This is the Pre-Published Version.
The following publication C. Yang, Z. Deng, K. Choi and S. Wang, "Takagi–Sugeno–Kang Transfer Learning Fuzzy Logic System for the Adaptive Recognition of Epileptic Electroencephalogram Signals," in IEEE Transactions on Fuzzy Systems, vol. 24, no. 5, pp. 1079-1094, Oct. 2016 is available at http://dx.doi.org/10.1109/TFUZZ.2015.2501438.

Takagi-Sugeno-Kang Transfer Learning Fuzzy Logic System for the Adaptive Recognition of Epileptic Electroencephalogram Signals

Changjian Yang[1]; Zhaohong Deng[1,2*], *IEEE Senior Member*; Kup-Sze Choi[3], *IEEE Member*; Shitong Wang[1]

[1] School of Digital Media, Jiangnan University, Wuxi, Jiangsu, P. R. China

[2] Department of Biomedical Engineering, University of California, Davis, CA, USA

[3] Centre for Smart Health, School of Nursing, Hong Kong Polytechnic University, Hong Kong

*Corresponding author: dzh666828@aliyun.com, zhdeng@ucdavis.edu

Abstract—The intelligent recognition of electroencephalogram (EEG) signals has become an important approach to the detection of epilepsy. Among existing intelligent identification methods, fuzzy logic systems (FLSs) have shown a distinctive advantage in identifying epileptic EEG signals because of their strong learning abilities and interpretability. Like many conventional intelligent methods for recognizing EEG signals, in the training of FLS it is assumed that the training dataset and test dataset are drawn from data that are identically distributed. However, this assumption is not necessarily valid in practice as it is not uncommon for the two datasets to have different distributions. To overcome this problem, a strategy is presented in this paper to construct a Takagi-Sugeno-Kang (TSK) FLS based on transductive transfer learning for identifying epileptic EEG signals. Two novel objective functions, achieved by integrating the transductive transfer learning mechanism, are proposed for the training of the TSK FLS. As regression and binary classification are two common approaches to multi-class classification, the TSK transfer learning FLS algorithms for regression (TSK-TL-FLS (Reg)) and binary classification (TSK-TL-FLS (BC)) are developed respectively to construct the corresponding TSK FLS. Both algorithms are further used to perform a multi-class classification to recognize epileptic EEG signals. Their performance in the epileptic

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

EEG datasets indicates promise in dealing with situations where the training and test datasets differ with regard to data distribution.

Key words—electroencephalogram, epilepsy detection, distribution diversity, TSK fuzzy logic system, transfer learning, feature extraction.

I INTRODUCTION

Epilepsy is a transient brain dysfunction syndrome caused by the sudden abnormality of brain neurons [1]. It is a relatively common brain disorder that is characterized by repeated seizures. Studies show that abnormalities are detected in the electroencephalogram (EEG) signals of 80% of epileptic patients. Hence, the identification of epileptic EEG signals plays an important role in the diagnosis of epilepsy, especially for those whose condition is difficult to diagnose clinically, such as those with atypical epileptic seizures, masked epilepsy, and other unusual forms of epilepsy.

Identifying epileptic EEG signals involves two main steps: feature extraction and classification. There are three classic methods for extracting features in EEG signals: time-domain analysis methods [12], frequency-domain analysis methods [13], and time-frequency analysis methods [14-16]. Many intelligent classification methods have also been adopted to identify epileptic EEG signals [2-11, 26-33]. Representative methods include the naive Bayes algorithm (NB) [2], linear discriminant analysis (LDA)[3], the decision tree algorithm (DT) [4-5], the K-neighbor algorithm (KNN) [6], the support vector machine (SVM) [7], and fuzzy logic systems (FLSs) [8-11]. This study focuses on the development of classification methods that are adaptive for recognizing epileptic EEG signals. Among existing classification models, the superior interpretability and strong learning abilities of FLS make it a favorable choice for various practical applications.

Epileptic EEG signals are usually collected under different conditions to facilitate a clinical diagnosis. Typically, the signals are experimentally obtained under five different situations and categorized into the following *class*es: Class 1 – EEG signals obtained from healthy people with their eyes open; Class 2 – EEG signals obtained from healthy people with their eyes closed; Class 3 – EEG signals acquired from the hippocampus formation of the opposite hemisphere of patients with epilepsy during a seizure-free interval (preictal); Class 4 – EEG signals acquired from the epileptogenic zone of patients with epilepsy during a seizure-free interval (preictal); and Class 5 – EEG signals obtained from patients with epilepsy during a seizure (ictal) [25]. Although FLS and other methods of classification have been shown to be effective at recognizing epileptic EEG signals, unsolved challenges remain that affect the performance of intelligent epileptic EEG recognition systems. One of these is the inevitable variations in EEG signals. Such variations can be caused by: 1) the blinking of eyes, respiration, muscle movements, or other physiological effects; 2) intra-subject variability resulting from changes in the subject's state of mind and mood during the EEG recording; and 3) measurement variations resulting from changes in electrode positions or contact impedance.

As a result, the data distribution of the epileptic EEG signals that are obtained can differ greatly. This presents a great challenge to conventional intelligent model construction, where identical data distribution is assumed. For instance, suppose that a conventional recognition method is used to build a classification model using the labeled data of Class 1 and Class 5 EEG signals (the training data, from healthy subjects and patients respectively). If this model is then used to identify Class 2 and Class 5 signals (the test data, from healthy subjects and patients respectively), the accuracy of the identification will decline drastically because the signals were collected under different conditions and had different data distributions. Existing intelligent recognition methods are thus inappropriate in such cases. Hence, alternative approaches that are more adaptable to drifting in data distribution between the training and test data are needed to meet the challenge of coming up with practical method of identifying epileptic EEG signals.

