



ARTICLE

# Cross-Domain TSK Fuzzy System Based on Semi-Supervised Learning for Epilepsy Classification

Zaihe Cheng<sup>1</sup>, Yuwen Tao<sup>2</sup>, Xiaoqing Gu<sup>3</sup>, Yizhang Jiang<sup>2</sup> and Pengjiang Qian<sup>2,\*</sup>

<sup>1</sup>School of Internet of Things, Wuxi Institute of Technology, Wuxi, 214121, China

<sup>2</sup>School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, 214122, China

<sup>3</sup>School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, 213164, China

\*Corresponding Author: Pengjiang Qian. Email: qianpjiang@jiangnan.edu.cn

Received: 10 November 2022 Accepted: 07 March 2023 Published: 28 June 2023

## ABSTRACT

Through semi-supervised learning and knowledge inheritance, a novel Takagi-Sugeno-Kang (TSK) fuzzy system framework is proposed for epilepsy data classification in this study. The new method is based on the maximum mean discrepancy (MMD) method and TSK fuzzy system, as a basic model for the classification of epilepsy data. First, for medical data, the interpretability of TSK fuzzy systems can ensure that the prediction results are traceable and safe. Second, in view of the deviation in the data distribution between the real source domain and the target domain, MMD is used to measure the distance between different data distributions. The objective function is constructed according to the MMD distance, and the distribution distance of different datasets is minimized to find the similar characteristics of different datasets. We introduce semi-supervised learning to further explore the relationship between data. Based on the MMD method, a semi-supervised learning (SSL)-MMD method is constructed by using pseudo-tags to realize the data distribution alignment of the same category. In addition, the idea of knowledge dissemination is used to learn pseudo-tags as additional data features. Finally, for epilepsy classification, the cross-domain TSK fuzzy system uses the cross-entropy function as the objective function and adopts the back-propagation strategy to optimize the parameters. The experimental results show that the new method can process complex epilepsy data and identify whether patients have epilepsy.

## KEYWORDS

Takagi-Sugeno-Kang fuzzy systems; back propagation; semi-supervised learning; inheritance mechanism; transfer learning

## 1 Introduction

Epilepsy is a disease caused by the abnormal discharge of neurons and can cause brain dysfunction. Electroencephalogram (EEG) detection is an effective way to identify epilepsy. At present, traditional EEG signal analysis is mostly completed by human experts according to their experience and is evaluated by means of the amplitude, frequency, and transient distribution of EEG. With the development of machine learning and the introduction of time-domain and frequency-domain analysis



into EEG signal processing methods, machines have gradually become smart tools that can assist doctors in diagnosing epilepsy [1,2].

For original EEG signals, due to a large amount of data and the changeable characteristics, it is unrealistic to directly use machine learning methods to identify them. Therefore, it is necessary to use feature extraction methods to extract effective features for machine learning. In this regard, short-time Fourier transform (STFT), wavelet packet decomposition (WPD), and kernel principal component analysis (KPCA) are commonly used. Fuzzy systems refer to a category of inference models that use fuzzy rules and offer interpretability. The membership function of fuzzy systems can match linguistic variables, and each process of data transmission in these models can be depicted by artificially designated linguistic variables. Such a white box model meets the interpretability requirement in the field of medicine well. Consequently, we use the Takagi-Sugeno-Kang (TSK) fuzzy system as a basis to design our model in this study [3,4]. TSK fuzzy systems usually optimize parameters via three pathways: 1) by the genetic algorithm, 2) by the least square method, and 3) by back propagation (BP)-based gradient descent. However, these methods have respective limitations. Genetic algorithms evaluate numerous candidate solutions for identifying the optimal solutions, which leads to high computing and memory costs [5]. The basic idea of the least squares method is to minimize the sum of squares between measured data and estimated data. However, when facing the classification problem, the label value is used as the estimation data, which cannot reflect its actual physical significance [6]. The back-propagation method can use cross-entropy and the Softmax function to reflect the probability that the data belong to a certain class, but the time complexity is high when dealing with large amounts of data; moreover, gradient disappearance may occur, which easily leads to inaccurate modelling results [7].

As deep learning has developed, its high-precision recognition rate makes deep learning attractive for epilepsy EEG detection, but because the process of deep learning is unexplainable, we proposed a back-propagation TSK fuzzy system. The backpropagation method can make a TSK fuzzy system more flexible because deep learning applies iterative training, so it provides the possibility of combining excellent deep learning algorithms with machine learning [8,9]. In addition, backpropagation can use cross-entropy, which is more suitable for classification tasks, as the loss function. This is the natural advantage of gradient descent TSK fuzzy systems compared with the first two solutions.

Two assumptions are made in traditional classification learning: 1) The learned training sample and the new test sample must satisfy independent and identical distributions. 2) There must be enough available training samples to learn a good classification model [10]. However, in practical applications, we find that these two conditions are often not met. In this study, we use a parameter-free method, called maximum mean discrepancy (MMD), to initially make the training data have the same distribution as the future testing data [11]. Then, the model learns and obtains the predicted value as a pseudolabel. The label information is used to further adjust the distribution. Thus, the model can meet the conditions of assumption 1. On this basis, we use pseudolabels as extended features to increase the data characterization information. To make the model satisfy the conditions of assumption 2, the dataset information is increased horizontally, the model learns effective features, and the model fully learns potential knowledge.

Our main contributions are as follows:

- 1) To reflect the probability distribution of the epilepsy EEG scenario, we firstly use the cross-entropy function as the loss function of the TSK fuzzy system to achieve parameter optimization.

- 2) A flexible back-propagation TSK fuzzy system is proposed to handle complex EEG data. The backpropagation method will make the TSK fuzzy system more flexible, because deep learning is also iterative training, so it provides the possibility of combining excellent algorithms in deep learning in future work.
- 3) Combining the concept of semi-supervised learning, a cross-domain TSK fuzzy system is proposed, and the ability of the TSK fuzzy system is optimized from the perspective of data characteristics and data distribution. It makes the proposed TSK model become a data-driven learning model.

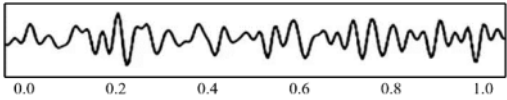
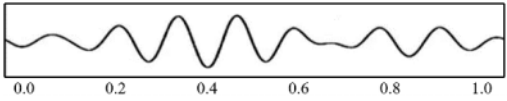
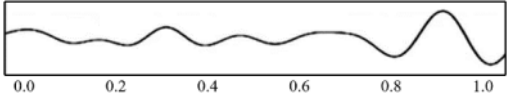
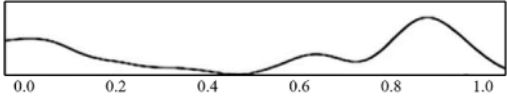
The detailed chapters are arranged as follows: [Section 2](#) introduces epilepsy signal data, the feature extraction method, the machine learning methods, and the back-propagation TSK fuzzy system. [Section 3](#) explores the proposed SSL-BP-TSK fuzzy system. [Section 4](#) presents the experiments on real-world epilepsy EEG signals. Section 6 is the conclusion.

