



Identification of Epileptic EEG Signals Through TSK Transfer Learning Fuzzy System

Zhaoliang Zheng¹, Xuan Dong², Jian Yao¹, Leyuan Zhou³, Yang Ding³ and Aiguo Chen^{1*}

¹ School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China, ² Department of Tuberculosis and Respiratory Diseases, Wuhan Jinyintan Hospital, Wuhan, China, ³ Department of Radiotherapy, Affiliated Hospital, Jiangnan University, Wuxi, China

OPEN ACCESS

Edited by:

Yuanpeng Zhang,
Nantong University, China

Reviewed by:

Yanhui Zhang,
Hebei University of Chinese Medicine,
China

Min Shi,

Fuzhou University of International
Studies and Trade, China

*Correspondence:

Aiguo Chen
agchen@jiangnan.edu.cn

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 08 July 2021

Accepted: 16 August 2021

Published: 10 September 2021

Citation:

Zheng Z, Dong X, Yao J, Zhou L,
Ding Y and Chen A (2021)
Identification of Epileptic EEG Signals
Through TSK Transfer Learning Fuzzy
System. *Front. Neurosci.* 15:738268.
doi: 10.3389/fnins.2021.738268

We propose a new model to identify epilepsy EEG signals. Some existing intelligent recognition technologies require that the training set and test set have the same distribution when recognizing EEG signals, some only consider reducing the marginal distribution distance of the data while ignoring the intra-class information of data, and some lack of interpretability. To address these deficiencies, we construct a TSK transfer learning fuzzy system (TSK-TL) based on the easy-to-interpret TSK fuzzy system the transfer learning method. The proposed model is interpretable. By using the information contained in the source domain and target domains more effectively, the requirements for data distribution are further relaxed. It realizes the identification of epilepsy EEG signals in data drift scene. The experimental results show that compared with the existing algorithms, TSK-TL has better performance in EEG recognition of epilepsy.

Keywords: epilepsy EEG signals, TSK fuzzy system, transfer learning, knowledge learning, interpretability

INTRODUCTION

Epilepsy is a disease caused by the sudden discharge of cerebral neurons. EEG technology (Suk et al., 2018) can monitor the changes of brain electrical signals, so we often use EEG intelligent recognition technology to detect epilepsy (Litt et al., 2001; Iasemidis et al., 2003; Dorai and Ponnambalam, 2010). Nowadays, many machines learn algorithms (Wang et al., 2015; Zhang et al., 2021) have been used to recognize epileptic signals, such as Decision Tree (Wang et al., 2015), nearest neighbor (KNN) (Iskan et al., 2011), Naive Bayes Algorithm (NB; Iskan et al., 2011), support vector machines (SVM; Yang et al., 2014), and fuzzy system (Aarabi et al., 2009; Rabbi and Fazel-Rezai, 2012; Deng et al., 2014a,b, 2018; Jiang et al., 2015). It has been proved that these algorithms can detect epilepsy faster and more accurately than doctors. However, as shown in **Figure 1**, only when the training set and test set obey the similar distribution can they show good classification performance. However, in most cases, as shown in **Figure 2**, the distribution characteristics of EEG data are not exactly the same. In order to make full use of their similar information, some researchers have proposed to use transfer learning algorithm, such as LMPROJ (Yang et al., 2014) and STL (Wang et al., 2018), applying the old knowledge we have gained to new fields. Different from traditional machine learning, which acquires knowledge from data and applies it to new problems, transfer learning focuses on transferring the learned knowledge to new

problems. Although the problem of different data distribution has been solved to a certain extent, these transfer learning algorithms only consider reducing the marginal distribution probability or conditional distribution probability (Deng et al., 2018) of data, without comprehensive balance, and these algorithms lack of interpretability.

To solve these problems, we propose a new method of EEG recognition based on transfer learning and a fuzzy system. The traditional method has a single model structure, so it cannot achieve good results in the face of complex scenes. Different from the traditional method, we pay attention to how to make full use of the previously marked data while ensuring the accuracy of the model on the new task. We not only minimize the marginal distribution or conditional distribution probability, but also combine them to minimize the joint probability distribution (Deng et al., 2018), and reach the best balance between marginal distribution and conditional distribution. In terms of ensuring interpretability, we use the TSK fuzzy system. Its IF-THEN rules can help us understand the rules of model operation more clearly. It has been widely used in data flow modeling, mining tasks, metacognitive learning, and multi-task learning. In order to realize this system, a TSK fuzzy system construction method based on transfer learning (TSK-TL) was developed. The experimental results show that compared with the existing algorithms, TSK-TL has better performance in EEG recognition of epilepsy.

Our contributions are mainly reflected on: (1) The introduction of transfer learning technology (Wang and Mahadevan, 2011; Quanz et al., 2012; Xiao and Guo, 2012), the proposed model in ensuring the accuracy of recognition At the same time, it has higher interpretability; (2) It has stronger robustness and can handle more complex data scenes; (3) It realizes the more accurate identification of epileptic EEG signals in data drift scenarios.

The rest of the manuscript is organized as follows. In the section “Backgrounds,” we briefly introduced the EEG data set, the classical TSK model and the related contents of transfer learning. In section “Identification of Epileptic EEG Signals Through TSK Transfer Learning fuzzysystem,” we first introduced the framework based on transfer learning, and then proposed the objective function of the TSK-TL. In section “Experimental Process and Result Analysis,” we introduce the details of our experiment to test the performance of TSK-TL. The conclusion is given in the last section.

BACKGROUNDS

This section introduces the data sets and their processing methods used in the research, the classical TSK fuzzy system and the related content of transfer learning.

Epilepsy EEG Signal Dataset

The original epileptic EEG data set used in this study is divided into five groups, i.e., Group A to Group E, each group contains 100 single-channel signal segments, and the sampling rate of all samples is adjusted to 173.6 Hz. Among them, the data from

healthy people are divided into groups A and B. The difference is that the eyes of group A are opened and group B is closed. The data of groups C, D, and E are obtained from volunteers with epilepsy in different states (Li, 2021). **Figure 3** shows the five groups of original epilepsy EEG signals. **Table 1** presents these five groups in detail.

The distribution law of EEG signals changes with time. Its amplitude is very small, and it is easy to be affected by other human biological currents such as ECG, EOG, and EMG. At the same time, it has strong randomness, and the noise in the signal is very complicated. Therefore, the experimental results obtained by using the original EEG signal directly are not ideal. According to previous work (Jiang et al., 2017; Tsujikawal et al., 2018), WPD, STFT, and KPCA are three typical feature extraction methods to process epileptic EEG signals (Blanco et al., 1997; Zhang et al., 2000; Srinivasan et al., 2005; Vivaldi and Bassi, 2006; Tzallas et al., 2009; Tang and Durand, 2012; Teng et al., 2017). **Figure 4** shows the sample of group A processed by the three feature extraction methods.

Classical TSK Fuzzy System

Because of its unique interpretability, fuzzy systems has been widely used in modeling and intelligent control. In addition, the output of TSK fuzzy system is more concise. The training process can be transformed into a linear regression problem or a quadratic programming problem, which makes the training process more efficient (Deng et al., 2012; Jiang et al., 2017).

The inference rules of TSK fuzzy system are usually defined as:

$$R^k : \text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_d \text{ is } A_d^k, \tag{1}$$

$$\text{Then } f^k(\mathbf{x}) = p_0^k + p_1^k x_1 + \dots + p_d^k x_d, k = 1, \dots, K$$

K is the number of fuzzy rules. Each rule is premised on the input vector $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$ and maps the fuzzy set in the input space $A^k \subset \mathbb{R}^d$ to the change single case represented by $f^k(\mathbf{x})$. A_i^k is the fuzzy subset of the i th dimension of the input vector \mathbf{x} under the k th rule. \wedge is a fuzzy conjunction operator. According to previous work (Jiang et al., 2017) the result of the TSK fuzzy model can be expressed as

$$y^0 = \frac{\sum_{k=1}^K \mu^k(\mathbf{x}) f^k(\mathbf{x})}{\sum_{k'=1}^K \mu^{k'}(\mathbf{x})} = \sum_{k=1}^K \tilde{\mu}^k(\mathbf{x}) f^k(\mathbf{x}) \tag{2.a}$$

where

$$\mu^k(\mathbf{x}) = \prod_{i=1}^d \mu_{A_i^k}(x_i) \tag{2.b}$$

and

$$\tilde{\mu}^k(\mathbf{x}) = \mu^k(\mathbf{x}) / \sum_{k'=1}^K \mu^{k'}(\mathbf{x}) \tag{2.c}$$

we use the Gaussian membership function, i.e.,

$$\mu_{A_i^k}(x_i) = \exp\left(\frac{-(x_i - c_i^k)^2}{2\delta_i^k}\right) \tag{2.d}$$

