

Research Article

Multifeature Metric Learning Based on Enhanced Equidistance Embedding for Electroencephalogram Recognition of Epilepsy

Bing Lu ¹, Haipeng Lu ^{1,2}, Guohua Zhou ^{1,2,3}, Xinchun Yin ², Xiaoqing Gu ³,
and Tongguang Ni ³

¹School of Information Engineering, School of Artificial Intelligence, Changzhou Institute of Industry Technology, Changzhou, 213164 Jiangsu, China

²College of Information Engineering, Yangzhou University, Yangzhou, 225127 Jiangsu, China

³School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, 213164 Jiangsu, China

Correspondence should be addressed to Tongguang Ni; ntg@cczu.edu.cn

Received 20 October 2021; Revised 19 November 2021; Accepted 22 November 2021; Published 6 December 2021

Academic Editor: Weiwei Cai

Copyright © 2021 Bing Lu 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.

Mobile edge computing (MEC) has the ability of pattern recognition and intelligent processing of real-time data. Electroencephalogram (EEG) is a very important tool in the study of epilepsy. It provides rich information that can not be provided by other physiological methods. In the automatic classification of EEG signals by intelligent algorithms, feature extraction and the establishment of classifiers are both very important steps. Different feature extraction methods, such as time domain, frequency domain, and nonlinear dynamic feature methods, contain independent and diverse specific information. Using multiple forms of features at the same time can improve the accuracy of epilepsy recognition. In this paper, we apply metric learning to epileptic EEG signal recognition. Inspired by the equidistance constrained metric learning algorithm, we propose multifeature metric learning based on enhanced equidistance embedding (MMLE³) for EEG recognition of epilepsy. The MMLE³ algorithm makes use of various forms of EEG features, and the feature weights are adaptively weighted. It is a big advantage that the feature weight vector can be adjusted adaptively, without manual adjustment. The MMLE³ algorithm maximizes the distance between the samples constrained by the cannot-link, and the samples of different classes are transformed into equidistant; meanwhile, MMLE³ minimizes the distance between the data constrained by the must-link, and the samples of the same class are compressed to one point. Under the premise that the various feature classification tasks are consistent, MMLE³ can fully extract the associated and complementary information hidden between the features. The experimental results on the CHB-MIT dataset verify that the MMLE³ algorithm has good generalization performance.

1. Introduction

Mobile edge computing (MEC) converges cloud computing capabilities and Internet service environment to the edge of the network, which can provide services to users nearby, and effectively makes up for the deficiencies of cloud computing [1]. The combination of MEC and artificial intelligence technology is a research hotspot in recent years. MEC has rich application scenarios in the field of intelligent medicine. Electroencephalogram (EEG) analysis is widely used in neuroscience, especially in the diagnosis and seizure of epilepsy [2, 3]. In clinical practice, the diagnosis of epilepsy is mainly based on the patient's history of seizures, and further

examination and diagnosis are made concerning the EEG signals. EEG-based epilepsy detection mainly relies on the personal experience of the doctor. With the gradual development of intelligent medical treatment, the automatic recognition and detection of epileptic EEG signals have become an important auxiliary detection means. How to extract effective features from EEG and design appropriate classification algorithms is the key task to epilepsy detection.

At present, the most commonly used feature extraction methods include the following: time domain analysis, frequency domain analysis, time-frequency analysis, nonlinear dynamics, and model-based methods [4]. The time domain feature method regards EEG as a time series, calculates the

```

Input: Multi-feature matrix  $\mathbf{X}^l$ , its cannot-link and must-link sets;
Output: The metric matrix  $\mathbf{Q}$  and  $\Delta$ .
Initialization:  $\Delta = [1/L, 1/L, \dots, 1/L]$ ,  $\mathbf{Q} = \mathbf{I}$ ,
Repeat
 $t = t+1$ 
Step 1: fixed  $\mathbf{Q}(t)$ , compute  $\Delta(t)$  using Eq.(13);
Step 2: fixed  $\Delta(t)$ , compute  $\mathbf{Q}(t)$  using Eqs.(9)-(11);
Step 3: compute the value of objective function  $J(t)$ ;
Until  $J(t)$  is convergence or  $t \geq t_{\max}$ 
Step 4: obtain the optimal  $\mathbf{Q}$  and  $\Delta$ .

```

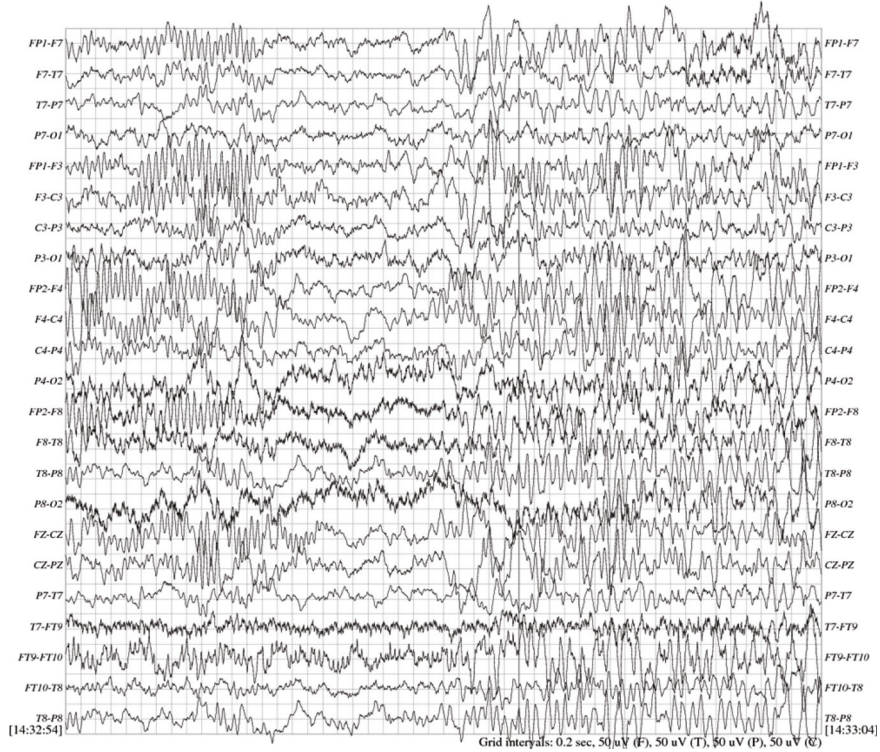
ALGORITHM 1: MMLE³.

