





Research Article

Anomaly Detection in EEG Signals: A Case Study on Similarity Measure

Guangyuan Chen ¹, Guoliang Lu ¹, Zhaohong Xie ^{2,3} and Wei Shang ^{2,3}

¹Key Laboratory of High-Efficiency and Clean Mechanical Manufacture of MOE, National Demonstration Center for Experimental Mechanical Engineering Education, School of Mechanical Engineering, Shandong University, Jinan 250061, China

²Institute of Neurology, Shandong University, Jinan, China

³Department of Neurology, Second Hospital of Shandong University, Jinan, China

Correspondence should be addressed to Guoliang Lu; luguoliang@sdu.edu.cn and Wei Shang; wshang85@aliyun.com

Received 22 July 2019; Accepted 13 December 2019; Published 10 January 2020

Academic Editor: Mario Versaci

Copyright © 2020 Guangyuan Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motivation. Anomaly EEG detection is a long-standing problem in analysis of EEG signals. The basic premise of this problem is consideration of the similarity between two nonstationary EEG recordings. A well-established scheme is based on sequence matching, typically including three steps: feature extraction, similarity measure, and decision-making. Current approaches mainly focus on EEG feature extraction and decision-making, and few of them involve the similarity measure/quantification. Generally, to design an appropriate similarity metric, that is compatible with the considered problem/data, is also an important issue in the design of such detection systems. It is however impossible to directly apply those existing metrics to anomaly EEG detection without any consideration of domain specificity. *Methodology.* The main objective of this work is to investigate the impacts of different similarity metrics on anomaly EEG detection. A few metrics that are potentially available for the EEG analysis have been collected from other areas by a careful review of related works. The so-called power spectrum is extracted as features of EEG signals, and a null hypothesis testing is employed to make the final decision. Two indicators have been used to evaluate the detection performance. One is to reflect the level of measured similarity between two compared EEG signals, and the other is to quantify the detection accuracy. *Results.* Experiments were conducted on two data sets, respectively. The results demonstrate the positive impacts of different similarity metrics on anomaly EEG detection. The Hellinger distance (HD) and Bhattacharyya distance (BD) metrics show excellent performances: an accuracy of 0.9167 for our data set and an accuracy of 0.9667 for the Bern-Barcelona EEG data set. Both of HD and BD metrics are constructed based on the Bhattacharyya coefficient, implying the priority of the Bhattacharyya coefficient when dealing with the highly noisy EEG signals. In future work, we will exploit an integrated metric that combines HD and BD for the similarity measure of EEG signals.

1. Introduction

In recent years, we have witnessed significant improvements of using electroencephalogram (EEG) measurement for data acquisition in a wide range of clinical applications. It has also led to the development of data mining methods that discover potential patterns in the data, aiming at characterization of dynamic EEG behaviours. Representative examples include early detection of epileptic seizure [1–3], sleep process monitoring [4–7], and many other neurological disordering related health assessment and surgery problems [8–10].

Time series is an important class of EEG data. One of its mining tasks is to detect potential anomaly event(s)/pattern(s) at an early stage in a long-term EEG monitoring process, which is highly required by change detection [11–13], seizure prediction [14, 15], etc. Hence, the notion of “*anomaly EEG detection*” is defined in the following sections.

The basic premise of anomaly EEG detection is consideration of the similarity between two nonstationary EEG recordings. A well-established scheme is based on sequence matching. Figure 1 illustrates the computation process of this scheme. The continuously monitored EEG signal is first

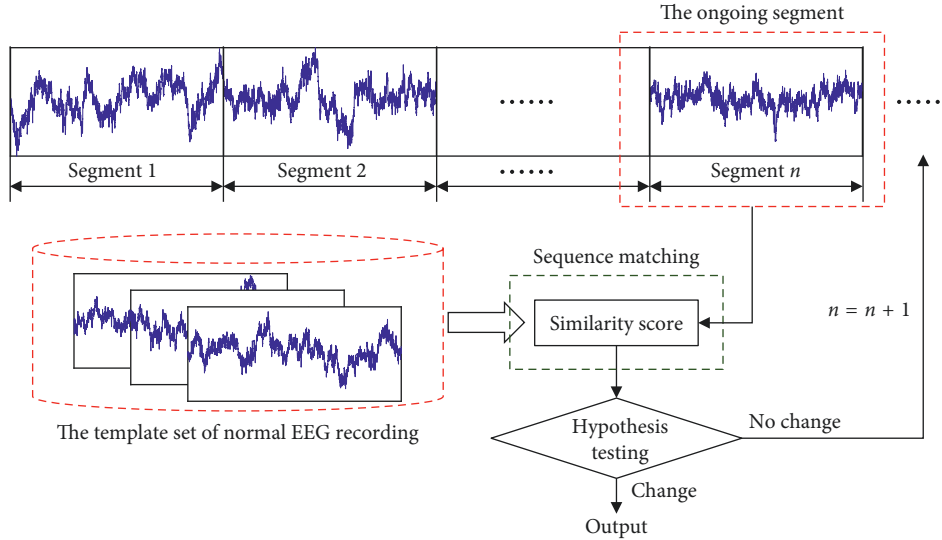


FIGURE 1: The basic premise of anomaly EEG detection.

divided into nonoverlapping (or overlapping) segments; then, the ongoing segment under inspection is compared with those ones that are usual under normal states. It is worth noting that these normal EEG segments can be collected with a prior collection phase or directly taken from the past within the signal itself. The resulting comparison results, i.e., the similarity scores, allow for a change detection by testing a null hypothesis, $H_0: \theta = \theta_0$ against $H_A: \theta \neq \theta_0$ on the parameters θ of an assumed distribution. The Gaussian distribution is the most typical assumption, and some other quantifiers, e.g., a direct threshold, can be also applicable to achieve this end. To summarize, three techniques are crucial to the success of anomaly detection, described as follows:

- (i) *Feature Extraction*. To extract explanatory parameters from the raw EEG data in order to reduce data redundancy
- (ii) *Similarity Measure*. To employ a specific metric to measure/quantify the similarity between two data recordings, i.e., individual EEG segments
- (iii) *Decision-Making*. To make a decision by testing a null hypothesis based on the resulting similarity scores

Along this line of research, many efforts have been made to enhance the feature extraction as seen in [16–18], and some of them also involve the decision-making [4, 19, 20]. Nonetheless, we should be aware that it is also an important aspect to design an appropriate similarity metric, that is compatible with the considered data, when designing such an anomaly detection system [21]. Here, one can note that although the design of similarity metric has been an important problem in the context of statistics and data mining [22–24], the metric used for EEG signal processing still needs to be clarified due to the domain specificity. However, to the best of our knowledge, few of existing studies associated with the EEG signal processing takes into account this issue in the design of anomaly EEG detection systems.

The main objective of this work is to investigate the impacts of different similarity metrics on anomaly EEG detection based on a sequential matching scheme, which uses similarity measure coupled with a null hypothesis testing. Thus, we collect a variety of most popular and state-of-the-art metrics from other areas that would be potentially available for our problem and modify/extend them if necessary to incorporate with the anomaly EEG detection. Impacts of different metrics on anomaly detection results are evaluated based on two data sets. The experimental results reveal the different impacts of investigated metrics. Especially, the HD and BD are demonstrated outperforming performances than other competitors including PCCD, SKLD, KD, BD, and the typically used ED. This study therefore provides a preliminary basis for the EEG signal processing.

The organization of the rest of this paper is given as follows. Section 2 formulates the considered problem. Section 3 introduces several typical metrics that are potentially available for EEG signal analysis. Section 4 describes the testing data and the experimental implementation. Section 5 shows the results with some discussion. Section 6 finally concludes this paper and shows the future work.

2. Problem Formulation

In this section, we first assume that the collected EEG recordings have been already represented by employed features (the feature extraction will be given in the following Section 4.2.1). We then review the method of anomaly EEG detection in the following [25].

The anomaly detection is concerned with recognising new inputs that differ in some way from those that are usual under normal states [26]. Based on this, for a given query EEG recording x , it is a common practice to compare it with a set of normal templates $\{y_j\}$, $j = 1, \dots, M$, where y_j is a EEG recording template and M is the total number. This size of the templates is a trade-off between sensitiveness to EEG

status change and robustness to noise. If the size of the template is larger, it will be more robust to noise but less sensitive to change because the change often occurs instantaneously, and vice versa. In this paper, the size of the templates was set as 20 seconds empirically according to our clinical experience. The (anti-)similarity can be then quantified as the maximum similarity between the query recording and the templates using a similarity metric s . We denote it as $S(x) \leftarrow \max_j s\{\hat{P}, \hat{Q}_j\}$, where the \hat{P} and \hat{Q}_j are the features extracted from x and y_j . The x is inspected as an anomaly event if the resulting similarity score $S(x)$ exceeds a predefined threshold λ , i.e., $S(x) < \lambda$; otherwise, it is inspected as normal. Here, it is worth to mention that the detection can achieve a scalable and flexible detection result with using a different value of λ . However, since the focus of this paper is on the investigation of similarity metric, we do not make additional discussion on this issue. The interested reader can refer to [27, 28] for more discussions on this issue.

The similarity metric s is essential to report an accurate and reliable detection result, and its construction normally relies on a specific distance metric. A greater value of distance indicates a smaller level of similarity. More importantly, for the two given EEG recordings \hat{P} and \hat{Q}_j , the employed distance metric needs to satisfy several fundamental properties:

- (i) *Nonnegativity*, i.e., $s\{\hat{P}, \hat{Q}_j\} \geq 0$
- (ii) *Identity*, i.e., $s\{\hat{P}, \hat{Q}_j\} = 0$ if and only if $\hat{P} = \hat{Q}_j$
- (iii) *Symmetry*, i.e., $s\{\hat{P}, \hat{Q}_j\} = s\{\hat{Q}_j, \hat{P}\}$
- (iv) *Triangle inequality*, i.e., $s\{\hat{P}, \hat{Q}_j\} \leq s\{\hat{P}, \hat{R}\} + s\{\hat{R}, \hat{Q}_j\}$, where \hat{R} is a third EEG recording that is not equivalent to both \hat{P} and \hat{Q}_j

Here, one can note that, the distance metric for similarity quantification is not necessary to meet all of these properties especially the triangle inequality, under which such kinds of distance are called as non-metric distances [29].

