponding to the sample training set were $\{n_{1},\mathrm{n}_{1},\mathrm{L}\ ,\mathrm{n}_{k}\}$, the split Gini index is:

$$GINI(M^{*})=\frac{n_{1}}{n}GINI(M_{1})+\frac{n_{2}}{n}GINI(M_{2})+\frac{n_{3}}{n}GINI(M_{3}) \tag{8}$$

All the decision trees $\{T_{b}\}_{1}^{m}$ are aggregated. For an input sample, the decision trees of $\mathbf{m}$ have the recognition results of $\mathbf{m}$, and the RF recognition system inherits all the recognition voting results. Forecasting is performed on the new node, and the most recognized number of votes is the output $\{\hat{C}_{b}(x)\}_{1}^{m}$. The pseudo code for the RF-GSO is shown in Table 3.

Table 3 Random forest classifier in pseudo-code

ALGORITHM 2: Random Forest (RF)

1. **For** $i=1$ to m :
   (a) Draw a bootstrap sample Z^{*} of size P from the training data.
   (b) Grow a random forest tree T_{b} to the boost strapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size $n_{\min}$ is reached.
2. **Output** ensemble of tree $\{T_{b}\}_{1}^{m}$
   To make a prediction at a new point x

Classification: Let $\hat{C}_b(x)$ be the class prediction of the random forest tree.

Then $\hat{C}_{rf}^{m}(x) =$ majority vote $\{\hat{C}_b(x)\}_1^m$

3.1.3. Gradient Boosting machine

The GBM is a method for the gradual enhancement or improvement of error, it was designed by Jerome H. Friedman of Stanford University [50], who considered estimation of the functional dependence $y=\eta(\mathrm{s(t)})$. Then, the loss function $\psi(y,\eta)$ is minimized:

$$\hat{\eta}(\mathrm{s(t)}) = \hat{y} = \arg\min \psi(y,\eta) \tag{9}$$

The function estimate $\hat{y} = \sum_{i=1}^{M} \hat{y}_i$ is parametrized, with $\hat{y}_i$ defined as a boost. We can draw up a greedy strategy that estimates $\hat{y}_k = \hat{y}_{k-1} + \mathrm{V}_k \cdot \xi\left(\overline{s(\mathrm{t})}, \theta_k\right)$ at each recursive where $\xi\left(\overline{s(\mathrm{t}_i)}, \theta\right)$ is called a base learner, that is, a decision tree. The function is built as:

$$(\mathrm{V}_k, \theta_k) = \arg\min_{\mathrm{V},\theta} \sum_{i=1}^{N} \psi\left(y^{(\mathrm{i})}, \hat{\eta}_{k-1}\right) + \mathrm{V}\cdot\xi\left(\overline{s(\mathrm{t}_i)}, \theta\right) \tag{10}$$

While this optimization problem is hard for a general loss function and base learners, Friedman suggested a new function $\xi\left(\overline{s(\mathrm{t})}, \theta\right)$ to be the most parallel to the negative gradient along the observed data, whereby the optimization task becomes a classic least-square minimization. Table. 4 describes the pseudo code of GBM.

Table 4. Gradient boosting machine classifier in pseudo-code

ALGORITHM 3: Gradient Boosting Machine (GBM)

Data: n observed data features {T-F features, statistical features $\overline{s(\mathrm{t}_i)}$ }

Input: Calculate loss function $\psi(y,\eta)$ and base-learner classifier $\xi\left(\overline{s(\mathrm{t})}, \theta\right)$ to number of iterations M.

Process:

1. Build predicted classifier $\hat{\eta}(\mathrm{s(t)})$ for $\overline{s(\mathrm{t})}$.
2. Initialize $\hat{\eta}_0 = \arg\min_{\Delta_k} \sum_{i=1}^{N} \psi\left(\overline{s(\mathrm{t}_i)}, \Delta_k\right)$;

   for $m \in \{1, 2, \ldots, \mathrm{M}\}$
3. Compute the negative gradient $\zeta_k(\mathrm{s(t)})$;
4. Fit a new base-learner function $\xi\left(\overline{s(\mathrm{t})}, \theta_k\right)$;
5. Find the best gradient descent step-size V_k to get tree classifier:

   $$\mathrm{V}_k = \arg\min_{\mathrm{V}\theta} \sum_{i=1}^{N} \psi\left(y^{(\mathrm{i})}, \eta_{k-1} \hat{(\mathrm{s}(\mathrm{t}_i))}\right) + \mathrm{V}\cdot\xi\left(\overline{s(\mathrm{t}_i)}, \theta_k\right)$$
6. Update the function $\eta_k = \Delta_k \zeta_k(\mathrm{s(t)})$ gradient boosting machine classifier $\eta\left(\overline{s(\mathrm{t}_i)}\right) = \eta_k + \eta_{k-1}$

end

return $\eta\left(\overline{s(\mathrm{t}_i)}\right)$;

3.1.4. Comparative analysis of the methods

One of the purposes of this paper was to maintain robustness while solving the multi-level classification, thereby ensuring recognition accuracy. SVMs are widely used for classifying EEG, and RFs can achieve excellent performance in pattern recognition. We compare the proposed GBM with an SVM and RF, discussing them from three main aspects: the multi-level classification problem, the generalization ability, and the sensitivity of parameter selection. The conclusions obtained may provide valuable references for other researchers using pattern recognition systems.

- Multi-level classification problem

The SVM was initially used to solve problems with two classifications, by finding the optimal hyperplane that divides the data into two. The idea of maximizing the classification margin is the core of the SVM method. In the practical application of data mining, it is generally necessary to solve the classification problem of multiple classes. This can only be solved by constructing a combination of multiple two-level classifiers. Although a multi-level classification problem can be solved using this combination method, it is cumbersome and does not guarantee superior precision [38]. RF and GBM are decision tree models based on integration ideas, and they are better suited for solving multi-level classification problems.

- Sensitivity of parameter selection

The performance of a support vector machine depends mainly on the selection of the kernel function; therefore, a practical problem is how to choose the appropriate kernel function. At present, the more mature approach is to artificially choose the kernel function and its parameters on the basis of experience plus an element of randomness. Kernel functions should have different forms and parameters for different problem areas, and so domain knowledge should be introduced when making the selection. Currently, there is no good way to solve the problem of kernel function selection.