FIGURE 1: The example EEG signals in CHB-MIT dataset.

correlation statistics of the sequence, and extracts the corresponding epileptic EEG features. For example, Kaya et al. [5] used the histogram features based on local binary patterns (LBP) together with Bayesian networks to classify epileptic seizures. Another widely used strategy is to extract frequency domain features from a given EEG signal. Fourier transform is one of the most commonly used algorithms to extract frequency domain features from time series data. Frassinetti et al. [6] proposed a preprocessing step method. The signal is filtered by a fixed wavelet transform to reduce possible artifacts. Then, the support vector machine fine Gaussian method is used to detect epilepsy. Chandel et al. [7] proposed a combination of features based on ternary wavelet decomposition to predict the onset and termination of epilepsy. This method extracted standard deviation, variance, and high-order moments to represent the characteristics of

different EEG activities and used linear discriminant analysis and K -nearest neighbor (KNN) classifiers to classify EEG between seizure and interictal periods. In the extraction of nonlinear dynamic features, the method based on complexity analysis is widely used in epilepsy detection, and the most commonly used is the feature extraction method using entropy strategy. For example, Xiang et al. [8] developed a feature extraction algorithm using fuzzy entropy. This method first calculated the fuzzy entropy of EEG signals from different epileptic states, then performed feature selection, and finally used a support vector machine for prediction. Hussein et al. [9] identified EEG seizures by modifying the fuzzy entropy with minimum variance. Firstly, appropriate filtering and independent quantity analysis were carried out to remove noise and artifacts, and then, the proportional operation was carried out to obtain the

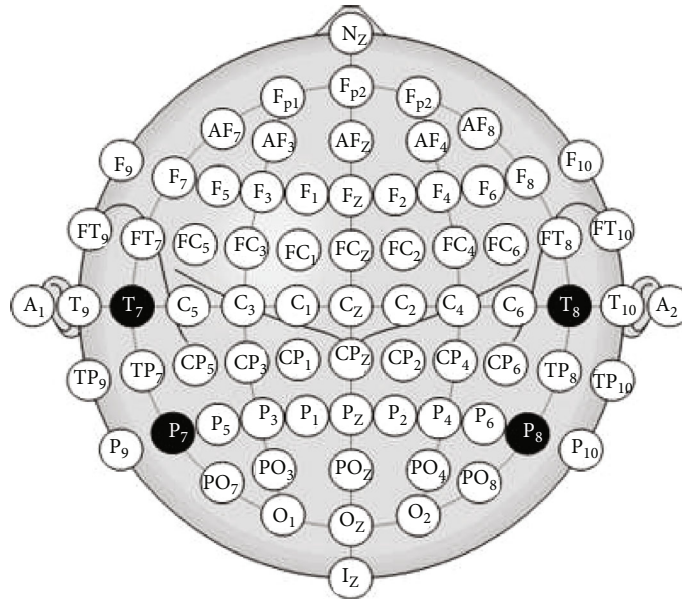


FIGURE 2: The international standard 10-20 system used in the CHB-MIT dataset.

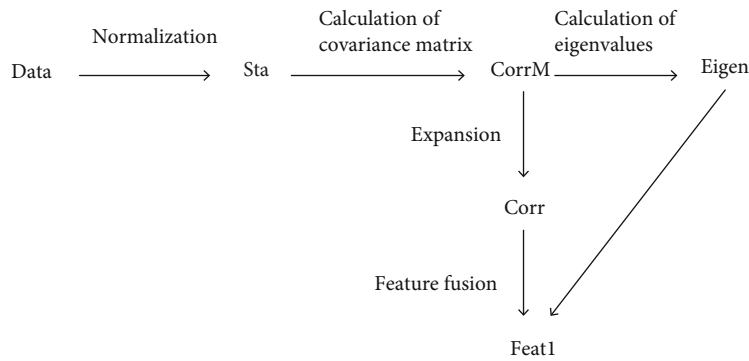


FIGURE 3: The time domain feature extraction process in this study.

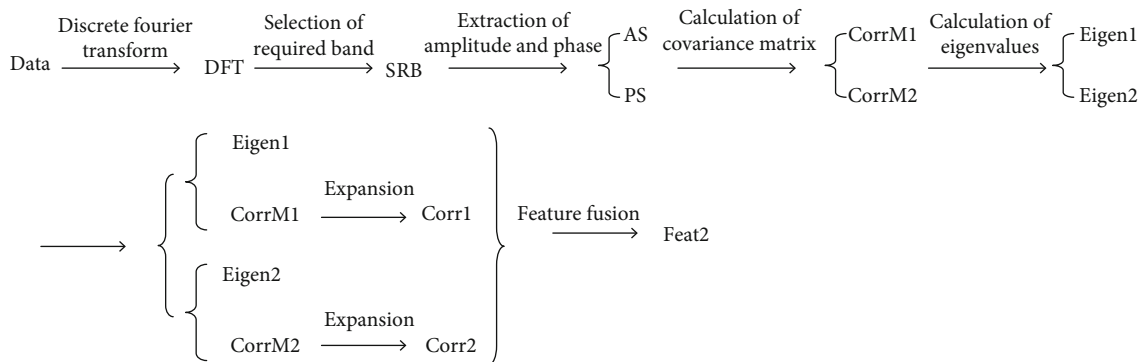


FIGURE 4: The frequency domain feature extraction process in this study.

optimal features. Empirical mode decomposition is an analysis method based on the Fourier transform. Recently, it is also widely used in epilepsy detection. Bajaj and Pachori [10] regarded the intrinsic mode function (IMF) as a group of amplitude frequency modulation signals and gave the analytical expression of IMF by using the Hilbert transform

form of IMF. IMF transform calculated two kinds of bandwidth, namely, a bandwidth and FM bandwidth. Kaleem et al. [11] used a new variant of empirical mode decomposition. This model allowed the detrending of the signal based on the time scale, which decomposed the signal into detrended components and nondetrended components