Based on the above definition, the similarity metric can also be confirmed as $S \in [0, 1]$ with value of 1 if two compared EEG recordings are identical and 0 if nonidentical at all. In the following, we identify some typical metrics with potentials to solving our problem by careful reviewing of the relevant literature. In particular, during the identification, two following issues were considered:

- (i) The metric should satisfy three properties of *scalability*, *sensitivity*, and *coverage*, according to [30]
- (ii) Among various metrics, we only pay attention to the ones which only calculate the similarity between two sequences with equal lengths

3. Common Metrics

This section introduces a variety of metrics from other areas that would be potentially available for our problem and modify/extend them if necessary to incorporate with the considered anomaly EEG detection problem.

Let us assume that we have two sequences, $P = \{p(k)\}$, $k = 1, 2, \dots, K$, and $Q = \{q(k)\}$, $k = 1, 2,$

\dots, K , where $p(k)$ and $q(k)$ are the observed values of P and Q at time k , respectively. A variety of typical metrics, that are potentially available for EEG analysing, are introduced to measure the similarity between P and Q .

3.1. Euclidean Distance (ED). ED is the most common metric that refers to the real distance between two points in space [31]. The ED between P and Q can be calculated by

$$d^{(ED)} = \sqrt{\sum_{k=1}^n (p(k) - q(k))^2}. \quad (1)$$

Taking into account the characteristics of similarity metric described in Section 2, we use the reciprocal of $d^{(ED)}$ to represent the similarity as

$$s^{(ED)} = \frac{1}{d^{(ED)}}. \quad (2)$$

3.2. Pearson Correlation Coefficient Distance (PCCD). PCCD, proposed by Pearson, is a statistic used to reflect the degree of linear correlation between two series, with values between -1 and 1 . A larger value of this metric implies a stronger correlation of the two compared series [32]. The PCCD between P and Q can be calculated by

$$d^{(PCCD)} = \frac{\sum_{k=1}^K (p(k) - \bar{p})(q(k) - \bar{q})}{\sqrt{\sum_{k=1}^K (p(k) - \bar{p})^2} \sqrt{\sum_{k=1}^K (q(k) - \bar{q})^2}}. \quad (3)$$

So, the similarity defined by PCCD is then calculated by

$$s^{(PCCD)} = |d^{(PCCD)}|. \quad (4)$$

3.3. Symmetric Kullback–Leibler Divergence (SKLD). SKLD can be used to measure the difference between two probability distributions, widely used in information retrieval and data science [33, 34]. The SKLD between P and Q can be calculated by

$$D(P\|Q) = \sum_{k=1}^K \left(p(k) \log \left(\frac{p(k)}{q(k)} \right) \right), \quad (5)$$

$$D(Q\|P) = \sum_{k=1}^K \left(q(k) \log \left(\frac{q(k)}{p(k)} \right) \right),$$

but it is not a distance metric because of its asymmetry. In order to solve the problem, symmetric Kullback–Leibler divergence is very popular in various statistical distance metrics [35] and is calculated by

$$d^{(SKLD)} = \frac{D(P\|Q) + D(Q\|P)}{2}. \quad (6)$$

Then, the similarity can be gotten as

$$s^{(SKLD)} = \frac{1}{d^{(SKLD)}}. \quad (7)$$

3.4. Hellinger Distance (HD). HD was first proposed by Hellinger in [36]. It is used in probability and statistics to measure the similarity between two probability distributions, which belongs to f-divergence [36]. The HD between P and Q can be calculated by

$$d^{(\text{HD})} = \frac{1}{\sqrt{2}} \|\sqrt{P} - \sqrt{Q}\|_2. \quad (8)$$

Thus, the similarity based on HD can be calculated as

$$s^{(\text{HD})} = \frac{1}{d^{(\text{HD})}}. \quad (9)$$

3.5. Kolmogorov Distance (KD). KD was introduced by Kolmogorov [37]. This statistical distance plays an important role in probability theory and hypothesis testing [38], and it is widely used to measure the difference between two probability distributions [39]. Therefore, the KD between P and Q can be calculated by

$$d^{(\text{KD})} = \|P - Q\|_{\infty}. \quad (10)$$

Thus, the similarity based on KD can be calculated as

$$s^{(\text{KD})} = \frac{1}{d^{(\text{KD})}}. \quad (11)$$

3.6. Bhattacharyya Distance (BD). In the statistics, BD which was proposed by Bhattacharyya in [40], also known as the Hellinger distance, measures the similarity of two discrete or continuous probability distributions. It is closely related to the Bhattacharyya coefficient, which measures the overlap between two statistical samples or populations [23]. The Bhattacharyya coefficient can be used to determine the separability of the class classification used in the measurement of two samples that are considered relatively close. The BD between P and Q is defined as

$$s^{(\text{BD})} = -\ln(BC(P, Q)), \quad (12)$$

where $BC(X, Y)$ is the Bhattacharyya coefficient.

$$BC(P, Q) = \sum_{k=1}^K (\sqrt{p(k)q(k)}). \quad (13)$$

In the above schemes of distance metric, the similarity by some of them does not satisfy the condition $s \in [0, 1]$, as summarized in Table 1. To cope with this problem, the similarity needs to be normalized for some of them, and the normalization will be given in Section 4.2.

4. Materials and Methods

This section introduces the testing data and the implementation of our experiments.

4.1. Testing Data. The testing data in this section are from two data sets:

TABLE 1: The range of the distance metrics.

	ED	PCCD	SKLD	HD	KD	BD
Range of distance value	$[0, +\infty]$	$[-1, 1]$	$[0, +\infty]$	$[0, +\infty]$	$[0, +\infty]$	—
Range of similarity score	$[0, +\infty]$	$[0, 1]$	$[0, +\infty]$	$[0, +\infty]$	$[0, +\infty]$	$[0, 1]$

- (i) The first data set is established based on our system setup. The process of data collection is depicted in Figure 2. Electrodes are placed in accordance with the International 10–20 Electrode Placement Method to collect EEG signals. The original multi-channel EEG signals are obtained using the data collector. The sampling rate of data collection used here is 512 Hz. The channel C4 was chosen for our testing. Three neurological experts are invited to check the original data and label the ground-truth according to their domain experiences, i.e., which part is normal and which part is abnormal. Here, it must be pointed out that the normal status represents that the EEG signal is in a stable status, and the abnormal status includes an unstable status of the EEG signal that might be caused by seizures or other abnormal physical activities. The data are divided into several samples using a 10,000 points non-overlapping window. Examples of tested data samples are shown in Figure 3(a).
- (ii) The second data set is taken from the public Bern-Barcelona EEG data set. They randomly select 3,750 pairs of simultaneously recorded signals from the pool of all signals measured at focal and nonfocal EEG channels, respectively, and divide the recordings into time windows of 20 seconds. The original data are recorded with a sampling rate of 1,024 Hz. Then, these EEG signals were downsampled to 512 Hz prior to further analysis so that each piece of EEG data contains 10,240 samples in length [41]. Examples of data in this data set are shown in Figure 3(b).

Additionally, for each data set, we first select 30 pieces of most table normal data segments to form a template set, and the stability and normality here are judged according to domain experts, and the residuals are as the test data. Moreover, the test data are further equally divided into two groups: one for optimizing threshold and one for final testing. Both groups contain 30 pieces of data segments, of which 15 pieces are normal data segments, and the other pieces are abnormal. The detection performance was evaluated with cross-validation of these two groups. We repeat the whole process of the evaluation twenty times, such that the final results can be obtained and analysed.

4.2. Experimental Implementation. Consistent with the mechanism of anomaly EEG detection introduced previously in this paper, we perform three steps, i.e., feature

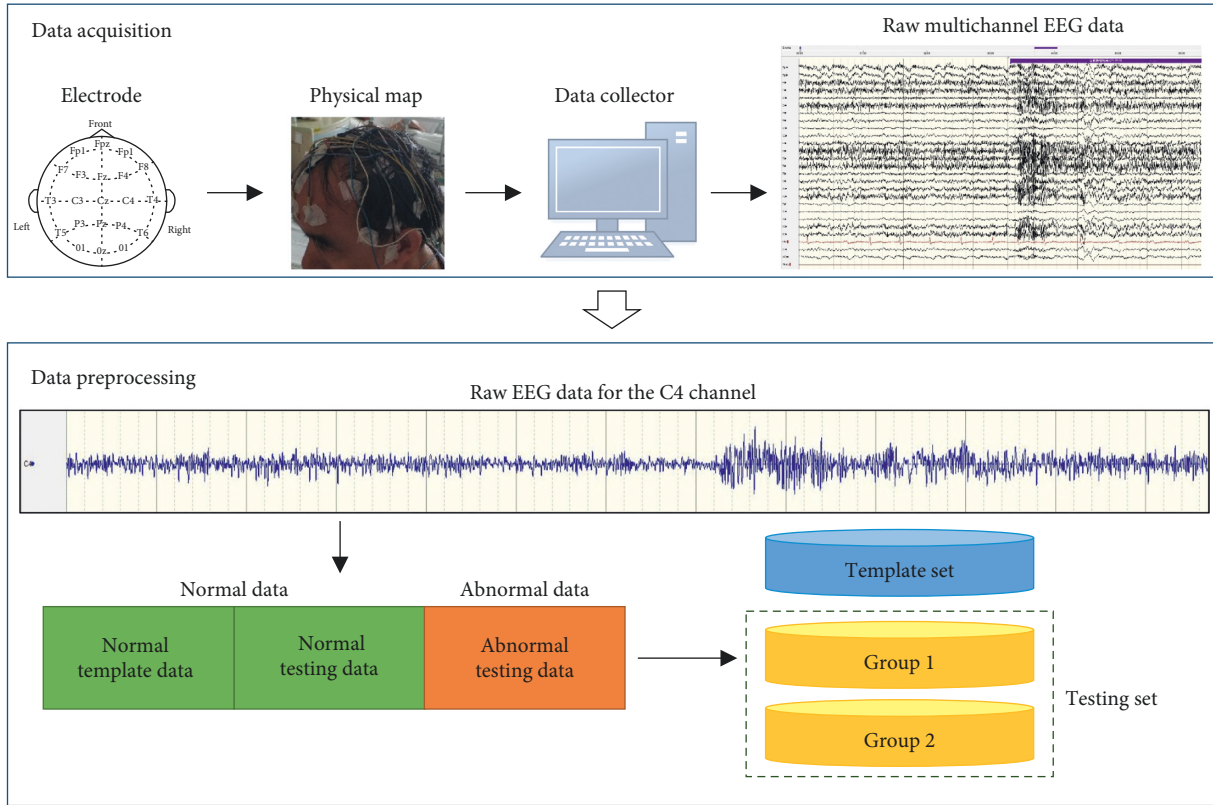


FIGURE 2: Collection of testing data based on our setup.