- Generalization ability

The main characteristic of the RF method is the selection of features using the principle of minimization of the Gini index. Because of the random selection of samples and features, it is not easy to fall into overfitting. In the bagging algorithm, the tree growing obtains an average predictive power across all decision trees using a parallel boosting method. Each tree is constructed on a sample of raw data and the results of the trees are voted on to achieve the final result, with the training results of different trees not being further optimized.

The essential difference between GBM and RF is that each tree in GBM learns the residuals of all previous tree conclusions. The residual is the true value minus the predicted value. GBM is superior to RF in that it is not based on decision trees built in parallel. The construction of a GBM classifier involves moving along the direction in which the gradient drops the fastest. The gradient generates a completely new decision tree at each iteration. To make up for the lack of an original recognition system, the partial derivative of the loss function at each training sample point is used to construct a weak learner. Therefore, the GBM classification system has stronger generalizing abilities and better adaptability to new data than RF and SVM techniques.

The above comparative analysis indicates that the GBM classifier is the most suitable for the three-level classification problem with the epilepsy EEG.

3.2. Parameter optimization and cross-validation

3.2.1. Parameter optimization

The gradient boosting machine identification algorithm generates decision tree and boosting parameters during the training process. Although the GBM classifier does not result in much over-fitting as the decision tree grows, the high learning rate still causes over-fitting of the classification model. If we reduce the learning rate and increase the decision tree blindly, the calculations can be very expensive and take a long time to run. This paper proposes an improved grid search algorithm to optimize and configure the parameters of the GBM model to improve the classification performance of the gradient boosting machine classifier.

The GSO algorithm resorts to meshing the variable regions and then traversing all the grid points, solving the objective function values to satisfy the constraints, and thereby selecting the

Peer-reviewed version available at *Sensors* **2019**, *19*, 219; doi:10.3390/s19020219

optimal values. It takes a lot of training time to traverse all the parameters on the grid, and in this paper, the GSO algorithm is improved to reduce the training time. The specific steps are as follows.

First, we use a long-distance step size for a rough search over a large range. Second, the mesh is built on the coordinate system, with its mesh nodes being the corresponding parameter pairs of decision trees and boosting. The optimal parameters and recognition accuracy are output when there is a set of parameters that meet the requirements; we selected the parameter with the smallest penalty parameter as a more selective object when multiple sets of parameters met the requirements. Next, a second accurate search is performed in small steps on the set of parameters: the above steps are repeated with the step set to 0.1 to find the global optimal hyper-parameters. A flowchart of this parameter optimization of the GBM model based on the improved grid search algorithm is shown in Fig. 3.

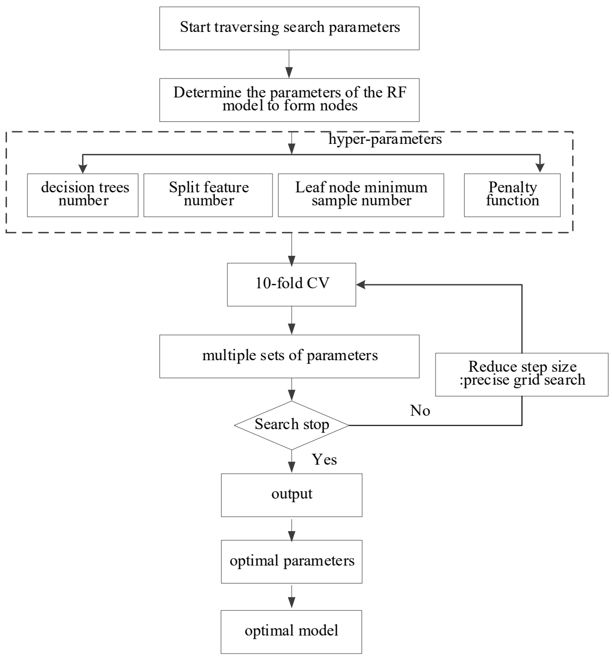


Figure 3. Parameters optimization flow in the GSO algorithm

Generally, the default value for the learning rate is 0.1; however, for different problems, values between 0.05 and 0.2 can determine the optimal number of decision trees at the current learning rate. In this paper, the optimal learning rates determined by the GSO algorithm is 0.06.

3.2.2. K-fold cross-validation

To reduce the influence of the selected training and testing data on the model verification, k-fold CV was used. This involves the training data being divided into subsets without repetition.

$$\{V_1, V_2, \ldots, V_k\}, (V_i \cap V_j = \varnothing) \tag{14}$$

k-1 subsets were used for training, with the remaining subset being used for testing. This process was repeated k times to obtain k accuracy values, which were then averaged to provide a mean value for the evaluation. The automatic seizure detection systems of Guo et al.[51], Nicolaou et al.[52], Samiee et al[53], and Yuanfa Wang et al[38]. did not use CV, while Qu et al. used the default 5-fold CV[32]. In this study, 10-fold CV was used to obtain more reliable and robust performance results. The training set was randomly divided into 10 subsets, with only one subset being used as the verification set. The other residual subsets were used to train the EEG classifier on data corresponding to different levels of epileptic seizure. The use of 10-fold CV reduces the over-fitting phenomenon and increases the credibility of the data classification. The pseudo code for the 10-fold CV is shown in Table 5.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 219; doi:10.3390/s19020219

Table 5 10-fold cross-validation in pseudo-code

ALGORITHM4: 10-Fold Cross-validation (s, GBM, L, 10)

Input: $s(t), s(t_i) \in s(t)$ Sample set

gradient boosting machine (GBM): Decision algorithm

L : Loss function 10: Fold number

Process:

1: Define: $V_1 \oplus V_2 \oplus \cdots \oplus V_{10} = s$

$$\Leftrightarrow \begin{array}{c} V_1 + V_2 + \cdots + V_{10} = s \\ V_i \cap (V_1 + \cdots + V_{i-1} + V_{i+1} + \cdots + V_{10}) = \{0\} \end{array}$$

2: **for** i from 1 to 10 do

3: $f_i = RF(\mathrm{s}/\mathrm{V}_i)$

4: **for** cc in V_i do

5: $e_j = L(f_i, \mathrm{s}(\mathrm{t}_i))$

6: **end for**

7: **end for**

8: **Return** e

4. Experimental results and discussion

The experiments were performed on an Acer PC with a 2.8 GHz Intel Core i5-6200U CPU, 8 GB of low voltage memory, 1 TB of storage, and a 64-bit operating system.