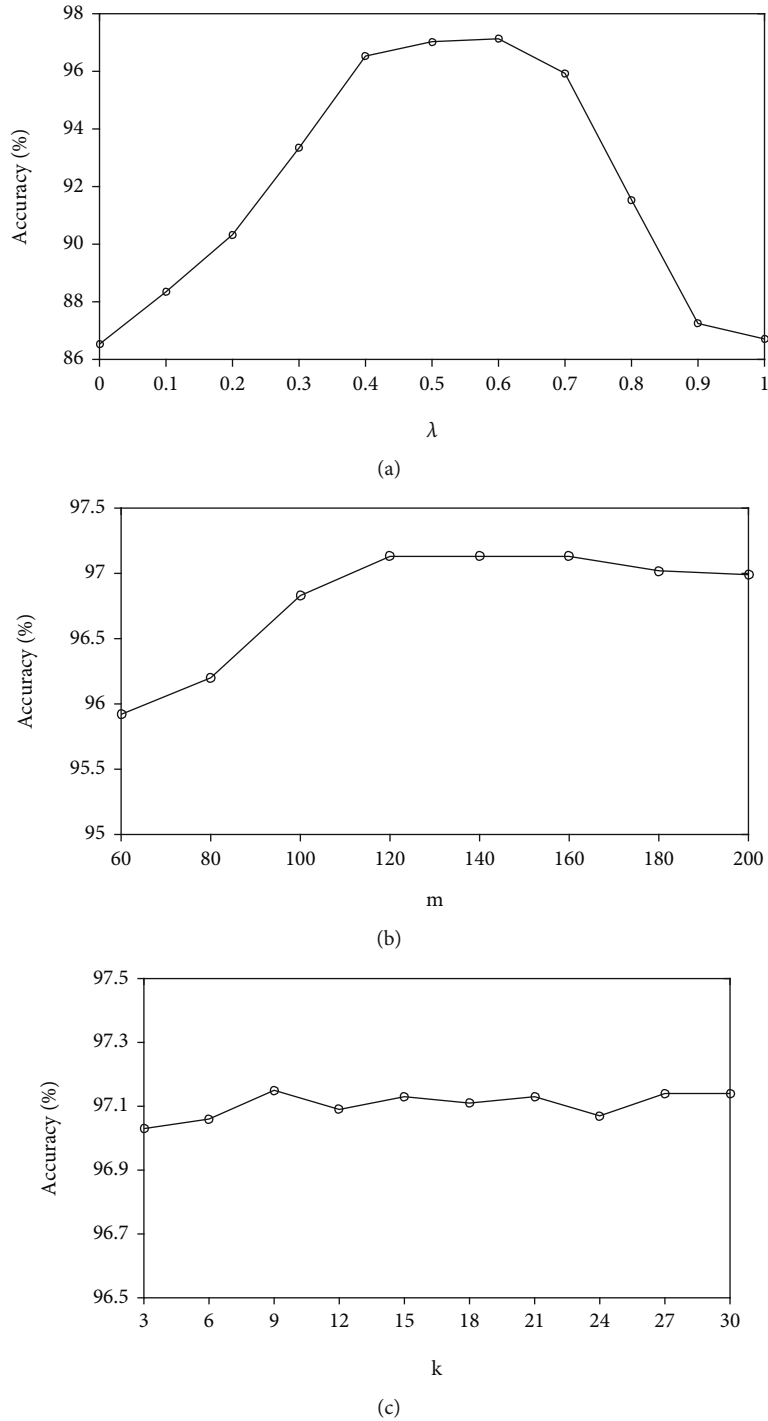


FIGURE 5: The classification accuracy of MMLE³ with different parameters: (a) λ ; (b) m ; (c) k .

according to the frequency separation standard, and then extracted features from the decomposed components. Usman et al. [12] converted the EEG data into a proxy channel and then used the empirical mode decomposition to improve the prediction results.

Supervised learning in machine learning is widely used in epilepsy detection. Some famous supervised learning algorithms, such as KNN, decision trees, and metric learning, have been successfully used for epilepsy detection. Metric

learning is aimed at learning a more suitable distance measurement criterion in the feature space, in order to more accurately represent the similarity between samples. Metric learning is widely used in face recognition, object detection, image recognition, and so on. Weinberger and Saul [13] developed a large margin nearest neighbor analysis algorithm based on a support vector machine. The obtained Mahalanobis distance had the advantages of maximum marginal and internal consistency. Liu et al. [14] developed a

TABLE 1: Results of specificity rates (%) of MMLE³ and the comparison algorithms.

Patient ID	LMNN	ITML	RDML-CCPVL	EquiDML	MV-TSK-FS	CMML	MvCVM	MMLE ³
1	93.68	94.12	95.07	95.41	96.09	96.17	96.08	98.09
2	93.97	94.03	94.65	95.23	96.10	96.28	95.30	97.60
3	91.20	91.16	92.05	93.18	93.18	94.14	93.18	95.59
4	94.37	93.70	95.15	95.61	96.21	96.63	95.49	98.14
5	93.41	93.24	94.35	94.64	95.75	95.49	94.91	97.29
7	93.07	92.45	93.00	94.66	94.86	95.05	94.29	96.59
8	94.32	94.35	94.99	95.83	96.85	96.86	95.75	98.35
9	93.32	94.27	94.48	95.25	96.27	96.20	95.52	97.71
10	92.73	93.31	94.00	94.49	94.97	95.43	94.94	96.73
11	93.14	93.18	93.76	94.64	95.31	95.24	94.78	96.97
13	93.02	93.30	93.74	95.03	95.67	95.89	95.17	97.33
14	92.91	92.84	93.12	94.53	95.23	95.33	94.37	96.85
15	93.38	93.10	93.90	94.84	95.66	95.29	95.26	97.40
17	93.30	94.16	94.10	95.37	95.73	96.16	95.07	97.78
18	93.29	93.94	94.13	95.59	96.02	96.07	94.99	97.75
19	93.88	94.37	95.05	95.86	96.20	96.93	95.71	98.25
20	93.61	93.61	94.14	95.25	95.83	95.71	95.16	97.75
21	93.63	93.63	94.69	95.21	96.09	95.81	95.89	97.83
22	93.67	93.68	93.72	94.43	95.88	95.61	95.56	97.23
23	93.25	93.72	94.49	95.34	96.23	96.28	94.84	97.62
24	93.08	93.67	94.49	94.85	95.68	95.54	95.17	97.31
Mean	93.34	93.52	94.15	95.01	95.71	95.81	95.12	97.44

The bold values mean that they are the best classification results in the experiments.

global metric learning algorithm, which made the separation of different categories of samples greater in EEG signal recognition. Phan et al. [15] developed a global metric learning framework using supervised information. The algorithm can directly process EEG data without preprocessing such as artifact removal. Alwasiti et al. [16] developed a depth metric learning algorithm. Different from the traditional deep learning model, a large amount of training data was required; this algorithm only required very little training data.