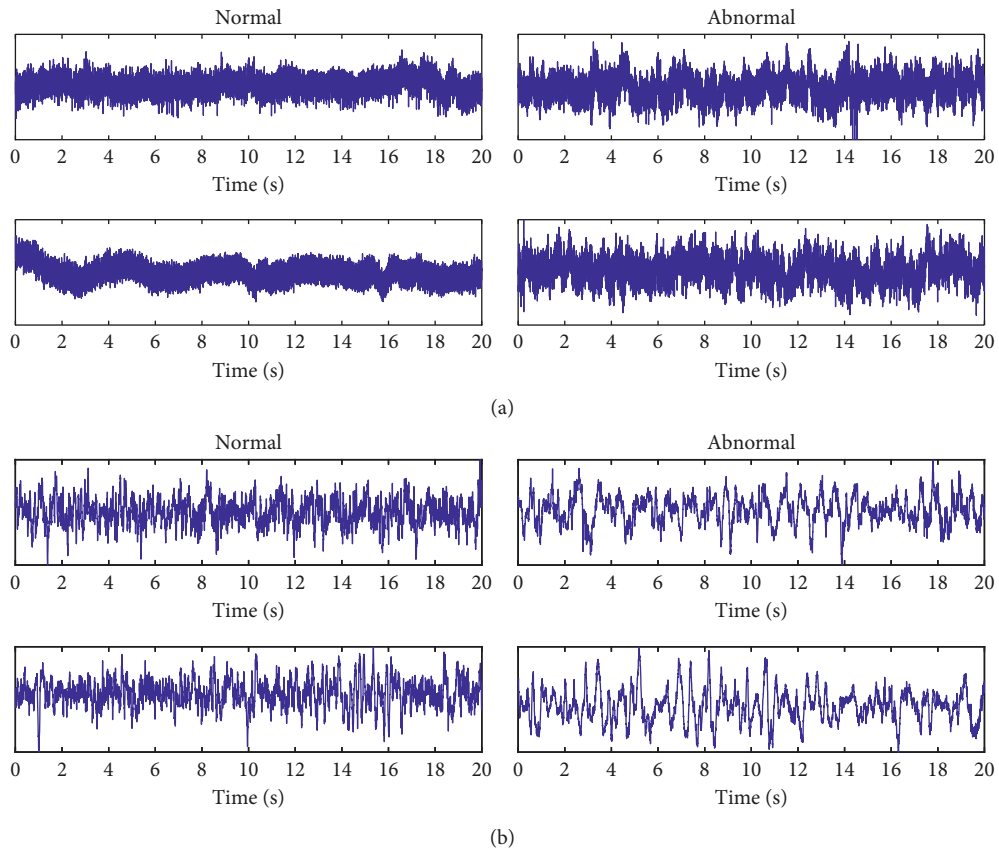


FIGURE 3: Examples of testing data: (a) collected data with our setup, and (b) data taken from Bern-Barcelona EEG database. From left to right: examples of normal data and examples of abnormal data.

extraction, similarity measure, and decision-making, to carry out our experiment. Let us first denote each i th piece of template data as $y_i(n)$, $n = 1, 2, \dots, N$, and denote each i th piece of testing data as $x_i(n)$, $n = 1, 2, \dots, N$. Main methodologies used in the experiments are then introduced in the following.

4.2.1. Feature Extraction. We extract the so-called power spectrum [21] from the raw EEG data as the feature. Let us assume that the observed value of a piece of the EEG signal at the n th point has been denoted as $x(n)$, $n = 1, 2, \dots, N$. The EEG signal was observed in discrete situation, where the transform is discrete in both time and frequency domains [42]. We may review the discrete Fourier transform (DFT) calculation, which is formulated as

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)kn}, \quad k = 1, 2, \dots, N, \quad (14)$$

where $X(k)$ is the output of the transform and k indicates the frequency index.

Recall that the main frequency components of EEG are δ -wave (< 4 Hz), θ -wave (4–8 Hz), α -wave (8–14 Hz), β -wave (14–30 Hz), and γ -wave (> 30 Hz) [43]. That is, if a neurological disorder happens, the amplitudes of these frequencies change accordingly. Thus, they are called characteristic frequencies; i.e., different disorders have different characteristic frequencies. Actually, many successful attempts have been reported using these frequencies to diagnose the neurological disorders [44, 45]. We hence use a subband of [0.1, 70] Hz covering these frequencies empirically for EEG inspection.

After a subband passing filtering (the resulting EEG data are denoted as $x'(n)$ after filtering), the power spectrum $\hat{P}(k)$ can be estimated using the Welch method, a typical power spectrum estimation method, by

$$\hat{P}(k) = \frac{1}{MUL} \sum_{i=1}^L \left| \sum_{n=0}^{M-1} x_n^i(n) d_2(n) e^{-j(2\pi/N)kn} \right|^2, \quad (15)$$

$$k = 1, 2, \dots, K,$$

where $U = (1/M) \sum_{n=0}^{M-1} d_2^2(n)$ and $d_2(n)$ is the window function. The resulting power spectrum $\hat{P}(k)$ allows for the quantitative inspection of EEG data. An example is shown in Figure 4. It can be found that the anomaly EEG signals have the disordering amplitude variations and are polluted with a high ratio of noise. As a result, it would be very difficult to judge whether the EEG signal is abnormal through time-domain analysis. In contrast, the difference between normal and abnormal EEG signals in the frequency domain is more clear, thus allowing for quantitative inspection, i.e., similarity measure, for EEG data inspection.

Based on the above calculation of power spectrum, the testing data $x_i(n)$ and the compared template $y_j(n)$ can be represented as their corresponding power spectrums \hat{P}_i and \hat{Q}_j , respectively.

4.2.2. Similarity Measure. $s\{\hat{P}_i, \hat{Q}_j\}$ is the similarity between \hat{P}_i and \hat{Q}_j , which is calculated through the metrics described in Section 3. The similarity $S(x_i)$ of x_i to a normal status is thought of as the minimum s among all templates, i.e., $S(x_i) \leftarrow \min_j s\{\hat{P}_i, \hat{Q}_j\}$.

Furthermore, in order to satisfy the requirement described in Section 2, $\{S(x_i)\}$ should be normalized as $[0, 1]$ by

$$S'(x_i) = \frac{S(x_i) - \min\{S(x_i)\}}{\max\{S(x_i)\} - \min\{S(x_i)\}}, \quad i = 1, 2, \dots, 60. \quad (16)$$

We still use $\{S(x_i)\}$ to represent the similarity for simplicity in the following.

4.2.3. Decision-Making. In order to inspect whether $x_i(n)$ is normal or not, a threshold λ should be predefined. The decision is subsequently made by testing the following hypothesis:

$$\begin{aligned} H_0: S(x_i) > \lambda, \\ H_A: S(x_i) \leq \lambda. \end{aligned} \quad (17)$$

If the similarity between of testing data $x_i(n)$ is greater than the threshold λ , the data are inspected as a normal data; otherwise, it is considered as abnormal. We first carry out a prior estimation to confirm the optimal value of λ with a number of EEG testing data and then use it to detect all other testing EEG signals in the experiment. The results shown in the following section are obtained by the optimal value of λ .

5. Results and Discussion

5.1. Experiment I: Investigation on Data Set I. As described in Section 4.1, the evaluation was repeated 20 times to obtain the final result. In the following, detailed results for one of evaluations are provided.

Figure 5 provides the detection results for all investigated metrics using the data of our database. In the left of each subfigure, we show the computed similarities of each training data including normal training data and abnormal training data. The similarities are gathered and then arranged in ascending order (normal testing data) or descending order (abnormal testing data). As such, two curves corresponding to normal testing data and abnormal testing data can be obtained, and they intersect at point O . The abscissa of point O (AOPO) can provide an overall evaluation for normal and abnormal testing data. A smaller AOPO means a greater difference between the normal recordings and the abnormal recordings, indicating that the similarity indicator is better; otherwise, the similarities between the two classes of recordings are not much low, meaning that the similarity indicator is not good enough. From these results, it can be clearly seen that HD and BD achieve the best result and the KD and SKLD have achieved *not-so-good* results, while the ED and PCCD have the worst results.