4.1 Real clinical EEG dataset

This paper used an open-source database available at the University of Bonn and extracted the key features to detect continuous ictal epilepsy patients, intermittent epilepsy patients, or healthy subjects from their EEGs. The datasets have been widely used to test methods proposed by many researchers, and can be considered as a benchmark for developing seizure detection schemes. The noninvasive EEG datasets were obtained from 25 subjects with medically intractable partial epilepsy.

The datasets were divided into five groups of ictal scalp EEG signals: {F, N, O, Z, and S}. Each group of data contained 100 samples from five subjects. The raw EEG data was recorded using a standard 10-20 system with a sampling frequency of 173.61 Hz. The age of the subjects ranged from 19 to 60 years, they were all right-handed, and the locations of the epileptogenic foci for each subject were identified by experienced epileptologists. The five EEG datasets {F, N, O, Z, S} were subjected to standard normalization procedures and were combined into three types {F/N}-{O/Z}-{S} according to the level of disease. More detailed information about the five EEG datasets {F, N, O, Z, S} are provided in Table 6.

Table 6. Dataset description

Data Sources	Parameter Description	Dataset category	Subject condition	Epileptogenic foci	Electrode collection area	Samples Number
	5 groups	{O/Z}	Health	Scalp surface	All brain areas	200
Bonn	173.6 Hz.	{F/N}	Intermittent	Intracranial	Lesionoutside	200

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 219; doi:10.3390/s19020219

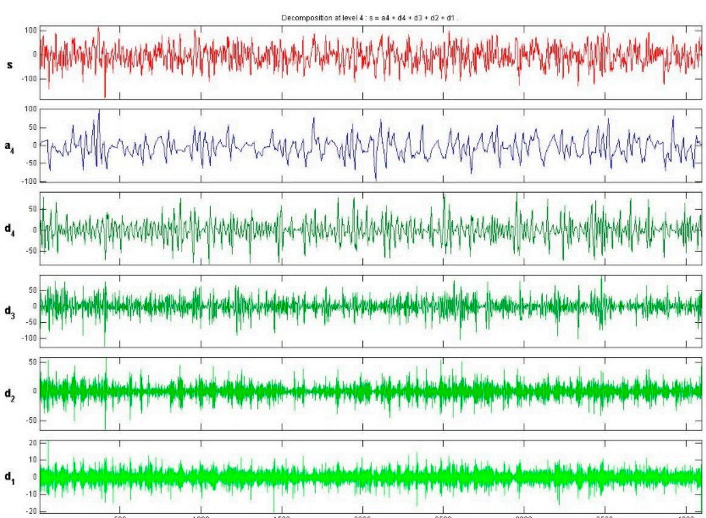
university	23.6 s.		epilepsy		inside area	
	4096data points.	{S}	Continuous Ictal epilepsy	Intracranial	Intra lesional area	100

A major goal of this paper is to classify the existing EEG signals {F/N}-{O/Z}-{S} into the three types. The datasets {O/Z} were from healthy subjects in an alert state and only used EEG signals acquired from the surface of the scalp. The datasets {F/N} were from epilepsy patients who did not suffer a seizure within the area covered by the intracranial EEG signals during the data acquisition period. The {S} dataset was from ill patients with epileptic episodes from lesions within the area of the intracranial EEG.

First, the raw EEG signals were preprocessed using the open source toolbox EEGlab running under Matlab. This involved several steps, including Butterworth filtering, removal of artifacts, baseline corrections, and cutting the data into segments [42].

4.2 Time frequency and statistical feature extraction

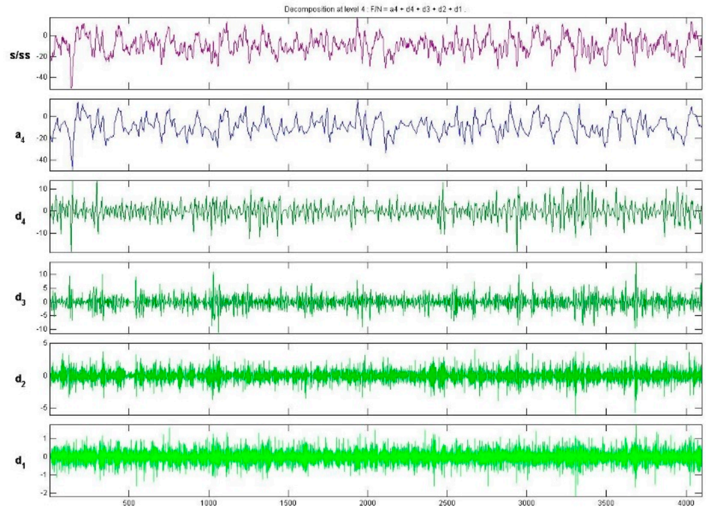
The joint time-frequency distribution is a power spectrum analysis able to accommodate the properties of non-stationary signals. The effective frequency range obtained after band pass filtering was 0.5 to 50 Hz. We used SW to perform a 4-layer decomposition into a spectrogram. One-dimensional EEG data are transformed into a two-dimensional time-frequency distribution, where for every time point on the x-axis, a distribution of instantaneous frequencies is estimated and plotted on the y-axis. Fig. 4 shows the SW visual decomposition process for the continuous epilepsy {S}, intermittent epilepsy {FN} and healthy subject {O/Z} datasets. The raw EEGs are expressed in the first column of Fig. 4. (a), (b), and (c). The EEGs are divided into several feature segments according to the frequency domains $\gamma(25\sim 50\,\text{Hz}),\beta(12\sim 25\,\text{Hz}),\alpha(6\sim 12\,\text{Hz}),$ $\theta(3\sim 6\,\text{Hz}),\delta(0\sim 3\,\text{Hz})$. In the first decomposition process, the detail coefficient d_1 and approximation coefficient a_1 are generated. Next, a_1 is injected into the SW to generate the detail coefficient d_2 and approximation coefficient a_2. The other wavelet coefficients are obtained in a similar way. The decompositions {S}, {FN}, and {O/Z} of the EEG datasets are shown in lines 2 to 6 of Fig. 4. (a), (b), (c).