## 2 Related Works

### 2.1 Epilepsy Signal Data

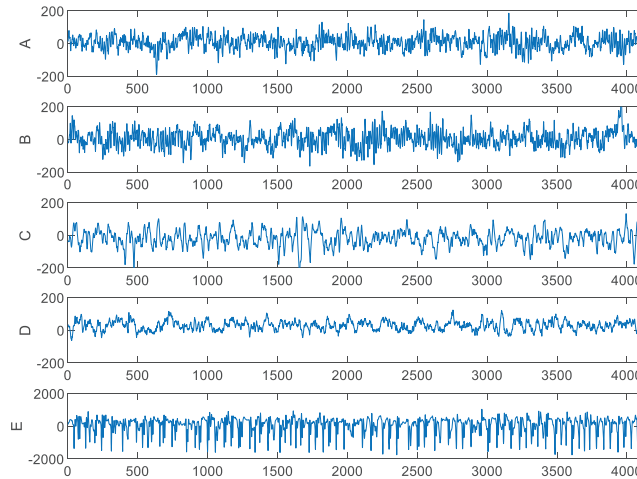
EEG records electrical signals from the cerebral cortex by measuring the electrical activity of a group of neurons. Because the patterns of EEG signals are often superimposed by the electrical signals from various parts of the brain in the direction of the electrode connections, different from ECG waveforms, it is very difficult for EEG waveforms to distinguish PQRST waves directly in the time domain and locate the specific excited region. Therefore, the analysis of brain waves in the frequency domain is slightly more important. The basic structural features of the alpha, beta, theta, and delta waves, which are roughly divided according to their frequency, can be summarized as follows: alpha (8–13 Hz), beta (18–30 Hz), theta (4–7 Hz), and delta (1–3.5 Hz). These bands are closely related to human behaviour, please see [Table 1](#).

**Table 1:** Details of the original EEG data

EEG type	Frequency	Basic characteristics	Waveform
$\beta$	18–30 Hz	People in an excited state	
$\alpha$	8–13 Hz	People in a waking state	
$\theta$	4–7 Hz	People in a tired state	
$\delta$	1–3.5 Hz	People in deep sleep	

Based on the structural features of the four waveforms in the table, we can quickly conclude that by using other feature extraction methods, such as the Fourier transform, we can derive the intensity of the EEG signal at different frequency components, and the existence of these waveforms and their intensity information are obtained.

A real-world epilepsy EEG signal [2] is shown in Fig. 1. Group A and Group B are the EEG signals of people without epilepsy with their eyes opened and closed, respectively. Group C and Group D are the EEG signals of patients with epilepsy (without seizures) when their eyes are opened and closed, respectively. Group E is the brain electrical signal of patients during epileptic seizures. From Fig. 1, it can be seen that the signals of Groups A and B are stable and dense, and the signals of Groups C, D and E have large amplitudes, and the intensity fluctuates from high to low, indicating that the neuron discharge is abnormal.



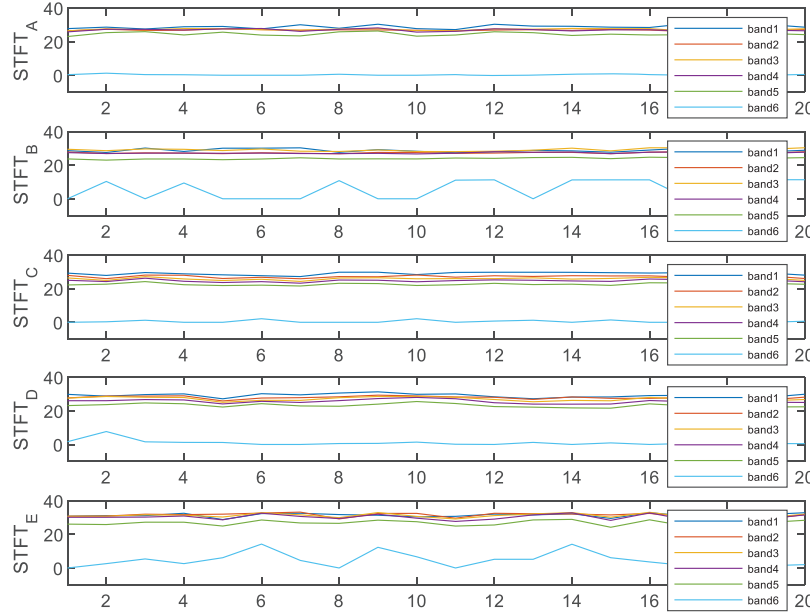
**Figure 1:** Raw epilepsy electroencephalogram

In addition to the differences in EEG signal waveforms of different subjects in different states, the EEG signals of subjects in a group are also different at different times. Therefore, we use the feature extraction method to extract artificial features of the signal to reduce the amount of data and noise data and facilitate model learning.

## 2.2 Extraction and Machine Learning Methods

As previously mentioned, on the original EEG signal, due to the large amount of data and changeable characteristics, effective feature extraction methods are needed to extract informative features for machine learning to reduce the computing cost. To this end, the traditional STFT, WPD and KPCA methods are enlisted in our study. Numerous traditional algorithms have been used for medical data analysis [12]. Nevertheless, one limitation of many well-established algorithms is that they can work well only when the data distributions of the training and testing data are the same or overall similar. Otherwise, they are commonly inefficient and even invalid. We briefly introduce three time-frequency methods commonly used in signal analysis [13]. 1) Wavelet packet decomposition (WPD). WPD is suitable for the analysis of nonstationary signals. It can be adapted according to the characteristics of the signal. The wavelet basis function can analyse signals well. 2) Short-time Fourier transform (STFT). Taking the STFT feature extraction method as an example, Fig. 2 shows an epilepsy EEG signal after the feature extraction method. The signal of each group is reduced from 1000 dimensions to 6 dimensions, which not only reduces the number of features but also makes the amplitude of the waveform more stable. Compared with Fig. 1, the amplitude decreases from  $-200\sim 200$  to  $0\sim 40$ . This greatly reduces the calculation cost. The reserved 6-dimensional features (especially the 6th dimension) have a unique signal structure, which is also conducive to machine

learning. 3) Kernel principal component analysis (KPCA). This method introduces the concept of a kernel function into PCA.



**Figure 2:** EEG data after STFT feature transformation

In the field of medicine, the interpretability of the employed machine learning models is indispensable. Doctors will not believe in the prediction results of machines unless they can well understand the inference mechanisms of machines. Compared with SVM, KNN, and NB, TSK fuzzy systems have the desired trait of good interpretability.

### 2.3 Maximum Mean Discrepancy Strategy

The maximum mean discrepancy (MMD) is a nonparametric method used to measure the distance between two distributions [14]. Here, we use MMD to measure the distribution difference between the training data (the source domain) and testing data (the target domain). Minimizing the MMD distance facilitates the two distributions being closer, so that the knowledge learned by the model is compatible with both the training and testing data [14].

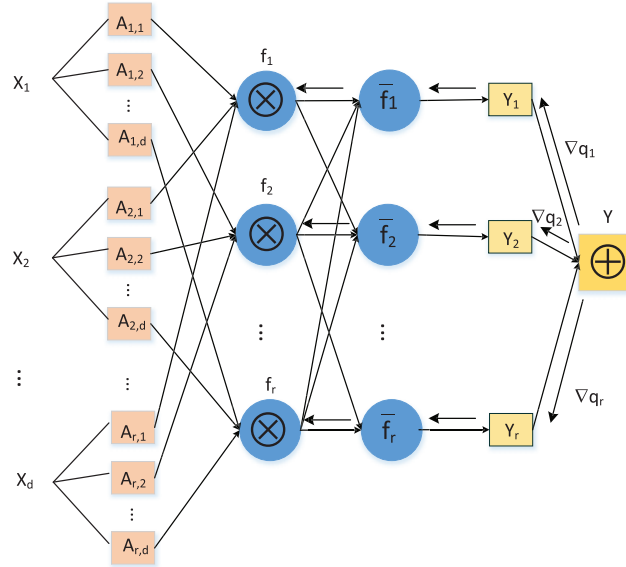
Given source domain data  $D_{Source} = \{x_i, y_i\}, i = 1, \dots, n_S$  and target domain data  $D_{Target} = \{x_i\}, i = 1, \dots, n_T$ , the MMD distance is defined as:

$$\begin{aligned}
 \text{distance}(D_S, D_T) &= \left\| \frac{1}{n_S} \phi(\mathbf{x}_S) - \frac{1}{n_T} \phi(\mathbf{x}_T) \right\|_H^2 \\
 &= \frac{1}{n_S^2} \phi(\mathbf{x}_S)^2 + \frac{1}{n_T^2} \phi(\mathbf{x}_T)^2 - \frac{2}{n_S n_T} \phi(\mathbf{x}_S) \phi(\mathbf{x}_T)
 \end{aligned} \tag{1}$$

where  $\phi(\cdot)$  represents the kernel function.

