

Research Article

A Novel Transfer Enhanced α -Expansion Move Learning Model for EEG Signals

Jiangwei Cai , Lu Zhao , and Anqi Bi 

Department of Computer Science and Engineering, Changshu Institute of Technology, Changshu, Jiangsu, China

Correspondence should be addressed to Anqi Bi; anqi_b@cslg.edu.cn

Received 16 March 2021; Revised 9 April 2021; Accepted 16 April 2021; Published 24 April 2021

Academic Editor: Chenxi Huang

Copyright © 2021 Jiangwei Cai 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.

In this paper, we focus on recognizing epileptic seizure from scant EEG signals and propose a novel transfer enhanced α -expansion move (TrEEM) learning model. This framework implants transfer learning into the exemplar-based clustering model to improve the utilization rate of EEG signals. Starting from Bayesian probability theory, by leveraging Kullback-Leibler distance, we measure the similarity relationship between source and target data. Furthermore, we embed this relationship into the calculation of similarity matrix involved in the exemplar-based clustering model. Then we sum up a new objective function and study this new TrEEM scheme earnestly. We optimize the proposed TrEEM model by borrowing the mechanism utilized in EEM. In contrast to other machine learning models, experiments based on synthetic and real-world EEG datasets show that the performance of the proposed TrEEM is very promising.

1. Introduction

Epilepsy is a kind of chronic disease, which is caused by the sudden abnormal discharge of brain neurons, resulting in transient brain dysfunction. Usually patients themselves have no obvious impression of the epileptic seizure process. For this reason, doctors can only diagnose the patient's condition according to the patient's family members or other personnel present during the epileptic seizure in the past. However, the accuracy of this manual diagnosis method is low. The pathogenesis of epilepsy is mainly manifested by abnormal neural discharge and abnormal brain waves. Although medical imaging, such as Computed Tomography (CT), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), Single-Photon Emission Computed Tomography (SPECT), Positron Emission Computed Tomography (PET), has made great progress over the years, and the major diagnostic method of epilepsy is based on electroencephalogram (EEG). More specifically, PET and fMRI cannot be used as common technical means because of their technical requirements and costs. In addition to the high cost, MRI cannot judge the nonstructural lesions as well. Invasive cortical electroencephalogram (ECoG) requires

craniotomy and implantation of electrodes, which has a high risk; and noninvasive EEG and MEG can provide functional and structural detection. Taking all these into account, EEG has been widely concerned in more and more theoretical researches and clinical practice because of its low cost, convenient signal acquisition, and noninvasiveness.

The research on diagnosis of epilepsy through EEG signals has been a hot topic in related fields, compared with manual diagnostic method, and machine learning methods are less time-consuming and more accurate [1–8]. Numerous machine learning models have been used to recognize epileptic EEG signals, such as support vector learning [9, 10], fuzzy system [1, 3], naïve Bayes [11], and exemplar-based clustering model [2, 12, 13]. The traditional machine learning process is usually divided into the three following steps, as shown in Figure 1: (1) EEG signal preprocessing improves signal to noise ratio and provides high-quality input signal for spike detection. (2) According to the characteristics of spikes, artificial design features can reduce the signal dimension and highlight the difference between spikes and background signals. (3) According to the obtained features, spike signals are detected by the machine learning mechanism involved.

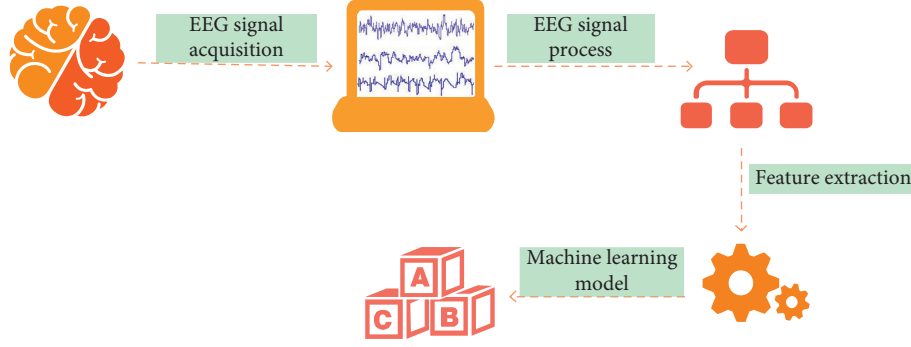


FIGURE 1: General steps of machine learning model processing EEG signals.

In summary, one of the significant issues in the field of processing EEG signals by machine learning technique is the insufficient training data. We briefly introduce some mechanisms for epileptic diagnosis through EEG signals here. Jiang [1] integrates transductive transfer learning, semisupervised learning, and Takagi-Sugeno-Kang (TSK) fuzzy system to take full advantage of the scant training data. Zhu [5] proposed dic-mv-fcm, which automatically evaluates the importance and weights of each view and then performs weighted multiview fuzzy clustering based on FCM framework to achieve accurate fuzzy partition. Bi [2] proposed a novel model called FEEM for incomplete EEG signal, which first compresses the potential exemplar list and thus reduces the scale of pairwise similarity matrix. However, to make better use of training data, we still need to do a lot of work, and we focus on this issue in this paper as well. Specifically, this paper aims at recognizing epileptic seizure from scant EEG signals.

Transfer learning is believed to be an effective strategy to solve problems caused by insufficient training data [1, 5, 13, 14]. Assume that there are two datasets from similar source: one has plenty of features and details and is easily to be learned, while the other one lacks details and is hard to be learned. Transfer learning offers an idea of leveraging the description of former data to study the latter data. The sufficient well-described data is called source data, while the insufficient rough data is named as target data. Accordingly, transfer learning utilizes source data to improve the learning result of target data. Under this framework, effectively measuring the relationship between source data and target data is an important part and has a great influence on the efficiency of relevant study model. Thus, starting from Bayesian probability theory, this paper first extends the concept of similarity matrix in the exemplar-based clustering model; and this strategy also broadens the application range of the algorithm to transfer learning scenario. By leveraging Kullback-Leibler distance, we propose a new transfer enhanced α -expansion move learning model called TrEEM. The detailed contributions of this paper are listed as follows:

- (i) According to the transfer learning theory [1, 5, 13, 14], considering the similarity between source and target data, the proposed model TrEEM should keep the target data close enough to the

source data. Theoretically supported by the information theory, based on the Bayesian probability framework, TrEEM utilizes KL distance to measure the similarity between source data and target data and minimizes this KL distance in the optimization process.

- (ii) In the scenario of recognizing epileptic seizure, we aim at diagnosing the actual patient. As TrEEM is built on graph theory and pairwise similarity matrix and is an exemplar-based clustering model, this model selects exemplar from actual data. This advantage fits the requirements in the relevant scenario here.
- (iii) TrEEM embeds KL distance between target data and source data into the calculation of similarity matrix. Thus, the optimization mechanism utilized in EEM can be directly used to solve the new target function of TrEEM. In detail, we leverage α -expansion move optimization algorithm which performs better than LBP [15, 16] algorithm.