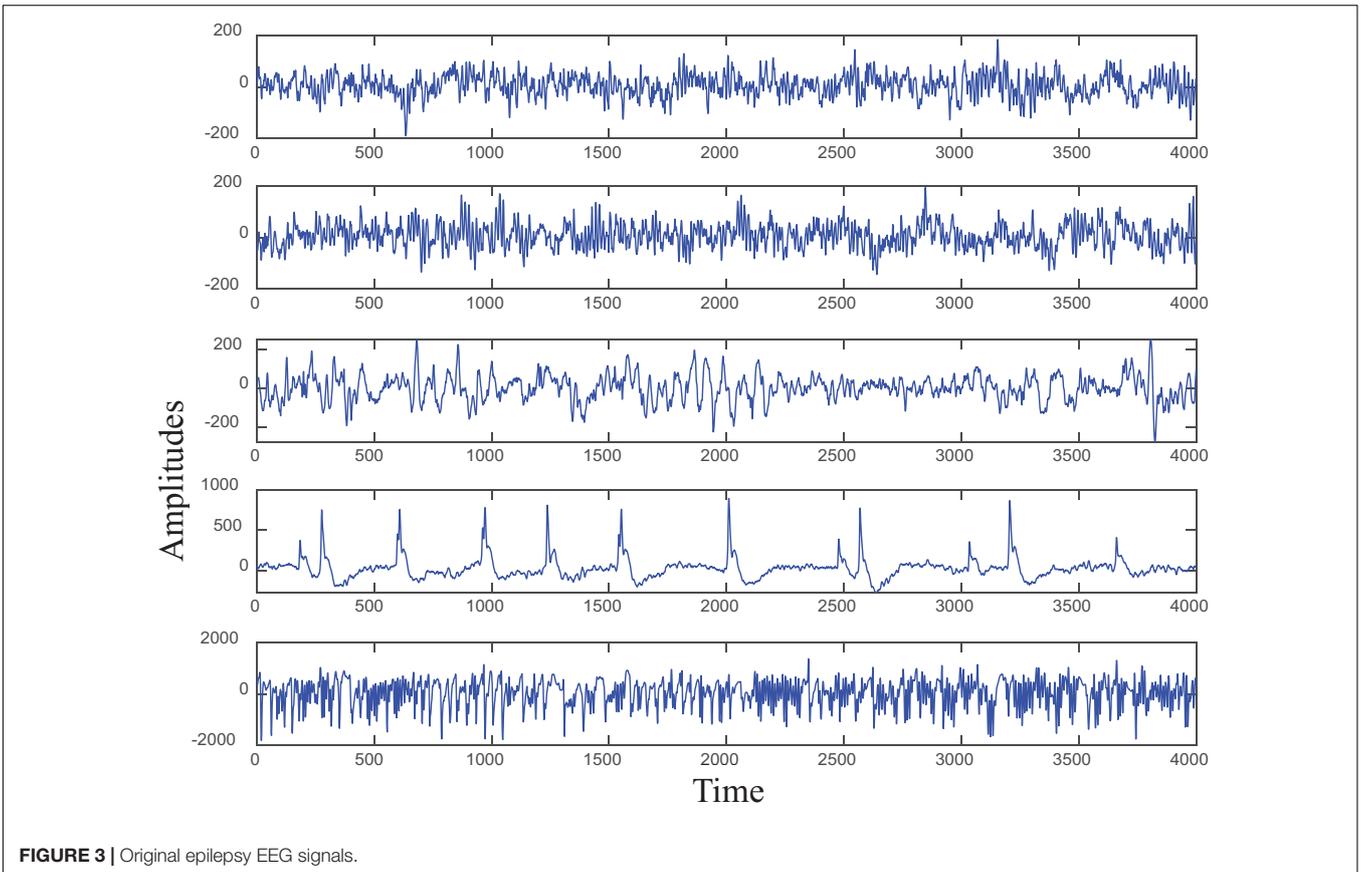
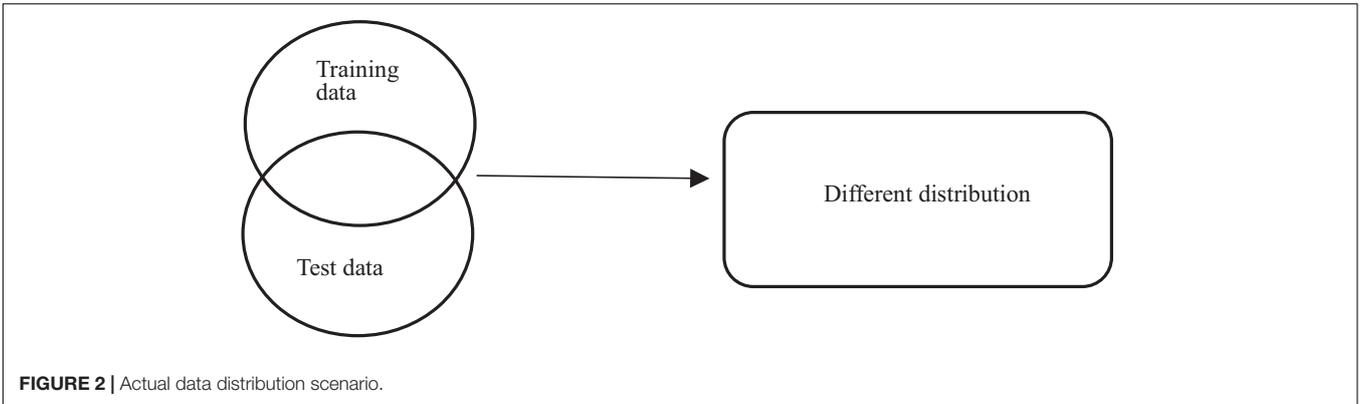
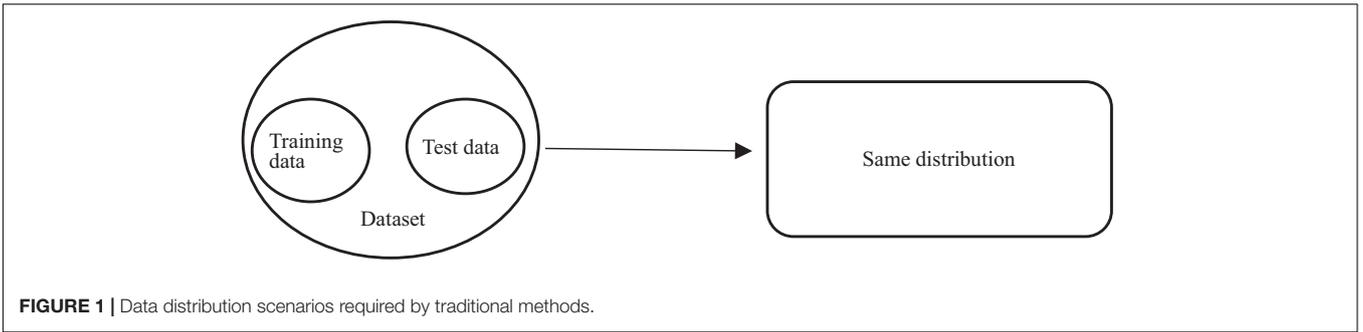


TABLE 1 | Description of the dataset.

Volunteer	Group	Size of group	Description
Healthy volunteer	A	100	EEG signals in healthy volunteers with eyes open
	B	100	EEG signals in healthy volunteers with eyes closed
Volunteer with epilepsy	C	100	EEG signals of the hippocampal structure of the brain during the intermittent period of epilepsy
	D	100	EEG signals in epileptic areas of the brain during intermittent epilepsy
	E	100	EEG signals measured in epileptic patients during the seizure

as the fuzzy membership function. In the paper, we use the FCM algorithm to obtain c_i^k and δ_i^k . They can be estimated by the following expressions

$$c_i^k = \frac{\sum_{j=1}^N u_{jk} x_{ji}}{\sum_{j=1}^N u_{jk}} \tag{2.e}$$

$$\delta_i^k = h \cdot \frac{\sum_{j=1}^N u_{jk} (x_{ji} - c_i^k)^2}{\sum_{j=1}^N u_{jk}} \tag{2.f}$$

u_{jk} is the fuzzy membership corresponding to the j th sample in the k th cluster. And h is the artificially adjustable scale parameter. After determining these antecedent parameters, let

$$\mathbf{x}_e = (\mathbf{1}, \mathbf{x}^T)^T \tag{3.a}$$

$$\tilde{\mathbf{x}}^k = \tilde{\mu}^k(\mathbf{x}) \mathbf{x}_e \tag{3.b}$$

$$\mathbf{x}_g = ((\tilde{\mathbf{x}}^1)^T, (\tilde{\mathbf{x}}^2)^T, \dots, (\tilde{\mathbf{x}}^K)^T)^T \tag{3.c}$$

$$\mathbf{p}^k = (p_0^k, p_1^k, \dots, p_d^k)^T \tag{3.d}$$

$$\mathbf{p}_g = ((\mathbf{p}^1)^T, (\mathbf{p}^2)^T, \dots, (\mathbf{p}^K)^T)^T \tag{3.e}$$

According to the above transformation, Eq. 2a be converted to the following linear regression problem (Jiang et al., 2017).

$$\mathbf{y}^o = \mathbf{p}_g^T \mathbf{x}_g \tag{3.f}$$

A well-performing algorithm is proposed (Jiang et al., 2017) to train the classic TSK-FS. The objective function of this algorithm is

$$\min_{\mathbf{p}_g} J_{TSK-FS}(\mathbf{p}_g) = \frac{1}{2} \mathbf{p}_g^T \mathbf{p}_g + \frac{\lambda_1}{2} \|\mathbf{p}_g^T \mathbf{X} - \mathbf{Y}\|^2 \tag{4}$$

where $\frac{1}{2} \mathbf{p}_g^T \mathbf{p}_g$ is the regularization term, which can effectively promote the generalization ability of the TSK-FS; \mathbf{p}_g is a consequent parameter; \mathbf{X} is the matrix obtained by (5c); λ_1 is a regularization parameter. It can adjust the balance between model complexity and error tolerance; $[\mathbf{Y} = y_0, y_1, \dots, y_n]$ is the label matrix.