### 2.4 Back-Propagation TSK Fuzzy System

The rules in a back propagation TSK (BP-TSK) fuzzy system are equivalent to the neurons in neural networks. The network structure of the BP-TSK fuzzy system is also similar to that of a neural network, as shown in Fig. 3.



**Figure 3:** Structure of the BP-TSK fuzzy system.  $f_r$  is calculated by (3.3),  $\bar{f}_r$  is calculated by (3.4),  $\nabla q_r$  is included in  $\nabla q_g$ , which is calculated by (8), and  $\otimes$  is the multiplication operator

Given training dataset  $DTrain = \{x_n, y_n\}, n = 1, \dots, N$ , where  $x \in R^{n \times d}, d = 1, \dots, D$ ,  $D$  is the feature dimension,  $y \in R^{n \times c}, c = 1, \dots, C$ , and  $C$  is the total number of categories. Assuming that the BP-TSK fuzzy system has  $R$  rules, it is constructed as follows:

Rule <sub>$r$</sub> : IF  $x_1$  is  $A_{r,1}$  and  $x_2$  is  $A_{r,2}$  and  $\dots$  and  $x_D$  is  $A_{r,D}$   
 THEN  $y_r^c(x) = q_{r,0}^c + q_{r,1}^c x_1 + \dots + q_{r,D}^c x_D$  (2)

where  $A_{r,d}$  is the membership function, including the antecedent parameters, and  $q_{r,d}^c$  denotes the consequent parameters of class  $c$ .

The output of the BP-TSK fuzzy system is the weighted results of each rule:

$$y^c(x) = \frac{\sum_{r=1}^R f_r(x) y_r^c(x)}{\sum_{r=1}^R f_r(x)} = \sum_{r=1}^R \bar{f}_r(x) y_r^c(x) \quad (3.1)$$

The membership degree of  $d$ -dimensional features in rule  $r$  is:

$$\mu_{A_{r,d}}(x_d) = \exp\left(-\frac{(x_d - m_{r,d})^2}{2\sigma_{r,d}^2}\right) \quad (3.2)$$

where  $m_{r,d}$  and  $\sigma_{r,d}$  represent the centre and standard deviation of the Gaussian kernel function, respectively, which are often obtained using fuzzy C-means (FCM) clustering or other similar algorithms. In the formulation of rules, we assign semantic words at different levels according to the level of the antecedent parameter  $\sigma_{r,d}$  in each rule.

Trigger value of rule  $r$ :

$$f_r(x) = \prod_{d=1}^D \mu_{A_{r,d}}(x_d) \tag{3.3}$$

where  $\bar{f}_r(x)$  represents the normalization trigger level of the  $r$ -th rule:

$$\bar{f}_r(x) = \frac{f_r(x)}{\sum_{r=1}^R f_r(x)} \tag{3.4}$$

The following formula is constructed according to [12]:

$$x_e = (1_{dx1}, x^T)^T \tag{4.1}$$

$$\bar{x}_r = \bar{f}_r(x) x_e \tag{4.2}$$

$$x_g = ((\bar{x}_1)^T, (\bar{x}_2)^T, \dots, (\bar{x}_r)^T)^T \tag{4.3}$$

$$q_r^c = (q_{r,1}^c, q_{r,2}^c, \dots, q_{r,D}^c)^T \tag{4.4}$$

$$q_g^c = ((q_1^c)^T, (q_2^c)^T, \dots, (q_r^c)^T)^T \tag{4.5}$$

Then, (3.1) can be expressed as [15]:

$$y^c = x_g q_g^c \tag{4.6}$$

For classification tasks, cross-entropy can be used to constitute the loss function of the BP-TSK fuzzy system through the Softmax function:

$$p(y^c(x)) = \frac{\exp(y^c(x))}{\sum_{c=1}^C \exp(y^c(x))} \tag{5}$$

$$Loss = - \sum_{n=1}^N y_n p(y^c(x)) \tag{6}$$

The objective function of the BP-TSK fuzzy system can be formulated for all attributes  $x_{gi}$  as follows:

$$\begin{aligned} \min_{q_g^c} L &= - \sum_{n=1}^N y_n \ln \alpha_i + \frac{\lambda_1}{2} \|q_g^c\|^2 \\ &= - \sum_{n=1}^N y_n \ln \frac{\exp(z_i)}{\sum_{c=1}^C \exp(z_i)} + \frac{\lambda_1}{2} \|q_g^c\|^2 \\ &= - \sum_{n=1}^N y_n \ln \frac{\exp(x_{gi}^T q_g^c)}{\sum_{c=1}^C \exp(x_{gi}^T q_g^c)} + \frac{\lambda_1}{2} \|q_g^c\|^2 \end{aligned} \tag{7}$$

where the former item is an empirical risk item and the latter is a structural risk item.  $q_g^c$  is the subsequent parameter of type  $c$ , and  $\lambda_1$  is the regularization parameter controlling the tolerance of the classifier to errors and can be adjusted by cross-validation.

Minimizing the objective function  $L$  requires obtaining the partial derivative of  $q_g^c$  and setting the derivative equal to 0:

$$\frac{\partial L}{\partial q_g^c} = \frac{\partial L}{\partial \alpha} \frac{\partial \alpha}{\partial z} \frac{\partial z}{\partial q_g^c} = \sum_i^N x_{gi} \left( \frac{\exp(x_{gi}^T q_g^c)}{\sum_{c=1}^C \exp(x_{gi}^T q_g^c)} - y_i^c \right) + \lambda_1 q_g^c \quad (8)$$

We can use the AdaBelief optimizer [15] or other optimization methods [16] to update the parameters and obtain the optimal parameters  $q_g^*$  after iterations.

The algorithm of the BP-TSK fuzzy system is as follows:

---

**Algorithm 1:** BP-TSK fuzzy system

---

**Initialization:** Training set  $D_{train}(x_i, y_i)$ , testing set  $D_{test}(x_j)$ , number of rules  $R$ , number of iterations  $T$ , regularization parameters of  $\lambda_1$ .

**Parameter optimization**

Step 1: Take the rule number  $R$  as the centre of the cluster, perform FCM clustering on the training set  $x_n$ , and obtain the antecedent parameters.

Step 2: Construct dataset  $D_{train} = \{x_{gi}, y_i\}$ ,  $D_{test} = \{x_{gj}\}$  according to (3.1)–(3.3).

Step 3: Calculate the gradient of the subsequent parameters according to (7).

Step 4: Pass the gradient to the optimizer and update the parameters.

**Output:** Obtain the prediction result of the model according to (3.6).

---

### 3 SSL-BP-TSK Fuzzy Systems

This chapter introduces the MMD transfer learning items in the BP-TSK fuzzy system and adds two semi-supervised learning items on this basis. One is used to adjust the data distribution, and the other is used to expand the data characteristics.