Many classification algorithms largely rely on the distance measurement of the input data. In EEG classification, the key problem is to find a good distance measurement, to classify the test EEG into the class of the nearest EEG samples. Many researches have shown that an appropriate distance measure can significantly improve classification accuracy. In metric learning, EEG recognition depends on the similarity measurement between the input EEG data samples, and the similarity measurement between EEG data samples is realized by the distance measurement of the input feature vector of EEG data samples [17, 18]. Therefore, it is crucial to find a good distance metric in the sample feature space.

Due to the rhythm of EEG signals and the collection of EEG signals of multiple channels, EEG data samples have rich feature information. The contribution of different forms of features to EEG recognition is different, some play a decisive role, some play an auxiliary role, and some play a small

or no role. Metric learning measures the similarity of EEG data samples and treats all features equally. Obviously, this strategy cannot accurately measure the similarity between EEG data samples, which will affect EEG signal recognition. In addition, various types of EEG signal features can be obtained from different feature extraction algorithms. Based on the principle of consistency and complementarity, the features of each type will contain specific information, and the use of multiple forms of features at the same time will improve the accuracy of epilepsy recognition.

In this paper, we apply metric learning to the recognition of epilepsy EEG signals. We make full use of various forms of EEG features and assign their different weights automatically. We try to find an appropriate distance measure for EEG data samples, so as to measure the similarity between EEG data samples more accurately, and finally achieve the purpose of improving the accuracy of epilepsy recognition. To achieve this goal, we propose multifeature metric learning based on enhanced equidistance embedding (MMLE³) for EEG recognition of epilepsy. We learn from the techniques of the EquiDML algorithm [19] to maximize the distance between the samples constrained by the cannot-link, so that the samples belonging to different classes are transformed into equidistant. At the same time, the distance between the data constrained by the must-link is minimized, so that the samples belonging to the same class are compressed to one point. In the process of metric matrix learning, feature weight vectors are introduced, and various

TABLE 2: Results of sensitivity rates (%) of MMLE³ and the comparison algorithms.

Patient ID	LMNN	ITML	RDML-CCPVL	EquiDML	MV-TSK-FS	CMML	MvCVM	MMLE ³
1	92.72	93.46	93.64	94.52	95.89	95.32	95.20	97.28
2	92.40	93.04	93.73	94.71	94.58	95.46	95.03	97.13
3	90.76	90.98	91.12	92.67	92.94	92.61	92.87	94.64
4	93.05	93.91	93.98	95.03	95.66	96.18	95.02	97.28
5	92.69	92.84	92.82	93.98	94.99	94.72	93.90	96.49
7	91.32	91.77	92.80	92.89	94.55	94.20	93.85	95.95
8	93.57	94.11	94.01	94.82	95.91	95.50	94.67	97.47
9	92.44	92.65	93.66	94.54	95.35	95.49	94.75	96.81
10	92.18	91.68	92.70	93.77	94.35	94.20	93.35	95.94
11	92.10	91.83	92.98	93.56	95.01	94.27	94.50	96.23
13	92.41	92.36	92.87	93.95	94.53	94.66	94.60	96.43
14	91.76	92.45	92.36	93.63	94.23	94.66	93.38	96.00
15	92.95	92.19	93.35	94.24	95.22	94.87	93.90	96.51
17	93.34	92.83	93.27	94.54	95.50	95.47	94.91	96.95
18	92.78	93.31	93.88	94.07	95.49	95.16	95.08	97.01
19	93.07	93.55	94.54	95.37	95.50	96.20	94.73	97.44
20	93.00	93.44	93.70	94.09	95.12	94.95	94.31	96.99
21	93.02	93.14	93.57	94.94	95.66	95.09	94.99	96.96
22	92.09	92.91	92.85	93.61	94.68	94.58	94.16	96.52
23	93.01	92.48	93.10	94.06	94.82	94.78	94.18	96.85
24	92.65	92.98	92.76	93.48	94.38	94.96	94.16	96.56
Mean	92.54	92.76	93.22	94.12	94.97	94.92	94.36	96.64

The bold values mean that they are the best classification results in the experiments.

features are adaptively weighted to effectively adjust the weight relationship between various features. Under the premise of the consistency of various feature classification tasks, the MMLE³ algorithm can effectively mine the hidden and complementary information between the features and highlight the role of the optimal feature, and it has a stronger discriminative ability. We conduct experiments on the CHB-MIT dataset, and the experimental results validate the effectiveness of MMLE³.

2. Related Work

Metric learning uses a given pair of samples to calculate the similarity between pairs of feature vectors. Metric learning generally uses distance metrics. Taking the commonly used Mahalanobis distance as an example, the distance metric between the two samples \mathbf{z}_i and \mathbf{z}_j can be written as

$$\|\mathbf{z}_i - \mathbf{z}_j\|_{\mathbf{Q}}^2 = (\mathbf{z}_i - \mathbf{z}_j)^T \mathbf{Q} (\mathbf{z}_i - \mathbf{z}_j), \quad (1)$$

where \mathbf{Q} is a positive semidefinite matrix. \mathbf{Q} can be decomposed as $\mathbf{Q} = \mathbf{H}\mathbf{H}^T$, where the matrix $\mathbf{H}^{d \times m}$ ($m \leq d$) is metric matrix (or projection matrix). Therefore, Equation (1) can be expressed as

$$\|\mathbf{z}_i - \mathbf{z}_j\|_{\mathbf{Q}}^2 = (\mathbf{H}\mathbf{z}_i - \mathbf{H}\mathbf{z}_j)^T (\mathbf{H}\mathbf{z}_i - \mathbf{H}\mathbf{z}_j). \quad (2)$$

Therefore, the essence of metric learning is to learn a mapping space. In classification tasks, the commonly used strategy for metric learning is to output a positive value close to zero for pairs of samples of the same class and output larger values for pairs of samples of different classes.