In the field of EEG-based brain-computer interfaces (BCIs), those researching the subject of the intelligent recognition of brain waves to identify the intention of users also face the same problem of disparity in data distribution [56-60]. Here, much attention has been paid to recognition methods based on transfer learning, which have been used to overcome the problem with some degree of success. Naturally, it is also anticipated that recognition methods based on transfer learning will be useful for recognizing the epileptic EEG signals discussed in this study. The existing transductive transfer learning method – the large-margin-projected transductive support vector machine (LMPROJ) – has indeed been adopted and its performance is promising when compared with conventional methods [24].

Transfer learning has been studied extensively in the past decades for training models by simultaneously using the data in the target domain (the current scene) and the information of the source domains (related scenes) [24, 37-39]. Depending on whether or not the target domain contains labeled samples, transfer learning techniques can be divided into two categories: (i) the inductive transfer learning method and (ii) the transductive transfer learning method. The inductive transfer learning method is mainly used in scenes where a large amount of labeled data are available in the source domain, with only a small amount of labeled data in the target domain. On the other hand, the transductive transfer learning method is mainly applied to those scenes where labeled data are only available in the source domain and the target domain only contains unlabeled data. Clearly, the transductive transfer learning method has broader applicability than its inductive counterpart. The concern in this paper is the problem of how to identify epileptic EEG signals without any labeled data in the target domain. Therefore, the inductive transfer learning method is not appropriate; rather, methods based on transductive transfer learning [24] are investigated here for recognizing epileptic EEG signals.

Furthermore, the Takagi-Sugeno-Kang (TSK) FLS, which is based on transfer learning, is also investigated in this study for its ability to identify epileptic EEG signals. The TSK FLS, as a classic intelligent model, has been studied broadly and used in many fields [41-51]. Although many training algorithms have been proposed for the construction of a TSK FLS, algorithms suitable for transfer learning are still being investigated. To develop a TSK FLS that is more effective for those scenarios where drifting occurs in the distribution of data between the source domain and the target domain, several methods for constructing a TSK FLS based on transfer learning have been proposed [36, 52, 53]. However, these methods were all developed for inductive transfer learning, and thus are not suitable for the recognition of epileptic EEG signals, which is the concern in this study, and which is in essence a transductive learning task. In this paper, we focus on studying methods for constructing a TSK FLS based on transductive transfer learning.

In particular, by incorporating the transductive transfer learning mechanism, it is possible to propose two novel objective functions for training a TSK FLS to carry out regression and binary classification, respectively. The corresponding algorithms are called TSK-TL-FLS (Reg) and TSK-TL-FLS (BC). Both algorithms are further used to perform a multi-class classification for identifying epileptic EEG signals. Experimental studies are conducted to validate whether the proposed methods can achieve better recognition results than the classic methods when there is a difference in data distribution between the training dataset and the test dataset. The experiments also show that even if the data distribution in the two datasets is identical, the performance of the proposed methods is also comparable to that of the classic methods.

The main contributions of our work are as follows: (1) the transductive transfer learning strategy is introduced into the learning procedure for constructing a TSK FLS for recognizing epileptic EEG signals; (2) two methods for constructing a TSK FLS based on transductive transfer learning are proposed for regression and binary classification, respectively, which are then further applied to the multi-class classification of epileptic EEG signals; (3) comprehensive experiments comparing the proposed methods s with related methods are conducted to substantiate the significance of the proposed methods in advancing the recognition of epileptic EEG signals.

The proposed methods are novel in two ways. First, they are distinct from existing FLS construction methods based on transfer learning [36, 52, 53]. The latter were all developed for inductive transfer learning and are not suitable for recognizing epileptic EEG signals because of the absence of labeled data in the target domain. By using the transductive transfer learning mechanism, the proposed methods have solved this problem and achieved promising results. Second, the proposed methods take advantage of methods based on fuzzy rules, including better interpretability, a good ability to learn, and a stronger ability to model uncertainty.

The rest of this paper is organized as follows. The related work, including a review of classic methods for extracting features of epileptic EEG signals, the classification techniques, and the concept of the TSK FLS, are briefly described in section II. In section III, the proposed methods for constructing a TSK FLS based on transductive transfer learning are presented and discussed in detail. Experimental studies on the performance of the proposed methods for recognizing epileptic EEG signals are reported in section IV. Finally, the conclusions are given in section V.

II RELATED WORK

The identification of epileptic EEG signals generally includes two steps. First, feature extraction methods are employed to extract useful features from the original EEG signals. Second, an intelligent model is constructed based on the available datasets. The whole process is shown in Fig. 1. In this section, classic methods of extracting and classifying the features of EEG signals are reviewed briefly. A discussion of the TSK FLS model is also included.