The paper is organized as follows. The related works are discussed in Section 2. We illustrate the target function and optimization mechanism of the proposed TrEEM in Section 3. The simulation experimental results and analysis are shown in Section 4. We make a conclusion in Section 5.

2. Related Works

Many researchers are committed to using machine learning technology to classify EEG signals, including SVM, fuzzy system, naïve Bayes, and exemplar-based clustering model. In this section, we illustrate two popular learning frameworks, namely, Enhanced α -Expansion Move (EEM) and TSK fuzzy system. EEM is a widely used exemplar-based learning model, and TSK fuzzy system is a typical fuzzy-rule-based clustering model.

2.1. Enhanced α -Expansion Move. Consider a dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times D}$; N is the total number of D -dimensional data points. E is the output, whereas the element $E(\mathbf{x}_i)$ refers to the exemplar for each \mathbf{x}_i .

The target function of a typical exemplar-based clustering model is defined as follows [12, 15]:

$$\max_E \sum_{p=1}^N s(\mathbf{x}_p, \mathbf{x}_{E(p)}) - \sum_{p=1}^N \sum_{q>p}^N \theta_{p,q}(E(p), E(q)), \quad (1)$$

where S is the similarity matrix of the dataset, and the elements are defined as $S(i, j) = -\|\mathbf{x}_i - \mathbf{x}_j\|^2$; $\theta_{p,q}(E(p), E(q))$ is shown as follows:

$$\theta_{p,q}(E(p), E(q)) = \begin{cases} M, & E(p) = q, \quad E(q) \neq p, \\ \text{or } E(q) = p, & E(p) \neq q, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

In [15], the authors regard the above target function as the energy function of Markov random field (MRF) and verifies the below theorem.

Theorem 1. When, for $\forall p, q, E(p), E(q), \alpha \in \{1, 2, \dots, N\}$, equation (3) is verified, the graph-theory based framework can be used to optimize the target function of the exemplar-based clustering model as shown in equation (1).

$$\theta_{p,q}(E(p), E(q)) + \theta_{p,q}(\alpha, \alpha) \leq \theta_{p,q}(E(p), \alpha) + \theta_{p,q}(\alpha, E(q)). \quad (3)$$

Enhanced α -Expansion Move (EEM) framework optimizes the above target function by an improved algorithm [15]. In more detail, theoretically supported by Theorem 1 and graph-cuts [16] algorithm, EEM expands the active region of candidate exemplar from a single data to the whole dataset. EEM defines the second optimal candidate exemplar $S(i)$ for \mathbf{x}_i as below, which is selected from the whole dataset as mentioned above.

$$S(i) = \arg \max_{s \in (E/I)} s(\mathbf{x}_i, \mathbf{x}_s), \quad \forall \mathbf{x}_i \in \mathbf{X}_I, \quad (4)$$

where $\mathbf{X}_I = \{\mathbf{x}_i | E(i) = I\}$ is the dataset among which the exemplar is I , and $s \in (E/I)$ represent other exemplars in E expect for I . In this way, the optimization mechanism behaves more rapidly and effectively.

EEM algorithm is one of the most popular exemplar-based clustering models, and it performs effectively and steadily in numerous simulation experiments involved [2, 12–15]. Scientists have applied this model for data stream, constrained supervised learning, and EEG signal processing.

2.2. TSK Fuzzy System. TSK fuzzy system is a rule-based system and it is widely used as a typical fuzzy system model for both classification and clustering. Generally, the k th TSK fuzzy rule for K fuzzy rules can be described as R^k .

$$\begin{aligned} \text{IF } \mathbf{x}_1 \text{ is } A_1^k \wedge \mathbf{x}_2 \text{ is } A_2^k \wedge \dots \wedge \mathbf{x}_N \text{ is } A_N^k, \\ \text{THEN } f^k(\mathbf{x}) = p_0^k + p_1^k \mathbf{x}_1 + \dots + p_N^k \mathbf{x}_N, \end{aligned} \quad (5)$$

where A_i^k is a fuzzy set subscribed by the input \mathbf{x}_i for the k th fuzzy rule and \wedge is a fuzzy conjunction operator. Each rule is premised on the input data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times D}$ which is mapped to a singleton $f^k(\mathbf{x})$. Thus, the output of the TSK fuzzy system is defined as

$$y^0 = \sum_{k=1}^K \frac{\mu^k(\mathbf{x}) f^k(\mathbf{x})}{\sum_{m=1}^K \mu^m(\mathbf{x})} = \sum_{k=1}^K \tilde{\mu}^k(\mathbf{x}) f^k(\mathbf{x}), \quad (6)$$

where

$$\begin{aligned} \mu^k(\mathbf{X}) &= \prod_{i=1}^N \mu_{A_i^k}(\mathbf{x}_i), \\ \tilde{\mu}^k(\mathbf{X}) &= \frac{\mu^k(\mathbf{X})}{\sum_{m=1}^K \mu^m(\mathbf{X})}, \end{aligned} \quad (7)$$

where $\mu_{A_i^k}(\mathbf{x}_i)$ is the membership grade that can be obtained using Gaussian membership function, and the other involved parameters also could be estimated using clustering techniques and other partition methods [1, 3–5].

Accordingly, based on the relevant theory of TSK fuzzy system, the target model above in equation (6) converts to a parameter learning process of the corresponding linear regression model. In line with recent achievements, TSK fuzzy model has strong interpretability and robustness. For this reason, this TSK fuzzy model is widely used among numerous intelligent medical diagnosis systems, including recognizing epileptic seizure from EEG signals.

In this section, we briefly introduce two popular machine learning clustering frameworks used in the recognition of EEG signals, namely, EEM and TSK fuzzy system. The detailed descriptions are shown in Table 1. Considering the scenario of diagnosing epileptic patients from some healthy patients based on their EEG signals, we focus on EEM clustering model in the rest of this paper.

3. Transfer Enhanced α -Expansion Move Learning Model

In this section, we first analyze the theoretical basis of TrEEM from Bayesian probabilistic framework. Second, we induce the novel algorithm TrEEM in detail. Then, considering the optimization algorithm utilized in EEM algorithm, we optimize target function as well. Generally, the structure of this novel model is shown in Figure 2. See Figure 2; on the basis of source-data-based exemplar set, starting from Bayesian probability framework, TrEEM first imbeds the distance between source data and target data in the calculation of similarity matrix. This distance is measured by Kullback-Leibler distance. Then we induce the novel target function for TrEEM. Finally, TrEEM directly calls the optimization algorithm in EEM to solve this model and obtain the target-data-based exemplar set.

Besides, we list the frequently used notations in Table 2.