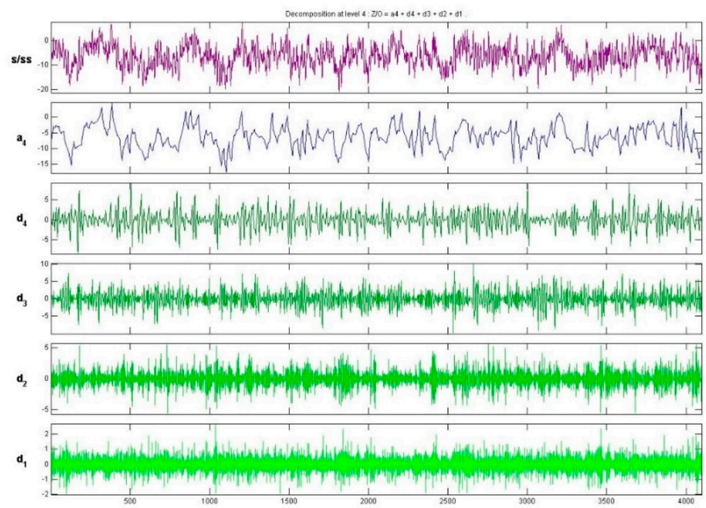
(a)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at Sensors 2019, 19, 219; doi:10.3390/s19020219



(b)



(c)

Figure 4. (**a**) Raw {S} data and corresponding wavelet decomposition; (**b**) Raw {F/N} data and corresponding wavelet decomposition; (**c**) Raw {O/Z} data and corresponding wavelet decomposition.

The absolute value of the data is taken to avoid negative energy. To ensure the credibility of the test results, arithmetic average processing was performed for the above three groups of data and they were compressed into single column matrices. The energy mean, number of cases, and variance of the datasets {F/N}-{O/Z}-{S} are shown in Table 7. Dataset {S} was observed to have the largest standard deviation and the highest mean energy.

Table 7. Statistical features of the data

Datasets	{FN}	{O/Z}	{S}
Mean	-5.94	-6.31	-4.74
Number of cases	4097	4097	4097
Standard deviation	13.10	4.56	38.55

Peer-reviewed version available at Sensors 2019, 19, 219; doi:10.3390/s19020219

4.3 Classification implementation

The architectures of three types of classifiers used for epilepsy detection are shown in Fig. 5. The training data and its corresponding labels are included in the prepared category dataset. The EEGs {S}-{F/N}-{O/Z} are decomposed into five frequency sub-bands by four levels of SW. The mean and standard deviation values of the wavelet coefficients are then calculated to create a ten-dimension feature vector. The training sets {S}, {F/N}, and {O/Z} are labeled with "1", "0" and "-1", respectively. The ten-dimensional feature vector and pre-trained SVM, RF, and GBM classifiers act on the feature recognition module of the scheme. In the practical applications, we use 10-fold CV in the process of training the classifier, because of the number of epileptic datasets. After 10 operations, the average is used as the final CV error $CVe = \frac{1}{10}\sum_{q=1}^{10} e_q$ for selecting the classifier, where $e_q = \frac{1}{m}\sum_{n=1}^{m}\left(\hat{y}_n - y_n\right)^2$ is the average error of the qth test set and m is the number of samples in the qth test set. During the training process, the GSO searches for the optimal values for the generated parameters.

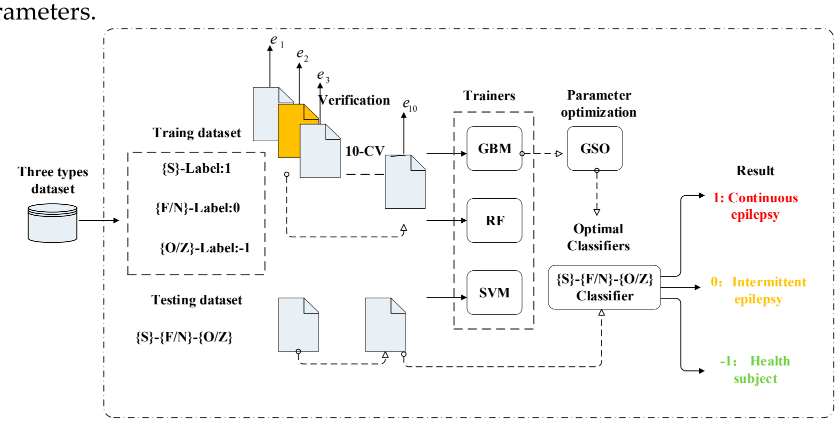


Figure 5. Classification implementation

4.4 Multiple-index evaluation and comparisons

It is also essential to conduct a multiple-index verification of the program after completing the design of the epilepsy detection scheme. We investigated the validity of the proposed method through several experiments using the Bonn University data. The signification costs for the different categories are not equal. Some of the performance indicators, such as accuracy, sensitivity, specificity, and the confusion matrix for the three-level classification of the epilepsy EEG, are defined in Table 8.

Table 8. Definition of the classification multiple-index

Test\Real type	{O/Z} (1)	{F/N} (2)	{S} (3)	Sensitivity (SEN)	Specificity (SPE)	Accuracy (ACC)
{O/Z} (1)	A_{11}	A_{12}	A_{13}	$\frac{A_{11}}{A_1}$	$\frac{A_{22}+A_{23}+A_{32}+A_{33}}{A_2+A_3}$	$\frac{A_{11}+A_{22}+A_{33}}{\text{All}}$
{F/N} (2)	A_{21}	A_{22}	A_{23}	$\frac{A_{22}}{A_2}$	$\frac{A_{11}+A_{13}+A_{31}+A_{33}}{A_1+A_3}$	

Peer-reviewed version available at Sensors 2019, 19, 219; doi:10.3390/s19020219

{S} (3)	A_{31}	A_{32}	A_{33}	$\frac{A_{33}}{A_3}$	$\frac{A_{11}+A_{12}+A_{21}+A_{22}}{A_1+A_2}$

The parameters $A_{ij}\ (i=j)$ are defined as the correct classification probability of sub-dataset $\{i\}$ in the five datasets. Similarly, $A_{ij}\ (i \neq j)$ represents the incorrect classification probability. The parameters $A_i = \sum_{i=1}^{3} A_{1i}$ are the sum of all classification rates of sub-datasets $\{i\}$ $(i, j = 1,2,3)$.