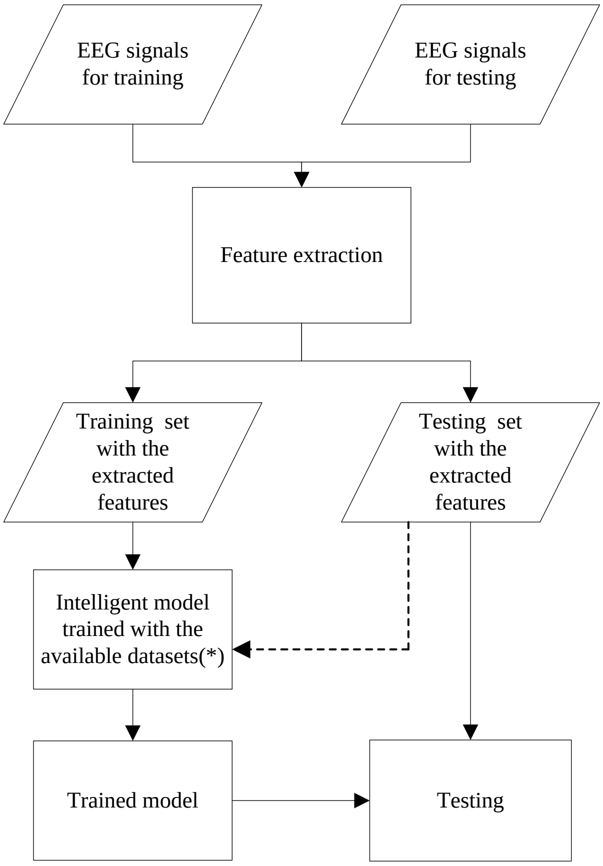

Fig. 1 The process of recognizing EEG signals. (*) denotes that for some methods, e.g. a transductive learning based method, a test dataset is also needed in the model training procedure.

A. Feature Extraction

A time-domain analysis is a common method for extracting features of EEG signals. It characterizes EEG signals based on the properties of the waveform, e.g., wave amplitude and wave kurtosis. Time-domain analysis methods are intuitive in that the features have an explicit physical meaning [12]. Frequency-domain analysis methods make use of the frequency characteristics of the signals to extract information from the brain waves. For example, a frequency-domain analysis method was used to estimate the EEG power spectrum and measure the variations in the typical rhythms of EEG signals [13]. A power spectrum analysis of EEG signals can be performed to transform changes in signal amplitude into changes in signal power so that variations in the brain waves at different frequencies can be directly observed. On the other hand, since EEG signals are non-stationary and highly stochastic in nature, it is insufficient to extract features from either the time domain or the frequency domain [14]. A time-frequency analysis is thus proposed. A popular approach of this kind is wavelet transforms [15, 16], which have been widely applied to analyzing epileptic EEG signals.

In this study, wavelet packet decomposition (WPD) [15] and kernel principal component analysis (KPCA) [34] are adopted for feature extraction. For WPD, a series of binary wavelets are obtained by decomposing the original EEG signals with Daubechies4 (db4) wavelet coefficients. The signals are split into six bands, namely, Band 1 to Band 6, and the corresponding frequency ranges are 0-2 Hz, 2-4 Hz, 4-8 Hz, 8-15 Hz, 16-30 Hz, and 31-60 Hz. The characteristic frequency ranges for a spike wave, a slow wave, and a sharp wave, which are useful characteristics for detecting epilepsy, are 1-2.5 Hz, 5-12.5 Hz, and 13.5-50 Hz, respectively. The WPD method based on db4 wavelet coefficients can exactly split these waves into different bands. An example of decomposed EEG signals is shown in Fig. 2.

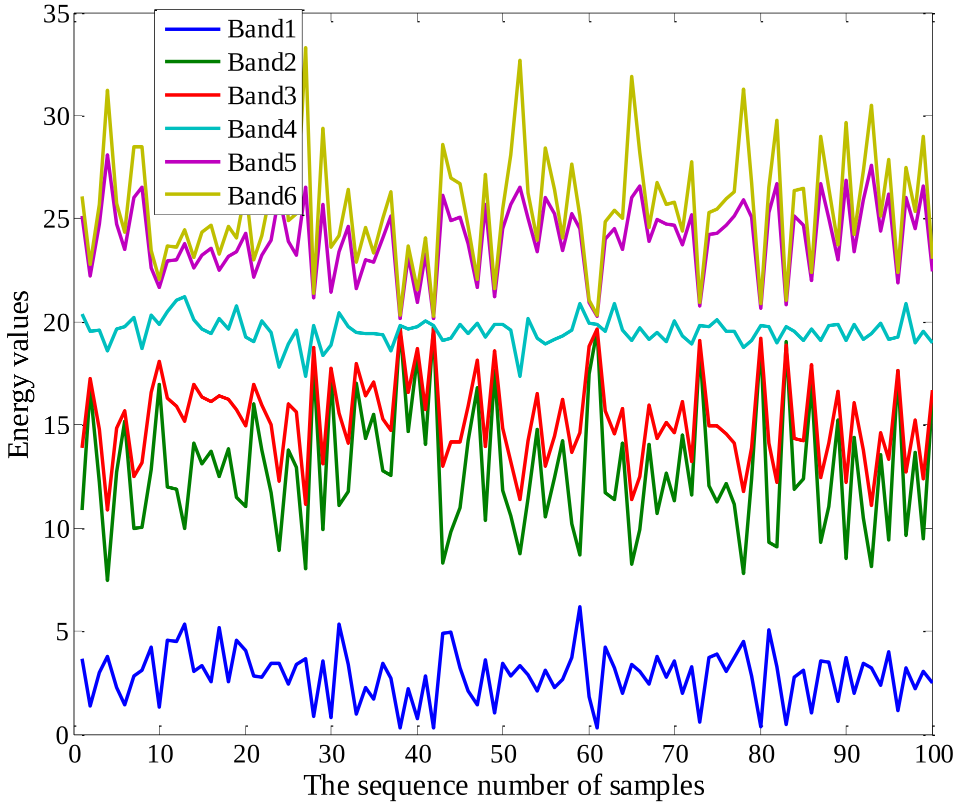

Fig. 2 Waveforms of features extracted using WPD.