Given a labeled dataset $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n]$ with dimensionality d and n number samples, the label matrix \mathbf{Y} is composed of all class labels of \mathbf{X} . The sets of must-link \mathbf{M} and cannot-link \mathbf{C} are defined as

$$\begin{aligned} \mathbf{M} &= \left\{ (\mathbf{z}_i, \mathbf{z}_j) \mid y_i = y_j \right\}, \\ \mathbf{C} &= \left\{ (\mathbf{z}_i, \mathbf{z}_j) \mid y_i \neq y_j \right\}. \end{aligned} \quad (3)$$

According to the classification principle of minimum intraclass distance and maximum interclass distance, a supervised metric learning framework can be represented as

$$\begin{aligned} \|\mathbf{z}_i - \mathbf{z}_j\|_{\mathbf{Q}}^2 &< \delta_1 \quad (\mathbf{z}_i, \mathbf{z}_j) \in \mathbf{M}, \\ \|\mathbf{z}_i - \mathbf{z}_j\|_{\mathbf{Q}}^2 &> \delta_2 \quad (\mathbf{z}_i, \mathbf{z}_j) \in \mathbf{C}, \end{aligned} \quad (4)$$

where δ_1 and δ_2 are thresholds for sets of must-link and cannot-link, respectively, and $\delta_1 < \delta_2$.

In the EquiDML algorithm [19], the sample pairs in set \mathbf{M} are gathered directly to a signal point. The distances of

TABLE 3: Results of accuracy rates (%) of MMLE³ and the comparison algorithms.

Patient ID	LMNN	ITML	RDML-CCPVL	EquiDML	MV-TSK-FS	CMML	MvCVM	MMLE ³
1	93.50	93.77	94.56	95.49	95.98	96.52	95.34	97.74
2	93.15	93.95	93.62	95.15	95.94	95.37	94.90	97.35
3	91.30	91.73	91.39	93.11	93.26	93.65	93.26	95.19
4	93.87	94.08	94.51	95.71	96.48	95.84	95.47	97.84
5	92.90	93.40	93.31	94.35	95.03	94.89	94.86	97.05
7	91.87	92.46	93.28	93.86	94.69	94.50	94.00	96.38
8	93.84	93.90	94.75	95.25	96.36	96.35	95.86	97.98
9	93.46	93.04	94.20	94.45	95.24	95.62	95.44	97.33
10	92.11	92.54	92.72	94.18	94.27	94.75	93.70	96.47
11	92.85	92.99	93.14	94.54	94.58	94.74	94.26	96.75
13	92.78	93.62	94.03	94.06	95.15	95.79	94.56	97.03
14	92.08	92.94	93.48	93.78	94.19	95.10	94.34	96.48
15	93.38	93.10	93.50	94.50	95.63	95.70	94.61	97.09
17	93.59	93.44	94.28	95.10	95.87	95.81	95.42	97.53
18	93.47	93.22	94.16	95.37	95.27	95.27	95.58	97.39
19	93.77	94.36	95.08	95.85	96.06	96.17	96.00	97.98
20	93.18	93.83	94.24	95.39	95.47	96.01	95.29	97.46
21	93.08	93.67	94.03	94.87	95.34	96.20	95.19	97.54
22	92.92	92.93	93.91	94.49	94.78	95.75	95.06	97.01
23	92.71	92.98	93.52	94.29	94.95	95.26	95.34	97.23
24	92.68	92.72	93.37	94.65	95.45	95.16	94.79	96.97
Mean	92.98	93.27	93.77	94.69	95.24	95.45	94.92	97.13

The bold values mean that they are the best classification results in the experiments.

sample pairs in set \mathbf{C} are forced to have the same constant value. The constraints in the EquiDML algorithm are expressed as

$$\begin{aligned} \|\mathbf{z}_i - \mathbf{z}_j\|_{\mathbf{Q}}^2 &= 0 \quad (\mathbf{z}_i, \mathbf{z}_j) \in \mathbf{M}, \\ \|\mathbf{z}_i - \mathbf{z}_j\|_{\mathbf{Q}}^2 &= \mu \quad (\mathbf{z}_i, \mathbf{z}_j) \in \mathbf{C}, \end{aligned} \quad (5)$$

where μ is a positive value. The equidistance constraint indicates that the distance between classes must be greater than the distance within classes. In the metric space, the samples in the \mathbf{C} set correspond to different classes of samples, and the distances of any different pairs will have the same constant value.

3. Multifeature Metric Learning Based on Enhanced Equidistance Embedding

3.1. The Objective Function of MMLE³. We try to learn a metric space with distinguishing ability. In this metric space, the samples belonging to the same class with different feature forms are as close as possible, and the samples belonging to different classes with different feature forms are as far away as possible. That is, there are more compact intra-class distances and more separable interclass distances in the metric space. The Mahalanobis distance of the l -th feature of samples \mathbf{z}_i and \mathbf{z}_j can be written in the following form

$$\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2. \quad (6)$$

Based on the EquiDML algorithm [19], the proposed MMLE³ algorithm makes use of the correlation and difference between multiple forms of features and makes the proposed algorithm more distinguishable by learning the complementary information of different types of features. Then, the MMLE³ algorithm can be represented by

$$\begin{aligned} L(\mathbf{Q}, \Delta) &= \min \sum_{l=1}^L \Delta_l \left(\frac{\lambda}{2|\mathbf{M}^l|} \sum_{(\mathbf{z}_i, \mathbf{z}_j) \in \mathbf{M}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 + 1 \right)^2 \right. \\ &\quad \left. + \frac{1-\lambda}{2|\mathbf{C}^l|} \sum_{(\mathbf{z}_i, \mathbf{z}_j) \in \mathbf{C}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 - \mu \right)^2 + \frac{1}{2} \|\mathbf{Q} - \mathbf{I}\|_F^2 \right), \\ \text{s.t.} \quad &\sum_{l=1}^L \Delta_l = 1, \quad \Delta_l \geq 0, \end{aligned} \quad (7)$$

where \mathbf{M}^l and \mathbf{C}^l are the sample sets of must-link and cannot-link with the l -th feature expression, respectively. $|\mathbf{M}^l|$ and $|\mathbf{C}^l|$ are the size of \mathbf{M}^l and \mathbf{C}^l , respectively. λ is the trade-off parameter, and Δ is the feature weight vector, and