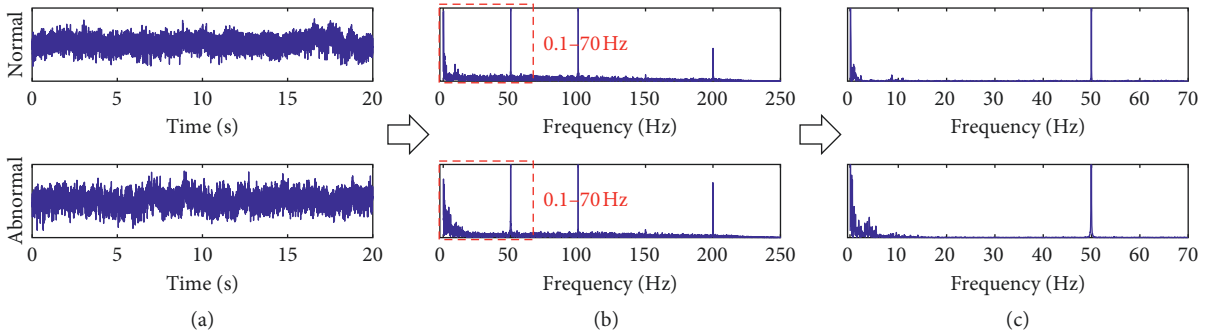


FIGURE 4: An example of feature extraction: (a) original signal; (b) DFT signal; (c) power spectrum.

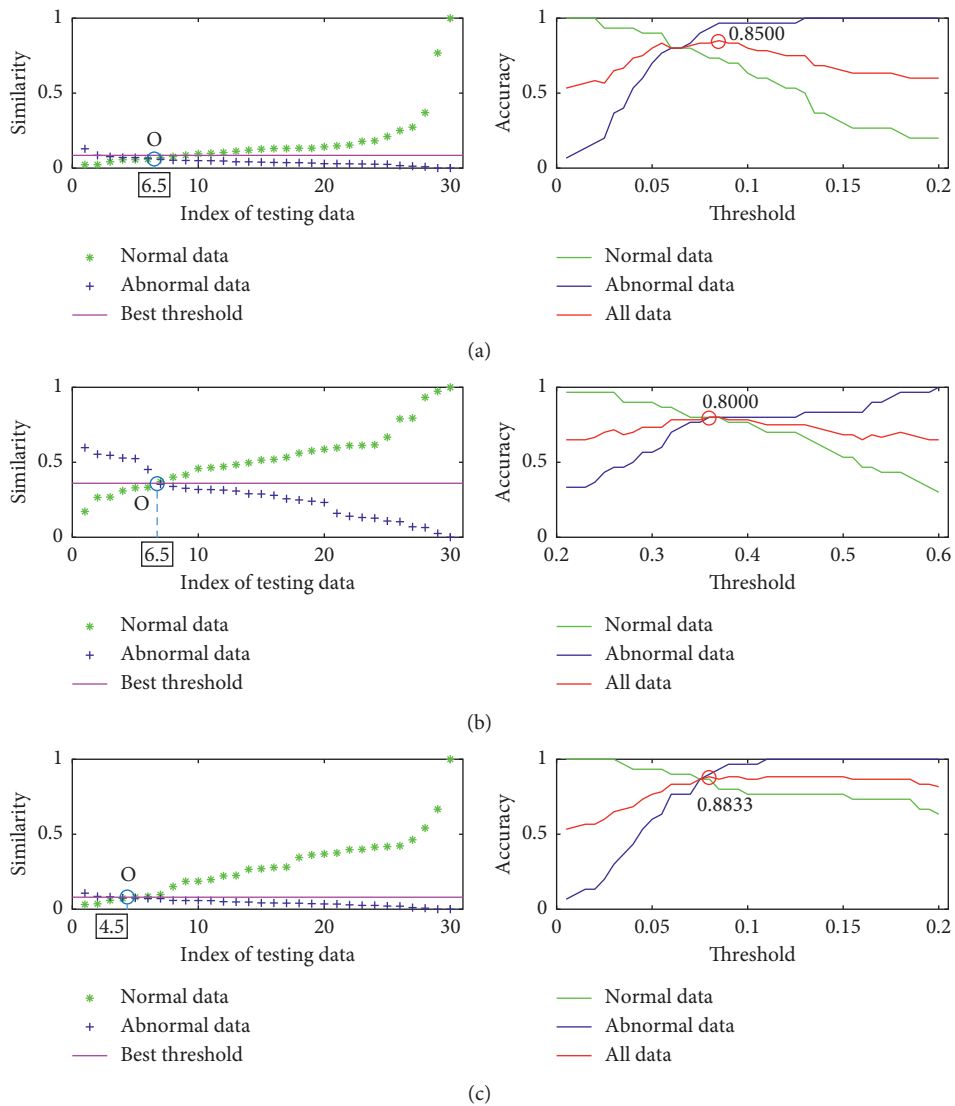


FIGURE 5: Continued.

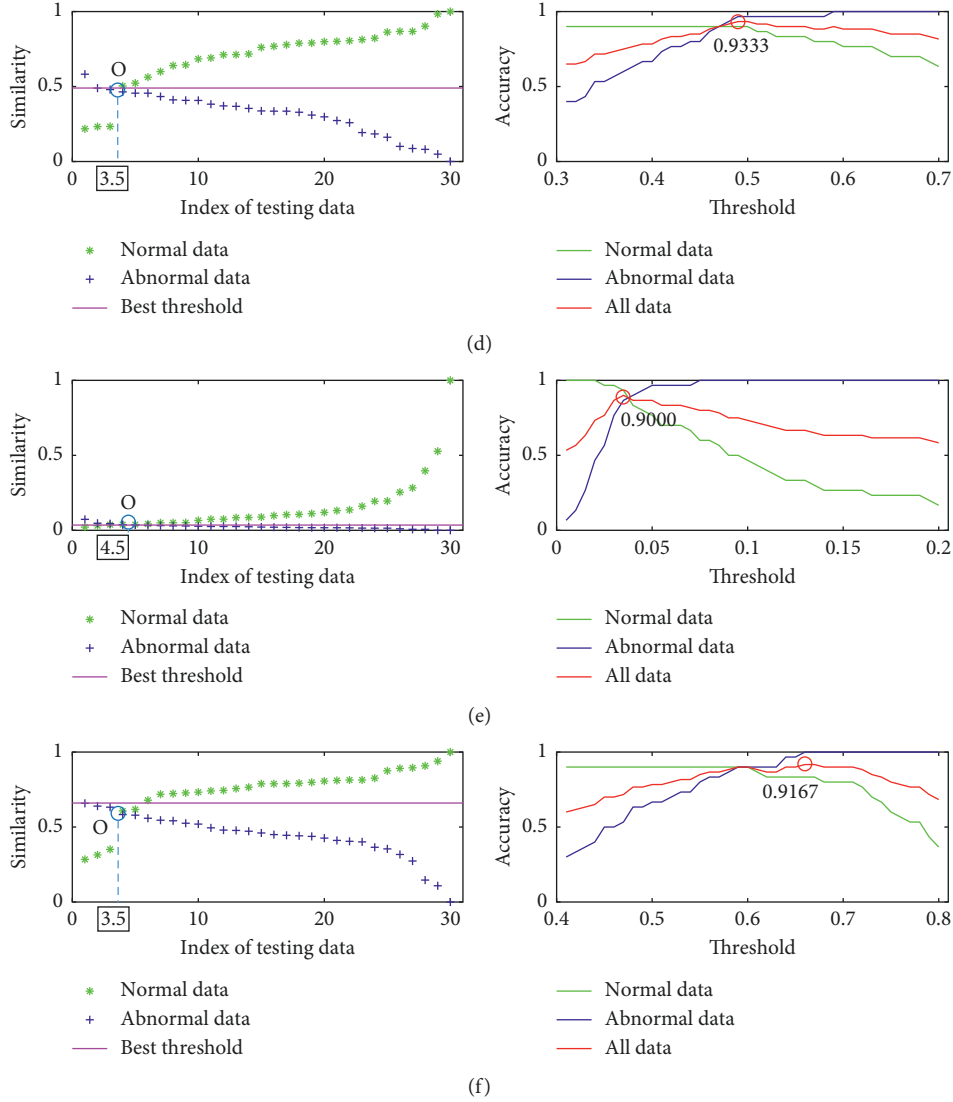


FIGURE 5: Results of six metrics using the training data of our database. From left to right: the similarity between each piece of data in the training data set and the template set and the accuracy of the metric for the normal training data, abnormal training data, and all training data. Detection result by using (a) ED, (b) PCCD, (c) SKLD, (d) HD, (e) KD, and (f) BD.

The other indicator of *accuracy* is also used to quantify the detection performance, which is defined as

$$accuracy = \frac{TP}{TP + FN}, \quad (18)$$

where TP is true positive indicating the number of data that are inspected correctly and FN is false negative indicating the number of data that are inspected incorrectly. The right of each subfigure in Figure 5 shows the results of all metrics in term of *accuracy*. The hypothesis testing described in Section 4.2.3 is used to classify the group 1 of testing data using all investigated metrics with different threshold λ values. Therefore, the higher the *accuracy*, the better the metric. And it can be seen that, for each metric, as λ increases, *accuracy* increases first and then decreases. The values of λ corresponding to the highest *accuracy* are used to calculate the *accuracy* of the group 2 data set. Two examples are given in

Figures 6 and 7, in which we show the similarity scores of all investigated metrics (using their optimal λ) for a normal testing recording and an abnormal testing recording. It can be found that PCCD and KD output wrong results for the abnormal testing data, while the others output the right results. It can be seen that the HD achieves the best performance outperforming other metrics.

We summarize the results of investigated metrics by combining their results in two terms of AOPO and *accuracy* in Table 2. It can be seen that (1) HD and BD are the best metrics in terms of AOPO and (2) HD works best in terms of *accuracy*.