3.1. Theoretical Preliminary of TrEEM Scheme. As mentioned before, transfer learning considers two datasets from similar source, namely, source data and target data; and the relationship between source data and target data is considered as a significant factor in this model (see Table 2); in the following part, we define the sufficient well-described source data as $\tilde{\mathbf{X}}$. After study, we obtain the source-data-based exemplar set denoted as L_s in the above table. Then the insufficient target

TABLE 1: Descriptions of two popular machine learning algorithms used in recognition of EEG signals.

Algorithms	Theoretical basis	Descriptions	Optimization frameworks
EEM	Graph theory	Select exemplar from actual data, do not need to preset the cluster number	Enhanced graph-cuts optimization algorithm, expand the candidate region
TSK	Fuzzy system	Rule-based learning model, strong interpretability and robustness	Parameter learning process of corresponding linear regression model

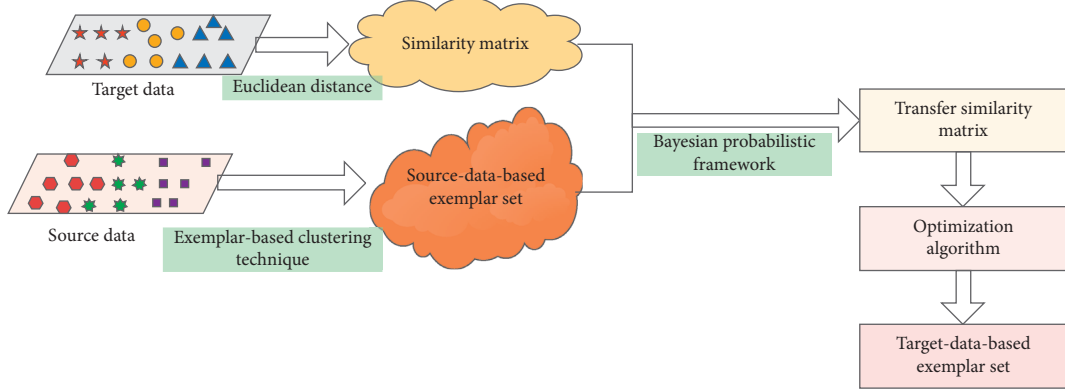


FIGURE 2: Structure of TrEEM algorithm.

TABLE 2: Involved notations and descriptions.

Notations	Descriptions
$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times D}$	Target data
$\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_{N_s}\} \in \mathbb{R}^{N_s \times D}$	Source data
$\mathbf{S} = (s(\mathbf{x}_i, \mathbf{x}_j))$	Pairwise similarity matrix
L_s	Source-data-based exemplar set
$L_s(i)$	Source exemplar for target sample x_i
E	Target-data-based exemplar set
$E(i)$	Exemplar for target sample x_i

data is defined as \mathbf{X} above. Moreover, probabilistic framework contributes to measuring this relationship as well. Therefore, supported by Gaussian probability hypothesis and exemplar-based cluster mechanism, we built the pairwise probabilistic relationship of target data by leveraging the corresponding similarity as follows:

$$p(x_i, x_{E(i)}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{s(x_i, x_{E(i)})}{2\sigma^2}\right), \quad (8)$$

where $s(x_i, x_{E(i)})$ is the similarity between x_i and its current exemplar $x_{E(i)}$ and parameter σ is a standard deviation from Gaussian probability hypothesis.

As to the exemplar set, we should exclude the situation when an exemplar appoints other exemplars among current exemplar set except for itself as its own exemplar. Consequently, Bayesian posterior probability of an exemplar set is defined as follows:

$$p(E) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\sum_{m=1}^N \sum_{n=1}^N \frac{\theta_{m,n}(E(m), E(n))}{2\sigma^2}\right), \quad (9)$$

and $\theta_{m,n}(E(m), E(n))$ here is the same as the definition shown in equation (2).

Accordingly, under Bayesian probabilistic framework and the discussion of EEM algorithm in Section 2, the objective function in equation (1) is equal to the following function:

$$\begin{aligned} \max_E Q(E) &= \prod_{i=1}^N \ln p(x_i, x_{E(i)}) \ln p(E) \\ &= \sum_{i=1}^N \ln p(x_i, x_{E(i)}) + \ln p(E). \end{aligned} \quad (10)$$

In conclusion, equation (10) defines another form of the target function of EEM by introducing Bayesian probabilistic framework and Gaussian probability hypothesis. Starting from this target function, we would be able to design TrEEM for recognition of epileptic EEG signals in the next subsections.

3.2. Design of TrEEM Scheme. According to information theory, the Kullback-Leibler distance (KL distance) is a natural distance between two real probability distributions and it has been widely applied to solve numerous issues [17–19]. The definition of KL distance is shown below.

Definition 1. Consider two probability distributions as P and Q ; the KL distance from P to Q is as follows:

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) \ln \frac{P(x)}{Q(x)}, \quad (11)$$

where $\mathbf{X} = \{x_1, x_2, \dots, x_N\}$ is the input data.

What is worth mentioning is the fact that KL distance is an asymmetric measurement, namely, $D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$, according to Definition 1.

Furthermore, given L_s as a possible exemplar set, $L_s(i)$ is the best exemplar for x_i among current exemplar set L_s . As discussed above, we also define $L_s(x_i)$ under Bayesian probabilistic framework as follows:

$$L_s(i) = \arg \max_{l \in L_s} p(x_i, \tilde{x}_l), \quad (12)$$

where $p(x_i, \tilde{x}_l)$ is obtained from equation (8). In transfer learning, actually two datasets are involved, that is, source data and target data. In equation (12), note that the first x_i is from the target data, and the second \tilde{x}_l is from the possible exemplar set, namely, from the source data. Thus, see Table 2; to make the distinction clear, the symbol \tilde{x}_i represents the source data, while x_i stands for the target data in the rest of this paper.

Although the target data is not exactly same as source data, according to those theoretical analyses of transfer learning, the source-data-based learning model and results should contribute to the learning of new target data as well [3, 4, 20–22]. Otherwise, it will become negative transfer

learning, which is not under discussion in this paper. Accordingly, we believe the target-data-based exemplar that is set to be evaluated is supposed to be similar to the source-data-based exemplar set. In this section, we measure the difference between target-data-based exemplar set and source-data-based exemplar set by the aforementioned KL distance with the definition shown in Definition 1. To be specific, in the process of designing the TrEEM learning model, we minimize the difference of target exemplar E and source exemplar set L_s by controlling the KL distance between them. The structure of TrEEM is shown in Figure 2 in detail. In view of this goal, on the basis of the probabilistic target function in equation (10) of EEM, we build the novel target function for the proposed TrEEM model as follows:

$$\max Q(E) = \frac{\prod_{i=1}^N \ln p(x_i, x_{E(i)}) \ln p(E)}{\lambda D_{KL}(L_s E)}, \quad (13)$$

where E is the target-data-based exemplar set to be obtained and L_s represents the source-data-based exemplar set, as shown in Table 2. λ is the regularization parameter. In terms of maximum a priori (MAP) principle and combining Definition 1 and equation (12), (13) becomes

$$\max Q(E) = \sum_{i=1}^N \ln p(x_i, x_{E(i)}) + \ln p(E) - \lambda \sum_{i=1}^N p(x_i, \tilde{x}_{L_s(i)}) [\ln p(x_i, \tilde{x}_{L_s(i)}) - \ln p(x_i, x_{E(i)})]. \quad (14)$$

Observing equation (14), we can find that the values of the second and third terms belong to the same magnitudes; hence, the value of λ will not be large and the specific determination strategy will be discussed in Section 4.