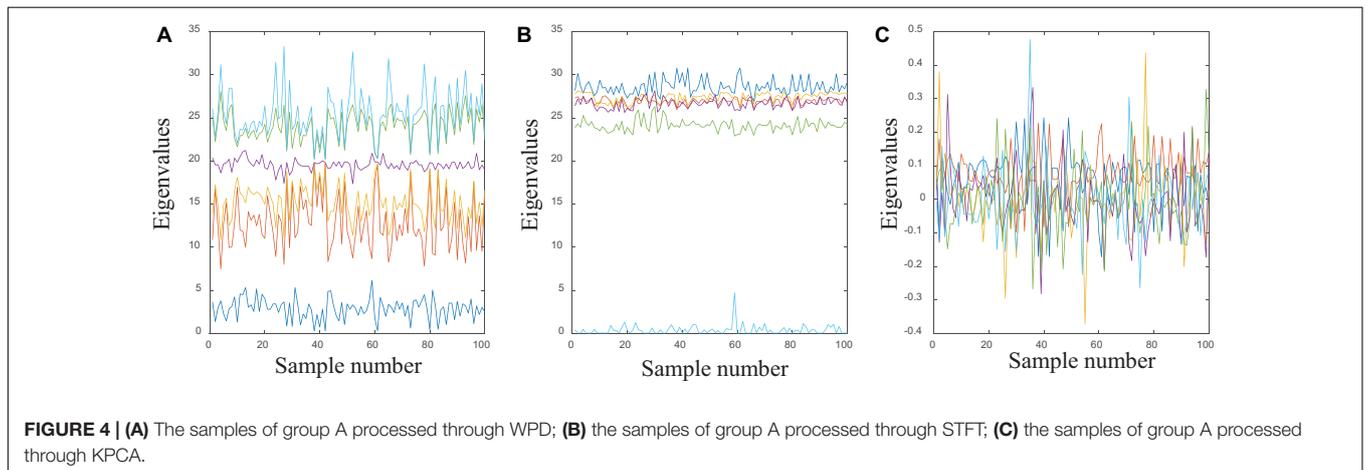
In order to obtain the optimal \mathbf{p}_g , the derivative of $J_{TSK-FS}(\mathbf{p}_g)$ with respect to \mathbf{p}_g can be set to 0, and then the optimal solution of \mathbf{p}_g can be obtained as follows:

$$\mathbf{p}_g = (\mathbf{I} + \lambda_1 \cdot \mathbf{X}\mathbf{X}^T)^{-1} \cdot (\lambda_1 \cdot \mathbf{X}\mathbf{Y}) \tag{5}$$

Through the optimal prior and posterior parameters, we can establish a classic TSK fuzzy system.

Preparatory Knowledge of Transfer Learning

There are three methods for transfer learning: data distribution adaptation, feature selection, and subspace learning (Shi et al., 2013; Zheng, 2021). The basic concept of data distribution adaptation is to make the probability distribution of the data of the source domain (\mathcal{D}_s) and the target domain (\mathcal{D}_t) the same or similar through some transformations. Feature selection method considers that the source domain and the target domain contain some common features, and their data distribution is similar. Then, the common features are extracted through the machine learning method, and the model can be built based on these features. The subspace learning method usually assumes that the data of the source domain and target domain



will have similar distributions in the transformed subspace, and then learn through statistical feature transformation or manifold transformation.

The joint distribution adaptation (Wang et al., 2018) adopted in the paper belongs to the data distribution adaptive methods. Specifically, the core of joint distribution adaptation is to simultaneously minimize the marginal probability distribution and the conditional probability distribution of the two domains. The distance in machine learning has various forms. Here we use the MMD (Long et al., 2012) as a distance measurement. It can be calculated as follows:

$$MMD_H^2(D_s, D_t) = \left\| \frac{1}{n} \sum_{i=1}^n \varnothing(\mathbf{x}_i) - \frac{1}{m} \sum_{j=n+1}^{n+m} \varnothing(\mathbf{x}_j) \right\|^2 \quad (6)$$

where \varnothing is a feature mapping, n is the number of samples in the source domain and m indicates the number of samples in the target domain.

IDENTIFICATION OF EPILEPTIC EEG SIGNALS THROUGH TSK TRANSFER LEARNING FUZZY SYSTEM

In this section, we will introduce in detail the transfer learning techniques we use. Combined with the analysis and research on the rules and parameter learning strategies of the classical TSK fuzzy system, a TSK-TL method for detecting epileptic signals is proposed.

Framework Based on Transfer Learning

The transfer learning strategy used in this study is divided into two parts: joint distribution adaptation (Zheng, 2021) and historical knowledge learning mechanism. As shown in Figure 5, the framework of epilepsy EEG signals recognition based on transfer learning theory is given. In order to make

full use of the information of source domain and target domain, our work mainly includes three steps: (1) minimizing the marginal probability distribution, (2) minimizing the conditional probability distribution, and (3) further learning with historical knowledge.

(1) Minimize the marginal probability distribution

The marginal distributions of the source domain and target domain are represented by P_s and P_t , respectively. The paper builds a new model based on TSK-FS, \mathbf{P}_g can be taken as a projected vector, \mathbf{x}_g is a projected vector. And in order to make MMD a proper regularization for the classifier, we adopt the projected MMD (Long et al., 2013). Then the marginal probability distribution can be obtained by

$$D(P_s, P_t) = \left\| \frac{1}{n} \sum_{i=1}^n \mathbf{P}_g^T \mathbf{x}_{si} - \frac{1}{m} \sum_{j=1}^m \mathbf{P}_g^T \mathbf{x}_{tj} \right\|^2 \quad (7)$$

where \mathbf{x}_{si} is the i th sample of the source domain, \mathbf{x}_{tj} is the j th sample of the target domain, n and m indicate the number of samples in the source domain and target domain, respectively.

(2) Minimize the conditional probability distribution

Reducing the distance of conditional distribution is actually to achieve intra-class migration, but we don't know the label of the target domain. In the paper, we use some traditional classification algorithms (such as SVM) to obtain the pseudo-label of the target domain. At the same time, we assume that the calculated pseudo-class centroid may be located not far from the real class centroid (Long et al., 2012). Therefore, we can calculate the distance of conditional probability distribution by using both true label and pseudo label. The conditional probability distribution of source domain and target domain are represented by Q_s and Q_t , and then the conditional distribution can be calculated as follows:

$$D^{(c)}(Q_s, Q_t) = \left\| \frac{1}{n^{(c)}} \sum_{\mathbf{x}_i \in \mathcal{D}_s^{(c)}} \mathbf{P}_g^T \mathbf{x}_i - \frac{1}{m^{(c)}} \sum_{\mathbf{x}_j \in \mathcal{D}_t^{(c)}} \mathbf{P}_g^T \mathbf{x}_j \right\|^2 \quad (8)$$

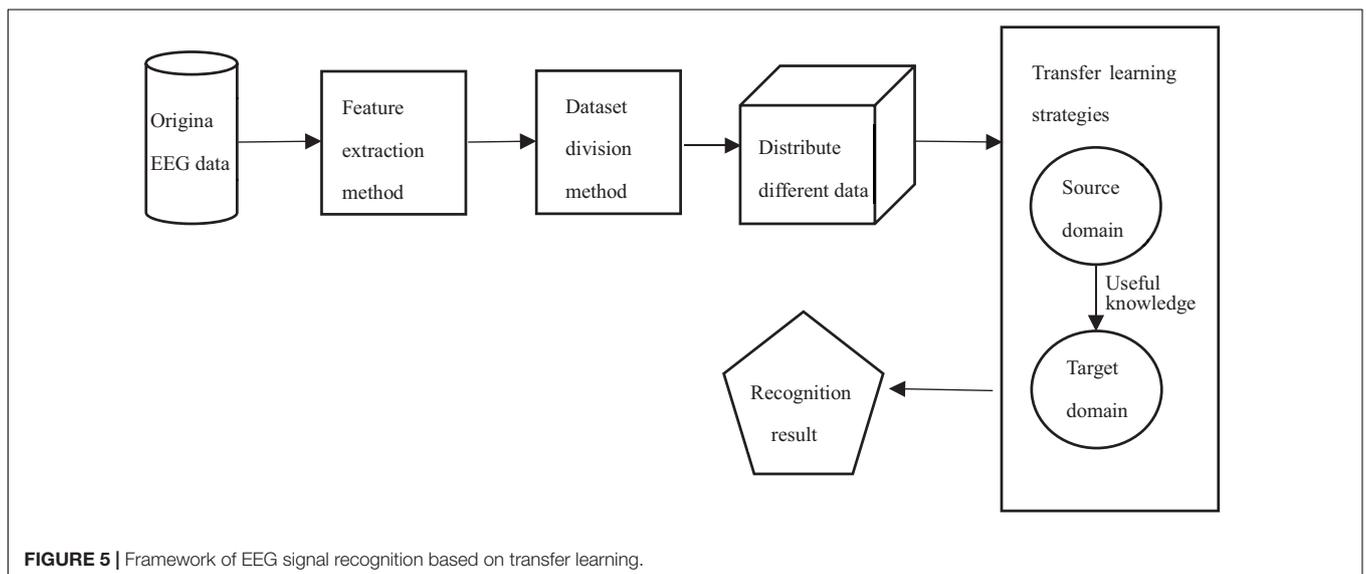


FIGURE 5 | Framework of EEG signal recognition based on transfer learning.

where $c \in \{0, 1, \dots, C\}$ is the category tag. $\mathcal{D}_s^{(c)}$ is a set of the data belonging to class c in the source domain and $n^{(c)} = |\mathcal{D}_s^{(c)}|$. Correspondingly, $\mathcal{D}_t^{(c)}$ is a set of the data belonging to class c in the target domain and $m^{(c)} = |\mathcal{D}_t^{(c)}|$.