In KPCA [34], which is an extension of the well-known principal component analysis (PCA) method commonly used for reducing dimensions, the kernel method is used to transform the original EEG signals into feature space through nonlinear mapping, in order to perform a PCA in high-dimensional space [35]. Given the nonlinearity of EEG signals, the nonlinear mapping of KPCA significantly enhances the ability to extract the features of the signals. An example of the features of the EEG signals extracted by KPCA is shown in Fig. 3.

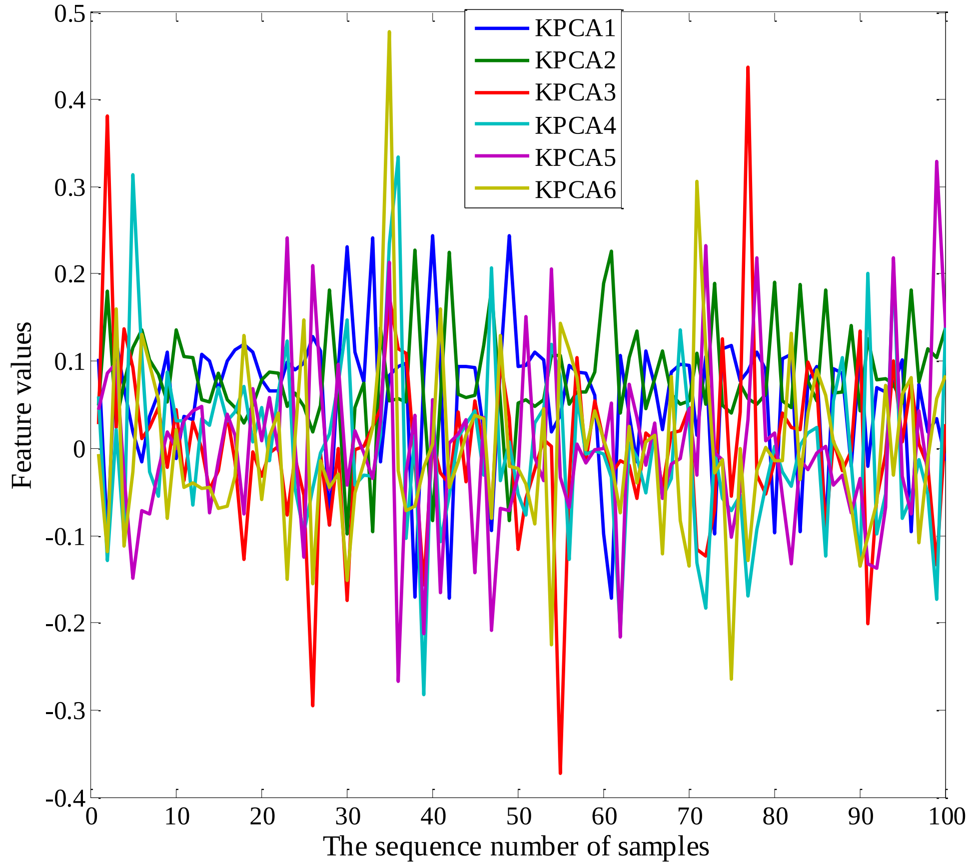

Fig. 3 Waveforms of features extracted using KPCA.

B. Classification Methods

Since the 1990s, many intelligent classification methods have been used to identify epileptic EEG signals [26]-[33]. Examples include LDA [26,27], DT [28], NB[29], the nearest mean algorithm (NM) [30,31,32], and SVM [33]. They have been successfully used in applications concerning the detection of epilepsy. For example, SVM, an effective tool for classification involving small datasets of high dimensionality, has been used for the intelligent detection of epileptic EEG signals [33].

C. *TSK FLS*

1) Concept and principle

The proposed transfer learning FLS is based on the classic TSK FLS model, which is indeed the most widely used FLS model. The fuzzy inference rules of a TSK FLS [17] take the following form:

$$\text{IF } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \ldots \wedge x_d \text{ is } A_d^k$$
$$\text{Then } f^k(\mathbf{x}) = p_0^k + p_1^k x_1 + p_2^k x_2 + \cdots + p_d^k x_d, k = 1, 2, \cdots, K, \tag{1}$$

where x_j, $(j = 1, 2, ..., d)$ is the jth dimension of the input variable and d is the number of the dimensions of the inputs. K is the number of fuzzy rules in the rule base. The expression $f^k(\boldsymbol{x})$ in Eq. (1) represents the output of the kth rule of the TSK FLS. The fuzzy sets in the input space $A^k \subset R^d$ are mapped to the fuzzy sets in the output space $f^k(x) \subset R$. Here, A_i^k is the fuzzy set of the ith dimension of the input variable under the kth rule, and $\wedge$ is a fuzzy conjunction operation. p_j^k, $(j = 0, 1, \ldots, d)$ are the coefficients of the linear functions in the consequents. When multiplicative conjunction, multiplicative implication, and additive combination are employed, the output of the TSK FLS can be formulated as [20]

$$y = \sum_{k=1}^{K} \frac{\mu^k(\mathbf{x})}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x})} f^k(\mathbf{x}) = \sum_{k=1}^{K} \tilde{\mu}^k(\mathbf{x}) f^k(\mathbf{x}) = \sum_{k=1}^{K} \tilde{\mu}^k(x)(p_0^k + p_1^k x_1 + \cdots + p_d^k x_d), \tag{2}$$

with

$$\mu^k(\mathbf{x}) = \prod_{i=1}^{d} \mu_{A_i^k}(x_i), \tag{3}$$

$$\tilde{\mu}^k(\mathbf{x}) = \frac{\mu^k(\mathbf{x})}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x})}. \tag{4}$$