Δ_l is the feature weight of the l -th sample features. It is worth emphasizing that

- (1) in order to reduce the convergence time, MMLE³ uses the shifted squared loss $(\mathbf{z} + 1)^2 - 1$ for the set \mathbf{M}^l
- (2) Δ_l is not a parameter that needs to be adjusted manually. It can be obtained in a closed-form solution

3.2. The Optimization Procedure. According to the Lagrange multiplier method, the Lagrangian function of Equation (7) can be represented as

$$L = \sum_{l=1}^L \Delta_l \left(\frac{\lambda}{2|\mathbf{M}^l|} \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{M}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 + 1 \right)^2 + \frac{1-\lambda}{2|\mathbf{C}^l|} \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{C}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 - \mu \right)^2 + \frac{1}{2} \|\mathbf{Q} - \mathbf{I}\|_F^2 \right) - \Lambda \left(\sum_{l=1}^L \Delta_l - 1 \right), \quad (8)$$

where Λ is the Lagrange parameter.

There are two tune parameters \mathbf{Q} and Δ in the MMLE³. We use the alternating updating strategy to obtain their optimal parameters. When Δ is fixed in Equation (8), the opti-

mization of \mathbf{Q} is equal to solve the following problem:

$$\frac{\partial L(\mathbf{Q})}{\partial \mathbf{Q}} = \sum_{l=1}^L \Delta_l \left(\frac{\lambda}{|\mathbf{M}^l|} \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{M}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 + 1 \right) (\mathbf{z}_i^l - \mathbf{z}_j^l) (\mathbf{z}_i^l - \mathbf{z}_j^l)^T + \frac{1-\lambda}{|\mathbf{C}^l|} \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{C}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 - \mu \right) (\mathbf{z}_i^l - \mathbf{z}_j^l) (\mathbf{z}_i^l - \mathbf{z}_j^l)^T + (\mathbf{Q} - \mathbf{I}) \right). \quad (9)$$

Using the simplest projected gradient method, \mathbf{Q} can be updated by

$$\mathbf{Q}_{h+1} = \mathbf{Q}_h + \eta_h (\mathbf{Q}_h^* - \mathbf{Q}_h), \quad (10)$$

where $\mathbf{Q}_h^* = [\mathbf{Q}_h - \varepsilon_h \nabla L(\mathbf{Q}_h)]^+$. $[\cdot]^+$ denotes projection on $\Omega = \{\mathbf{Q} \mid \mathbf{Q} \succ 0\}$. $\eta_h \in (0, 1]$ and ε_h are the step and regulation parameters, respectively.

At the h -th iteration, denote $\mathbf{Q}_h^* = [\Gamma_h]^+ = \arg \min_{\mathbf{Q}^* \in \Omega} [\mathbf{Q}^* - \Gamma_h]_F^2$, where $\Gamma_h = \mathbf{Q}_h - \varepsilon_h \nabla L(\mathbf{Q}_h)$. Using the positive semidefinite matrix approximation method [19], we can obtain \mathbf{Q}_h^* as

$$\mathbf{Q}_h^* = \mathbf{P}_h \Sigma_h^+ \mathbf{P}_h, \quad (11)$$

where $\mathbf{P}_h \mathbf{P}_h^T = \mathbf{I}$. $\Sigma_h^+ = \max\{0, \Sigma_h\}$ is a positive diagonal matrix.

When \mathbf{Q} is fixed in Equation (8), the optimization of Δ is equal to solve $\partial L / \partial \Delta_l = 0$. We can obtain

$$\begin{cases} \Delta_l \left(\frac{\lambda}{2|\mathbf{M}^l|} \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{M}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 + 1 \right)^2 + \frac{1-\lambda}{2|\mathbf{C}^l|} \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{C}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 - \mu \right)^2 + \frac{1}{2} \|\mathbf{Q} - \mathbf{I}\|_F^2 \right) - \Lambda = 0, \\ \sum_{l=1}^L \Delta_l - 1 = 0. \end{cases} \quad (12)$$

Then, the solution of Δ_l is

$$\Delta_l = \frac{1 / \left((\lambda / |\mathbf{M}^l|) \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{M}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 + 1 \right)^2 + (1 - \lambda / |\mathbf{C}^l|) \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{C}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 - \mu \right)^2 + \|\mathbf{Q} - \mathbf{I}\|_F^2 \right)}{\sum_{l=1}^L \left(1 / \left((\lambda / |\mathbf{M}^l|) \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{M}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 + 1 \right)^2 + (1 - \lambda / |\mathbf{C}^l|) \sum_{(\mathbf{z}_i^l, \mathbf{z}_j^l) \in \mathbf{C}^l} \left(\|\mathbf{z}_i^l - \mathbf{z}_j^l\|_{\mathbf{Q}}^2 - \mu \right)^2 + \|\mathbf{Q} - \mathbf{I}\|_F^2 \right) \right)}. \quad (13)$$

Obviously, different from the manual parameter adjustment strategy, the weight parameter in the MMLE³ algorithm is adaptive and it can converge to the extreme value at any initial value.

Based on the above analysis, we give the MMLE³ algorithm as follows.

4. Experiments

4.1. Dataset and Feature Extraction. The used dataset called CHB-MIT is from Boston Children's Hospital [20]. The signal data is recorded by the international standard 10-20

system, and the sampling frequency is 256 Hz. The example EEG signals and the used international standard 10-20 system are shown in Figures 1 and 2, respectively. The dataset includes the cortical EEG data of 23 patients with epilepsy. Among the 23 patients, 5 are males, aged between 3 and 22 years, and 17 females, aged 1.5-19 years. The data No. 21 is collected again by patient No. 1, one and a half years later. The gender of patient No. 24 is unknown. In our experiment, 21 out of 24 patients are selected, excluding Nos. 6, 12, and 16, since some channel data of these patients can not be read.

We use two forms of EEG features in the experiment. The first form of features is the time domain features of EEG signals. We extract the correlation coefficient matrix and its eigenvalues of the original EEG signal and fuse them. The detailed time domain feature extraction process is shown in Figure 3. The “Date” is the original EEG signals. The “Sta” is the standardized matrix of EEG signals. The “CorrM” is the correlation coefficient matrix of “Sta,” and the “Eigen” is the eigenvalue corresponding to the “CorrM” matrix. The “Corr” is the expansion of “CorrM.” The “Feat1” is the experimental time domain feature by feature fusion of “Corr” and “Eigen.”