By integrating Equations 9 and 10, the joint probability distribution distance can be calculated as follows:

$$D(J_s, J_t) = D(P_s, P_t) + \sum_{c=1}^C D^{(c)}(Q_s, Q_t) \tag{9}$$

(3) Combine the historical knowledge to further learn

The parameter \mathbf{P}_{g_0} obtained through classic TSK-FS is used to further guide the learning, and then the complete transfer learning item is

$$D(\mathcal{D}_s, \mathcal{D}_t) = D(J_s, J_t) + \lambda_3 D(\mathbf{P}_{g_0}, \mathbf{P}_g) \tag{10}$$

where

$$D(\mathbf{P}_{g_0}, \mathbf{P}_g) = \|\mathbf{P}_{g_0} - \mathbf{P}_g\|^2 \tag{11}$$

The Objective Function of TSK-TL

We design the objective function of TSK-TL as

$$\min_{\mathbf{P}_g} J_{TSK-TL}(\mathbf{P}_g) = g + \lambda_2 D(J_s, J_t) + \lambda_3 D(\mathbf{P}_{g_0}, \mathbf{P}_g) \tag{12.a}$$

where

$$g = \frac{1}{2} \mathbf{P}_g^T \mathbf{P}_g + \frac{\lambda_1}{2} \|\mathbf{P}_g^T \mathbf{X}_s - \mathbf{Y}\|^2 \tag{12.b}$$

$$\begin{aligned} D(J_s, J_t) &= D(P_s, P_t) + \sum_{c=1}^C D(Q_s, Q_t) \\ &= \left\| \frac{1}{n} \sum_{i=1}^n \mathbf{P}_g^T \mathbf{x}_{si} - \frac{1}{m} \sum_{j=1}^m \mathbf{P}_g^T \mathbf{x}_{tj} \right\|^2 \\ &+ \left\| \frac{1}{n^{(c)}} \sum_{\mathbf{x}_i \in \mathcal{D}_s^{(c)}} \mathbf{P}_g^T \mathbf{x}_i - \frac{1}{m^{(c)}} \sum_{\mathbf{x}_j \in \mathcal{D}_t^{(c)}} \mathbf{P}_g^T \mathbf{x}_j \right\|^2 \end{aligned} \tag{12.c}$$

$$D(\mathbf{P}_{g_0}, \mathbf{P}_g) = \|\mathbf{P}_{g_0} - \mathbf{P}_g\|^2 \tag{12.d}$$

where \mathbf{P}_g is the expected projection of TSK-TL, \mathbf{P}_{g_0} is the consequent parameter of the classical TSK model. \mathbf{X}_s is a data matrix from the source domain. $\mathbf{Y} = [y_0, y_1, \dots, y_n]$ is the label matrix, if the i th sample belongs to healthy volunteers, then y_i is 1, otherwise y_i is -1. And $\lambda_1, \lambda_2, \lambda_3$ are the regularization parameters.

Then we further explain the above formula as follows:

- (1) Equation 12b is the training model of the classic TSK fuzzy system, so the TSK-TL we proposed inherits all its advantages.
- (2) When experimenting with the classical TSK fuzzy system, the result is poor because of the distribution difference of data. By Equation 12c, the joint probability distribution distance is minimized to optimize the experimental results.

- (3) In Equation 12d, \mathbf{P}_g is further optimized by measuring the distance between \mathbf{P}_{g_0} and \mathbf{P}_g . If λ_3 is infinite and the term is optimal, then \mathbf{P}_{g_0} and \mathbf{P}_g are equal.
- (4) The regularization parameters $\lambda_1 = 0, \lambda_2 = 0,$ and $\lambda_3 = 0$ are used to control the balance between different terms. We use the grid search method to determine their values.

Solution of TSK-TL

In Equation 12c, the first term can be converted as follows:

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{i=1}^n \mathbf{P}_g^T \mathbf{x}_{si} - \frac{1}{m} \sum_{j=1}^m \mathbf{P}_g^T \mathbf{x}_{tj} \right\|^2 \\ &= \left\| \frac{1}{n} \mathbf{P}_g^T [\mathbf{x}_{s1} \ \mathbf{x}_{s2} \ \dots \ \mathbf{x}_{sn}]_{1 \times n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1} \right. \\ &\quad \left. - \frac{1}{m} \mathbf{P}_g^T [\mathbf{x}_{t1} \ \mathbf{x}_{t2} \ \dots \ \mathbf{x}_{tm}]_{1 \times m} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{m \times 1} \right\|^2 \\ &= \text{tr} \left(\frac{1}{n^2} \mathbf{P}_g^T \mathbf{X}_s \mathbf{1} (\mathbf{P}_g^T \mathbf{X}_s \mathbf{1})^T + \frac{1}{m^2} \mathbf{P}_g^T \mathbf{X}_t \mathbf{1} (\mathbf{P}_g^T \mathbf{X}_t \mathbf{1})^T \right. \\ &\quad \left. - \frac{1}{nm} \mathbf{P}_g^T \mathbf{X}_s \mathbf{1} (\mathbf{P}_g^T \mathbf{X}_s \mathbf{1})^T - \frac{1}{nm} \mathbf{P}_g^T \mathbf{X}_t \mathbf{1} (\mathbf{P}_g^T \mathbf{X}_t \mathbf{1})^T \right) \\ &\quad + \text{tr} \left[\mathbf{P}_g^T \left(\frac{1}{n^2} \mathbf{1} \mathbf{1}^T \mathbf{X}_s^T \mathbf{X}_s + \frac{1}{m^2} \mathbf{1} \mathbf{1}^T \mathbf{X}_t^T \mathbf{X}_t \right. \right. \\ &\quad \left. \left. - \frac{1}{nm} \mathbf{1} \mathbf{1}^T \mathbf{X}_s^T \mathbf{X}_t - \frac{1}{nm} \mathbf{1} \mathbf{1}^T \mathbf{X}_t^T \mathbf{X}_s \right) \mathbf{P}_g \right] \\ &= \text{tr} \left(\mathbf{P}_g^T \mathbf{X} \mathbf{M}_0 \mathbf{X}^T \mathbf{P}_g \right) \end{aligned}$$

Similarly, the second term can be converted as:

$$\text{tr} \left(\mathbf{P}_g^T \mathbf{X} \mathbf{M}_c \mathbf{X}^T \mathbf{P}_g \right)$$

where \mathbf{X} is a matrix composed of the source domain and target domain data. \mathbf{M}_c is MMD matrix computed as:

$$(\mathbf{M}_c)_{ij} = \begin{cases} \frac{1}{n^{(c)}n^{(c)}}, & \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_s^{(c)} \\ \frac{1}{m^{(c)}m^{(c)}}, & \mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_t^{(c)} \\ \frac{-1}{n^{(c)}m^{(c)}}, & \begin{cases} \mathbf{x}_i \in \mathcal{D}_s^{(c)}, \mathbf{x}_j \in \mathcal{D}_t^{(c)} \\ \mathbf{x}_j \in \mathcal{D}_s^{(c)}, \mathbf{x}_i \in \mathcal{D}_t^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

So

$$D(J_s, J_t) = \text{tr} \left(\mathbf{P}_g^T \mathbf{X} \mathbf{M} \mathbf{X}^T \mathbf{P}_g \right) \quad \mathbf{M} = \sum_{c=0}^C \mathbf{M}_c$$

Then,

$$\min_{\mathbf{P}_g} J_{TSK-TL}(\mathbf{P}_g) = \frac{1}{2} \mathbf{P}_g^T \mathbf{P}_g + \frac{\lambda_1}{2} \|\mathbf{P}_g^T \mathbf{X}_s - \mathbf{Y}\|^2 + \lambda_2 \text{tr}(\mathbf{P}_g^T \mathbf{X} \mathbf{M} \mathbf{X}^T \mathbf{P}_g) + \lambda_3 \|\mathbf{P}_{g_0} - \mathbf{P}_g\|^2$$

By setting the derivative of J_{TSK-TL} with respect to \mathbf{P}_g to zero, i.e., $\frac{\partial J_{TSK-TL}}{\partial \mathbf{P}_g} = 0$, we can get the optimal solution of \mathbf{P}_g as follows:

$$\mathbf{P}_g = \left(\mathbf{X} \mathbf{M} + \mathbf{X} \mathbf{M}^T \right) \mathbf{X}^T + \lambda_1 \mathbf{X}_s \mathbf{X}_s^T + \lambda_3 \mathbf{I} \Big)^{-1} (\lambda_1 \mathbf{X}_s \mathbf{Y}^T + \lambda_3 \mathbf{P}_{g_0}) \tag{14}$$

By (7), (15), and (16), we can obtain the optimal posterior parameter \mathbf{P}_g . Then based on \mathbf{P}_g , the final decision function can be obtained as follows

$$f(\mathbf{x}_{gi}) = \text{sign}(\mathbf{P}_g \mathbf{x}_{gi}) \tag{15}$$

The details of the proposed TSK-TL algorithm are as follows:

Algorithm of TSK-TL.