Here, $\mu^k(x)$ is the fuzzy membership value associated with the fuzzy set A^k, which is normalized to $\tilde{\mu}^k(x)$. The commonly used membership function is the Gaussian-like membership function, i.e.,

$$\mu_{A_i^k}(x_i) = \exp(\frac{-(x_i - c_i^k)^2}{2\delta_i^k}), \quad (5)$$

where the center parameters c_i^k and the width parameters δ_i^k can be estimated using different methods, such as by employing classic clustering algorithms or other fuzzy space partition methods. When the fuzzy c-means (FCM) clustering algorithm [18, 19] is adopted to evaluate the above parameters, c_i^k and δ_i^k are given by

$$c_i^k = \frac{\sum_{j=1}^{N} u_{jk} x_{ji}}{\sum_{j=1}^{N} u_{jk}}, \quad (6)$$

$$\delta_i^k = \frac{h \cdot \sum_{j=1}^{N} u_{jk}(x_{ji} - c_i^k)^2}{\sum_{j=1}^{N} u_{jk}}, \quad (7)$$

where u_{jk} is the membership degree of the input vector $\mathbf{x}_j$ for the *k*th class, which can be obtained from the clustering results of the FCM algorithm on the input dataset. N is the number of the training data. In Eq. (7), h is an adjustable parameter, which can be set manually or determined by employing a certain learning technique, such as the commonly used cross-validation strategy.

Once the antecedent parameters of the TSK FLS are evaluated, the output of this model can be expressed in a linear form [20, 21], i.e.,

$$y = \mathbf{p}_\mathbf{g}^T \mathbf{x}_\mathbf{g} \quad (8)$$

with

$$\mathbf{x}_e = (1, \mathbf{x}^T)^T, \quad (9)$$

$$\tilde{\mathbf{x}}^k = \tilde{\mu}^k(\mathbf{x})\mathbf{x}_e, \quad (10)$$

$$\mathbf{x}_g = ((\tilde{\mathbf{x}}^1)^T, (\tilde{\mathbf{x}}^2)^T, \cdots, (\tilde{\mathbf{x}}^K)^T)^T, \quad (11)$$

$$\mathbf{p}^k = (p_0^k, p_1^k, \cdots, p_d^k)^T, \quad (12)$$

$$\mathbf{p}_g = ((\mathbf{p}^1)^T, (\mathbf{p}^2)^T, \cdots, (\mathbf{p}^K)^T)^T. \quad (13)$$

Here, $\mathbf{p}_g^T$ is the transposition of $\mathbf{p}_g$; K is the number of rules; k is the serial number of the kth rule; d is the number of dimensionalities of the input data; $\tilde{\mu}^k(\mathbf{x})$ is a normalized fuzzy membership value computed by Eqs. (3) and (4); $\mathbf{x}_g$ represents the new data in the mapping feature space mapped using fuzzy rules; $\mathbf{p}_g$ is the combined vector of the consequent parameters of all of the fuzzy rules, i.e. $\mathbf{p}^k (k=1,\dots,K)$, in the trained TSK FLS. Using Eqs. (9)-(13), Eq. (2) can be transformed into Eq. (8). Comparing Eq. (2) with Eq. (8), we can see that the final output of the TSK FLS can be viewed as the output of a linear model in the new feature space, mapped by using the fuzzy inference rules. For example, the input vector $\mathbf{x}$ in the original space just corresponds to the new vector $\mathbf{x}_g$ in the mapping feature space, where $\mathbf{x}_g$ is constructed by Eqs. (9)-(11).

2) TSK FLS Training Method based on the L2 Norm Penalty and the ε-Insensitive Loss

Many training methods have been proposed for the model training of the TSK FLS. Among them, an important method is the one based on the L2 norm penalty and the ε-insensitive loss (L2-TSK-FLS) [20]. A brief review of this method is given here because it is closely related to the proposed method of constructing a TSK FLS based on transfer learning.

The L2-TSK-FLS considers the training of the TSK FLS as a linear regression problem in a feature space mapped by fuzzy inference rules, as shown in Eq. (8). For a given regression dataset $D=\{\mathbf{x}_i, y_i \mid \mathbf{x}_i \in R^d, y_i \in R\}$, where $\mathbf{x}_i$ and y_i are the input and output respectively, the objective function of the L2-TSK-FLS is constructed as follows:

$$\min_{\mathbf{p}_g,\xi^+,\xi^-,\varepsilon} L(\mathbf{p}_g,\xi^+,\xi^-,\varepsilon)=\frac{1}{\tau N}\sum_{i=1}^{N}((\xi_i^+)^2+(\xi_i^-)^2)+\frac{1}{2}(\mathbf{p}_g^T\mathbf{p}_g)+\frac{2}{\tau}\varepsilon$$
$$s.t.\ \begin{cases} y_i-\mathbf{p}_g^T\mathbf{x}_{gi}<\varepsilon+\xi_i^+ \\ \mathbf{p}_g^T\mathbf{x}_{gi}-y_i<\varepsilon+\xi_i^- \end{cases},\ \forall i \tag{14}$$