This paper summarizes the processing results of the Bonn University data over recent years, including the techniques used, the number of classification levels, and the results of multiple-index evaluations. As listed in Table 10, almost all researchers have classified the data into two levels, {Z}-{S}, {O}-{S}, {N}-{S}, {F}-{S}, {ZO}-{S}, {NF}-{S}, {ZO}-{NF}, or {FNOZ}-{S}[52, 54-58]; although Wang et al. conducted three-category classification according to {FN}-{OZ}-{S} in 2017, and achieved an accuracy rate of 93.9% [38]. With our method, we achieved better results on the three-category problem, with an accuracy of 96.5%. Many different indicators of machine learning have been evaluated for the SM-GBM-GSO approach, as shown in Table 10. From the above experimental results, we infer that our proposed approach exhibits potential for automated three-level classification of Epilepsy EEG.

Table 9. Comparison of ACC of the two and three-level classifications

Authors	Techniques	10-fold CV	Dataset	ACC%	AUC	CM/PRC
Guo et al. [54] (2010)	DWT and line length, ANN	No	{Z}-{S} {FNOZ}-{S}	100 97.7	No	No
Gandhi et al. [55] (2011)	DWT, energy and std, SVM, NN	Yes	{FNOZ}-{S}	95.4	No	No
Nicolaou et al. [52] (2012)	Permutation entropy, SVM	No	{Z}-{S} {O}-{S} {N}-{S} {F}-{S} {FNOZ}-{S}	93.5 82.8 88.0 79.94 86.1	No	No
Samiee et al. [53] (2015)	STFT Spectral coefficients with their statistical, values,Bayes,LR, SVM,KNN,ANN	No	{Z}-{S} {O}-{S} {N}-{S} {F}-{S} {FNOZ}-{S}	99.8 99.3 98.5 94.9 98.1	No	No
Swami et al. [56] (2016)	DTCWT, energy an std, Shannon entropy features, RNN	Yes	{Z}-{S} {O}-{S} {N}-{S} {F}-{S} {ZO}-{S} {NF}-{S} {FNOZ}-{S}	100 98.89 98.72 93.3 99.1 95.1 95.2	No	No
P. Li et al. [57] (2016)	Distribution entropy and sample entropy Statistical analysis	No	for sample entropy distribution entropy for short length data	mean	Yes 2-level classification 0.93–0.97 0.66–0.87	No
Manish et al. [58] (2017)	ATFFWT and FD, LS-SVM	Yes	{Z}-{S} {O}-{S} {N}-{S} {F}-{S}	100 100 99 98.5	No	No

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at Sensors 2019, 19, 219; doi:10.3390/s19020219

			{ZO}-{S}	100		
			{NF}-{S}	98.6		
			{ZO}-{NF}	92.5		
			{FNOZ}-{S}	99.2		
Yuanfa Wang et al. [38] (2017)	DWT, SVM	No	**{FN}-{OZ}-{S}**	93.9	No	No
This work	Symlets wavelets, statistical mean energy std and PCA, GBM, RF, and SVM GSO	Yes	{Z}-{S}	100	Yes	Yes
			{O}-{S}	100	3-level	
			{N}-{S}	98.4	classification	
			{F}-{S}	98.1	GBM –GSO	
			{ZO}-{S}	100	0.9695	
			{NF}-{S}	98.1	RF –GSO	
			{ZO}-{NF}	93.2	0.9586	
			{FNOZ}-{S}	98.4	SVM –GSO	
			{FN}-{OZ}-{S}	96.5	0.9538	

Furthermore, we compared the confusion matrix (CM) for the three EEG dataset categories {S}-{F/N}-{O/Z} labeled as continuous ictal epilepsy patients, intermittent epilepsy patients, and healthy subjects by the GBM, RF, and SVM classifiers. The GBM classifier achieved higher performance than the RF and SVM classifiers in Fig. 6. This not only guarantees high prediction of true positives and true negatives, as can be seen on the main diagonal line, but also avoids errors from false positives and true negatives, represented by the off-diagonal line.

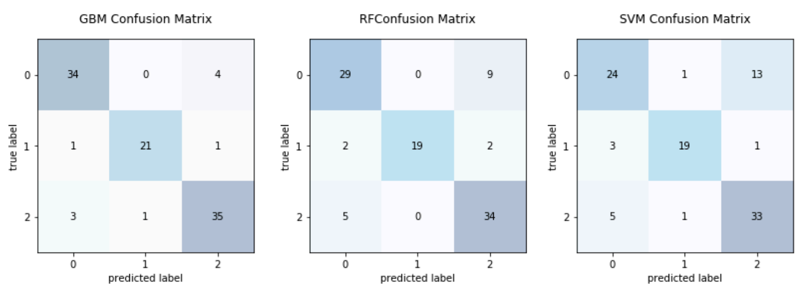
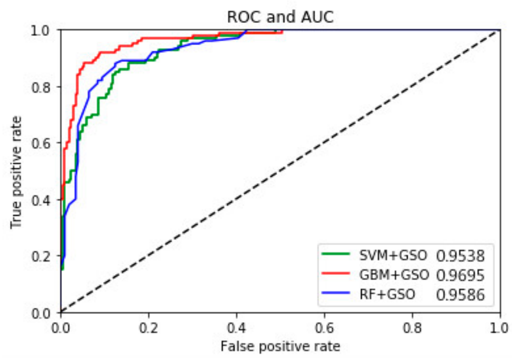


Figure 6. Confusion matrices comparing GBM, RF, and SVM with GSO for {FN}-{OZ}-{S}