The second form of features is the frequency domain features of EEG signals. The detailed frequency domain feature extraction process is shown in Figure 4. The amplitude spectrum and phase spectrum in the frequency domain are two important features related to the time domain information. After being extracted the amplitude and phase features (called “AS” and “PS” in Figure 4), the correlation coefficients (called “CorrM1” and “CorrM2” in Figure 4) and eigenvalues of the spectrum (called “Eigen1” and “Eigen2” in Figure 4) are further extracted. The Feat2 is the experimental frequency domain feature by feature fusion of “Eigen1,” “Eigen2,” “Corr1,” and “Corr2.” The third form of features is the nonlinear features of EEG signals. In the experiment, the Shannon entropy, spectral entropy, and differential entropy of each delta (1-4 Hz), theta (4-7 Hz), alpha (7-13 Hz), and beta (13-30 Hz) band of EEG signals are calculated. Then, the nonlinear feature Feat3 is obtained by three entropies.

We compare MMLE³ with seven algorithms. The comparison algorithms include LMNN [13], ITML [21], RDML-CCPVL [22], EquiDML [19], CMML [23], MV-TSK-FS [4], and MvCVM [24]. The slack variable in ITML is selected in {0.01, 0.1, 1, 10}. In CMML, the tradeoff parameter, learning rate, and parameter p are set to 1, 10^{-6} , and 5, respectively. The number of fuzzy rules in MV-TSK-FS is selected in {5, 10, ..., 30}, and three regulation parameters are set in $\{10^{-2}, 10^{-1}, \dots, 10^2\}$. In MV-TSK-FS and MvCVM, the penalty parameter for each view is selected in {1, 10, 10^2 , 10^3 }, and the Gaussian kernel parameter is selected in $\{10^{-2}, 10^{-1}, \dots, 10^2\}$. In MMLE³, the parameter λ is selected in {0.1, 0.2, ..., 0.9}, and the parameter μ is set to be 2. The KNN is used as the classifier in MMLE³. We use the grid search and 5-fold cross strategy to select the best variables. The running environment of all algorithms is CPU i7-8700k, 3.2GHz, and 32GB RAM, and software is Matlab 2016. The evaluation index adopts the specificity, sensitivity,

and classification accuracy rate. The experiment is executed 10 times.

4.2. MMLE³ Performance Verification. The classification accuracy of MMLE³ with different parameters is shown in Figure 5. The first parameter is the balance parameter λ . The parameter λ is between [0,1] to balance the proportion of minimizing the distance term of the same class samples and maximizing the distance term of different class samples in the objective function. The accuracy of MMLE³ with different λ is shown in Figure 5(a). When the balance parameter is 1, the MMLE³ algorithm only optimizes the must-link constraint and ignores the cannot-link constraint, so its classification accuracy is low. When the balance parameter is close to 0, the objective function of MMLE³ ignores the optimization of the must-link constraint, so the classification accuracy of MMLE³ is also unsatisfactory. From Figure 5(a), when the balance parameter is between 0.4 and 0.6, the two optimization terms can be balanced, so that the EEG data samples in the metric space have the highest discriminative ability, and the classification accuracy of MMLE³ is the highest.

Second, we evaluate the dimension m in matrix \mathbf{Q} . The dimension of each form of features is 200. The MMLE³ algorithm obtains the Mahalanobis matrix by metric learning on the multiple forms of EEG features and projects the features into the projection space. The dimension m takes an important role in MMLE³. The classification accuracy of MMLE³ with different m is shown in Figure 5(b). When m is very small, MMLE³ will ignore most of the discriminative feature information, which leads to low classification accuracy. With the increase of m value, the discriminative feature information increases, and this improves the classification accuracy. When m increases to a certain value, all the discriminative feature information has been obtained, and the remaining small or ineffective feature information has little contribution to the EEG signals recognition. Therefore, the classification accuracy of MMLE³ keeps stable.

Thirdly, we evaluate the KNN classifier parameter in MMLE³. The k parameter is selected in {3, 6, ..., 30}. The accuracy of MMLE³ with different k is shown in Figure 5(c). The value of k has little effect on classification accuracy. Regardless of the value of k , the fluctuation of classification accuracy is very small.

4.3. Algorithm Performance Comparison. The proposed algorithm MMLE³ is compared with several algorithms on 4CHB-MIT dataset. During the experiment, every algorithm runs 10 times and the specificity, sensitivity, and accuracy of all algorithms are recorded in Tables 1–3. CMML, MV-TSK-FS, MvCVM, and MMLE³ can make use of various forms of EEG features. In the experiment, three forms of EEG features: time domain, frequency domain, and nonlinear features, are used. When analyzing the time domain characteristics, the mode of the eigenvalue of the correlation coefficient matrix of EEG signal will change before and after the seizure, which shows that the time domain correlation coefficient matrix and its eigenvalue can predict the seizure

and termination of epilepsy to a certain extent. The amplitude and phase in the frequency domain are effective features, which can directly reflect the difference between the seizure period and seizure interval. Entropy can describe the uncertainty of information source and plays an important role in nonstationary EEG signals. MMLE³ obtains the best classification performance and has the highest generalization ability, which is 2.43%, 2.52%, and 2.44% higher than baseline algorithm EquiDML in specificity, sensitivity, and accuracy. It can be seen that the comprehensive use of multi-feature information can promote the accuracy of epilepsy recognition.

In addition, the MMLE³ algorithm uses the constraint forms of must link and cannot link to project the samples into a low-dimensional space, in which the distance between the samples constrained by cannot link is maximized, and the samples of different classes are transformed into equidistant; meanwhile, the distance between the samples constrained by must link is minimized, and the samples of different classes are compressed to a point. We introduce the feature weight vector to adaptively weigh various features and effectively adjust the weight relationship between various features in the process of metric matrix learning. On the premise that all kinds of feature classification tasks are consistent, the MMLE³ algorithm can effectively mine the association and complementary information hidden among features and highlight the role of optimal features. The MMLE³ algorithm has stronger discrimination ability. Therefore, the results in Tables 1–3 indicate that various forms of EEG features can be treated differently in the MMLE³ algorithm and the similarity between EEG data samples can be measured more accurately. The MMLE³ algorithm shows superiority and effectiveness for EEG recognition of epilepsy.