#### 3.1 MMD for the BP-TSK Fuzzy System

By introducing MMD into the TSK system, the distribution distance can be further defined as:

$$\begin{aligned} MMD_{(D_S, D_T)} &= \left\| \frac{1}{n_S} (\mathbf{x}_{gS}^T \mathbf{q}_g) - \frac{1}{n_T} (\mathbf{x}_{gT}^T \mathbf{q}_g) \right\|_H^2 \\ &= \frac{1}{n_S^2} \mathbf{q}_g^T \mathbf{x}_{gS} \mathbf{x}_{gS}^T \mathbf{q}_g + \frac{1}{n_T^2} \mathbf{q}_g^T \mathbf{x}_{gT} \mathbf{x}_{gT}^T \mathbf{q}_g - \frac{2}{n_S n_T} \mathbf{q}_g^T \mathbf{x}_{gS}^T \mathbf{x}_{gT} \mathbf{q}_g \\ &= \mathbf{q}_g^T \left( \frac{1}{n_S^2} \mathbf{x}_{gS}^T \mathbf{x}_{gS} + \frac{1}{n_T^2} \mathbf{x}_{gT}^T \mathbf{x}_{gT} - \frac{2}{n_S n_T} \mathbf{x}_{gS}^T \mathbf{x}_{gT} \right) \mathbf{q}_g \end{aligned} \quad (9)$$

Let

$$M = \frac{1}{n_S^2} \mathbf{x}_{gS}^T \mathbf{x}_{gS} + \frac{1}{n_T^2} \mathbf{x}_{gT}^T \mathbf{x}_{gT} - \frac{2}{n_S n_T} \mathbf{x}_{gS}^T \mathbf{x}_{gT} \quad (10)$$

Then, (9) can be rewritten as:

$$MMD_{(D_S, D_T)} = \mathbf{q}_g^T M \mathbf{q}_g \quad (11)$$



The objective function of TSK after adding the MMD item is:

$$\min_{q_g^c} L = - \sum_{i=1}^N y_i \ln \alpha_i + \frac{\lambda_1}{2} \|q_g^c\|^2 + \lambda_2 (q_g^c)^T M \quad q_g^c \quad (12)$$

The derivative corresponding to  $q_g$  is:

$$\frac{\partial L}{\partial q_g^c} = \sum_i x_{gi} \left( \frac{\exp(x_{gi}^T q_g^c)}{\sum_{c=1}^C \exp(x_{gi}^T q_g^c)} - y_i^c \right) + (\lambda_1 E + \lambda_2 M) q_g^c \quad (13)$$

### 3.2 SSL-MMD for BP-TSK Fuzzy Systems

The label information of the source domain and the target domain can provide more data distribution information, so many experts have begun to study domain adaptation.

After the BP-TSK fuzzy system is added to the MMD item, it can learn the similar characteristics of two-domain data, so the prediction result has a certain degree of reliability. Based on the MMD-BP-TSK fuzzy system, we save the prediction results as pseudolabels. Then, we use pseudolabels to make the data distributions of the two domains more similar according to categories so that the model can learn more refined knowledge features.

Fig. 4 shows the three steps of data distribution processing: a) the unprocessed initial data distribution, b) the data distribution after processing by the MMD method, and c) the data distribution after processing by the SSL-MMD method.

In Fig. 4a, when the model has learned the data with labels, it is difficult for the model to adjust its decision boundary when the distribution is different from the previous data distribution, that is, when there is a chaotic data distribution.

In Fig. 4b, we can see that the chaotic data distribution has been improved, and the model can find the appropriate decision boundary between the two domains.

In Fig. 4c, a pseudotag is added on the basis of Fig. 4b to guide the model to determine a more appropriate decision boundary.

The objective function of SSL-MMD is:

$$SSL - MMD_{(D_S, D_T)} = \sum_{c=1}^C \left\| \sum_{i=1}^{n_S} w_i^c (x_{gi,S}^T q_g) - \sum_{j=1}^{n_T} w_j^c (x_{gj,T}^T q_g) \right\|_H^2 \quad (14)$$

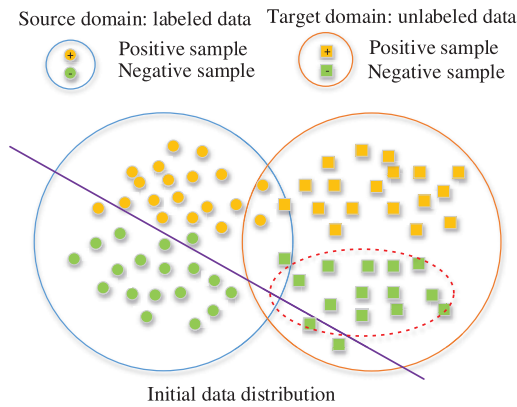
where  $\begin{cases} w_i^c = y_i/n_S \\ w_j^c = \hat{y}_j/n_T \end{cases}$ ,  $y_i$  is the known source domain label.  $\hat{y}_j$  is an unknown target domain label, which is replaced by a pseudolabel.

Let

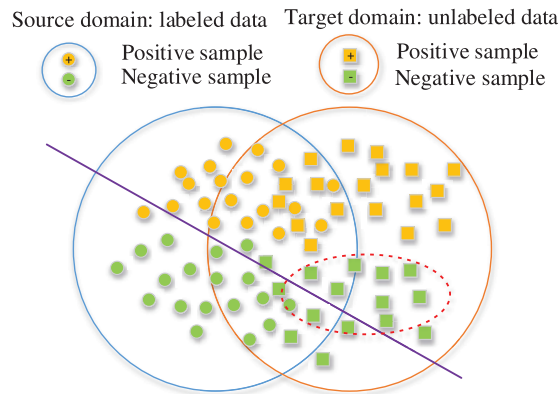
$$M_C = \frac{1}{n_S^2} w_i^{cT} x_{gS} x_{gS}^T w_i^c + \frac{1}{n_T^2} w_j^{cT} x_{gT} x_{gT}^T w_j^c - \frac{1}{n_S n_T} w_i^{cT} x_{gS} x_{gT}^T w_j^c - \frac{1}{n_S n_T} w_j^{cT} x_{gT} x_{gS}^T w_i^c \quad (15)$$

Similarly, (14) can be rewritten as:

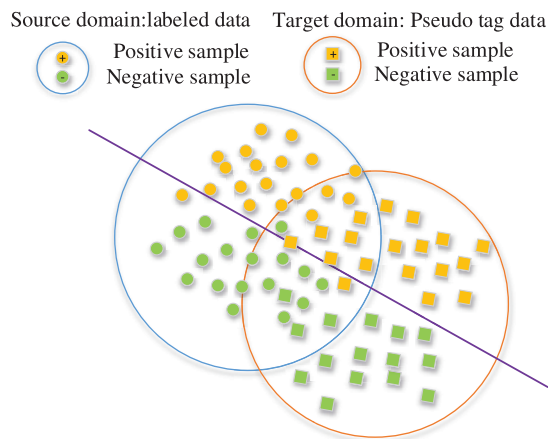
$$SSL - MMD_{(D_S, D_T)} = (q_g^c)^T M_C q_g^c \quad (16)$$



(a): Because the data distributions of the source and target domains are fairly different, the classifier trained in the source domain cannot predict the target domain data well



(b): MMD method narrows the overall gap between two data distributions



(c): SSL-MMD method uses label information to narrow the distribution of the same type of data

**Figure 4:** (a): Because the data distributions of the source and target domains are fairly different, the classifier trained in the source domain cannot predict the target domain data well. (b): MMD method narrows the overall gap between two data distributions. (c): SSL-MMD method uses label information to narrow the distribution of the same type of data

The objective function of SSL-MMD-BP-TSK is:

$$\min_{q_g^c} L = - \sum_{i=1}^N y_i \ln \alpha_i + \frac{\lambda_1}{2} \|q_g^c\|^2 + \lambda_2 (q_g^c)^T M \quad q_g^c + \lambda_3 (q_g^c)^T M_C \quad q_g^c \quad (17)$$

The derivative corresponding to  $q_g^c$  is:

$$\frac{\partial L}{\partial q_g^c} = \sum_i^N x_{gi} \left( \frac{\exp(x_{gi}^T q_g^c)}{\sum_{c=1}^C \exp(x_{gi}^T q_g^c)} - y_i^c \right) + (\lambda_1 E + \lambda_2 M + \lambda_3 M_C) q_g^c \quad (18)$$

### 3.3 Semi-Supervised Inherited Learning

We use the idea of deep learning to try to transfer the knowledge learned in the previous round to the next round. Therefore, the concept of a semi-supervised inheritance (SSL-I) mechanism is introduced to increase the information interaction between different domains. Therefore, we proposed a TSK fuzzy system for cross-domain learning, which includes SSL-MMD and SSL-I mechanisms.