Initialization: Set the number of fuzzy rules K , the regularization parameters λ_1 , λ_2 , and λ_3 .

Stage 1: Construction of datasets for linear regression

Step 1: Through classical FCM or other partition techniques to divide the input space of training data to determine the premise of TSK-FS.

Step 2: Construct the new training dataset and test dataset through (3a)–(3c).

Stage 2: Computation of historical knowledge parameter \mathbf{P}_{g_0}

Step 3: Obtain the \mathbf{P}_{g_0} by (5).

Stage 3: Obtain the parameter \mathbf{P}_g of TSK-TL

Step 4: Compute the MMD matrix by (13).

Step 5: Compute the consequent parameter \mathbf{P}_g of TSK-TL through (14).

Step 6: Generate the TSK-TL by (15).

EXPERIMENTAL PROCESS AND RESULTS ANALYSIS

In this section, the proposed TSK-TL method is evaluated by classifying EEG signals of epilepsy patients and healthy people. In addition, a comparative study of five traditional machine learning algorithms and two transfer learning algorithms is carried out. Details of the experiments are as follows.

Experimental Setup

In this paper, we use three classical feature extraction methods, namely WPD, STFT, and KPCA, to obtain EEG datasets from three different views and perform experiments on them, respectively. In each experiment, we set up 10 experimental datasets, and every dataset is composed of part of the data from two or three different groups in this view. The details of these datasets are shown in **Table 2**. The structure of this dataset ensures that there is no or only a part of the data coming from the same group. In short, there is no or only a part of the data in the source domain and the test domain has the same distribution.

TABLE 2 | Structure of experimental datasets.

Datasets	Source domain	Target domain
D1	Group A 50, Group E 50	Group A 50, Group C 50
D2	Group A 50, Group E 50	Group A 50, Group D 50
D3	Group B 50, Group E 50	Group B 50, Group C 50
D4	Group B 50, Group E 50	Group B 50, Group D 50
D5	Group A 50, Group E 50	Group B 50, Group C 50
D6	Group A 50, Group E 50	Group B 50, Group D 50
D7	Group B 50, Group E 50	Group A 50, Group D 50
D8	Group B 50, Group E 50	Group A 50, Group C 50
D9	Group A 50, Group B 50, Group E 50	Group A 50, Group B 50, Group C 50
D10	Group A 50, Group B 50, Group E 50	Group A 50, Group B 50, Group D 50

TABLE 3 | Experimental setup.

Methods for comparison	Classical recognition methods	Transfer learning recognition methods
	1. KNN 2. NB 3. SVM 4. C4.5 5. TSK-FS	1. LMPROJ 2. STL 3. TSK-TL
Data label	1. EEG signals from healthy volunteers are labeled “1” 2. EEG signals from volunteers with epilepsy are labeled “-1”	
Performance evaluation	1. Accuracy: The ratio of the number of test data with correct prediction results to the total number of test data 2. Friedman test and Holm’s post hoc test	
Method-specific settings	1. For TSK-FS, the number of fuzzy rules is taken the set {5, 10, 15, 20, 25, 30} and the regularization parameter λ_1 from the set $\{10^{-5}, 10^{-4}, \dots, 10^4, 10^5\}$ 2. For the proposed TSK-TL, the number of fuzzy rules is taken from the set {5, 10, 15, 20, 25, 30}, and the regularization parameters λ_1 , λ_2 and λ_3 from the set $\{10^{-5}, 10^{-4}, \dots, 10^4, 10^5\}$	
	Note: Each parameter is determined by the grid search	

During the experiment, the experiment of each experimental dataset is repeated 10 times. And take the average result of 10 times to evaluate the performance. The optimal hyperparameters of each experimental model are determined by the grid search. All the algorithms are implemented using MATLAB on a computer with Intel i5-4590 3.3 GHz CPU, 12 GB of RAM. The details of the experimental setup are shown in **Table 3**.

Recognition Performance

In our experiment, five traditional machine learning algorithms and two transfer learning algorithms, i.e., LMPROJ and STL are used for comparative experiments. The results are shown in **Tables 4–6**. By observing these results, three conclusions can be drawn as follows.

- (1) The proposed TSK-TL method not only reduces the difference of data distribution, but also optimizes the formation of new knowledge parameters through the prior knowledge parameters obtained from the classical TSK-FS model. The results show that our method can improve the accuracy of EEG signal recognition.

TABLE 4 | Experimental results of eight classifiers in the WPD view.

EEG datasets		Classifiers							
		KNN	NB	SVM	C4.5	TSK-FS	STL	LMPR-OJ	TSK-TL
D1	Mean (SD)	0.8771 (0.0275)	0.5636 (0.0180)	0.9086 (0.0135)	0.8845 (0.0254)	0.9157 (0.0079)	0.8529 (0.0160)	0.9286 (0.0157)	0.9586 (0.0168)
D2	Mean (SD)	0.8857 (0.0223)	0.6127 (0.0119)	0.9300 (0.0153)	0.8818 (0.0218)	0.9329 (0.0170)	0.8843 (0.0321)	0.9357 (0.0127)	0.9571 (0.0111)
D3	Mean (SD)	0.9623 (0.0113)	0.8373 (0.0385)	0.9571 (0.0160)	0.9645 (0.0104)	0.9614 (0.0069)	0.9789 (0.0038)	0.9773 (0.0049)	0.9971 (0.0049)
D4	Mean (SD)	0.9819 (0.0049)	0.8345 (0.0550)	0.9600 (0.0100)	0.9655 (0.0082)	0.9586 (0.0090)	0.9717 (0.0049)	0.9835 (0.0071)	0.9957 (0.0053)
D5	Mean (SD)	0.7529 (0.0472)	0.2655 (0.0151)	0.5900 (0.0183)	0.7936 (0.0566)	0.7271 (0.0287)	0.8629 (0.0522)	0.6829 (0.0801)	0.9543 (0.0199)
D6	Mean (SD)	0.7286 (0.0313)	0.3164 (0.0234)	0.5943 (0.0140)	0.7745 (0.0591)	0.7571 (0.0298)	0.8457 (0.0544)	0.6000 (0.0870)	0.9329 (0.0359)
D7	Mean (SD)	0.9157 (0.0230)	0.7927 (0.0447)	0.7914 (0.0313)	0.7727 (0.0174)	0.9043 (0.0190)	0.9171 (0.0298)	0.8900 (0.0258)	0.9614 (0.0234)
D8	Mean (SD)	0.8914 (0.0267)	0.8309 (0.0411)	0.7843 (0.0113)	0.8136 (0.0344)	0.8986 (0.0302)	0.9297 (0.0168)	0.8757 (0.0172)	0.9571 (0.0256)
D9	Mean (SD)	0.9143 (0.0254)	0.7152 (0.0233)	0.7439 (0.0101)	0.8994 (0.0212)	0.9217 (0.0112)	0.9367 (0.0159)	0.9286 (0.0171)	0.9714 (0.0120)
D10	Mean (SD)	0.9057 (0.0190)	0.7382 (0.0161)	0.7305 (0.0127)	0.9006 (0.0232)	0.9189 (0.0071)	0.9133 (0.0154)	0.9190 (0.0141)	0.9657 (0.0071)

TABLE 5 | Experimental results of eight classifiers in the STFT view.