The above experimental process was implemented 20 times. In order to analyse all the experimental results, we calculated the average of the AOPO and *accuracy* values obtained from all experiments based on a global mean measure and show the results in Table 3. It is noticed that the

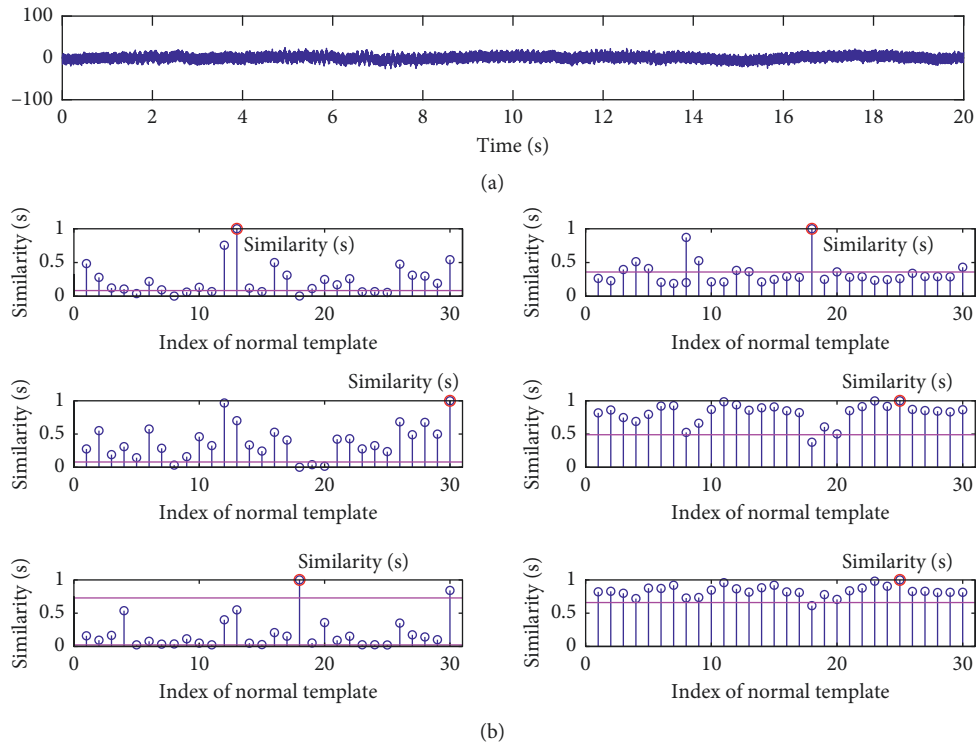


FIGURE 6: An example of similarity scores for a normal testing data in our database metrics. The similarities S between the testing data and the template set are labelled by red circles. From top to bottom: (a) the original data; (b) the similarities obtained by ED, PCCD, SKLD, HD, KD, and BD.

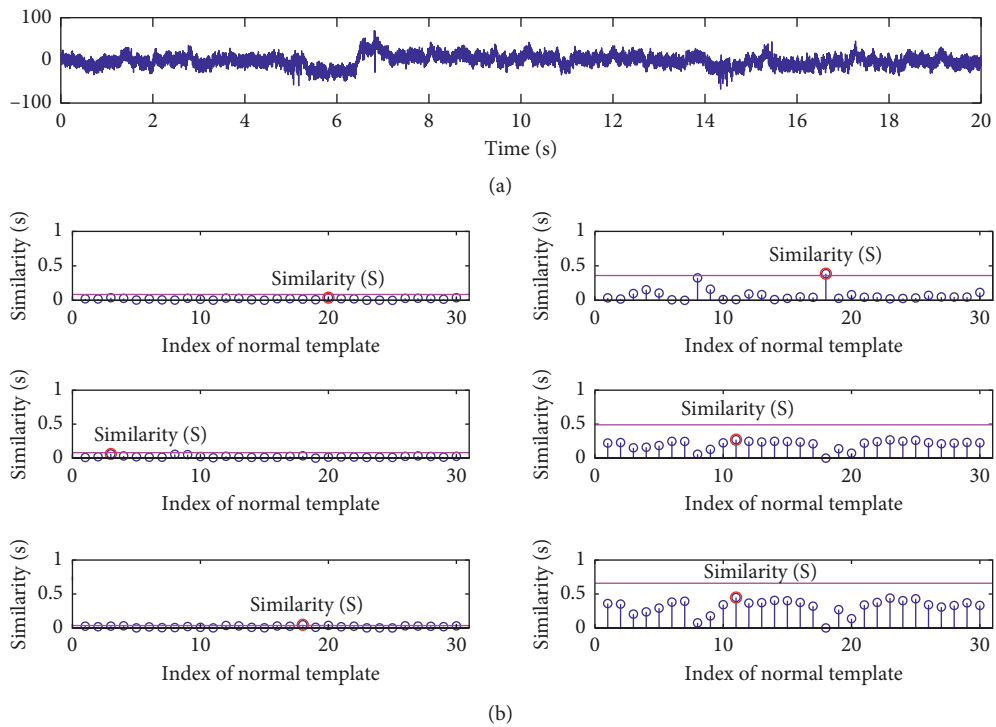


FIGURE 7: An example of the similarity scores for an abnormal testing data in our database. The similarities S between the testing data and the template set are labelled by red circles. From top to bottom: (a) the original data; (b) the similarities obtained by ED, PCCD, SKLD, HD, KD, and BD.

TABLE 2: The compared results of the six metrics in two ways using the testing data of our database for one experiment.

	ED	PCCD	SKLD	HD	KD	BD
AOPO	6.5	6.5	4.5	3.5	4.5	3.5
<i>Accuracy</i>	0.8167	0.7500	0.8500	0.9167	0.8667	0.9000

TABLE 3: The mean results using our database for all experiments.

	ED	PCCD	SKLD	HD	KD	BD
AOPO	6.75	6.35	4.85	3.65	4.75	3.70
<i>Accuracy</i>	0.8033	0.7583	0.8350	0.9133	0.8600	0.8933

metrics of HD achieve the best performance in terms of AOPO, i.e., 3.65; in terms of *accuracy*, the HD outperforms others. Based on these results, the investigated metrics can be ranked as $HD > BD > KD > SKLD > ED = PCCD$.

5.2. Experiment II: Investigation on Bern-Barcelona Data Set.

The result of one repetitive evaluation on the Bern-Barcelona data set is also shown. Figure 8 gives the detection results for all investigated metrics using the training data of the public Bern-Barcelona EEG database. In the left of each subfigure, we show the computed similarities of each testing data. And the similarities are also arranged in ascending order (normal testing data) or descending order (abnormal testing data). Therefore, the AOPOs in this experiment can be gotten. From these results, it can be clearly seen that HD, KD, and BD achieve the best result, the ED and PCCD have achieved *not-so-good* results, while the SKLD has the worst results. The right of each subfigure in Figure 8 shows the results of all metrics in term of *accuracy*. It is clear that, for each metric, as λ increases, *accuracy* increases first and then decreases too. The values of λ corresponding to the highest *accuracy* which is marked as λ_0 are also used to calculate the *accuracy* of the group 2. Two examples are given in Figures 9 and 10, in which we show the similarity scores of all investigated metrics (using their λ_0) for a normal testing recording and an abnormal testing recording. It can be found that all the metrics output the right result for the normal testing data. But for the abnormal testing data, only ED and HD output the correct result. In terms of *accuracy*, BD, HD, and HD are also better than the others.

The results of investigated metrics are also summarized in Table 4. It can be clearly seen that, in this experiment, HD, KD, and BD have achieved the best results in terms of AOPO; in terms of *accuracy*, BD works best.

The above experimental procedure was also implemented 20 times. The averages of the AOPO and *accuracy* values obtained from all experiments are shown in Table 5. Therefore, for the Bern-Barcelona EEG database, the metrics of BD achieves the best performance in terms of AOPO, i.e., 1.55; in terms of *accuracy*, the BD outperforms others. Based on these results, the investigated metrics can be ranked as $BD > HD > KD > PCCD > ED > SKLD$.

5.3. Experiment III: Investigation on Effect of Feature Extraction.

In order to investigate the effect of feature

extraction on detection performance, five representative features including mean, root mean square (RMS), empirical mode decomposition (EMD), discrete wavelet transform (DWT), and artifact subspace reconstruction (ASR) that are used in EEG signal analysis, are investigated in this section. Their operations are provided in Table 6. The processes of similarity measure and decision-making stated in Section 4.2 are also implemented to classify the testing data. The results of AOPO and accuracy of our database are shown in Tables 7 and 8, respectively. The results of the Bern-Barcelona EEG database are shown in Tables 9 and 10, respectively.