Unlike the SSL-MMD method, which reduces the difference in data distributions, the SSL-I method makes full use of pseudolabels to allow the model to learn additional feature representations. The SSL-I method merges the existing tags and pseudo tags with the source domain data and target domain data, respectively, to construct a new dataset.

Experiments show that pseudolabels can be used for effective feature representation, allowing the model to learn deeper knowledge.

Suppose a given dataset is  $D_{train}(x_i, y_i), D_{test}(x_j)$ ; after learning by the MMD-BP-TSK fuzzy system, the prediction result is obtained. The prediction result is used as a pseudolabel  $\hat{y}$  of the target domain data to reconstruct the datasets  $ExD_{Source}\{x_i, y_i\}, x_i \in \mathfrak{R}^{n \times (d+y_i)}, i = 1, \dots, N_s$  and  $ExD_{Target}\{x_j, \hat{y}_j\}, x_j \in \mathfrak{R}^{n \times (d+\hat{y}_j)}, j = 1, \dots, N_T$ .

where  $y_i$  and  $\hat{y}_j$  are labels in the form of one-hot vectors. The new dataset is learned again through the SSL-MMD-BP-TSK fuzzy system. The structure of the CD-BP-TSK fuzzy system is shown in Fig. 5.

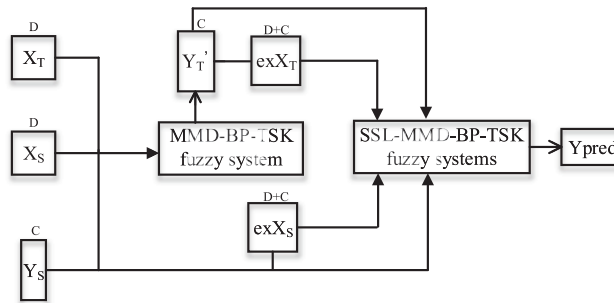


Figure 5: Structure of the CD-BP-TSK fuzzy system

The proposed new classification algorithm is summarized as follows:

**Algorithm 2:** CD-BP-TSK fuzzy system

**Initialization:** Training set  $D_{train}(x_i, y_i)$ , testing set  $D_{test}(x_j)$ , number of rules  $R$ , number of iterations  $T$ , regularization parameters  $\lambda_1, \lambda_2$ , and  $\lambda_3$ , learning rate  $\alpha = 0.01$ , optimizer parameters  $\beta_1 = 0.9, \beta_2 = 0.999$  and  $\varepsilon = 10^{-8}$ .

**Parameter optimization**

Step 1: Take the rule number  $R$  as the centre of the cluster, perform FCM clustering on the training set  $x_n$ , and obtain the antecedent parameters.

Step 2: Construct dataset  $D_{train} = \{x_{gi}, y_i\}, D_{test} = \{x_{gj}\}$  according to (4.1)–(4.3).

//MMD-BP-TSK-FS

Step 3: Use (10) to calculate the data distribution matrix  $M$ .

Step 4: Calculate the gradient of the subsequent parameters according to (13).

Step 5: Use (4.6) to obtain the prediction result and save it as a pseudolabel  $\hat{y}$ .

//SSL-MMD and SSL-I

Step 6: Use pseudolabels to construct a new dataset  $EX_{Source}, EX_{Target}$ , and calculate the weight  $w^c$ .

Step 7: Use (18) to calculate the subsequent parameter  $q_g^c$ .

Step 8: Pass the gradient to the optimizer and update the parameters.

**Output:** Obtain the prediction result of the model according to (4.6)

## 4 Experiment Analysis

In this section, we introduce the EEG datasets and perform performance tests. We compare the new method with other machine learning methods and observe whether the experimental results are significantly improved. Finally, the interpretability of the new model is discussed.

### 4.1 Experimental Setup

In this section, we build datasets in different scenarios to comprehensively test our methods. The experiment was completed in the following working environment: Processor: Intel(R) Core(TM) i5-8500 CPU@3.00 GHz; Memory: 12.0 GB; Software: MATLABR2021A.

#### 1) Build the datasets

The datasets are divided into two categories. As shown in Table 2, one is a dataset with a similar data distribution, and the other is a dataset with a chaotic data distribution.

#### 2) Performance index

In classification problems, accuracy is used to judge whether a model's performance is excellent.

### 4.2 Analysis of the Results

#### 1) Comparison algorithm

We compared 6 traditional machine learning algorithms and 3 semi-supervised learning algorithms, as shown in Table 3.

Among the compared methods, traditional methods, such as SVM, NB, KNN, and LDA, are set according to [17,18], and the semi-supervised method, S4VM, is set according to [19]. The rule number  $R$  in the TSK fuzzy system is selected from  $\{5, 6, 7, 8, 9, 10\}$ , and the regularization parameter  $\lambda_1, \lambda_2, \lambda_3$  is obtained from  $\{10^{-3}, 10^{-2}, 10^{-1}, 0, 10^1, 10^2, 10^3\}$ . The number of iterations of the BP-TSK fuzzy systems is selected from  $\{1000, 1500, 2000, 2500, 3000\}$ .

**Table 2:** Details of the constructed EEG dataset

Scenes	Datasets	Description
Similar distribution	AC-AC	50 positive classes and 50 negative classes
	AE-AE	
	ADE-ADE	25 positive classes and 50 negative classes
	ACD-ACD	
Chaotic distribution	AC-AD	50 positive classes and 50 negative classes
	AC-AE	
	AE-AC	
	ABE-ABD	50 positive classes and 25 negative classes
	ABD-ABE	
ACE-ADE	25 positive classes and 50 negative classes	

**Table 3:** Abbreviations for the comparison algorithms

Support vector machine	SVM
Naive bayes	NB
K-nearest neighbour	KNN
Linear discriminant analysis	LDA
Least squares TSK fuzzy system	LS-TSK-FS
Backpropagation TSK fuzzy system	BP-TSK-FS
Semi-supervised SVM	S4VM
Our method 1	SSL-MMD-BP-TSK-FS
Our method 2	CD-BP-TSK-FS

## 2) Experimental results

The accuracies of each algorithm on the dataset after the three feature extraction methods are shown in [Tables 4–6](#). Obviously, all methods can achieve good accuracy for datasets with similar distributions. However, the accuracies of the three semi-supervised learning methods are much higher than those of the other algorithms. Although all data in the dataset are taken from the same defined group, there is still a deviation in the data distribution.

Second, for datasets with chaotic data distributions, the performance of traditional methods decreases significantly, and the minimum accuracy is only 51.4%, which is because they cannot be classified correctly. There is also a case in which the traditional algorithm is effective for datasets with chaotic data distributions, which reflects that although scenes with chaotic data distributions are constructed according to different defined groups, the results are not scientific.