With the technological developments in machine learning over recent years, the identification accuracy and confusion matrix can be considered insufficient to judge the accuracy of a classification. We can construct a classifier with high accuracy or recall, but it is difficult to ensure both at the same time. Therefore, we used the ROC and AUC to assess the performance of the classifiers [59]. To allow an ROC curve to be drawn the classifier must provide a confidence value that is judged as positive or negative for each sample. The AUC defines a natural measure for overall performance assessment of a classifier based on the ROC. Li et al. also used the AUC index for their results on the same dataset, but their values of 0.66–0.87 as shown in Table 9 are not very satisfactory [57]. Fig. 7. summarizes the AUC comparisons between the proposed GBM, RF, and SVM identifiers with GSO using subsets {F/N}-{O/Z}-{S}, with values of 0.9695, 0.9586, and 0.9538, respectively. In medical detection, a high true-positive rate is more desirable for a fixed lower-false positive rate. By definition, we consider the higher true-positive value to be the better one.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 November 2018 doi:10.20944/preprints201811.0493.v1

Peer-reviewed version available at Sensors 2019, 19, 219; doi:10.3390/s19020219



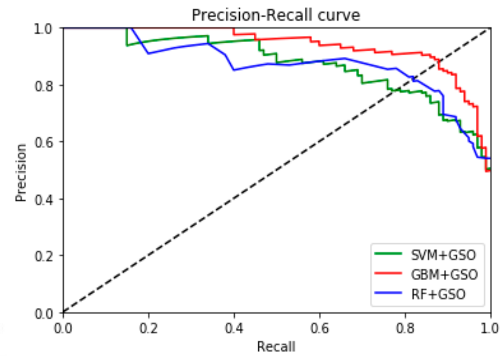


Figure 7. Comparison of ROCs for the three-level classification. **Figure 8.** Comparison of PRCs space for the three-level classification

The precision recall curve (PRC) has a wide range of applications in the field of classification and retrieval; it represents the relationship between precision and recall. The precision values of the vertical axis represent the correct predictions as the ratio of positive samples to all positive samples, while the recall of the horizontal axis represents the correctly predicted ratio of positive samples to true samples. When the precision and recall are high, we can be assured that the classification performance is good. It can be seen in Fig. 8 that the GBM-GSO classifier has the best performance in the three-level classification according to the multiple indicators of accuracy, CM, ROC, AUC, and PRC.

5. Conclusions

The use of EEG signals has changed the method of monitoring epileptic seizures. In this study, the proposed integrative SW-GBM-GSO methods of auxiliary medical diagnostic system for Epilepsy EEG presented excellent performance in a three-level classification of healthy subjects, intermittent epilepsy, and continuous ictal epilepsy. Symlet wavelets were used to decompose the EEG data into five time-frequency sub-bands, while the mean and standard deviation of statistical features were calculated. Subsequently, a modified grid search optimizer was used to search for the optimal parameters using a variable-step method. The use of 10-fold CV avoided overfitting of the classifier. We then compared GBM with SVM and RF in the classification of the EEG data. Considering that most other schemes have only been concerned with classification accuracy, we focused on multiple indicators to illustrate the misclassification factors. These indicators are essential in medical screening. According to the experimental results and multiple co-verification indicators, we conclude that the proposed Symlet wavelet processing, a gradient boosting machine, and a grid search optimizer integrative methods obtain the highest performance in the three-level classification.

In the future, we intend to optimize our detection approach to improve its running speed and achieve higher recognition rates for multiple levels of epileptic seizure. We also hope to transfer the technology out of the laboratory and plan to develop a smart mobile application such as "UMindSleep" to assist medical diagnosis of the epilepsy patient. The EEG signals would be transmitted to a mobile terminal through a wireless sensor network [60] [61-62]. The scheme could assist medical diagnosis and be used to alert medical professionals to an epileptic occurrence. It should be especially useful for people or infants who suffer paroxysmal epilepsy and who could be monitored at home in the evening.

Funding: This research was funded by National Science Foundation of China, grant number 61773032.

References

1. Moshé, S. L.; Perucca, E.; Ryvlin, P.; Tomson, T., Epilepsy: new advances. *Lancet* **2015,** *385* (9971), 884-898.
2. Leijten, F., Multimodal seizure detection: A review. *Epilepsia* **2018,** *59*.