Introducing the definitions of $p(x_i, x_{E(i)})$, $p(x_i, \tilde{x}_{L_s(i)})$, and $p(E)$ in equations (8) and (9) and discarding the constant terms, equation (14) can be simplified into the following equation:

$$\max Q(E) = \ln p(E) + \sum_{i=1}^N [1 + \lambda p(x_i, \tilde{x}_{L_s(i)})] \ln p(x_i, x_{E(i)}). \quad (15)$$

Comparing equations (15) and (10), we conclude that they are similar in structure. According to Theorem 1 in Section 2, TrEEM's target function also can be solved by graph-cuts mechanism. Consequently, we will discuss the optimization mechanism step by step in the next subsection.

3.3. Optimization Mechanism of the TrEEM Scheme. As mentioned before, the novel target function in equation (15) is similar to that of EEM algorithm under Bayesian probabilistic framework, so the optimization mechanism utilized in the EEM algorithm is supposed to be helpful in solving the novel target function. However, we need to deal with the difference between these two models firstly.

In detail, we redefine the similarity relationship of target data by imbedding source-data-based exemplar set L_s .

Specifically, we single out the suitable exemplar from L_s for target sample x_i by equation (12) and build the new pairwise transfer similarity matrix $S_t = (s_t(x_i, x_j))$ according to the new measurement in the following equation:

$$s_t(x_i, x_j) = -\left[1 + \lambda p(x_i, \tilde{x}_{L_s(x_i)})\right] d(x_i, x_j), \quad (16)$$

where $d(x_i, x_j) = \|x_i - x_j\|^2$ is the Euclidean distance between samples x_i and x_j , λ is the regularization parameter, $\tilde{x}_{L_s(x_i)}$ refers to the exemplar singled out from source data. By introducing this new definition of similarity relationship, the target function equation (15) of TrEEM is equal to equation (10) in structure. Meanwhile, the constraint condition in Theorem 1 is true for TrEEM model as well. Therefore, the optimization mechanism of EEM algorithm is also suitable for the proposed TrEEM model; and the novel model TrEEM is described in detail in Algorithm 1.

EEM utilizes α -expansion move to optimize its learning model. As discussed above, the mechanism is also suitable for the proposed TrEEM model. We analyze this Enhanced α -Expansion Move optimization mechanism step by step here. Firstly, as the target functions shown in both equations (15) and (10) also can be defined as the energy function of MRF, we consider this optimization process as an energy reduction process of the MRF. In general, we start from the change values of energy to decide whether to accept new exemplar for a sample. Secondly, the improved optimization mechanism is designed to broaden the effective field when changing the sample's exemplar. That is to say, assume that a

Input: Target dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times D}$, source data $\tilde{\mathbf{X}} = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_{N_s}\} \in \mathbb{R}^{N_s \times D}$, source-data-based exemplar set L_s , self similarity $d(\mathbf{x}_i, \mathbf{x}_j)$, regularization factor λ, σ .

Output: Valid target-data-based exemplar set $E(N)$.

- (1) **for** $\mathbf{x}_i \in \mathbf{X}$ **do**
- (2) single out the nearest exemplar $L_s(i)$ for \mathbf{x}_i from source-data-based exemplar set L_s based on equation (12).
- (3) compute probabilistic Euclidean similarity $p(\mathbf{x}_i, \tilde{\mathbf{x}}_{L_s(i)})$ between \mathbf{x}_i and $L_s(i)$.
- (4) **end**
- (5) **for** $\mathbf{x}_i \in \mathbf{X}$ **do**
- (6) calculate transfer similarity matrix $\mathbf{S}_t = (s_t(\mathbf{x}_i, \mathbf{x}_j))$ by new probabilistic similarity $p(\mathbf{x}_i, \tilde{\mathbf{x}}_{L_s(i)})$ according to equation (16).
- (7) **end**
- (8) call the optimization process of EEM as shown in Algorithm 2.

ALGORITHM 1: Transfer Enhanced α -Expansion Move learning model.

Input: similarity matrix \mathbf{S} , dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times D}$, maximum number of iteration M

Output: valid exemplar set $E(N)$.

- (1) Randomly generate expansion order o .
- (2) Let $t = 1$;
- (3) **for** $e \in o$ **do**
- (4) **if** $e \in E$ **then**
- (5) compute R_{in}, R_{in}^l by equations (18), (19)
- (6) **if** $R_{in}^l > 0$ **then**
- (7) for $\forall \mathbf{x}_i \in \mathbf{X}_e$, set $E(\mathbf{x}_i) = A(i)$
- (8) **end**
- (9) **else**
- (10) compute $R_{out}, R_{out}^e, R_{out}^l$ by equations (20), (21), (22)
- (11) **if** $R_{out}^l > R_{out}^e$ **then**
- (12) for $\forall \mathbf{x}_i \in \mathbf{X}_{e/l}^e$, set $E(\mathbf{x}_i) = l$
- (13) **else**
- (14) for $\forall \mathbf{x}_i \in \mathbf{X}_e$, set $E(\mathbf{x}_i) = A(i)$
- (15) **end**
- (16) **if** $R_{out} > 0$ **then**
- (17) Accept the new exemplar l
- (18) **end**
- (19) **end**
- (20) $t = t + 1$
- (21) **end**
- (22) Until convergence

ALGORITHM 2: Optimization model of Enhanced α -Expansion Move algorithm.

sample's current exemplar is abandoned; it will search all the rest exemplars for a new exemplar. This new alternative exemplar is defined as follows:

$$A(i) = \arg \max_{a \in (E/l)} s(\mathbf{x}_i, \mathbf{x}_a), \quad \forall \mathbf{x}_i \in \mathbf{X}_l, \quad (17)$$

where l is the original exemplar for \mathbf{x}_i , E is current exemplar set, and $a \in (E/l)$ is the obtained alternative exemplar. By introducing this alternative exemplar $A(i)$ for \mathbf{x}_i , we enhance the optimization efficiency.

Note that TrEEM model redefines the similarity matrix as equation (16). So, the following discussion is based on the similarity matrix $\mathbf{S}_t = (s_t(\mathbf{x}_i, \mathbf{x}_j))$. Apparently, the optimization mechanism would consider two cases; namely, \mathbf{x}_i is among current exemplar set or is not among

current exemplar set. We analyze these two cases step by step in the next subsections.

3.3.1. Case I. \mathbf{x}_i is a current exemplar.

Obviously, in the process of optimization, this current exemplar \mathbf{x}_i may be abandoned. As previously analyzed, whether to keep \mathbf{x}_i in the ultimate exemplar set is decided by the reduction values of energy function calculated by the target function in equation (15).