EEG datasets		Classifiers							
		KNN	NB	SVM	C4.5	TSK-FS	STL	LMPROJ	TSK-TL
D1	Mean (SD)	0.6390 (0.0265)	0.5682 (0.0227)	0.9729 (0.0111)	0.8336 (0.0150)	0.8343 (0.1409)	0.8871 (0.0138)	0.8828 (0.0138)	0.9800 (0.0100)
D2	Mean (SD)	0.6575 (0.0341)	0.6309 (0.0187)	0.9429 (0.0137)	0.8373 (0.0142)	0.8368 (0.0548)	0.9014 (0.0107)	0.8771 (0.0143)	0.9857 (0.0079)
D3	Mean (SD)	0.5823 (0.0322)	0.5964 (0.0211)	0.9243 (0.0113)	0.7427 (0.1144)	0.9543 (0.0190)	0.8997 (0.0117)	0.9157 (0.0121)	0.9976 (0.0032)
D4	Mean (SD)	0.6060 (0.0481)	0.6409 (0.0489)	0.9647 (0.0053)	0.6927 (0.0827)	0.9171 (0.0496)	0.9167 (0.0026)	0.9529 (0.0049)	0.9986 (0.0053)
D5	Mean (SD)	0.62300 (0)	0.5473 (0.0233)	0.6029 (0.0150)	0.6727 (0.0236)	0.5943 (0.0812)	0.8014 (0.0324)	0.6314 (0.0564)	0.8583 (0.0676)
D6	Mean (SD)	0.5147 (0.0379)	0.4209 (0.0130)	0.5829 (0.0236)	0.6891 (0.0329)	0.5786 (0.1289)	0.7335 (0.0191)	0.6134 (0.0437)	0.8371 (0.0399)
D7	Mean (SD)	0.6015 (0.0452)	0.5909 (0.0212)	0.8400 (0.0163)	0.7127 (0.0163)	0.9071 (0.0160)	0.9046 (0.0107)	0.9171 (0.0160)	0.9786 (0.0079)
D8	Mean (SD)	0.6390 (0.0335)	0.6145 (0.0273)	0.8629 (0.0221)	0.6455 (0.0281)	0.9171 (0.0206)	0.9217 (0.0121)	0.9083 (0.0163)	0.9657 (0.0111)
D9	Mean (SD)	0.6837 (0.0136)	0.6739 (0.0020)	0.8990 (0.0186)	0.8133 (0.0149)	0.8805 (0.0214)	0.9049 (0.0076)	0.9129 (0.0065)	0.9857 (0.0067)
D10	Mean (SD)	0.7130 (0.0183)	0.7048 (0.0052)	0.9162 (0.0076)	0.7818 (0.0085)	0.8967 (0.0144)	0.9123 (0.0233)	0.9076 (0.0081)	0.9771 (0.0115)

TABLE 6 | Experimental results of eight classifiers in the KPCA view.

EEG datasets		Classifiers							
		KNN	NB	SVM	C4.5	TSK-FS	STL	LMPROJ	TSK-TL
D1	Mean (SD)	0.8813 (0.0336)	0.7760 (0.0219)	0.8167 (0.0462)	0.8955 (0.0121)	0.8700 (0.0394)	0.8414 (0.0308)	0.7781 (0.0164)	0.9120 (0.0311)
D2	Mean (SD)	0.9050 (0.0225)	0.8940 (0.0477)	0.9067 (0.0321)	0.9789 (0.0083)	0.9557 (0.0336)	0.9214 (0.0090)	0.8940 (0.0351)	0.9820 (0.0130)
D3	Mean (SD)	0.9295 (0.0206)	0.9120 (0.0152)	0.8500 (0.0361)	0.8382 (0.0325)	0.9360 (0.0152)	0.9200 (0.0100)	0.8120 (0.0130)	0.9500 (0.0187)
D4	Mean (SD)	0.9450 (0.0176)	0.9260 (0.0114)	0.8367 (0.0252)	0.9591 (0.0138)	0.9680 (0.0045)	0.9343 (0.0127)	0.8127 (0.0158)	0.9820 (0.0084)
D5	Mean (SD)	0.7034 (0.0526)	0.5560 (0.0195)	0.6067 (0.0115)	0.5945 (0.0543)	0.7240 (0.0658)	0.8217 (0.0492)	0.7020 (0.0327)	0.8289 (0.0430)
D6	Mean (SD)	0.7045 (0.0331)	0.6137 (0.0122)	0.6833 (0.0451)	0.7700 (0.0300)	0.7380 (0.0311)	0.8700 (0.0336)	0.7423 (0.0268)	0.8743 (0.0217)
D7	Mean (SD)	0.8827 (0.0239)	0.8109 (0.0138)	0.8700 (0.0400)	0.9218 (0.0271)	0.8943 (0.0292)	0.9186 (0.0135)	0.8343 (0.0151)	0.9420 (0.0228)
D8	Mean (SD)	0.8945 (0.0331)	0.8064 (0.0234)	0.8467 (0.0058)	0.7682 (0.0282)	0.8780 (0.0277)	0.8471 (0.0275)	0.8000 (0.0200)	0.9200 (0.0158)
D9	Mean (SD)	0.9063 (0.0158)	0.7358 (0.0414)	0.6689 (0.0038)	0.8345 (0.0333)	0.8687 (0.0197)	0.8771 (0.0148)	0.8667 (0.0054)	0.9280 (0.0179)
D10	Mean (SD)	0.9533 (0.0169)	0.7655 (0.0409)	0.6667 (0.0017)	0.9810 (0.0054)	0.9520 (0.0073)	0.9210 (0.0178)	0.9371 (0.0143)	0.9827 (0.0060)

- (2) Compared with the two transfer learning classifiers LMPROJ and STL, TSK-TL is superior to LMPROJ in recognition of epileptic EEG signals, and its performance is better than or at least comparable to the STL method. STL method uses affinity within the class to transfer knowledge within the class, and then to learn. The LMPROJ method only reduces the marginal distribution of data based on the SVM algorithm.
- (3) The TSK-TL method we proposed has a good recognition effect in the three views, and the average recognition accuracy in the WPD view is the highest, which points out the direction for us to select an appropriate feature

extraction method in practical applications. At the same time, it can be observed that when processing the D5 and D6 data sets, the recognition accuracy is greatly improved compared with other methods, which further verifies that our proposed method has a better effect in the face of data with large differences in data distribution.

Statistical Analysis

We evaluated the experimental results from a statistical point of view through the Friedman test (Demšar, 2006; Garcia and Herrera, 2008) and the Holm's *post hoc* test (Demšar, 2006; Garcia and Herrera, 2008). Friedman test is used to calculate the average

TABLE 7 | Comparison of average performance of TSK-TI and other seven methods using Friedman test ($\alpha = 0.05$).

Friedman Test for TSK-TL			
Algorithms	Friedman Rank	p-Value	H ₀
KPCA view			
KNN	3.8	0.000000006179	Rejected
NB	7.05		
SVM	6.4		
C4.5	4.2		
TSK-FS	3.4		
STL	3.8		
LMPROJ	6.35		
TSK-TL	1		
STFT view			
KNN	7	0.00000001016	Rejected
NB	7.8		
SVM	3.6		
C4.5	5.3		
TSK-FS	4.6		
STL	3.2		
LMPROJ	3.5		
TSK-TL	1		
WPD view			
KNN	4.5	0.000000004808	Rejected
NB	7.5		
SVM	6.5		
C4.5	5.4		
TSK-FS	4.2		
STL	3.3		
LMPROJ	3.6		
TSK-TL	1		

ranking of the compared methods and to determine whether the observed difference are statistically significant. We set the significance level of the test to 0.05. If the *p*-value is less than 0.05, the null hypothesis H₀ is rejected and we can confirm that there are significant difference. The Holm's *post hoc* test is

TABLE 8 | Holm's *post hoc* comparison of the results of Friedman procedure of TSK-TL and other seven methods

Holm's <i>post hoc</i> comparison for TSK-TL					
i	Algorithm	Z = (R ₀ -R _i)/SE	p	Holm = /i	H ₀
KPCA view					
7	NB	5.522869	0	0.007142	Rejected
6	SVM	4.929503	0.000001	0.008333	Rejected
5	LMPROJ	4.883859	0.000001	0.01	Rejected
4	C4.5	2.921187	0.003487	0.0125	Rejected
3	KNN	2.556039	0.010587	0.016667	Rejected
2	STL	2.556039	0.010587	0.025	Rejected
1	TSK-FS	2.19089	0.02846	0.05	Rejected
STFT view					
7	NB	6.207522	0	0.007142	Rejected
6	KNN	5.477226	0	0.008333	Rejected
5	C4.5	3.925345	0.000087	0.01	Rejected
4	TSK-FS	3.286335	0.001015	0.0125	Rejected
3	SVM	2.373464	0.017622	0.016667	Rejected
2	LMPROJ	2.282177	0.022479	0.025	Rejected
1	STL	2.008316	0.04461	0.05	Rejected
WPD view					
7	NB	5.933661	0	0.007142	Rejected
6	SVM	5.02079	0.000001	0.008333	Rejected
5	C4.5	4.016632	0.000059	0.01	Rejected
4	KNN	3.195048	0.001398	0.0125	Rejected
3	TSK-FS	2.921187	0.003487	0.016667	Rejected
2	LMPROJ	2.373464	0.017622	0.025	Rejected
1	STL	2.099603	0.035764	0.05	Rejected

then performed to evaluate the statistical differences between the control (i.e., the method that achieves the best Friedman rank) and other methods. The results of the Friedman test are shown in **Table 7**, and the results of the Holm's *post hoc* test are shown in **Table 8**.

As shown in **Table 7**, the Friedman test results reveal that the TSK-TL method performs better than the other seven classification methods in classification accuracy. And the results

TABLE 9 | TSK-TL Model with five rules trained on the dataset D2 for the KPCA view.