where ε is the ε-insensitive parameter; ξ_i^+ and ξ_i^- are the slack variables; N is the number of data in the training dataset; $\mathbf{x}_g$ represents the new data in the mapping feature space mapped with the fuzzy rules; and $\mathbf{p}_g$ defined in (13) is the combined vector with the consequent parameters of all the fuzzy rules in the trained TSK FLS. Furthermore, Eq. (14) can be transformed into a dual problem as follows, based on the optimization theory [20]:

$$\min_{\lambda_i'^{+},\lambda_i'^{-}} L(\lambda_i'^{+},\lambda_i'^{-}) = -\frac{2}{\tau^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\lambda_i'^{+}-\lambda_i'^{-})(\lambda_i'^{+}-\lambda_i'^{-})\mathbf{x}_{gi}^{T}\mathbf{x}_{gi}$$

$$-\frac{N}{\tau}\sum_{i=1}^{N}[(\lambda_i'^{+})^2+(\lambda_i'^{-})^2]+\sum_{i=1}^{N}\frac{2}{\tau}(\lambda_i'^{+}-\lambda_i'^{-})y_i\ , \tag{15}$$

$$s.t.\quad \sum_{i=1}^{N}(\lambda_i'^{+}+\lambda_i'^{-})=1,\ \lambda_i'^{+}\geq 0,\ \lambda_i'^{-}\geq 0,\ \forall i$$

where $\lambda_i'^{+}$ and $\lambda_i'^{-}$ are the Lagrangian multipliers, i.e., the solution variables of the dual problem. $\mathbf{x}_{gi}$ is defined as in (14). Eq. (15) is a classic quadratic programming (QP) problem, which can be easily extended to develop a scalable and fast algorithm for large datasets [22-24].

III TSK FLS FOR RECOGNIZING EPILEPTIC EEG SIGNALS BASED ON TRANSDUCTIVE TRANSFER LEARNING

The task in this study of detecting epileptic EEG signals is to identify the EEG signals of healthy people or patients with epilepsy under different conditions. This is essentially a classic problem of multi-class classification, and the task can be implemented with different strategies using existing multi-class classification methods. There are three major types of multi-class classification methods.

(1) Multi-class classifiers: This category of method can be directly implemented for the task of multi-class classification. For example, KNN is a typical multi-class classifier.

(2) Regression model based classifier: A regression model can be extended to perform classification tasks. For example, a simple way of doing this is to take the class labels as the model output in the model training procedure. The label of a test sample can then be obtained with the label nearest to the real output of the regression model. Fuzzy logic systems and radial basis neural networks are typical examples of

classic regression models that have been used for classification.

(3) Binary classification based classifier: For this type of method, a multi-class classification problem is first decomposed into multiple binary (two-class) classification problems. The binary classifiers for these individual two-class classification problems are then obtained using the one-versus-one strategy, for example. The multi-class classification problem is finally solved by applying a voting strategy to the results obtained from multiple binary classifiers. The SVM is a well-known binary classifier.

In this study, the TSK FLS construction methods for regression and binary classification respectively, i.e. TSK-TL-FLS (Reg) and TSK-TL-FLS (BC), are first investigated using transductive transfer learning. These two methods are then applied to the multi-class classification of epileptic EEG signals.

A. The Maximum Mean Distance and the Projected Maximum Mean Distance

Among the existing transductive transfer methods, LMPROJ is very simple and easy to implement. More importantly, it shows promise for use in recognizing epileptic EEG signals [24, 55]. LMPROJ is a transductive learning method based on the large margin mechanism in the feature space. This method makes use of the maximal mean distance (MMD) between the training domain and the testing domain to learn a projective vector. In order to obtain a desirable projective vector for the classification task in the target domain, the projected MMD (PMMD) between the projected training data in the source domain and the projected test data in the target domain is introduced into LMPROJ. In this paper, LMPROJ is used as the baseline classifier [24] for comparisons with the proposed methods.

MMD, as a measure for estimating the distance between two distributions, was discussed in [24]. Given a set of N training samples $D_S=\{\mathbf{x}_1,,\ldots,\mathbf{x}_N\}$ and a set of M test samples $D_T=\{\mathbf{z}_1,\ldots,\mathbf{z}_M\}$, the squared MMD between the two distributions, estimated on the two datasets, can be expressed as follows:

$$\mathrm{MMD}^2((D_S,D_T)=\left\|\frac{1}{N}\sum\nolimits_{j=1}^{N}\phi(\mathbf{x}_i)-\frac{1}{M}\sum\nolimits_{j=1}^{M}\phi(\mathbf{z}_j)\right\|^2, \tag{16.a}$$

where $\phi()$ is a function to map the vectors $\mathbf{x}_i$ and $\mathbf{z}_j$ respectively to $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{z}_j)$ in a new feature space. In the new feature space, the projected values of the vectors $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{z}_j)$ under a projective vector $\mathbf{w}^T$ can be obtained as follows:

$$f(\mathbf{x}_j) = \mathbf{w}^T \phi(\mathbf{x}_j), \quad j = 1, \cdots, N \ , \tag{16.b}$$

$$f(z_j) = \mathbf{w}^T \phi(z_j), \quad j = 1, \cdots, M \ . \tag{16.c}$$

Based on Eqs. (16.b) and (16.c), the squared PMMD between two distributions is given by

$$\begin{aligned} PMMD^2(\mathbf{w}, D_S, D_T) &= \left\| \frac{1}{N} \sum\nolimits_{i=1}^{N} \mathbf{w}^T \phi(\mathbf{x}_i) - \frac{1}{M} \sum\nolimits_{i=1}^{M} \mathbf{w}^T \phi(\mathbf{z}_i) \right\|^2 \\ &= \frac{1}{N^2} (\sum\nolimits_{i=1}^{N} \mathbf{w}^T \phi(\mathbf{x}_i))^2 + \frac{1}{M^2} (\sum\nolimits_{i=1}^{M} \mathbf{w}^T \phi(\mathbf{z}_i))^2 - \frac{2}{NM} \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{M} \mathbf{w}^T \phi(\mathbf{x}_i) \mathbf{w}^T \phi(\mathbf{z}_i) \end{aligned} . \tag{16.d}$$