3. Sukumaran, D.; Enyi, Y.; Sun, S.; Basu, A.; Zhao, D.; Dauwels, J., *A low-power, reconfigurable smart sensor system for EEG acquisition and classification*. 2012; p 9-12.
4. Nijsen, T. M. E.; Aarts, R. M.; Cluitmans, P. J. M.; Griep, P. A. M., Time-Frequency Analysis of Accelerometry Data for Detection of Myoclonic Seizures. *IEEE Transactions on Information Technology in Biomedicine A Publication of the IEEE Engineering in Medicine & Biology Society* **2010,** *14* (5), 1197-1203.
5. Milosevic, M.; Van, d. V. A.; Bonroy, B.; Ceulemans, B.; Lagae, L.; Vanrumste, B.; Van, H. S., Automated Detection of Tonic-Clonic Seizures using 3D Accelerometry and Surface Electromyography in Pediatric Patients. *IEEE Journal of Biomedical & Health Informatics* **2016,** *20* (5), 1333-1341.
6. Chen, H.; Xue, M.; Mei, Z.; Oetomo, S. B.; Chen, W., A Review of Wearable Sensor Systems for Monitoring Body Movements of Neonates. *Sensors* **2016,** *16* (12), 2134.
7. Coosemans, J.; Hermans, B.; Puers, R. In *Integrating wireless ECG monitoring in textiles*, The International Conference on Solid-State Sensors, Actuators and Microsystems, 2005. Digest of Technical Papers. Transducers, 2006; pp 228-232 Vol. 1.
8. Yang, B.; Yu, C.; Dong, Y., Capacitively Coupled Electrocardiogram Measuring System and Noise Reduction by Singular Spectrum Analysis. *IEEE Sensors Journal* **2016,** *16* (10), 3802-3810.
9. Jung, H. C.; Moon, J. H.; Baek, D. H.; Lee, J. H.; Choi, Y. Y.; Hong, J. S.; Lee, S. H., CNT/PDMS composite flexible dry electrodes for long-term ECG monitoring. *IEEE Transactions on Biomedical Engineering* **2012,** *59* (5), 1472-1479.
10. Yokus, M. A.; Jur, J. S., Fabric-Based Wearable Dry Electrodes for Body Surface Biopotential Recording. *IEEE transactions on bio-medical engineering* **2016,** *63* (2), 423.
11. Lei, R.; Jiang, Q.; Chen, K.; Chen, Z.; Pan, C.; Jiang, L., Fabrication of a Micro-Needle Array Electrode by Thermal Drawing for Bio-Signals Monitoring. *Sensors* **2016,** *16* (6), 908.
12. Spencer, S. S., MRI, SPECT, and PET imaging in epilepsy: their relative contributions. *Epilepsia* **1994,** *35 Suppl 6* (Suppl.6), S72-89.
13. Szabó, C. Á.; Morgan, L. C.; Karkar, K. M.; Leary, L. D.; Lie, O. V.; Girouard, M.; Cavazos, J. E., Electromyography - based seizure detector: Preliminary results comparing a generalized tonic - clonic seizure detection algorithm to video - EEG recordings. *Epilepsia* **2015,** *56* (9), 1432-1437.
14. Gu, Y.; Cleeren, E.; Dan, J.; Claes, K.; Paesschen, W. V.; Huffel, S. V.; Hunyadi, B., Comparison between Scalp EEG and Behind-the-Ear EEG for Development of a Wearable Seizure Detection System for Patients with Focal Epilepsy. *Sensors* **2018,** *18* (1), 29.
15. Gotman, J.; Gloor, P., Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. *Electroencephalography & Clinical Neurophysiology* **1976,** *41* (5), 513-529.
16. Adeli, H.; Zhou, Z. N., Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods* **2003,** *123* (1), 69-87.
17. Song, J.; Park, I. C., Pipelined Discrete Wavelet Transform Architecture Scanning Dual Lines. *IEEE Transactions on Circuits & Systems II Express Briefs* **2009,** *56* (12), 916-920.
18. Qinghua, H. E.; Peng, C.; Baoming, W. U.; Wang, H. E.; Zhu, L., *VEP SIGNAL EXTRACTION USING WAVELET IN BRAIN-COMPUTER INTERFACE RESEARCH*. 2014; p 937-942.
19. Tzallas, A. T.; Tsipouras, M. G.; Fotiadis, D. I., Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans Inf Technol Biomed* **2009,** *13* (5), 703-710.
20. Whitmer, D.; Worrell, G.; Stead, M.; Lee, I. K.; Makeig, S., Utility of Independent Component Analysis for Interpretation of Intracranial EEG. *Frontiers in Human Neuroscience* **2010,** *4* (44), 184.

21. Subasi, A.; Ismail Gursoy, M., EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications* **2010,** *37* (12), 8659-8666.
22. Bajaj, V.; Pachori, R. B., Classification of seizure and non-seizure EEG signals using empirical mode decomposition. *IEEE Transactions on Information Technology in Biomedicine A Publication of the IEEE Engineering in Medicine & Biology Society* **2012,** *16* (6), 1135-42.
23. Kovacs, P.; Samiee, K.; Gabbouj, M. In *On application of rational Discrete Short Time Fourier Transform in epileptic seizure classification*, IEEE International Conference on Acoustics, Speech and Signal Processing, 2014; pp 5839-5843.
24. Wang, X.; Gong, G.; Li, N., Multimodal fusion of EEG and fMRI for epilepsy detection. *International Journal of Modeling Simulation & Scientific Computing* **2018,** (2).
25. Boashash, B.; Boubchir, L.; Azemi, G. In *Time-frequency signal and image processing of non-stationary signals with application to the classification of newborn EEG abnormalities*, IEEE International Symposium on Signal Processing and Information Technology, 2011; pp 120-129.
26. Boashash, B.; Ouelha, S., Automatic signal abnormality detection using time-frequency features and machine learning. *Knowledge-Based Systems* **2016,** *106* (C), 38-50.
27. Li, D.; Xie, Q.; Jin, Q.; Hirasawa, K., A sequential method using multiplicative extreme learning machine for epileptic seizure detection. *Neurocomputing* **2016,** *214*, 692-707.
28. Sharma, M.; Pachori, R. B.; Rajendra Acharya, U., A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters* **2017,** *94*.
29. Swami, P.; Gandhi, T. K.; Panigrahi, B. K.; Bhatia, M.; Santhosh, J.; Anand, S., A comparative account of modelling seizure detection system using wavelet techniques. *International Journal of Systems Science Operations & Logistics* **2016,** *4* (1), 41-52.
30. He, Q.; Wu, B.; Wang, H.; Zhu, L. In *VEP Feature Extraction and Classification for Brain-Computer Interface*, 8th international conference on signal processing, 2006.
31. Boser, B. E.; Guyon, I. M.; Vapnik, V. N., A Training Algorithm for Optimal Margin Classiiers. *Proceedings of Annual Acm Workshop on Computational Learning Theory* **1992,** *5*, 144--152.
32. Fu, K.; Qu, J.; Chai, Y.; Zou, T., Hilbert marginal spectrum analysis for automatic seizure detection in EEG signals. *Biomedical Signal Processing & Control* **2015,** *18*, 179-185.
33. Brabanter, K. D.; Karsmakers, P.; Ojeda, F.; Alzate, C.; Brabanter, J. D.; Pelckmans, K.; Moor, B. D.; Vandewalle, J.; Suykens, J. A. K., LS-SVMlab Toolbox User's Guide : version 1.7. *Ku Leuven Leuven* **2010,** *66* (8), xix.
34. Liu, Y. C.; Lin, C. C. K.; Jing-Jane, T.; Sun, Y. N., Model-Based Spike Detection of Epileptic EEG Data. *Sensors* **2013,** *13* (9), 12536-12547.
35. Isa, R. M.; Pasya, I.; Taib, M. N.; Jahidin, A. H.; Omar, W. R. W.; Fuad, N.; Norhazman, H. In *EEG brainwave behaviour due to RF Exposure using kNN classification*, IEEE International Conference on System Engineering and Technology, 2016; pp 385-388.
36. Chai, R.; Tran, Y.; Naik, G. R.; Nguyen, T. N.; Ling, S. H.; Craig, A.; Nguyen, H. T.; Chai, R.; Tran, Y.; Naik, G. R. In *Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network*, Engineering in Medicine and Biology Society, 2016; p 4654.
37. Amo, C.; Santiago, L. D.; Barea, R.; López-Dorado, A.; Boquete, L., Analysis of Gamma-Band Activity from Human EEG Using Empirical Mode Decomposition. *Sensors* **2017,** *17* (5).
38. Wang, Y.; Li, Z.; Feng, L.; Bai, H.; Wang, C., Hardware design of multiclass SVM classification for epilepsy and epileptic seizure detection. *Iet Circuits Devices & Systems* **2018,** *12* (1), 108-115.