5. Conclusion

In clinical research, EEG is a basic tool for diagnosing and studying brain diseases, especially in the field of epilepsy diagnosis. This study explores how to improve the classification accuracy of epileptic EEG based on various feature extraction methods and metric learning algorithm. We propose the MMLE³ algorithm for EEG recognition of epilepsy. In the process of metric matrix learning, MMLE³ uses various forms of EEG features to effectively adjust the weight relationship between various features. Experiments show that the classification performance of comprehensive utilization of multiple features is significantly better than single feature, and multifeature metric learning has better stability and generalization ability. In the future, we will embed the proposed algorithm into the deep network for new latent representations. We will apply the proposed algorithm to clinical diagnosis in the next stage. In addition, with the development of computer-aided technology, visualized operating systems are one of the development trends of future medical care. We will also try to design the MMLE³ algorithm into a visual operating system to facilitate the application of clinical diagnosis.

Data Availability

Copies of the used data can be obtained free of charge from <https://physionet.org/content/chbmit/1.0.0/>.

Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Science and Technology Project of Changzhou City under Grant No. CE20215032, the Natural Science Foundation of Jiangsu Province under Grant BK 20211333, and the Future Network Scientific Research Fund Project under Grant No. FNSRFP-2021-YB-36.

References

- [1] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: a survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2018.
- [2] T. G. Ni, Q. X. Gu, and C. Zhang, "An intelligence EEG signal recognition method via noise insensitive TSK fuzzy system based on interclass competitive learning," *Frontiers in Neuroscience*, vol. 14, article 8372020, 2020.
- [3] Q. X. Gu, C. Zhang, and T. G. Ni, "A hierarchical discriminative sparse representation classifier for EEG signal detection," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 5, pp. 1679–1687, 2021.
- [4] Y. Z. Jiang, Q. X. Gu, D. C. Ji et al., "Smart diagnosis: a multiple-source transfer TSK fuzzy system for EEG seizure identification," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 25, Article ID 59, 2020.
- [5] Y. Kaya, M. Uyar, R. Tekin, and S. Yıldırım, "1D-local binary pattern based feature extraction for classification of epileptic EEG signals," *Applied Mathematics and Computation*, vol. 243, no. 15, pp. 209–219, 2014.
- [6] L. Frassinetti, C. Barba, F. Melani, F. Piras, R. Guerrini, and C. Manfredi, "Automatic detection and sonification of nonmotor generalized onset epileptic seizures: preliminary results," *Brain Research*, vol. 1721, no. 10, article 146341, 2019.
- [7] G. Chandel, P. Upadhyaya, O. Farooq, and Y. U. Khan, "Detection of seizure event and its onset/offset using orthonormal _triadic_ wavelet based features," *IRBM*, vol. 40, no. 2, pp. 103–112, 2019.
- [8] J. Xiang, C. Li, H. Li et al., "The detection of epileptic seizure signals based on fuzzy entropy," *Journal of Neuroscience Methods*, vol. 243, no. 3, pp. 18–25, 2015.
- [9] R. Hussein, M. Elgendi, Z. J. Wang, and R. K. Ward, "Robust detection of epileptic seizures based on L1-penalized robust regression of EEG signals," *Expert Systems with Applications*, vol. 104, no. 8, pp. 153–167, 2018.
- [10] V. Bajaj and R. B. Pachori, "Classification of seizure and non-seizure EEG signals using empirical mode decomposition," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1135–1142, 2012.

- [11] M. Kaleem, A. Guergachi, and S. Krishnan, "EEG seizure detection and epilepsy diagnosis using a novel variation of empirical mode decomposition," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4314–4317, Osaka, Japan, 2013.
- [12] S. M. Usman, M. Usman, and S. Fong, "Epileptic seizures prediction using machine learning methods," *Computational and Mathematical Methods in Medicine*, vol. 2017, Article ID 9074759, 10 pages, 2017.
- [13] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, no. 2, pp. 207–244, 2009.
- [14] Q. Liu, X. Zhao, and Z. Hou, "Metric learning for event-related potential component classification in EEG signals," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, pp. 2005–2009, Lisbon, Portugal, 2014.
- [15] H. Phan, Q. Do, T. L. Do, and D. L. Vu, "Metric learning for automatic sleep stage classification," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5025–5028, Osaka, Japan, 2013.
- [16] H. Alwasiti, M. Z. Yusoff, and K. Raza, "Motor imagery classification for brain computer interface using deep metric learning," *IEEE Access*, vol. 8, pp. 109949–109963, 2020.
- [17] T. G. Ni, C. Zhang, and X. Q. Gu, "Transfer model collaborating metric learning and dictionary learning for cross-domain facial expression recognition," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 5, pp. 1213–1222, 2021.
- [18] J. Xue, X. Gu, and T. Ni, "Auto-weighted multi-view discriminative metric learning method with Fisher discriminative and global structure constraints for epilepsy EEG signal classification," *Frontiers in Neuroscience*, vol. 14, article 586149, 2020.
- [19] J. Wang, Z. Wang, C. Liang, C. Gao, and N. Sang, "Equidistance constrained metric learning for person re-identification," *Pattern Recognition*, vol. 74, no. 2, pp. 38–51, 2018.
- [20] A. L. Goldberger, L. Amaral, L. Glass et al., "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [21] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information theoretic metric learning," in *Proceeding of the 24th International Conference of Machine Learning*, pp. 209–216, New York, NY, USA, 2007.
- [22] T. Ni, Z. Ding, F. Chen, and H. Wang, "Relative distance metric learning based on clustering centralization and projection vectors learning for person re-identification," *IEEE Access*, vol. 6, pp. 11405–11411, 2018.
- [23] J. Hu, J. Lu, J. Yuan, and Y. P. Tan, "Collaborative multi-view metric learning for visual classification," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*, Seattle, WA, USA, 2016.
- [24] C. Huang, F. Chung, and S. Wang, "Multi-view L2-SVM and its multi-view core vector machine," *Neural Networks*, vol. 75, no. 3, pp. 110–125, 2016.