Fuzzy rules base		
TSK Fuzzy Rule R ^k : IF x_1 is $A_1^k(c_1^k, \delta_1^k) \wedge x_2$ is $A_2^k(c_2^k, \delta_2^k) \wedge \dots \wedge x_d$ is $A_d^k(c_d^k, \delta_d^k)$, Then $f_x(x) = p_k^0 + p_k^1x_1 + \dots + p_k^dx_d$		
No. of rules	Antecedent parameters	Consequent parameters
K	$c^k = (c_1^k, \dots, c_d^k)^T, \delta^k = (\delta_1^k, \dots, \delta_d^k)^T$	$P_k = (P_k^0, P_k^1, \dots, P_k^d)^T$
1	$c^1 = [-0.1589, 0.0712, 0.0076, 0.0517, -0.000356, 0.0097]$ $\delta^1 [0.0029, 0.0037, 0.0052, 0.0063, 0.0039, 0.0032]$	$p_1 = [-2.7245, -0.5118, 0.9243, -0.295, -0.0378, -0.58, -1.1201]$
2	$c^2 = [0.0582, -0.0035, -0.0439, 0.1377, -0.0389, -0.0297]$ $\delta^2 [0.0033, 0.0021, 0.0025, 0.0056, 0.0017, 0.0042]$	$p_2 = [10.9842, -8.5028, -0.9881, 4.7443, -1.8692, 3.2596, -3.0533]$
3	$c^3 = [0.0035, 0.1507, 0.0516, -0.1698, -0.1519, 0.0789]$ $\delta^3 [0.0022, 0.0045, 0.0062, 0.0082, 0.0051, 0.0057]$	$p_3 = [2.045, 3.995, -4.3315, -1.4117, -0.9264, -0.0619, 0.4859]$
4	$c^4 = [0.1562, -0.0587, 0.139, 0.0092, 0.1228, -0.1636]$ $\delta^4 [0.0038, 0.0035, 0.005, 0.0054, 0.0033, 0.0072]$	$p_4 = [-3.8877, -2.2364, -0.7276, -1.1165, 0.1062, -1.3669, 0.9594]$
5	$c^5 = [-0.0667, -0.1566, -0.1623, -0.0287, 0.0347, 0.1368]$ $\delta^5 [0.0026, 0.0032, 0.0047, 0.0031, 0.0036, 0.0071]$	$p_5 = [2.0489, 4.0366, -4.2784, -1.4602, -1.0065, 0.0086, 0.9774]$

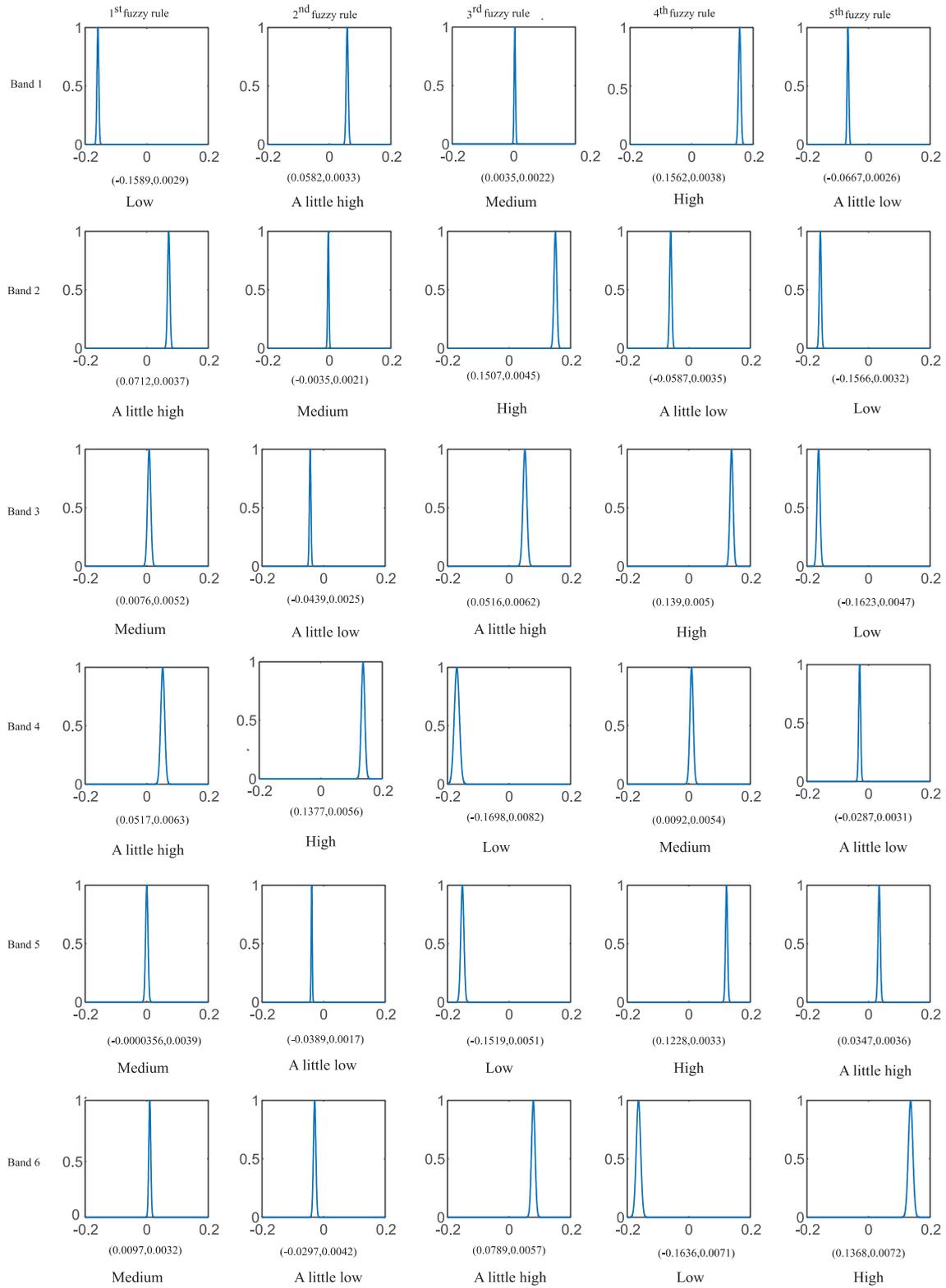


FIGURE 6 | The membership function of each fuzzy rule and the linguistic interpretation of each fuzzy subset of the TSK fuzzy system are obtained from the KPCA view.

of Holm's *post hoc* in **Table 8** also show that compared with other methods, TSK-TL—TL has better performance. This once again proves that our proposed TSK-TL has achieved better results in the detection of epileptic EEG signals.

Model Analysis

In this section, TSK-TL is analyzed through the model trained from the D2 dataset in the KPCA view. **Table 9** gives an example of the model, these model parameters prove that TSK-TL inherits the interpretability of the classical TSK fuzzy system. In this example, we set up five fuzzy rules. **Figure 6** shows the corresponding MF of each fuzzy set, where each MF has a fuzzy rule description, such as “the energy of a band of EEG signal is Low” (A little low, Medium, A little high or High). Because medical experts in different fields have different interpretations of fuzzy rules, the language description given is only a possible example.

According to the central value from low to high, these five MFs can be expressed as “Low,” “A little low,” “Medium,” “A little high,” and “High”. Finally, through the language expression of the IF-part of the fuzzy rule and the linear function corresponding to the THEN-part of the fuzzy rule, five fuzzy rules of the KPCA view are given:

The first fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is Low, A little high, Medium, A little high, Medium, and Medium, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^1(\mathbf{x}) = [-2.7245 - 0.5118x_1 + 0.9243x_2 - 0.295x_3 \\ - 0.0378x_4 - 0.58x_5 - 1.1201x_6]$$

The second fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is A little high, Medium, A little Low, High, A little Low and A little Low, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^2(\mathbf{x}) = [10.9842 - 8.5028x_1 - 0.9881x_2 + 4.7443x_3 \\ - 1.8692x_4 + 3.2596x_5 - 3.0533x_6]$$

The third fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is Medium, High, A little high, Low, Low and A little high, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^3(\mathbf{x}) = [2.045 + 3.995x_1 - 4.3315x_2 - 1.4117x_3 \\ - 0.9264x_4 - 0.0619x_5 + 0.4859x_6]$$

The fourth fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is High, A little low, High, Medium, High and Low, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^4(\mathbf{x}) = [-3.8877 - 2.2364x_1 - 0.7276x_2 \\ - 1.1165x_3 + 0.1062x_4 - 1.3669x_5 - 0.9594x_6]$$

The fifth fuzzy rule:

If the energy of EEG signal from band 1 to band 6 is A little Low, Low, Low, A little low, A little high, and High, respectively,

THEN the decision value under this rule is obtained by the following formula:

$$f^5(\mathbf{x}) = [2.0489 + 4.0366x_1 - 4.2784x_2 - 1.4602x_3 \\ - 1.0065x_4 + 0.0086x_5 + 0.9774x_6]$$

According to the final output value, i.e., $y = -1$ or $y = 1$, it can be judged whether the patient has epilepsy.