Specifically, if \mathbf{x}_i is accepted as an exemplar, the energy of the model remains unchanged, and the reduction value is equal to 0. Otherwise, if \mathbf{x}_i is not accepted, all samples whose exemplars are l would redetermine their exemplars; these samples are defined as $\mathbf{X}_l = \{\mathbf{x}_i | E(i) = l\}$. Theoretically supported by the related analysis in [2, 12, 14, 15], new exemplar for $\mathbf{x}_i \in \mathbf{X}_l$

would be $A(i)$ as shown in equation (17). Thus, the energy reduction R_{in}^l should be computed by the following equation:

$$R_{in}^l = \sum_{\mathbf{x}_i \in X_i} (s(\mathbf{x}_i, \mathbf{x}_{A(i)}) - s(\mathbf{x}_i, \mathbf{x}_l)). \quad (18)$$

Then, we take the greater value of 0 and R_{in}^l as the ultimate energy reduction for this case, as defined in the following equation:

$$R_{in} = \max\{0, R_{in}^l\}. \quad (19)$$

Namely, if R_{in}^l is the ultimate energy reduction, $\mathbf{x}_i \in X_i$ change their exemplars to $A(i)$. Otherwise, the current exemplar set is convincing and remains unchanged.

3.3.2. Case II. \mathbf{x}_l is not a current exemplar.

In this case, we define the current exemplar of \mathbf{x}_l as \mathbf{x}_e . When optimizing this situation, we firstly pretend to consider \mathbf{x}_l as a new alternative exemplar; namely, $E' = E, E'(\mathbf{x}_l) = l$. Then, similar to the analysis in case I, whether to accept \mathbf{x}_l as ultimate exemplar is decided by the reduction values of energy function. In detail, if \mathbf{x}_l is accepted as a feasible exemplar, some samples would change their exemplar from \mathbf{x}_e to \mathbf{x}_l . These samples are defined as $X_e^{l,e,l} = \{\mathbf{x}_i | s(\mathbf{x}_i, \mathbf{x}_l) > s(\mathbf{x}_i, \mathbf{x}_e)\}$. Thus, the corresponding energy reduction is defined as follows:

$$R_{out}^l = \sum_{\mathbf{x}_i \in X_e^{l,e,l}} (s(\mathbf{x}_i, \mathbf{x}_l) - s(\mathbf{x}_i, \mathbf{x}_e)). \quad (20)$$

On the other hand, may be current exemplar set E' is not convincing, so all samples would be certain to redetermine their exemplars including \mathbf{x}_l . As discussed before, the new exemplars for these samples are defined by equation (17), and the resulting energy reduction is listed as follows:

$$R_{out}^e = \sum_{\mathbf{x}_i \in X_e} (s(\mathbf{x}_i, \mathbf{x}_{A(i)}) - s(\mathbf{x}_i, \mathbf{x}_e)). \quad (21)$$

Remember that equations (20) and (21) are based on the assumption that $E' = E, E'(\mathbf{x}_l) = l$. Considering this, the energy reduction caused by \mathbf{x}_l which is not a current exemplar should be

$$R_{out} = s(\mathbf{x}_l, \mathbf{x}_l) - s(\mathbf{x}_l, \mathbf{x}_e) + \max_{l \in E} \{R_{out}^e, R_{out}^l\}. \quad (22)$$

To sum up, the optimization mechanism is shown below in detail.

3.4. Model Complexity. The similarity matrix is calculated according to the Euclidean distance; $s(\mathbf{x}_i, \mathbf{x}_j) = -\|\mathbf{x}_i - \mathbf{x}_j\|^2$. So, the scale of the similarity matrix is N^2 ; note that the amount of target data is not big. In the optimization process, we directly utilize the α -expansion move, which has $O(N^2)$ time complexity. For the proposed TrEEM, source-data-based exemplar set is actually one of the inputs and is out of the scope of the time complexity analysis of TrEEM here. Although we adopt EEM to obtain the source-data-based exemplar set $L(s)$, many other clustering models could be

helpful. TrEEM needs to select $L_s(i)$ from the source-data-based exemplar set in the first step, and this procedure has the time complexity of $O(N)$. In summary, the time complexity of TrEEM is $O(N^2)$ overall. Compared with other state-of-the-art transfer learning frameworks, it is very acceptable.

4. Experimental Results

To comprehensively evaluate the TrEEM model, we have conducted several experiments based on both synthetic and real-world datasets. For comparison, we also perform comparison with other different machine learning mechanisms, namely, EEM [15], multiclass SVM [23], TSK fuzzy system [24], and TSC [25] in the experiments. In this section, we will carefully analyze these experimental results.

4.1. Preparation. Before inputting the TrEEM model, we need to preprocess the original nonstationary EEG signals [1–3]. Usually, the features of EEG signals include time-domain features, frequency-domain features, and time-frequency features. In short, in time-domain analysis, statistics component features of the original EEG signals will be analyzed [26]. In frequency-domain analysis, power spectrum analysis and Short-Time Fourier Transforms (STFT) [27, 28] are commonly used. In time-frequency analysis, time domain and frequency domain are simultaneously extracted from high-dimensional and nonlinear EEG signals.

Various methods have been commonly used to extract EEG signals' features, including wavelet [29, 30], KPCA (Kernel Principal Component Analysis) [1, 2], and LDA (Linear Discriminant Analysis). In line with the experiments setting in [1–3], we use two feature extraction methods in this section, that is, KPCA and wavelet.

Besides, we use both synthetic and real-world datasets in this section. Firstly, we randomly generate 300 two-dimensional data points as 3 classes, shown in Figure 3. Then, we also choose Bonn EEG dataset [1, 2] as real-world data. The Bonn dataset is from the University of Bonn, Germany (<http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eeegdata.html>), and has five classes. Each class (A to E) contains 100 signal channel EEG segments of 23.6 s duration. The sampling rate of all the datasets was 173.6 Hz. Each sample has 6 attributes. Table 3 lists a brief description of this dataset.

In addition, we examine the involved experimental results from two performance indices, namely, RandIndex (RI) [2, 31] and Purity. Assume that N is the total number of data points; we give the definitions of them below. That is, RI is shown in the following equation:

$$RI = \frac{f_{00} + f_{11}}{(N(N-1)/2)}, \quad (23)$$

where f_{00} is the amount of data whose cluster is in line with their class, while f_{11} is the amount of those data whose cluster is inconsistent with their class. Also Purity is defined in the following equation:

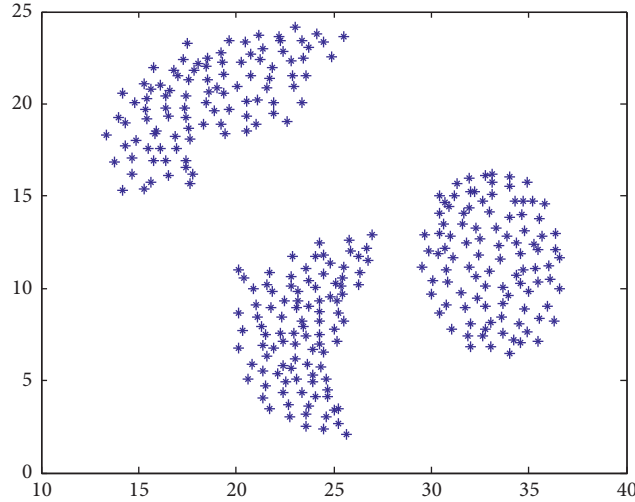


FIGURE 3: Synthetic dataset.