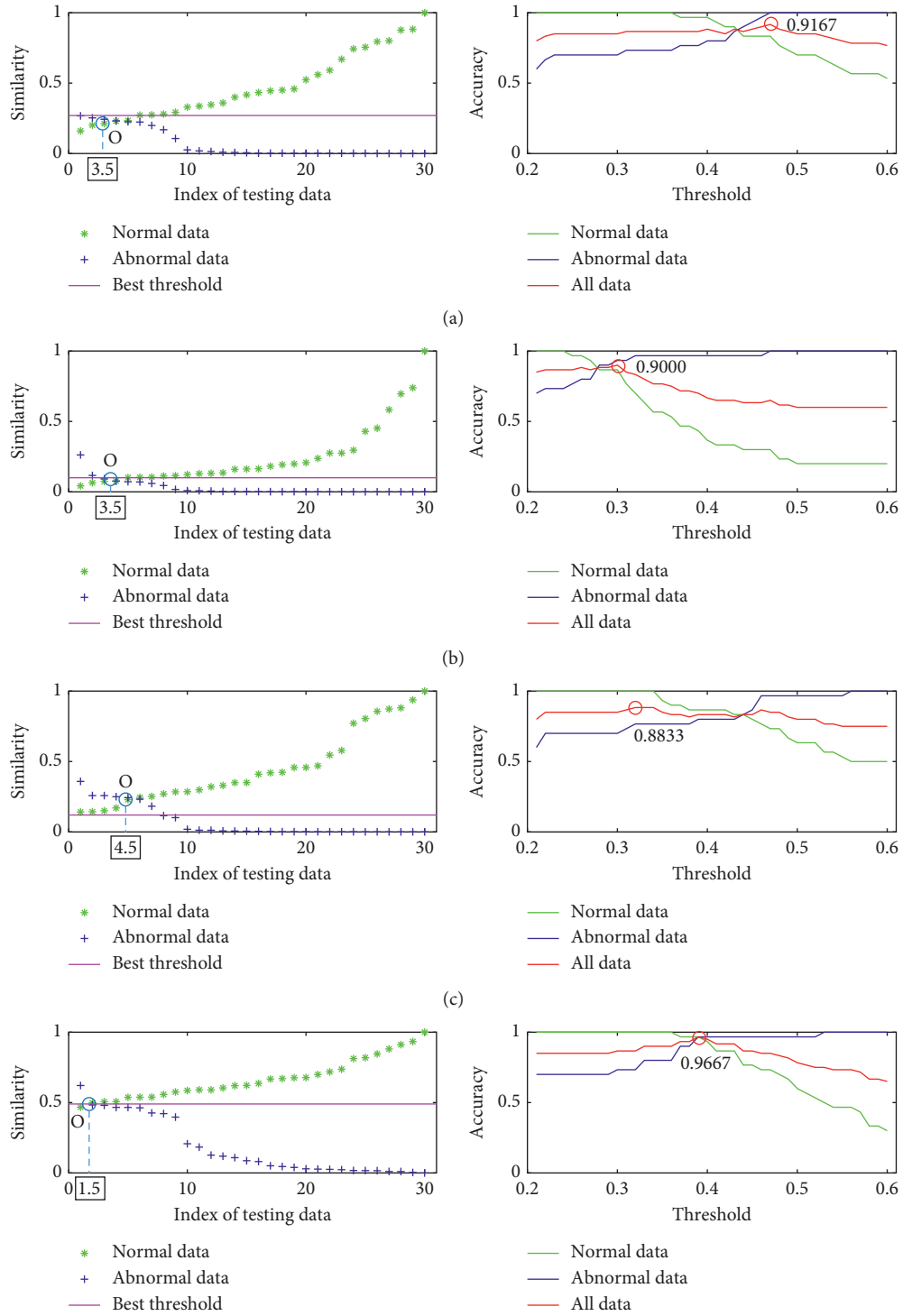
From the results shown in Table 7, we can see that, for our database, in terms of AOPO, the metrics of HD and BD perform better than others when using different features. Table 8 shows the results in term of accuracy. We see that the metrics of HD and BD performs better than others when using DFT, mean, RMS, and ASR; in comparison, PCCD also shows exciting results when using the features of EMD and DWT. Tables 9 and 10 show the detection results for the Bern-Barcelona EEG database. It can be clearly seen that the metrics of HD and BD perform better than other alternatives in both terms of AOPO and accuracy.

To summarize all these results, it can be also noted that, ED, as the most commonly used indicator, performs the worst in terms of AOPO and accuracy for both testing data sets. PCCD, SKLD, and KD have achieved *not-so-good* results. Among all investigated metrics, the metrics of HD and BD are more suitable for EEG signal analysis.

5.4. Result Summary and Discussion. Combining the results from two tested data sets, it is clear that HD and BD achieve a better performance than the other compared metrics. Recall that both BD and HD are obtained by certain transformations of the Bhattacharyya coefficient $BC(P, Q)$, i.e.,

$$\begin{aligned} s^{(HD)} &= 1 - BC(P, Q), \\ s^{(BD)} &= -\ln(BC(P, Q)). \end{aligned} \quad (19)$$

In this regard, HD and BD are thought of as an approximately equivalent measurement of two statistical samples. The difference between them is the sensitivity to noise, as discussed in [49]. However, it is very difficult to determine which of them is more appropriate for analysing the highly noisy EEG signals. As a potential solution of taking advantages of them, one can combine them using



(d)
FIGURE 8: Continued.

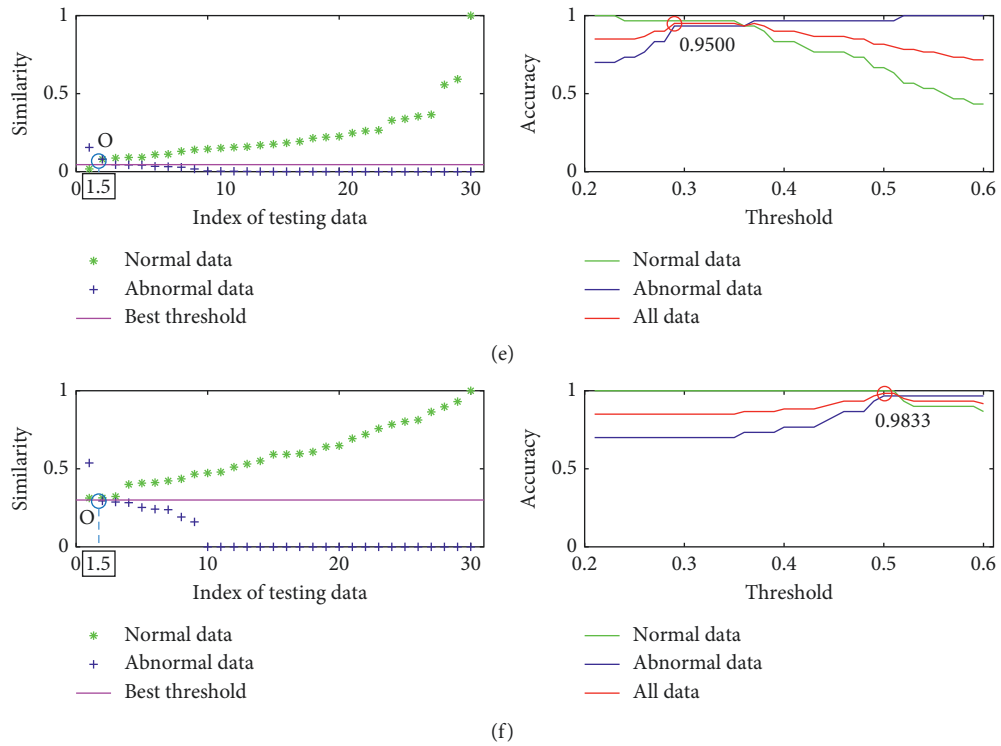


FIGURE 8: The detail results of six metrics using the training data of the Bern-Barcelona EEG database. From left to right: the similarity between each piece of data in the training data set and the template set; the accuracy of the metric for the normal training data, abnormal training data, and all training data. Detection result by using (a) ED, (b) PCCD, (c) SKLD, (d) HD, (e) KD, and (f) BD.

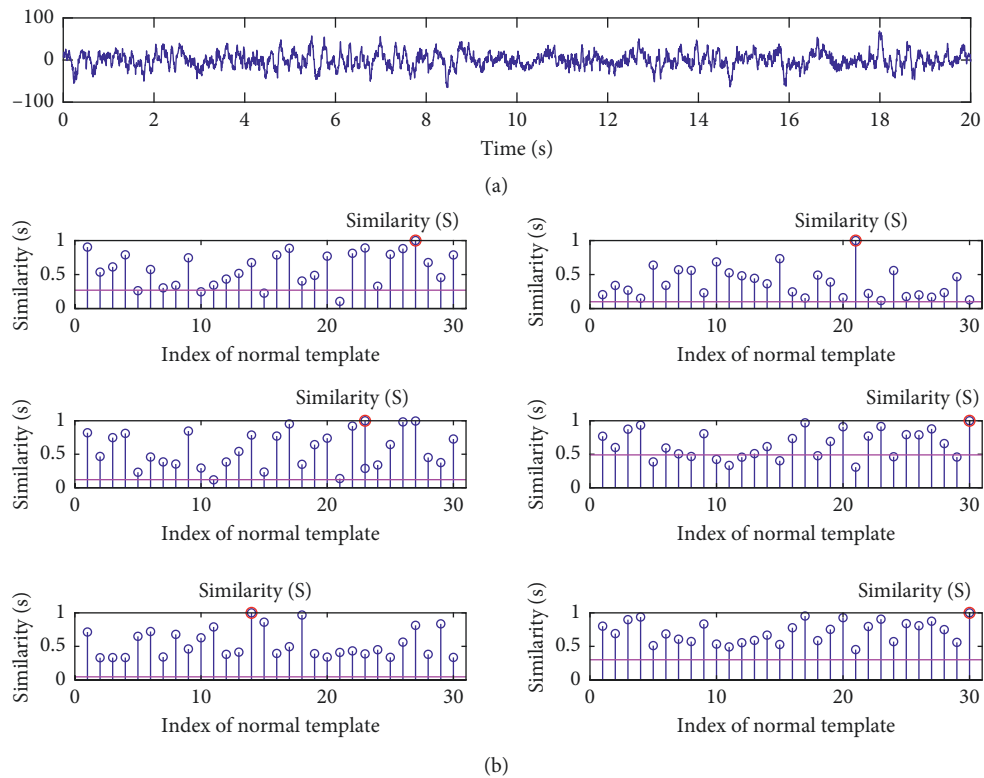


FIGURE 9: An example of similarity scores for a normal testing data of the Bern-Barcelona EEG database. The similarities S between the testing data and the template set are labelled by red circles. From top to bottom: (a) the original data; (b) the similarities obtained by ED, PCCD, SKLD, HD, KD, and BD.