Based on the above PMMD measure and the classic SVM binary classifier, the following objective criterion is introduced into LMPROJ to develop the binary classifier with transfer learning abilities.

$$\begin{aligned} &\min_{\mathbf{w},b} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{n} \xi_i^2 + \lambda PMMD^2(\mathbf{w}, D_S, D_T) \\ &s.t. \ y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \ , \ \ i = 1, 2, \ldots, n \end{aligned} \tag{17}$$

where the first two terms and constraint conditions are directly inherited from the SVM. In (17), $f(\mathbf{w}, b) = \mathbf{w}^T \phi(\mathbf{x}_i) + b$ is the decision function for the binary classification; the projective vector $\mathbf{w}^T$ and the bias b are the two parameters of the decision function in the original SVM classifier; ξ_i ($i = 1, 2, \ldots, n$) are the slack variables introduced into the original SVM for nonlinear classification tasks; and the parameters C and λ are used to balance the influence of the different terms in (17).

As discussed in [24], PMMD has been effectively used in transfer learning in order to obtain a better projective vector to reduce the discrepancy in data distributions between two different domains, such as the training and testing domains in transductive learning. By minimizing PMMD, an optimal projective vector $\mathbf{w}$ can be obtained for the subsequent task. This indicates that under the optimal project direction, denoted by $\mathbf{w}$, the discrepancy in the distribution of the projected data between the source domain and the target domain is minimum. Thus, the model trained by using the labeled data in the source domain under the above constraints can be effectively used for prediction in the target domain.

As shown in Eq. (8), the model output of the TSK FLS can be taken as the projected value of the vector $\mathbf{x}_g$ under the projective direction $\mathbf{p}_g$. Thus, it is natural to introduce PMMD to train the TSK FLS for transductive transfer learning, as in the case of LMPROJ [24]. According to Eqs. (9)-(11), the input data $D_s=\{\mathbf{x}_1,,\ldots,\mathbf{x}_N\}$ and $D_t=\{\mathbf{z}_1,\ldots,\mathbf{z}_M\}$ in the training domain and testing domain can first be mapped into the new feature space based on the fuzzy inference rules with fixed antecedents. In the mapping feature space, the two new datasets can be obtained from $D_s=\{\mathbf{x}_1,,\ldots,\mathbf{x}_N\}$ and $D_t=\{\mathbf{z}_1,\ldots,\mathbf{z}_M\}$ using Eqs. (9)-(11), which are given by

$$\begin{aligned}\tilde{D}_s&=\{\phi(\mathbf{x}_i)\,|\,\phi(\mathbf{x}_i)=\mathbf{x}_{gi},i=1,\cdots,N\}\\ \tilde{D}_t&=\{\phi(\mathbf{z}_i)\,|\,\phi(\mathbf{z}_i)=\mathbf{z}_{gi},i=1,\cdots,M\}\end{aligned}. \tag{18.a}$$

where $\mathbf{x}_{gi}$ and $\mathbf{z}_{gi}$ are obtained based on the input data $\mathbf{x}_i$ and $\mathbf{z}_i$ in the training domain and the test domain according to Eqs. (9)-(11), respectively. Since $\mathbf{x}_{gi}$ and $\mathbf{z}_{gi}$ are mapped from vectors $\mathbf{x}_i$ and $\mathbf{z}_j$ in the original space, Eq. (18.a) indicates that the mapping vectors $\mathbf{x}_{gi}$, $\mathbf{z}_{gi}$ are adopted as the mapping feature vector $\phi()$ defined in Eqs. (16. a) and (16.d). Thus, for the TSK FLS, we can define a PMMD between the training and testing domains for transductive transfer learning as follows,

$$\begin{aligned}&\text{PMMD}(\text{p}_g,\tilde{D}_S,\tilde{D}_T)^2=\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{p}_g^T\mathbf{x}_{gi}-\frac{1}{M}\sum_{i=1}^{M}\mathbf{p}_g^T\mathbf{z}_{gi}\right\|^2\\ &=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{p}_g^T\mathbf{x}_{gi}\mathbf{x}_{gj}\mathbf{p}_g+\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbf{p}_g^T\mathbf{z}_{gi}\mathbf{z}_{gj}\mathbf{p}_g-\frac{2}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbf{p}_g^T\mathbf{x}_{gi}\mathbf{z}_{gj}\mathbf{p}_g\end{aligned}, \tag{18.b}$$

where $\mathbf{p}_g$ is the vector of the combined consequent parameters defined in Eq. (13) and used as the projective vector here. $\tilde{D}_S$ and $\tilde{D}_T$ are defined in (18.a). By minimizing (18.b), the consequent parameters of the TSK FLS, i.e. $\mathbf{p}_g$, can be learned to reduce the discrepancy in data distribution between the training and testing domains as much as possible.