**Table 4:** Accuracy value of each model after STFT feature extraction

Datasets	SVM	NB	KNN	LDA	LS-TSK- FS	BP-TSK- FS	S4VM	SSL-MMD- BP-TSK-FS	CD-BP- TSK-FS
Similar distribution									
AC-AC 50-50	0.931	0.887	0.92	0.951	0.926	0.928	0.942	0.944	<b>0.955</b>
AE-AE 50-50	0.993	0.988	0.979	0.975	0.982	0.988	0.986	0.991	<b>0.994</b>
ADE-ADE 75-75	0.923	0.902	0.925	0.917	0.937	0.94	0.951	0.947	<b>0.967</b>
ACD-ACD 75-75	0.94	0.909	0.938	0.942	0.925	0.919	0.949	0.947	<b>0.959</b>
Chaotic distribution									
AC-AD 50-50	0.904	0.879	0.906	0.904	0.891	0.896	0.934	0.948	<b>0.96</b>
AC-AE 50-50	0.838	0.92	0.853	0.861	0.837	0.856	0.904	0.957	<b>0.961</b>
AE-AC 50-50	0.509	0.547	0.507	0.511	0.541	0.547	0.846	0.952	<b>0.969</b>
ABE-ABD 75-75	0.687	0.688	0.6823	0.728	0.718	0.706	0.831	0.941	<b>0.954</b>
ABD-ABE 75-75	0.921	0.894	0.779	0.918	0.861	0.872	0.84	0.954	<b>0.965</b>
ACE-ADE 75-75	0.916	0.907	0.915	0.899	0.935	0.928	0.953	0.961	<b>0.97</b>

**Table 5:** Accuracy value of each model after WPD feature extraction

Datasets	SVM	NB	KNN	LDA	LS-TSK- FS	BP-TSK- FS	S4VM	SSL-MMD- BP-TSK-FS	CD-BP- TSK-FS
Similar distribution									
AC-AC 50-50	0.902	0.877	0.885	0.934	0.914	0.912	0.916	0.955	<b>0.977</b>
AE-AE 50-50	0.878	0.848	0.839	0.904	0.885	0.891	0.862	0.939	<b>0.956</b>
ADE-ADE 75-75	0.826	0.741	0.857	0.898	0.905	0.905	0.901	0.931	<b>0.947</b>

(Continued)

**Table 5 (continued)**

Datasets	SVM	NB	KNN	LDA	LS-TSK-FS	BP-TSK-FS	S4VM	SSL-MMD-BP-TSK-FS	CD-BP-TSK-FS
ACD-ACD 75-75	0.93	0.881	0.911	0.93	0.932	0.919	0.917	0.949	<b>0.96</b>
Chaotic distribution									
AC-AD 50-50	0.915	0.819	0.885	0.909	0.88	0.884	0.876	0.951	<b>0.972</b>
AC-AE 50-50	0.679	0.463	0.556	0.706	0.583	0.611	0.750	0.93	<b>0.947</b>
AE-AC 50-50	0.514	0.523	0.531	0.664	0.781	0.799	0.823	0.944	<b>0.953</b>
ABE-ABD 75-75	0.739	0.691	0.707	0.803	0.87	0.879	0.781	0.953	<b>0.964</b>
ABD-ABE 75-75	0.811	0.598	0.777	0.876	0.752	0.77	0.762	0.932	<b>0.959</b>
ACE-ADE 75-75	0.812	0.726	0.856	0.878	0.897	0.895	0.893	0.924	<b>0.951</b>

**Table 6:** Accuracy value of each model after KPCA feature extraction

Datasets	SVM	NB	KNN	LDA	LS-TSK-FS	BP-TSK-FS	S4VM	SSL-MMD-BP-TSK-FS	CD-BP-TSK-FS
Similar distribution									
AC-AC 50-50	0.9021	0.829	0.8469	0.8579	0.855	0.855	0.896	0.948	<b>0.965</b>
AE-AE 50-50	0.9545	0.949	0.909	0.895	0.943	0.958	0.96	<b>0.998</b>	0.985
ADE-ADE 75-75	0.957	0.903	0.921	0.834	0.928	0.949	0.975	0.963	<b>0.999</b>
ACD-ACD 75-75	0.909	0.875	0.862	0.811	0.892	0.917	0.944	0.946	<b>0.981</b>
Chaotic distribution									
AC-AD 50-50	0.899	0.64	0.897	0.889	0.896	0.886	0.946	0.973	<b>0.995</b>
AC-AE 50-50	0.892	0.633	0.916	0.902	0.911	0.909	0.926	0.984	<b>0.999</b>
AE-AC 50-50	0.882	0.537	0.844	0.844	0.604	0.552	0.922	0.965	<b>0.983</b>
ABE-ABD 75-75	0.911	0.703	0.918	0.872	0.768	0.771	0.973	<b>0.984</b>	0.979

(Continued)

**Table 6 (continued)**

Datasets	SVM	NB	KNN	LDA	LS-TSK-FS	BP-TSK-FS	S4VM	SSL-MMD-BP-TSK-FS	CD-BP-TSK-FS
ABD-ABE 75-75	0.936	0.743	0.932	0.897	0.952	0.955	0.964	0.979	<b>0.996</b>
ACE-ADE 75-75	0.952	0.839	0.887	0.839	0.904	0.911	0.957	0.961	<b>0.994</b>

In conclusion, it is a very common situation that there are differences in data distributions. Although datasets are similar in different domains, there may still be differences in the data distribution. Furthermore, in datasets with chaotic data distributions, there may still be small differences in the data distribution.

It can be seen from [Tables 4-6](#) that the average accuracy of the proposed CD-BP-TSK-FS method reaches more than 95%, and the highest accuracy is 99.9% among the three feature extraction methods. This proves that the new method can solve the problem of performance degradation caused by different data distributions.

### 3) Statistical significance analysis

We use the newly proposed method and other methods to perform Friedman [20,21] and Holm tests [22,23] to observe whether there are statistically significant differences. The results are shown in [Tables 7 and 8](#). CD-BP-TSK-FS ranks first in the Friedman test. For the Holm test,  $p < 0.05$  represents a significant difference between our method and other methods. The experimental results show that, in addition to the comparison with the S4VM method under the KPCA characteristics, CD-BP-TSK-FS has statistically significant differences. Although there is no significant difference in performance with S4VM, our method is based on a fuzzy system, and its own explanatory nature can compensate for this shortcoming.

**Table 7: Friedman test**

Average rankings of the Friedman test			
	STFT	WPD	KPCA
SVM	5.75	6.5	4.8
NB	6.8	8.8	8.35
KNN	7.3	7.2	6.35
LDA	5.75	4.45	7.4
LS-TSK-FS	6.5	5.05	6.25
BP-TSK-FS	5.7	4.85	5.75
S4VM	3.8	5.5	3
SSL-MMD-BP-TSK-FS	2.4	2	1.9
CD-BP-TSK-FS	1	1	1.2



**Table 8:** Holm test

Post hoc comparisons ( $p$ values for $\alpha = 0.05$ )						
CD-BP-TSK-FS vs.	STFT		WPD		KPCA	
	z	p	z	p	z	p
SVM	3.878359	0.000105	4.204957	0.000026	2.939388	0.003289
NB	4.73568	0.000002	6.368673	0	5.837951	0
KNN	5.143928	0	5.062279	0	4.204957	0.000026
LDA	3.878359	0.000105	2.816913	0.004849	5.062279	0.001471
LS-TSK-FS	4.490731	0.000007	3.306811	0.000944	4.123308	0.000037
BP-TSK-FS	3.837534	0.000124	3.143512	0.001669	3.715059	0.000203
S4VM	2.28619	0.022243	3.0674235	0.000239	1.469694	<b>0.142</b>
SSL-MMD-BP-TSK-FS	1.143095	<b>0.253</b>	0.816497	<b>0.414</b>	0.571548	<b>0.568</b>

**4.3 Interpretability Analysis**

Fuzzy rules can explain the execution process within the entire model through rules [24,25]. However, if the number of rules is too large, we need to define many semantics to describe the rules. This will reduce the interpretation ability of the fuzzy system. Although our method is based on backpropagation, it does not require a large number of rules to be superimposed, and only 5–10 rules are needed to obtain good accuracy. Therefore, our method can completely replace rules with semantics to achieve interpretable functions.