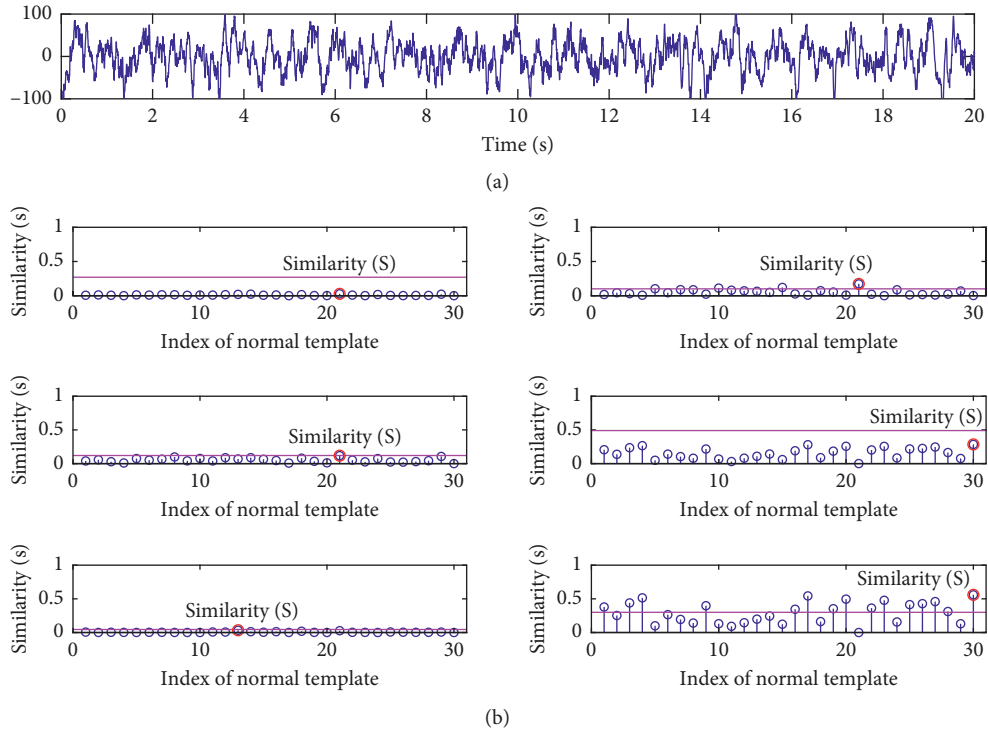


FIGURE 10: An example of similarity scores for an abnormal testing data of the Bern-Barcelona EEG database. The similarities S between the testing data and the template set are labelled by red circles. From top to bottom: (a) the original data; (b) the similarities obtained by ED, PCCD, SKLD, HD, KD, and BD.

TABLE 4: The compared results of the six metrics in two ways using the testing data of the Bern-Barcelona EEG database for one experiment.

	ED	PCCD	SKLD	HD	KD	BD
AOPO	3.5	3.5	4.5	1.5	1.5	1.5
Accuracy	0.9167	0.8833	0.8500	0.9500	0.9333	0.9667

TABLE 5: The mean results using the Bern-Barcelona EEG database for all experiments.

	ED	PCCD	SKLD	HD	KD	BD
AOPO	3.75	3.60	4.35	1.65	1.80	1.55
Accuracy	0.8833	0.8883	0.8583	0.9567	0.9250	0.9633

TABLE 6: Five compared feature extraction methods and the corresponding operations.

Name	Operation
Mean	$X_{\text{Mean}} = (1/N)\sum_{n=1}^N x(n)$
RMS	$X_{\text{RMS}} = \sqrt{(1/N)\sum_{n=1}^N x^2(n)}$
EMD	Empirical mode decomposition (EMD) is a method of signal decomposition based on the time-scale characteristics of the data itself, the detailed process of which can refer to [46].
DWT	Discrete wavelet transform (DWT) is a discrete wavelet transform method. Its detailed process can refer to [47].
ASR	Artifact subspace reconstruction (ASR) is relatively new technique, and it is based on new approach of signal reconstruction with the reference signal fragment. The detailed process of ASR can refer to [48].

TABLE 7: The AOPO results of the seven feature extraction methods using our database.

	ED	PCCD	SKLD	HD	KD	BD
DFT	6.5	6.5	4.5	3.5	4.5	3.5
Mean	14.5	13.5	10.5	7.5	10.5	8.5
RMS	10.5	8.5	7.5	4.5	6.5	5.5
EMD	9.5	7.5	7.5	5.5	7.5	4.5
DWT	9.5	6.5	8.5	4.5	6.5	6.5
ASR	7.5	5.5	4.5	3.5	5.5	3.5

machine learning-based optimization methods, such as *inputs selection* and *inputs weighting* [50–52], to form an integrated metric to measure the considered EEG recordings. This also comprises the direction of our future work.

TABLE 8: The accuracy results of the seven feature extraction methods using our database.

	ED	PCCD	SKLD	HD	KD	BD
DFT	0.8167	0.7500	0.8500	0.9167	0.8667	0.9000
Mean	0.5333	0.6333	0.6833	0.7833	0.7333	0.7500
RMS	0.6833	0.8833	0.7667	0.9000	0.8333	0.8833
EMD	0.7333	0.9500	0.8667	0.9000	0.8167	0.9167
DWT	0.7167	0.9833	0.7667	0.8833	0.8167	0.8500
ASR	0.8333	0.8833	0.8500	0.9000	0.8667	0.9167

TABLE 9: The AOPO results of the seven feature extraction methods using the Bern-Barcelona EEG database.

	ED	PCCD	SKLD	HD	KD	BD
DFT	3.5	3.5	4.5	1.5	1.5	1.5
Mean	13.5	9.5	8.5	7.5	8.5	5.5
RMS	7.5	6.5	6.5	4.5	5.5	2.5
EMD	10.5	6.5	4.5	3.5	5.5	3.5
DWT	9.5	5.5	7.5	4.5	6.5	3.5
ASR	7.5	3.5	4.5	1.5	6.5	2.5

TABLE 10: The accuracy results of the seven feature extraction methods using the Bern-Barcelona EEG database.

	ED	PCCD	SKLD	HD	KD	BD
DFT	0.9167	0.8833	0.8500	0.9500	0.9333	0.9667
Mean	0.5833	0.6333	0.7667	0.7833	0.7333	0.8333
RMS	0.7333	0.8833	0.8167	0.9167	0.8500	0.9333
EMD	0.6167	0.7833	0.8833	0.9167	0.8167	0.8833
DWT	0.7500	0.8333	0.8167	0.8833	0.8667	0.9000
ASR	0.7333	0.8667	0.8500	0.9333	0.7833	0.9167

6. Conclusions

Anomaly EEG detection is a long-standing problem in analysis of EEG signals. The basic premise of this problem is consideration of the similarity between two nonstationary EEG recordings, where a well-established scheme is based on sequence matching. Typically, this scheme includes three steps: feature extraction, similarity measure, and decision-making. Current approaches mainly focus on EEG feature extraction and decision-making, and few of them involve the similarity measure/quantification. Generally, to design an appropriate similarity metric, that is compatible with the considered problem/data, is also an important issue in the design of such detection systems. It is however impossible to directly apply those existing metrics to anomaly EEG detection without any consideration of domain specificity. The main objective of this work is to investigate the impacts of different similarity metrics on anomaly EEG detection. A few metrics that is potentially available for the EEG analysis have been collected from other areas by a careful review of related works, including Euclidean distance (ED), Hellinger distance (HD), Bhattacharyya distance (BD), Kolmogorov distance (KD), Pearson correlation coefficient distance (PCCD), and Symmetric Kullback–Leibler divergence (SKLD). Experiments were conducted on two data sets to

investigate them. Based on the results shown in Section 5, the following are found:

- (1) Experimental results demonstrate the positive impacts of different similarity metrics on anomaly EEG detection. Especially, the commonly used ED did not achieve satisfactory results when compared with other metrics. One main reason is that this metric does not consider the possibly different weight of each element in two compared EEG samples.
- (2) Among all investigated metrics, the HD and BD metrics, that are constructed based on the Bhattacharyya coefficient, show excellent performances. They achieved excellent performances for two inspected data sets: an AOPO value of 3.5 and an accuracy of 0.9167 for our data set and an AOPO value of 1.5 and an accuracy of 0.9667 for the Bern-Barcelona EEG data set. These findings reflect the priority of the Bhattacharyya coefficient when dealing with the highly noisy EEG signals. This study provides a preliminary basis for analysing the EEG data.

In order to take advantages of the Bhattacharyya coefficient, we will exploit an integrated metric combining HD and BD for similarity measure of EEG signals in the future work.

Abbreviations

AOPO:	The abscissa of point O
BD:	Bhattacharyya distance
DFT:	Discrete Fourier transform
ED:	Euclidean distance
EEG:	Electroencephalogram
HD:	Hellinger distance
KD:	Kolmogorov distance
PCCD:	Pearson correlation coefficient distance
SKLD:	Symmetric Kullback–Leibler divergence
Notations	
d :	The distance calculated through the metrics
P, Q :	Two given probability functions
$p(k)$:	The k th point of series P
$q(k)$:	The k th point of series Q
\hat{P} :	The power spectrum of x
\hat{Q} :	The power spectrum of y
\hat{R} :	An EEG recording that is not equivalent to both \hat{P} and \hat{Q}_j
$s\{\hat{P}_i, \hat{Q}_i\}$:	The similarity between \hat{P}_i and \hat{Q}_i
$S(x_i)$:	The similarity between x_i and the template set $\{y_i\}$
$x(n)$:	The n th point of a given EEG recording
$x'(n)$:	The n th point of resulting EEG data after filtering
x_i :	The i th of EEG recording of the testing data set
$X(k)$:	The k point of the frequency spectrum of $x(n)$
y_j :	The j th EEG recording of the template set
λ :	The threshold used for hypothesis testing.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Shandong Provincial Natural Science Foundation, China (ZR2019MEE063) and the Fundamental Research Funds of Shandong University (2018JC010).