TABLE 3: Description of Bonn EEG dataset.

Subjects	Groups	Descriptions
Healthy	A	Signals captured from volunteers with eyes open
	B	Signals captured from volunteers with eyes closed
	C	Signals captured from volunteers during seizure silence intervals
Epileptic	D	Signals captured from volunteers during seizure silence intervals
	E	Signals captured from volunteers during seizure activity

$$\text{Purity}(E, C) = \frac{1}{N} \sum_k \max_j |e_k \cap c_j|, \quad (24)$$

where $E = \{e_1, \dots, e_N\}$ is the cluster result obtained by the learning model, while $C = \{c_1, \dots, c_N\}$ is the real data label set.

In all, the experiments are implemented in 2010a Matlab on a PC with 64-bit Microsoft Window 10, an Intel (R) Core (TM) i7-4712MQ, and 8 GB memory.

4.2. Results Analysis. As mentioned before, four machine learning methods are involved in this section, namely, EEM, multiclass SVM, TSK-FS, and the proposed TrEEM algorithm. There is no need to preset the cluster number in advance for EEM and TrEEM. In fact, it is a huge advantage for all exemplar-based clustering frameworks, whereas cluster number is an important parameter for TSK-FS. Multiclass SVM and TSC are two typical classification methods. Both EEM and TrEEM need parameter self-similarity $d(\mathbf{x}_i, \mathbf{x}_j)$. For multiclass SVM, in line with [23, 32], we choose Gaussian kernel function. In TSK-FS, usually the number of clusters is set to be equal to the number of fuzzy rules. Also, TSC need to preset the number of clusters. We follow the parameter setting strategy in relevant papers here. Besides, 5-fold cross validation is used to search the optimal parameters; and Table 4 lists brief introductions of these involved methods and the parameter searching range.

To construct the transfer learning scenario, for both synthetic and real-world EEG signal datasets, we randomly choose 80% data as source data and the remaining 20% as target data. For statistical analysis, in the experiment procedure, each algorithm is repeatedly executed 10 times; and we record the average performance and the corresponding standard deviation of RI and Purity. Furthermore, to deeply observe different extraction methods of EEG signals, we use both KPCA and wavelet here. The detailed comparison in terms of RI and Purity of the proposed TrEEM model and other benchmark approaches is shown in Table 5.

Observing Table 5, in this experimental setting, especially considering the fact that Bonn EEG signal dataset has 6 attributes and 5 classes, the performance of TrEEM model is very promising. TrEEM model is capable of recognizing useful information from both synthetic and real-world EEG signal datasets. Moreover, compared with other benchmark machine learning models, the proposed approach TrEEM performs better in terms of RI and Purity in this scenario.

In the experiment procedure, we also find that parameter self-similarity $d(\mathbf{x}_i, \mathbf{x}_j)$ has important influence on the experimental results, especially on the obtained number of clusters. The finding is identified with other exemplar-based clustering models [2, 12–15], and the parameter selection method is also in line with these models. See Table 4; we multiply $d(\mathbf{x}_i, \mathbf{x}_j)$ with η , and the value of η is decided from $\{0.01, 0.1, 1, 10, 50, 100\}$. In detail, with large value of η , the TrEEM model would obtain a smaller number of clusters,

TABLE 4: Parameters settings of involved algorithms.

Algorithms	Parameter setting
EEM: a typical exemplar-based clustering model	Self-similarity $d(\mathbf{x}_i, \mathbf{x}_j) = \text{median value of similarities} \times \eta \eta \in \{0.01, 0.1, 1, 10, 50, 100\}$.
Multiclass SVM: a typical classification learning model	Kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-(\mathbf{x}_i - \mathbf{x}_j ^2 / 2\sigma^2))$, where $\sigma \in \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5\}$. Penalty parameter $C \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$.
TSK-FS: a widely used fuzzy-rule-based learning model	FCM [1] involved, the cluster number equals the number of fuzzy rules. Number of fuzzy rules $K \in \{5, 10, 15, 20, 25, 30\}$. Gaussian membership function $\sigma \in \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5\}$.
TSC: transfer spectral clustering model	Preset the cluster number $n \in \{5, 10, 15, 20, 25, 30\}$; for other parameter setting strategies, see [25].
TrEEM: the proposed transfer exemplar-based learning model	Self-similarity $d(\mathbf{x}_i, \mathbf{x}_j) = \text{median value of similarities} \times \eta \eta \in \{0.01, 0.1, 1, 10, 50, 100\}$. Regularization factor $\lambda \in \{0.01, 0.05, 0.1, 0.5, 0.5, 10, 15, 20, 25, 30\}$. $\sigma \in \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5\}$.
TSC: transfer spectral clustering model	Preset the cluster number $n \in \{5, 10, 15, 20, 25, 30\}$; for other parameter setting strategies, see [25].

TABLE 5: Comparison results of both synthetic and Bonn EEG datasets (the number in parentheses is the standard deviation).

Datasets (source data, target data, attributes, and classes)	Algorithms	Performance indices	
		RI	Purity
Synthetic dataset (240, 60, 2, 3)	EEM	0.8316 (0.0258)	0.7150 (0.1131)
	Multiclass SVM	0.8513 (0.0214)	0.8523 (0.0812)
	TSK-FS	0.8712 (0.0145)	0.8816 (0.0914)
	TSC	0.88230.0313	0.92360.0158
	TrEEM	0.8957 (0.0264)	0.9856 (0.0000)
Bonn EEG dataset (400, 100, 6, 5) (use KPCA to extract feature)	EEM	0.7754 (0.2146)	0.9800 (0.0000)
	Multiclass SVM	0.7827 (0.1834)	0.9643 (0.0023)
	TSK-FS	0.6819 (0.1579)	0.9623 (0.0000)
	TSC	0.72170.1241	0.9636 (0.0002)
	TrEEM	0.8323 (0.1652)	0.9600 (0.0012)
Bonn EEG dataset (400, 100, 6, 5) (use wavelet to extract feature)	EEM	0.7925 (0.0091)	0.7530 (0.0514)
	Multiclass SVM	0.7815 (0.0165)	0.9034 (0.0135)
	TSK-FS	0.7303 (0.0251)	0.9800 (0.0000)
	TSC	0.75010.1252	0.9600 (0.0021)
	TrEEM	0.8071 (0.0078)	0.9800 (0.0000)

while small η will bring in big cluster numbers. To fit with real data labels, in our experiments here, we set $\eta = 1$.