The AE-AE dataset under STFT feature extraction is used to illustrate the execution process of the fuzzy system. The antecedent and subsequent parameters are shown in Table 9.

**Table 9:** Antecedent and subsequent parameters in fuzzy systems

$A_{r,d}$	Antecedent parameters									
	$A_1(m, \sigma)$		$A_2(m, \sigma)$		$A_3(m, \sigma)$		$A_4(m, \sigma)$		$A_5(m, \sigma)$	
d1	31.344	0.551	28.427	0.537	31.325	0.489	29.527	0.552	31.122	0.583
	High		Low		Middle high		Middle low		Middle	
d2	32.244	0.461	27.123	0.626	32.095	0.727	27.320	0.912	29.778	0.876
	High		Low		Middle high		Middle low		Middle	
d3	32.739	0.843	27.562	0.452	31.657	0.741	27.785	0.580	29.920	0.692
	High		Low		Middle high		Middle low		Middle	
d4	31.812	1.151	26.815	0.445	31.101	0.998	27.000	0.589	29.291	0.865
	High		Low		Middle high		Middle low		Middle	
d5	28.255	0.995	24.173	0.304	27.618	0.828	24.356	0.440	26.031	0.551
	High		Low		Middle high		Middle low		Middle	

(Continued)

**Table 9 (continued)**

d6	11.917	7.391	0.405	0.961	5.137	3.163	0.283	1.373	1.464	2.838
	High		Low		Middle high		Middle low		Middle	
Subsequent parameters										
q1	c 1	0.005	-0.074	-0.034	-0.035	-0.027	-0.010			-0.001
	c 2	0.045	0.006	-0.006	0.038	0.022	0.010			0.001
q2	c 1	0.477	0.191	-0.184	-0.061	0.367	0.000			0.004
	c 2	0.014	-0.343	-0.077	-0.284	0.066	0.000			-0.004
q3	c 1	-0.030	-0.030	-0.030	-0.029	-0.026	-0.005			-0.001
	c 2	0.030	0.030	0.030	0.029	0.026	0.005			0.001
q4	c 1	0.264	-0.268	-0.016	-0.192	0.337	-0.489			0.009
	c 2	0.226	0.163	-0.134	0.023	-0.270	0.489			-0.009
q5	c 1	-0.105	0.020	0.021	0.037	0.012	0.000			0.000
	c 2	0.188	0.072	0.329	-0.091	-0.128	0.000			0.000

First, we use “High, Low, Middle, Middle High (MH), and Middle Low (ML)” to describe the strength of each feature element. Then, we use these strength terms in the IF-part to describe the strength of each feature element and use the corresponding linear expression in the Then-part to obtain the value. Finally, the final decision value of the model can be obtained. The decision-making process of the five rules is as follows:

1<sup>st</sup> rule:

IF: d1, d2, d3, d4, d5, and d6 are High,

$$\text{THEN: } f_1(x) = \begin{bmatrix} 0.005 - 0.074x_1 - 0.034x_2 - 0.035x_3 - 0.027x_4 - 0.010x_5 - 0.001x_6, \\ 0.045 + 0.006x_1 - 0.006x_2 + 0.038x_3 + 0.022x_4 + 0.010x_5 + 0.001x_6 \end{bmatrix}$$

2<sup>nd</sup> rule:

IF: d1, d2, d3, d4, d5, and d6 are Low,

$$\text{THEN: } f_2(x) = \begin{bmatrix} 0.477 + 0.191x_1 - 0.184x_2 - 0.061x_3 + 0.367x_4 + 0.000x_5 + 0.004x_6, \\ 0.014 - 0.343x_1 - 0.077x_2 - 0.284x_3 + 0.066x_4 + 0.000x_5 - 0.004x_6 \end{bmatrix}$$

3<sup>rd</sup> rule:

IF: d1, d2, d3, d4, d5, and d6 are MH,

$$\text{THEN: } f_3(x) = \begin{bmatrix} -0.030 - 0.030x_1 - 0.030x_2 - 0.029x_3 - 0.026x_4 - 0.005x_5 - 0.001x_6, \\ 0.030 + 0.030x_1 + 0.030x_2 + 0.029x_3 + 0.026x_4 + 0.005x_5 + 0.001x_6 \end{bmatrix}$$

4<sup>th</sup> rule:

IF: d1, d2, d3, d4, d5, and d6 are ML,

$$\text{THEN: } f_4(x) = \begin{bmatrix} 0.264 - 0.268x_1 - 0.016x_2 - 0.192x_3 + 0.337x_4 - 0.489x_5 + 0.009x_6, \\ 0.226 + 0.163x_1 - 0.134x_2 + 0.023x_3 - 0.270x_4 + 0.489x_5 - 0.009x_6 \end{bmatrix}$$

5<sup>th</sup> rule:

IF: d1, d2, d3, d4, d5, and d6 are Middle,

$$\text{THEN: } f_1(x) = \begin{bmatrix} -0.105 + 0.020x_1 + 0.021x_2 + 0.037x_3 + 0.012x_4 + 0.000x_5 + 0.000x_6, \\ 0.188 + 0.072x_1 + 0.329x_2 - 0.091x_3 - 0.128x_4 + 0.000x_5 + 0.000x_6 \end{bmatrix}$$

For ease of understanding, Fig. 6 shows the decision flow diagram of the model, where ‘ $\oplus$ ’ represents the value of the cumulative rules and ‘Max’ represents the principle of “winner takes all”.

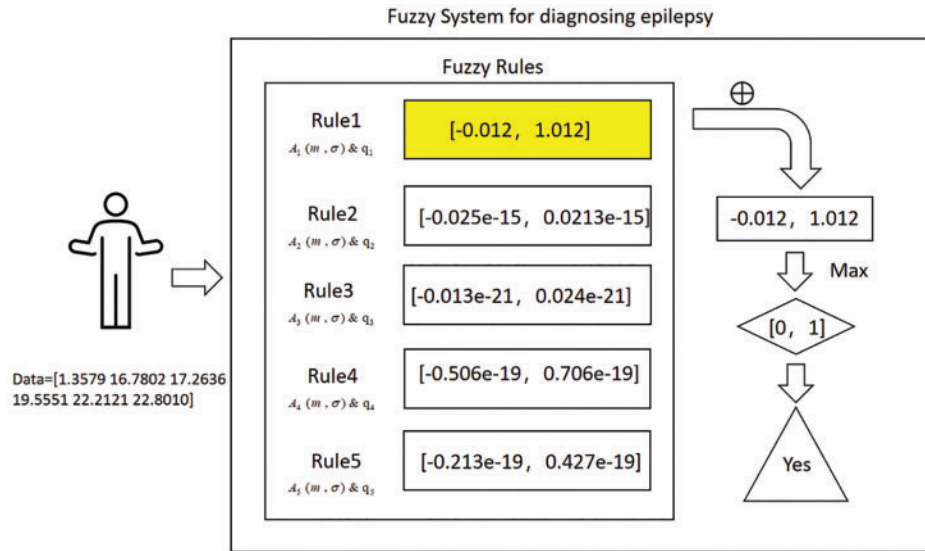


Figure 6: Decision flow diagram of the model

#### 4.4 Time Complexity Analysis

Taking the AC-AC dataset with STFT characteristics as an example, we show the time consumption of different optimization methods in the TSK fuzzy system. The results are shown in Table 10. In the experiment, the number of rules is set to 5, and the number of iterations in the BP solution method is set to 1000. The least-squares method is the fastest to optimize, while the back-propagation method requires more time. However, for medical data identification, it is acceptable that a more accurate diagnosis requires more time.