References

- [1] D. Gajic, Z. Djurovic, J. Gligorijevic, S. Di Gennaro, and I. Savic-Gajic, "Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis," *Frontiers in Computational Neuroscience*, vol. 9, p. 38, 2015.
- [2] D. Sikdar, R. Roy, and M. Mahadevappa, "Epilepsy and seizure characterisation by multifractal analysis of EEG subbands," *Biomedical Signal Processing and Control*, vol. 41, pp. 264–270, 2018.
- [3] 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.
- [4] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowledge-Based Systems*, vol. 128, pp. 115–124, 2017.
- [5] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel EEG," *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.
- [6] R. Ranjan, R. Arya, S. L. Fernandes, E. Sravya, and V. Jain, "A fuzzy neural network approach for automatic k-complex detection in sleep EEG signal," *Pattern Recognition Letters*, vol. 115, pp. 74–83, 2018.
- [7] S. Devuyt, T. Dutoit, P. Stenuit, and M. Kerkhofs, "Automatic k-complexes detection in sleep EEG recordings using likelihood thresholds," in *Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 4658–4661, IEEE, Buenos Aires, Argentina, August 2010.
- [8] B. R. O'Neill, M. H. Handler, S. Tong, and K. E. Chapman, "Incidence of seizures on continuous EEG monitoring following traumatic brain injury in children," *Journal of Neurosurgery: Pediatrics*, vol. 16, no. 2, pp. 167–176, 2015.
- [9] E. Amorim, J. C. Rittenberger, J. J. Zheng et al., "Continuous EEG monitoring enhances multimodal outcome prediction in hypoxic-ischemic brain injury," *Resuscitation*, vol. 109, pp. 121–126, 2016.
- [10] M. Bachmann, J. Lass, and H. Hinrikus, "Single channel EEG analysis for detection of depression," *Biomedical Signal Processing and Control*, vol. 31, pp. 391–397, 2017.
- [11] B. Boashash, G. Azemi, and N. Ali Khan, "Principles of time-frequency feature extraction for change detection in non-stationary signals: applications to newborn EEG abnormality detection," *Pattern Recognition*, vol. 48, no. 3, pp. 616–627, 2015.
- [12] B. Boashash and S. Ouelha, "Automatic signal abnormality detection using time-frequency features and machine learning: a newborn EEG seizure case study," *Knowledge-Based Systems*, vol. 106, pp. 38–50, 2016.
- [13] Z. Gao, G. Lu, P. Yan et al., "Automatic change detection for real-time monitoring of EEG signals," *Frontiers in Physiology*, vol. 9, p. 325, 2018.
- [14] M. L. Van Quyen, J. Martinerie, M. Baulac, and F. Varela, "Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings," *Neuroreport*, vol. 10, no. 10, pp. 2149–2155, 1999.
- [15] K. Gadhomi, J.-M. Lina, and J. Gotman, "Seizure prediction in patients with mesial temporal lobe epilepsy using EEG measures of state similarity," *Clinical Neurophysiology*, vol. 124, no. 9, pp. 1745–1754, 2013.
- [16] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi, and S. R. Alrshoud, "Epileptic seizure prediction using CSP and LDA for scalp EEG signals," *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 1240323, 11 pages, 2017.
- [17] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.
- [18] T. Zhang, W. Chen, and M. Li, "AR based quadratic feature extraction in the vmd domain for the automated seizure detection of EEG using random forest classifier," *Biomedical Signal Processing and Control*, vol. 31, pp. 550–559, 2017.
- [19] A. F. Rabbi and R. Fazel-Rezai, "A fuzzy logic system for seizure onset detection in intracranial EEG," *Computational Intelligence and Neuroscience*, vol. 2012, Article ID 705140, 12 pages, 2012.
- [20] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.
- [21] Q. Liu, X.-g. Zhao, and Z.-g. Hou, "Metric learning for event-related potential component classification in EEG signals," in *Proceedings of the 2014 22nd European Signal Processing Conference (EUSIPCO)*, pp. 2005–2009, IEEE, Lisbon, Portugal, September 2014.
- [22] B. K. Patra, R. Launonen, V. Ollikainen, and S. Nandi, "A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data," *Knowledge-Based Systems*, vol. 82, pp. 163–177, 2015.
- [23] G. Sidorov, A. Gelbukh, H. Gómez-Adorno, and D. Pinto, "Soft similarity and soft cosine measure: similarity of features in vector space model," *Computación y Sistemas*, vol. 18, no. 3, pp. 491–504, 2014.
- [24] L. Lin, G. Wang, W. Zuo, X. Feng, and L. Zhang, "Cross-domain visual matching via generalized similarity measure and feature learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1089–1102, 2017.
- [25] G. Chen, G. Lu, W. Shang, and Z. Xie, "Automated change-point detection of EEG signals based on structural time-series analysis," *IEEE Access*, vol. 7, pp. 180168–180180, 2019.
- [26] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014.
- [27] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.
- [28] G. Giannakakis, V. Sakkalis, M. Padiaditis, and M. Tsiknakis, "Methods for seizure detection and prediction: an overview," in *Modern Electroencephalographic Assessment Techniques*, pp. 131–157, Springer, Berlin, Germany, 2014.
- [29] M. Alamuri, B. R. Surampudi, and A. Negi, "A survey of distance/similarity measures for categorical data," in *Proceedings of the 2014 International Joint Conference on Neural*

- Networks (IJCNN)*, pp. 1907–1914, IEEE, Beijing, China, July 2014.
- [30] P. Papadimitriou, A. Dasdan, and H. Garcia-Molina, “Web graph similarity for anomaly detection,” *Journal of Internet Services and Applications*, vol. 1, no. 1, pp. 19–30, 2010.
- [31] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, “Euclidean distance matrices: essential theory, algorithms, and applications,” *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, 2015.
- [32] M. C. Whitlock and D. Schluter, *The Analysis of Biological Data*, Roberts and Company Publishers, Greenwood Village, CO, USA, 2009.
- [33] J. R. Hershey and P. A. Olsen, “Approximating the Kullback-Leibler divergence between Gaussian mixture models,” in *Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP’07*, Honolulu, HI, USA, April 2007.
- [34] S. Kullback, *Information Theory and Statistics*, Courier Corporation, Chelmsford, MA, USA, 1997.
- [35] S. Tabibian, A. Akbari, and B. Nasersharif, “Speech enhancement using a wavelet thresholding method based on symmetric Kullback-Leibler divergence,” *Signal Processing*, vol. 106, pp. 184–197, 2015.
- [36] E. Hellinger, “Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen,” *Journal für die Reine und Angewandte Mathematik (Crelles Journal)*, vol. 1909, no. 136, pp. 210–271, 1909.
- [37] A. Kolmogorov, “On the approximation of distributions of sums of independent summands by infinitely divisible distributions,” in *Contributions to Statistics*, pp. 159–174, Elsevier, Amsterdam, Netherlands, 1965.
- [38] C. A. Fuchs and J. van de Graaf, “Cryptographic distinguishability measures for quantum-mechanical states,” *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1216–1227, 1999.
- [39] H. Laurent and C. Doncarli, “Stationarity index for abrupt changes detection in the time-frequency plane,” *IEEE Signal Processing Letters*, vol. 5, no. 2, pp. 43–45, 1998.
- [40] A. Bhattacharyya, “On a measure of divergence between two statistical populations defined by their probability distributions,” *Bulletin of the Calcutta Mathematical Society*, vol. 35, pp. 99–110, 1943.
- [41] R. G. Andrzejak, K. Schindler, and C. Rummel, “Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients,” *Physical Review E*, vol. 86, no. 4, Article ID 046206, 2012.
- [42] A. R. Hassan, S. Siuly, and Y. Zhang, “Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating,” *Computer Methods and Programs in Biomedicine*, vol. 137, pp. 247–259, 2016.
- [43] P. Kellaway, “An orderly approach to visual analysis: characteristics of the normal EEG of adults and children,” *Current Practice: Clinical Electroencephalography*, pp. 139–199, Raven Press, New York, NY, USA, 1990.
- [44] A. Subasi and M. Ismail Gursoy, “EEG signal classification using PCA, ICA, LDA and support vector machines,” *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, 2010.
- [45] E. Başar and B. Güntekin, “Review of delta, theta, alpha, beta, and gamma response oscillations in neuropsychiatric disorders,” in *Supplements to Clinical Neurophysiology*, vol. 62, pp. 303–341, Elsevier, Amsterdam, Netherlands, 2013.
- [46] S. Li, W. Zhou, Q. Yuan, S. Geng, and D. Cai, “Feature extraction and recognition of ictal EEG using EMD and SVM,” *Computers in Biology and Medicine*, vol. 43, no. 7, pp. 807–816, 2013.
- [47] Y. Kumar, M. L. Dewal, and R. S. Anand, “Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network,” *Signal, Image and Video Processing*, vol. 8, no. 7, pp. 1323–1334, 2014.
- [48] M. Plechawska-Wojcik, M. Kaczorowska, and D. Zapala, “The artifact subspace reconstruction (ASR) for EEG signal correction. A comparative study,” in *Proceedings of the 2018 International Conference on Information Systems Architecture and Technology*, pp. 125–135, Springer, Nysa, Poland, September 2018.
- [49] A. Bhattacharya, P. Kar, and M. Pal, “On low distortion embeddings of statistical distance measures into low dimensional spaces,” in *Proceedings of the International Conference on Database and Expert Systems Applications*, pp. 164–172, Springer, Linz, Austria, August 2009.
- [50] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi, “Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1984–1996, 2016.
- [51] P. Du, S. Liu, J. Xia, and Y. Zhao, “Information fusion techniques for change detection from multi-temporal remote sensing images,” *Information Fusion*, vol. 14, no. 1, pp. 19–27, 2013.
- [52] S. Paul and S. Das, “Simultaneous feature selection and weighting—an evolutionary multi-objective optimization approach,” *Pattern Recognition Letters*, vol. 65, pp. 51–59, 2015.