The regularization factor λ has a big effect as well. As analyzed before, λ determines how the source data affects the clustering result, and the value should not be too large. Obviously, if λ is too large, the clustering result based on the target data will be very close to that based on the source data, which is not what we want. When $\lambda = 0$, it means that TrEEM does not take the source data into account and TrEEM degrades to the typical EEM framework. In particular, Figures 4–6 show the dependence of model results on the value of λ . When $\lambda > 0$, in terms of RI and Purity, source data improves the performance of TrEEM. Index Purity is more sensitive to λ , while RI changes slowly.

Table 6 shows the average running time of 10 times for each approach. Yet the time consumption of the proposed TrEEM model is a bit more than those of EEM and TSK-FS; it is still in the same magnitude. Considering the improvements in RI and Purity, we think that the time complexity is acceptable. The results also fit the discussion in Section 3.4.

Therefore, from experimental results in Tables 5 and 6 and the above analysis, we can conclude the following:

- (1) For both synthetic and real-world EEG signal datasets, TrEEM performs great. Thus, we believe that TrEEM can effectively absorb knowledge from scant target data when similar source data exists.
- (2) For time consumption, TrEEM takes source data into account, which will inevitably increase the time complexity. Remember that the scale of target data will not be big, and the time consumption is very acceptable especially when combined with the performance in Table 5.
- (3) Although TrEEM requires the most parameters shown in Table 6, λ and η have big effects. Observing Figures 4–6, the performance of RI and Purity depends more on the value of λ . Note that we can narrow the optimization range according to the discussion in Section 3. Thus, we believe that parameter setting would be easy.

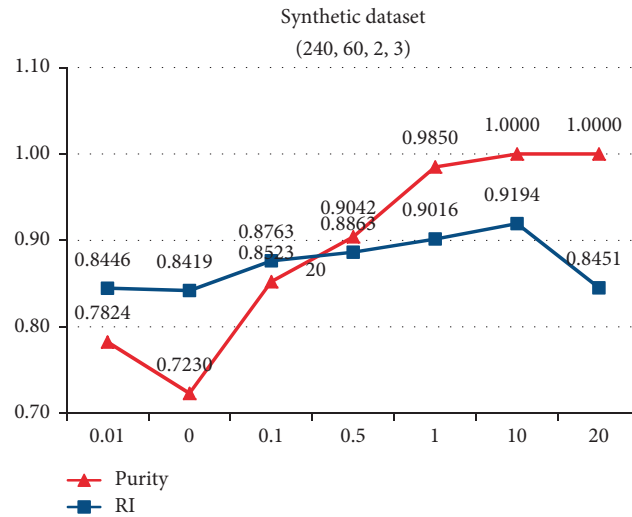
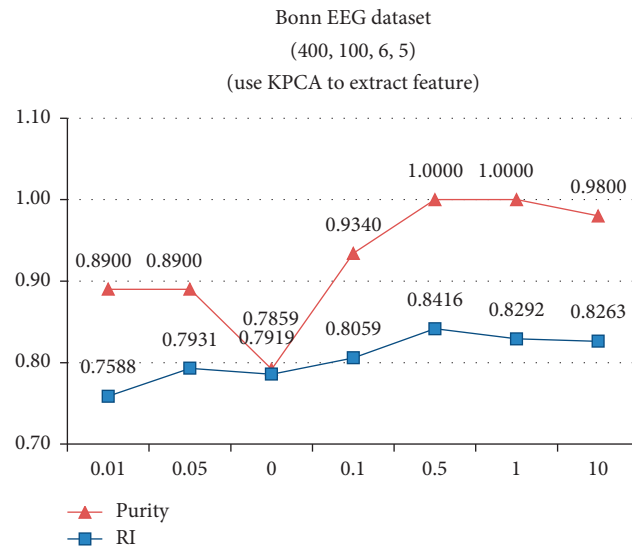
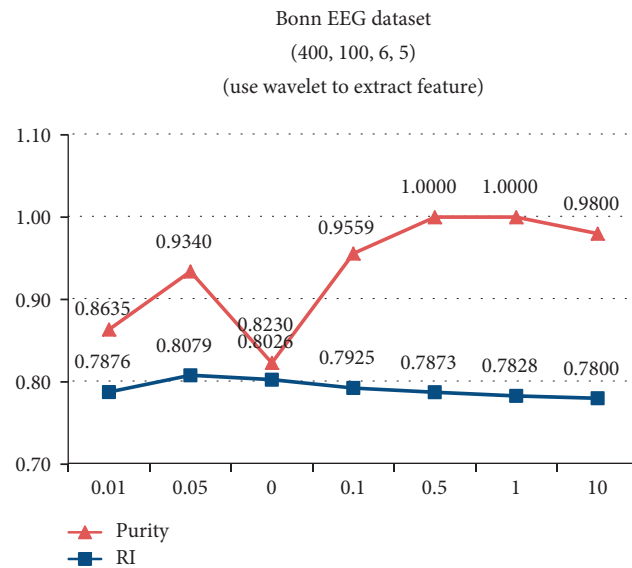
FIGURE 4: Effects of parameter λ on synthetic dataset.FIGURE 5: Effects of parameter λ on Bonn dataset with KPCA extract method.FIGURE 6: Effects of parameter λ on Bonn dataset with wavelet extract method.

TABLE 6: Average running time (Seconds) of the models on both synthetic and Bonn EEG datasets.

Datasets (source data, target data, attributes, and classes)	EEM	Multiclass SVM	TSK-FS	TSC	TrEEM
Synthetic dataset (240, 60, 2, 3)	0.1870	0.4562	0.0430	0.1520	0.2050
Bonn EEG dataset (400, 100, 6, 5) (use KPCA to extract feature)	0.6650	0.8934	0.3730	0.7923	0.6700
Bonn EEG dataset (400, 100, 6, 5) (use wavelet to extract feature)	0.4850	0.6327	0.1800	0.5331	0.5290

5. Conclusion

In conclusion, the contribution of this paper is providing a novel TrEEM framework to learn from few EEG signals when recognizing epileptic seizure. Starting from information theory, the proposed TrEEM method implants the similarity relationship between source and target data into the exemplar-based clustering model to improve the utilization rate of EEG signals, whereas this structure keeps all merits of the original optimization scheme. Therefore, without increasing the complexity of the model, TrEEM utilizes transfer learning method to learn from scant EEG signals. Yet our experimental results have shown promising performance of TrEEM, and several other perspectives should be considered as well. For instance, when each class contains unbalanced data, will this TrEEM method still work? And if we can provide multiple source data, what should we do to make them collaborate instead of bringing a negative effect? These are the problems that we should discuss in the future.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the 2018 Natural Science Foundation of Jiangsu Higher Education Institutions under Grant 18KJB5200001, by the Natural Science Foundation of Jiangsu Province under Grant no. BK20161268, and by the Humanities and Social Sciences Foundation of the Ministry of Education under Grant 18YJCZH229.