Table 10: Time consumption (seconds) for each model

LS-TSK-FS	0.00021*
BP-TSK-FS	0.05327
SVM	0.06481
S4VM	0.36716
CD-BP-TSK-FS	0.55132

Since the CD-BP-TSK-FS model is an optimization method using gradient descent to obtain the optimal parameters after processing, more time is required for the parameter optimization process. Furthermore, the comparison method uses the least square method, and they can obtain the analytic solution directly, which makes them faster.

In terms of time consumption, compared with the other methods mentioned above, our method is indeed at a disadvantage, but this is also an unavoidable problem of the backpropagation optimization

method. We will carry out future work to address this problem and to find a new optimization method to reduce the optimization time and improve the efficiency of the model.

## 5 Conclusion

The purpose of this paper is to develop an interpretable model for epileptic EEG signal recognition. This model is based on the TSK-fuzzy system, combines the ideas of transfer learning and semi-supervised learning, makes full use of false labels to enhance data features, and adjusts the data distribution to better mine hidden information in data. To achieve this goal, we proposed CD-TSK fuzzy systems. Experimental tests show that the new method can accurately predict epilepsy data in different scenarios with only a small number of samples. Our future work will concentrate on reducing computing costs and applying the model in related areas, such as brain-computer interfaces.

**Funding Statement:** This work is supported by the Fifth Key Project of Jiangsu Vocational Education Teaching Reform Research under Grant ZZZ13, and in part by the Science and Technology Project of Changzhou City under Grant CE20215032.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

1. Gu, X., Zhang, C., Ni, T. (2020). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(5), 1679–1687. <https://doi.org/10.1109/TCBB.2020.3006699>
2. Jiang, Y., Deng, Z., Chung, F. L., Wang, G., Qian, P. et al. (2016). Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system. *IEEE Transactions on Fuzzy Systems*, 25(1), 3–20. <https://doi.org/10.1109/TFUZZ.2016.2637405>
3. Jiang, Y., Xue, J., Wang, R., Xia, K., Gu, X. et al. (2019). Seizure recognition using a novel multitask radial basis function neural network. *Journal of Medical Imaging and Health Informatics*, 9(9), 1865–1870. <https://doi.org/10.1166/jmihi.2019.2807>
4. Gu, X., Shen, Z., Qu, J., Ni, T. (2022). Cross-domain EEG signal classification via geometric preserving transfer discriminative dictionary learning. *Multimedia Tools and Applications*, 81, 41733–41750. <https://doi.org/10.1007/s11042-021-11244-w>
5. Yu, G. R., Huang, L. W. (2019). Control of robotic Arm based on TSK fuzzy model and hierarchical genetic algorithms. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 1425–1430. Bari.
6. de los Angeles Hernandez, M., Melin, P., Méndez, G. M., Castillo, O., López-Juarez, I. (2015). A hybrid learning method composed by the orthogonal least-squares and the back-propagation learning algorithms for interval A2-c1 type-1 non-singleton type-2 TSK fuzzy logic systems. *Soft Computing*, 19(3), 661–678. <https://doi.org/10.1007/s00500-014-1287-8>
7. Wu, D., Yuan, Y., Huang, J., Tan, Y. (2019). Optimize TSK fuzzy systems for regression problems: Minibatch gradient descent with regularization, DropRule, and AdaBound (MBGD-RDA). *IEEE Transactions on Fuzzy Systems*, 28(5), 1003–1015. <https://doi.org/10.1109/TFUZZ.2019.2958559>
8. Tao, Y., Jiang, Y., Dong, X., Zhou, L., Ding, Y. et al. (2022). Domain adaptive extreme learning machine for epileptic EEG classification. *Journal of Intelligent & Fuzzy Systems*, 43(3), 3983–3992. <https://doi.org/10.3233/JIFS-212068>

9. Shi, Z., Wu, D., Guo, C., Zhao, C., Cui, Y. et al. (2021). FCM-RDpA: TSK fuzzy regression model construction using fuzzy C-means clustering, regularization, droprule, and powerball adabelief. *Information Sciences*, 574, 490–504. <https://doi.org/10.1016/j.ins.2021.05.084>
10. Lin, W. W., Mak, M. W., Chien, J. T. (2018). Multisource i-vectors domain adaptation using maximum mean discrepancy based autoencoders. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(12), 2412–2422. <https://doi.org/10.1109/TASLP.6570655>
11. Ren, W., Zhou, S., Yang, Z., Shi, Q., Sun, X. et al. (2021). Metric information matrix for maximum mean discrepancy for domain adaptation. *IEEE Access*, 9, 148017–148023. <https://doi.org/10.1109/ACCESS.2021.3123281>
12. Sharmila, A., Geethanjali, P. J. I. A. (2016). DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers. *IEEE Access*, 4, 7716–7727. <https://doi.org/10.1109/ACCESS.2016.2585661>
13. Ni, T., Gu, X., Zhang, C. (2020). An intelligence EEG signal recognition method via noise insensitive TSK fuzzy system based on interclass competitive learning. *Frontiers in Neuroscience*, 14, 837. <https://doi.org/10.3389/fnins.2020.00837>
14. Jiang, Y., Gu, X., Ji, D., Qian, P., Xue, J. et al. (2020). Smart diagnosis: A multiple-source transfer TSK fuzzy system for EEG seizure identification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 16(2s), 1–21. <https://doi.org/10.1145/3340240>
15. Zhuang, J., Tang, T., Ding, Y., Tatikonda, S. C., Dvornek, N. et al. (2020). Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. *Advances in Neural Information Processing Systems*, 33, 18795–18806.
16. Ali, M. N. Y., Sarowar, M. G., Rahman, M. L., Chaki, J., Dey, N. et al. (2019). Adam deep learning with SOM for human sentiment classification. *International Journal of Ambient Computing and Intelligence*, 10(3), 92–116. <https://doi.org/10.4018/IJACI>
17. Guerrero, M. C., Parada, J. S., Espitia, H. E. (2021). EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. *Heliyon*, 2021, e07258. <https://doi.org/10.1016/j.heliyon.2021.e07258>
18. Li, Y. F., Zhou, Z. H. (2014). Towards making unlabeled data never hurt. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1), 175–188.
19. Ding, S., Zhu, Z., Zhang, X. (2017). An overview on semi-supervised support vector machine. *Neural Computing and Applications*, 28(5), 969–978. <https://doi.org/10.1007/s00521-015-2113-7>
20. Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*, 7, 1–30.
21. Jiang, Y., Wu, D., Deng, Z., Qian, P., Wang, J. et al. (2017). Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12), 2270–2284. <https://doi.org/10.1109/TNSRE.7333>
22. Abe, F. (2022). Probability, *P* values, and statistical significance: Instructions for use by surgeons. *British Journal of Surgery*, 12, znac440. <https://doi.org/10.1093/bjs/znac440>
23. Tao, Y., Jiang, Y., Xia, K., Xue, J., Zhou, L. et al. (2021). Classification of EEG signals in epilepsy using a novel integrated TSK fuzzy system. *Journal of Intelligent & Fuzzy Systems*, 40(3), 4851–4866. <https://doi.org/10.3233/JIFS-201673>
24. Gu, X., Zhang, C., Ni, T. (2019). Feature selection and rule generation integrated learning for takagi-sugeno-kang fuzzy system and its application in medical data classification. *IEEE Access*, 7, 169029–169037. <https://doi.org/10.1109/Access.6287639>
25. Gu, X., Chung, F. L., Wang, S. (2017). Bayesian takagi–Sugeno–Kang fuzzy classifier. *IEEE Transactions on Fuzzy Systems*, 25(6), 1655–1671. <https://doi.org/10.1109/TFUZZ.2016.2617377>