39. Jouny, C. C.; Franaszczuk, P. J.; Bergey, G. K., Signal complexity and synchrony of epileptic seizures: is there an identifiable preictal period? *Clinical Neurophysiology* **2005,** *116* (3), 552-558.
40. Brinkmann, B. H.; Patterson, E. E.; Vite, C.; Vasoli, V. M.; Crepeau, D.; Stead, M.; Howbert, J. J.; Cherkassky, V.; Wagenaar, J. B.; Litt, B., Forecasting Seizures Using Intracranial EEG Measures and SVM in Naturally Occurring Canine Epilepsy. *Plos One* **2015,** *10* (8), e0133900.
41. Brinkmann, B. H.; Joost, W.; Drew, A.; Phillip, A.; Bosshard, S. C.; Chen, M.; Tieng, Q. M.; He, J.; Muñoz-Almaraz, F. J.; Paloma, B. R., Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. *Brain* **2016,** *139* (6), 1713-1722.
42. Tang, Y.; Durand, D., A tunable support vector machine assembly classifier for epileptic seizure detection. *Expert Systems with Applications* **2012,** *39* (4), 3925-3938.
43. Hefron, R.; Borghetti, B.; Schubert, C. K.; Christensen, J.; Estepp, J., Cross-Participant EEG-Based Assessment of Cognitive Workload Using Multi-Path Convolutional Recurrent Neural Networks. *Sensors* **2018,** *18* (5).
44. Liu, N. H.; Chiang, C. Y.; Chu, H. C., Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensors* **2013,** *13* (8), 10273-10286.
45. Iacoviello, D.; Petracca, A.; Spezialetti, M.; Placidi, G., A Classification Algorithm for Electroencephalography Signals by Self-Induced Emotional Stimuli. *IEEE Transactions on Cybernetics* **2016,** *46* (12), 3171-3180.
46. Siuly, S.; Li, Y., Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. *Comput Methods Programs Biomed* **2015,** *119* (1), 29-42.
47. Murugavel, A. S.; Ramakrishnan, S., Hierarchical multi-class SVM with ELM kernel for epileptic EEG signal classification. *Medical & Biological Engineering & Computing* **2016,** *54* (1), 149-161.
48. Kalyani, S.; Swarup, K. S., Classification and Assessment of Power System Security Using Multiclass SVM. *IEEE Transactions on Systems Man & Cybernetics Part C* **2011,** *41* (5), 753-758.
49. Archer, K. J.; Kimes, R. V., Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis* **2008,** *52* (4), 2249-2260.
50. Friedman, J. H., Greedy function approximation: A gradient boosting machine. *Annals of Statistics* **2001,** *29* (5), 1189-1232.
51. Ling, G.; Daniel, R.; Julián, D.; Rabu?Al, J. R.; Alejandro, P., Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *J Neurosci Methods* **2010,** *191* (1), 101-109.
52. Nicolaou, N.; Georgiou, J., Detection of epileptic electroencephalogram based on Permutation Entropy and Support Vector Machines. *Expert Systems with Applications* **2012,** *39* (1), 202-209.
53. Samiee, K.; Kovács, P.; Gabbouj, M., Epileptic Seizure Classification of EEG Time-Series Using Rational Discrete Short-Time Fourier Transform. *IEEE Trans Biomed Eng* **2015,** *62* (2), 541-552.
54. Guo, L.; Rivero, D.; Dorado, J.; Rabuñal, J. R.; Pazos, A., Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *J Neurosci Methods* **2010,** *191* (1), 101-109.
55. Gandhi, T.; Panigrahi, B. K.; Anand, S., *A comparative study of wavelet families for EEG signal classification*. Elsevier Science Publishers B. V.: 2011; p 3051-3057.
56. Swami, P.; Gandhi, T. K.; Panigrahi, B. K.; Tripathi, M.; Anand, S., A novel robust diagnostic model to detect seizures in electroencephalography. *Expert Systems with Applications* **2016,** *56* (C), 116-130.
57. Li, P.; Li, K.; Liu, C.; Zheng, D.; Li, Z. M.; Liu, C., Detection of Coupling in Short Physiological Series by a Joint Distribution Entropy Method. *IEEE transactions on bio-medical engineering* **2016,** *63* (11), 2231-2242.

58. Sharma, M.; Pachori, R. B.; Acharya, U. R., A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters* **2017,** *94*.

59. Saini, R.; Kaur, B.; Singh, P.; Kumar, P.; Roy, P. P.; Raman, B.; Singh, D., Don't Just Sign Use Brain Too: A Novel Multimodal Approach for User Identification and Verification. *Information Sciences* **2017,** *430*.

60. Li, G.; Chung, W. Y., Estimation of Eye Closure Degree Using EEG Sensors and Its Application in Driver Drowsiness Detection. *Sensors* **2014,** *14* (9), 17491-17515.

61. Chiang; Joyce; Ward; Rabab, K., Sensors, Vol. 14, Pages 2036-2051: Energy-Efficient Data Reduction Techniques for Wireless Seizure Detection Systems. *Sensors* **2014,** *14* (2), 2036-51.

62. Chiang, J.; Ward, R. K., Energy-Efficient Data Reduction Techniques for Wireless Seizure Detection Systems. *Sensors* **2014,** *14* (2), 2036-2051.