CONCLUSION

The study combines the classic TSK fuzzy system with transfer learning technology and proposes a TSK fuzzy system (TSK-TL) that is interpretable and can better adapt to scenarios of data distribution differences. It expands the application scenarios of the model and realizes the recognition of epileptic EEG signals with large data distribution differences in reality.

Although we have proved the effectiveness of TSK-TL, it can be further optimized. For example, there are several predefined parameters in the TSK-TL algorithm. When optimizing them, the optimization process of these parameters takes a lot of time. In the future, we will study the problem and develop more effective algorithms.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/benfulcher/hctsaTutorial_BonnEEG.

AUTHOR CONTRIBUTIONS

ZZ developed the theoretical framework and model in this work and drafted the manuscript. XD gave support for medical knowledge. ZZ, XD, JY, and AC implemented the algorithm and performed experiments and result analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grants U20A20228, 61772241, and 61873321, in part by the Science and Technology Demonstration Project of Social Development of Jiangsu Province under Grant BE2019631, and in part by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127.

REFERENCES

- Aarabi, A., Fazel-Rezai, R., and Aghakhani, Y. (2009). A fuzzy rule-based system for epileptic seizure detection in intracranial EEG. *Clin. Neurophysiol.* 120, 1648–1657. doi: 10.1016/j.clinph.2009.07.002
- Blanco, S., Kochen, S., Rosso, O. A., and Salgado, P. (1997). Applying time-frequency analysis to seizure EEG activity. *IEEE Engin. Med. Biol. Magaz.* 16, 64–71. doi: 10.1109/51.566156
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* 7, 1–30.
- Deng, Z., Choi, K. S., Jiang, Y., and Wang, S. (2014b). Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods. *IEEE Trans. Cybernet.* 44, 2585–2599. doi: 10.1109/tycb.2014.2311014
- Deng, Z., Jiang, Y., Chung, F. L., Choi, K. S., and Wang, S. (2014a). Transfer prototype-based fuzzy clustering. *IEEE Trans. Fuzzy Syst.* 24, 1210–1232. doi: 10.1109/TFUZZ.2015.2505330
- Deng, Z., Jiang, Y., Chung, F. L., Ishibuchi, H., and Wang, S. (2012). Knowledge-leverage-based fuzzy system and its modeling. *IEEE Trans. Fuzzy Syst.* 21, 597–609. doi: 10.1109/TFUZZ.2012.2212444
- Deng, Z., Peng, X., Xie, L., Choi, K. S., and Wang, S. (2018). Transductive joint-knowledge-transfer tsk-fs for recognition of epileptic eeg signals. *IEEE Trans. Neural Syst. Rehabil. Engin.* 26, 1481–1494.
- Dorai, A., and Ponnambalam, K. (2010). *Automated epileptic seizure onset detection. International Conference on Autonomous & Intelligent Systems.* New York, NY: IEEE.
- Garcia, S., and Herrera, F. (2008). An extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all pairwise comparisons. *J. Mach. Learn. Res.* 9, 2677–2694.
- Iasemidis, L. D., Shiau, D. S., Chaovalitwongse, W., Sackellares, J. C., Pardalos, P. M., Principe, J. C., et al. (2003). Adaptive epileptic seizure prediction system. *IEEE Trans. Biomed. Engin.* 50, 616–627.
- Iscan, Z., Dokur, Z., and Demiralp, T. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Syst. Appl.* 38, 10499–10505. doi: 10.1016/j.eswa.2011.02.110
- Jiang, Y., Deng, Z., Chung, F. L., and Wang, S. (2015). Multi-task TSK fuzzy system modeling using inter-task correlation information. *Inform. Sci.* 298, 512–533. doi: 10.1016/j.ins.2014.12.007
- Jiang, Y., Deng, Z., Chung, F. L., Wang, G., Qian, P., Choi, K. S., et al. (2017). Recognition of epileptic eeg signals using a novel multiview tsk fuzzy system. *IEEE Trans. Fuzzy Syst.* 25, 3–20. doi: 10.1109/TFUZZ.2016.2637405
- Li, L. (2021). DSP for condition monitoring equipment in real-time. *ASP Trans. Neural Inform. Comput.* 1, 22–25.
- Litt, B., Esteller, R., Echauz, J., D'Alessandro, M., Shor, R., Henry, T., et al. (2001). Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* 30, 51–64. doi: 10.1016/S0896-6273(01)00262-8
- Long, M., Wang, J., Ding, G., Cheng, W., Zhang, X., and Wang, W. (2012). *Dual transfer learning. In Proceedings of the 2012 SIAM International Conference on Data Mining.* Thailand: Society for Industrial and Applied Mathematics, 540–551.
- Long, M., Wang, J., Ding, G., Sun, J., and Yu, P. S. (2013). *Transfer feature learning with joint distribution adaptation. In Proceedings of the IEEE international conference on computer vision.* New York, NY: IEEE, 2200–2207.
- Quanz, B., Huan, J., and Mishra, M. (2012). Knowledge transfer with low-quality data: A feature extraction issue. *IEEE Trans. Knowl. Data Engin.* 24, 1789–1802. doi: 10.1109/TKDE.2012.75
- Rabbi, A. F., and Fazel-Rezai, R. (2012). A fuzzy logic system for seizure onset detection in intracranial EEG. *Comput. Intell. Neurosci.* 12, 7051401–7051402.
- Shi, X., Liu, Q., Fan, W., and Philip, S. Y. (2013). Transfer across completely different feature spaces via spectral embedding. *IEEE Trans. Knowl. Data Engin.* 25, 906–918. doi: 10.1109/TKDE.2011.252
- Srinivasan, V., Eswaran, C., and Sriraam, A. N. (2005). Artificial neural network based epileptic detection using time-domain and frequency-domain features. *J. Med. Syst.* 29, 647–660. doi: 10.1007/s10916-005-6133-1
- Suk, K. H., Daesub, Y., Soon, S. H., and Hee, P. C. (2018). Predicting the eeg level of a driver based on driving information. *IEEE Trans. Intell. Transport. Syst.* 20, 1215–1225. doi: 10.1109/ITS.2018.2848300
- Tang, Y., and Durand, D. (2012). A tunable support vector machine assembly classifier for epileptic seizure detection. *Expert Syst. Appl.* 39, 3925–3938. doi: 10.1016/j.eswa.2011.08.088
- Teng, T., Bi, L., and Liu, Y. (2017). Eeg-based detection of driver emergency braking intention for brain-controlled vehicles. *IEEE Trans. Intell. Transport. Syst.* 19, 1766–1773. doi: 10.1109/ITS.2017.2740427
- Tsujikawal, M., Onishil, Y., Kiuchil, Y., Ogatsul, T., and Hashimoto, S. (2018). *Drowsiness Estimation from Low-Frame-Rate Facial Videos using Eyelid Variability Features. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* New York, NY: IEEE.
- Tzallas, A. T., Tsipouras, M. G., and Fotiadis, D. I. (2009). Epileptic seizure detection in eegs using time–frequency analysis. *IEEE Trans. Inf. Technol. Biomed.* 13, 703–710. doi: 10.1109/ITB.2009.2017939
- Vivaldi, E. A., and Bassi, A. (2006). *Frequency domain analysis of sleep EEG for visualization and automated state detection. In 2006 International Conference of the IEEE Engineering in Medicine and Biology Society.* New York: IEEE, 3740–3743.
- Wang, C., and Mahadevan, S. (2011). *Heterogeneous domain adaptation using manifold alignment. In Twenty-second international joint conference on artificial intelligence.* Barcelona: IJCAI.
- Wang, G., Deng, Z., and Choi, K. S. (2015). *Detection of Epileptic Seizures in EEG Signals with Rule-Based Interpretation by Random Forest Approach. International Conference on Intelligent Computing.* Cham: Springer.
- Wang, J., Feng, W., Chen, Y., Yu, H., Huang, M., and Yu, P. S. (2018). *Visual domain adaptation with manifold embedded distribution alignment. In Proceedings of the 26th ACM international conference on Multimedia.* New York, NY: Association for Computing Machinery, 402–410.
- Xiao, M., and Guo, Y. (2012). *Semi-supervised kernel matching for domain adaptation. In Twenty-Sixth AAAI Conference on Artificial Intelligence.* Palo Alto, CA: AAAI Press.
- Yang, C., Deng, Z., Choi, K. S., Jiang, Y., and Wang, S. (2014). Transductive domain adaptive learning for epileptic electroencephalogram recognition. *Artific. Intell. Med.* 62, 165–177. doi: 10.1016/j.artmed.2014.10.002
- Zhang, Z., Fan, C., Zhang, L., and Kong, M. (2021). K Algorithm for microstrip peocessor design. *ASP Trans. Neural Inform. Comput.* 1, 15–21.
- Zhang, Z., Kawabata, H., and Liu, Z. Q. (2000). *EEG analysis using fast wavelet transform. In Smc 2000 conference proceedings. 2000 IEEE international conference on systems, man and cybernetics. cybernetics evolving to systems, humans, organizations, and their complex interactions (cat. no. 0, Vol. 4.* New York: IEEE, 2959–2964.
- Zheng, M. (2021). PCNN for power distribution network. *ASP Trans. Neural Inform. Comput.* 1, 9–14.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zheng, Dong, Yao, Zhou, Ding and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.