References

- [1] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.
- [2] Z. L. Bi Anqi and W. Ying, "Fast enhanced exemplar-based clustering for incomplete eeg signals," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 4147807, 11 pages, 2020.
- [3] Z. Jiang, F.-L. Chung, and S. Wang, "Recognition of multiclass epileptic eeg signals based on knowledge and label space inductive transfer," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 630–642, 2019.
- [4] Y. Jiang, Z. Deng, K.-S. Choi, F.-L. Chung, and S. Wang, "A novel multi-task TSK fuzzy classifier and its enhanced version for labeling-risk-aware multi-task classification," *Information Sciences*, vol. 357, pp. 39–60, 2016.
- [5] X. Zhu Jiaqi and L. Kang, "A novel double-index-constrained, multi-view, fuzzy-clustering algorithm and its application for detecting epilepsy electroencephalogram signals," *IEEE AC-CESS*, vol. 7, pp. 103 823–103 832, 2019.
- [6] D. Wen, P. Li, Y. Zhou et al., "Feature classification method of resting-state eeg signals from amnesic mild cognitive impairment with type 2 diabetes mellitus based on multi-view convolutional neural network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 8, pp. 1702–1709, 2020.
- [7] W. J. Xu Jiaca and H. Zheng, "Recognition of eeg signal motor imagery intention based on deep multi-view feature learning," *Sensors*, vol. 20, no. 12, 2020.
- [8] H. W. Liu ChienLiang and B. Xiao, "Epileptic seizure prediction with multi-view convolutional neural networks," *IEEE Access*, vol. 7, pp. 170 352–170 361, 2019.
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [10] L. L. Peng Yong, "Immune clonal algorithm based feature selection for epileptic eeg signal classification," in *Proceedings of the 2012 11th International Conference on Information Science, Signal Processing and Their Applications (ISSPA)*, Montreal, Canada, July 2012.
- [11] T. D. Z. Iscan and Z. Dokur, "Classification of electroencephalogram signals with combined time and frequency features," *Expert Systems With Applications*, vol. 38, no. 8, pp. 10499–10505, 2011.
- [12] S. W. Anqi Bi, "Incremental enhanced α -expansion move for large data: a probability regularization perspective," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 5, pp. 1–17, 2016.
- [13] K. Yizhang Jiang, "Anqi Bi, "Exemplar-based data stream clustering toward internet of things," *The Journal of Supercomputing*, vol. 76, no. 7, 2019.
- [14] A. Bi, F. Chung, S. Wang, Y. Jiang, and C. Huang, "Bayesian enhanced α -expansion move clustering with loose link constraints," *Neurocomputing*, vol. 194, pp. 288–300, 2016.
- [15] P. C. Y. Zheng, "Clustering based on enhanced α -expansion move," *IEEE Transaction on Knowledge and Data Engineering (TKDE)*, vol. 25, 2013.
- [16] W. T. F. M. F. Tappen, "Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters," in *Proceedings of the 9th IEEE International Conference on Computer Vision*, pp. 900–906, Nice, France, October 2003.
- [17] T. Kailath, "The divergence and bhattacharyya distance measures in signal selection," *IEEE Transactions on Communications*, vol. 15, no. 1, pp. 52–60, 1967.
- [18] T. M. Cover and T. Allen, *Elements of Information Theory*, Wiley Series in Telecommunications, Tsinghua University Press, Beijing, China, 1991.
- [19] Z. Liao, X. Lu, T. Yang, and H. Wang, "Missing data imputation: a fuzzy k-means clustering algorithm over sliding

- window,” in *Proceedings of the International Conference on Fuzzy Systems and Knowledge Discovery*, Tianjin, China, August 2009.
- [20] J. Zhang, Z. Yin, and R. Wang, “Pattern classification of instantaneous cognitive task-load through gmm clustering, laplacian eigenmap, and ensemble svms,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, no. 4, pp. 947–965, 2017.
 - [21] Z. Yin and J. Zhang, “Identification of temporal variations in mental workload using locally-linear-embedding-based eeg feature reduction and support-vector-machine-based clustering and classification techniques,” *Computer Methods and Programs in Biomedicine*, vol. 115, no. 3, pp. 119–134, 2014.
 - [22] A. P. Nguyen Chuong H. and K. Karavas George, “Inferring imagined speech using eeg signals: a new approach using riemannian manifold features,” *Journal of Neural Engineering*, vol. 15, no. 1, 2018.
 - [23] V. N. Vapnik, “An overview of statistical learning theory,” *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.
 - [24] Z. Deng, Y. Jiang, K.-S. Choi, F.-L. Chung, and S. Wang, “Knowledge-leverage-based tsf fuzzy system modeling,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 8, pp. 1200–1212, 2013.
 - [25] W. J. Jiang Yizhang and Z. Deng, “Transfer generalized fuzzy c-means clustering algorithm with improved fuzzy partitions by leveraging knowledge,” *Pattern Recognition and Artificial Intelligence*, vol. 26, no. 10, pp. 975–984, 2013.
 - [26] B. R. Greene, S. Faul, W. Marnane, G. Lightbody, I. Korotchikova, and G. Boylan, “A comparison of quantitative eeg features for neonatal seizure detection,” *Clinical Neurophysiology*, vol. 119, no. 6, pp. 1248–1261.
 - [27] W. Che Wan Fadzal and Mansor, “Short-time fourier transform analysis of eeg signal from writing,” in *Proceedings of the 2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, Malacca, Malaysia, March 2012.
 - [28] B. A. Vivaldi, “Frequency domain analysis of sleep eeg for visualization and automated state detection,” in *Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, NY, USA, September 2006.
 - [29] L. Z. Q. Zhang and H. Kawabata, “Eeg Analysis Using Fast Wavelet Transform,” in *Proceedings of the 2000 IEEE International Conference on Systems, Man and Cybernetics. “Cybernetics Evolving to Systems, Humans, Organizations, and Their Complex Interactions*, pp. 2959–2964, Nashville, TN, USA, 2000.
 - [30] X. C. D. Hu and W. Li, “Feature extraction of motor imagery eeg signals based on wavelet packet decomposition,” in *Proceedings of the 2011 IEEE/ICME International Conference on Complex Medical Engineering*, pp. 694–697, Harbin, China, May 2011.
 - [31] Q. W. Cao F, “Density-based clustering over an evolving data stream with noise,” in *Proceedings of the 2006 SIAM International Conference on Data Mining*, pp. 328–339, SIAM, Bethesda, MD, USA, April 2006.
 - [32] Z. J. Zhang Yuanpeng and X. Li, “Epileptic eeg signals recognition using a deep view-reduction tsf fuzzy system with high interpretability,” *IEEE Access*, vol. 7, no. 99, pp. 137 344–137 354, 2019.