B. TSK-TL-FLS (Reg)

The TSK-TL-FLS (Reg) method proposed for regression is presented in this subsection. For a given training dataset $D_s = \{(\mathbf{x}_1, y_1), , \ldots, (\mathbf{x}_N, y_N) \mid \mathbf{x}_i \in R^d, y_i \in R\}$ and a test dataset $D_t = \{\mathbf{z}_1, \ldots, \mathbf{z}_M \mid \mathbf{z}_i \in R^d\}$, the transfer learning objective function is given by

$$\min_{p_g, \xi^+, \xi^-, \varepsilon} L(\mathbf{p_g}, \xi^+, \xi^-, \varepsilon) = \frac{1}{\tau N}\sum_{i=1}^{N}((\xi_i^+)^2 + (\xi_i^-)^2) + \frac{1}{2}(\mathbf{p_g^T p_g}) + \frac{2}{\tau}\varepsilon + \lambda \mathrm{PMMD}^2(\mathrm{p}_g, \tilde{D}_s, \tilde{D}_t)^2 \quad (19)$$
$$s.t. \begin{cases} y_i - \mathbf{p_g^T x_{gi}} < \varepsilon + \xi_i^+ \\ \mathbf{p_g^T x_{gi}} - y_i < \varepsilon + \xi_i^- \end{cases}, \quad \forall i.$$

In Eq. (19), ε is the ε-insensitive parameter; ξ_i^+ and ξ_i^- are the slack variables; N is the number of the data in the training domain; $\mathbf{x}_{gi}$ represents the new data in the mapping feature space mapped from $\mathbf{x}_i$ with the fuzzy rules; and $\mathbf{p}_g$ is the vector of the combined consequent parameters defined in (13) and used here as the projective vector. The first three terms are directly inherited from the formulation of the L2-TSK-FLS [20] and the final term is used to realize transfer learning. Let

$$\mathbf{\Omega}_0 = \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{x}_{gi}\mathbf{x}_{gj}^T + \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbf{z}_{gi}\mathbf{z}_{gj}^T - \frac{2}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbf{x}_{gi}\mathbf{z}_{gj}^T\right); \quad (20.a)$$

$\mathrm{PMMD}^2(\mathbf{p}_g, \tilde{D}_s, \tilde{D}_t)$ can be represented as

$$\mathrm{PMMD}^2(\mathbf{p}_g, \tilde{D}_s, \tilde{D}_t) = \mathbf{p}_g^T \mathbf{\Omega}_0 \mathbf{p}_g . \quad (20.b)$$

Let $\mathbf{\Omega} = \frac{1}{2}\left(\mathbf{\Omega}_0 + \mathbf{\Omega}_0^T\right)$; Eq. (20.a) can be equivalently expressed as

$$\mathrm{PMMD}^2(\mathbf{p}_g, \tilde{D}_s, \tilde{D}_t) = \mathbf{p}_g^T \mathbf{\Omega} \mathbf{p}_g . \quad (20.c)$$

Then, Eq. (19) can be expressed as

$$\min_{p_g, \xi^+, \xi^-, \varepsilon} L(\mathbf{p}_g, \xi^+, \xi^-, \varepsilon) = \frac{1}{2}(\mathbf{p}_g^T\mathbf{p}_g) + \frac{1}{\tau N}\sum_{i=1}^{N}[(\xi_i^+)^2 + (\xi_i^-)^2] + \frac{2}{\tau}\varepsilon + \lambda \mathbf{p}_g^T\mathbf{\Omega}\mathbf{p}_g$$
$$s.t. \begin{cases} y_i - \mathbf{p}_g^T \mathbf{x}_{gi} < \varepsilon + \xi_i^+ \\ \mathbf{p}_g^T \mathbf{x}_{gi} - y_i < \varepsilon + \xi_i^- \end{cases}, \quad \forall i. \quad (21)$$

Here, the Lagrange function is used to solve Eq. (21), which is given by

$$L(\mathbf{p}_g,\xi^+,\xi^-,\varepsilon,\lambda_i^+,\lambda_i^-)=\frac{1}{2}(\mathbf{p}_g^T\mathbf{p}_g)+\frac{1}{\tau N}\sum_{i=1}^{N}[(\xi_i^+)^2+(\xi_i^-)^2]$$
$$+\frac{2}{\tau}\varepsilon+\lambda\mathbf{p}_g^T\boldsymbol{\Omega}\mathbf{p}_g+\sum_{i=1}^{N}\lambda_i^+(\mathrm{y}_i-\mathbf{p}_g^T\mathbf{x}_{gi}-\varepsilon-\xi_i^+)+\sum_{i=1}^{N}\lambda_i^-(\mathbf{p}_g^T\mathbf{x}_{gi}-\mathrm{y}_i-\varepsilon-\xi_i^-). \quad (22)$$

where λ_i^+ and λ_i^- are the Lagrangian multipliers. According to optimization theory, the necessary conditions for an optimal solution are given by

$$\frac{\partial L}{\partial \mathbf{p}_g}=\mathbf{p}_g+2\lambda\boldsymbol{\Omega}^T\mathbf{p}_g-\sum_{i=1}^{N}(\lambda_i^+-\lambda_i^-)\mathbf{x}_{gi}=\mathbf{0}\Rightarrow\mathbf{p}_g=(\mathbf{E}+2\lambda\boldsymbol{\Omega})^{-1}\sum_{i=1}^{N}(\lambda_i^+-\lambda_i^-)\mathbf{x}_{gi}, \quad (23a)$$

$$\frac{\partial L}{\partial \varepsilon}=\frac{2}{\tau}-\sum_{i=1}^{N}(\lambda_i^++\lambda_i^-)=0\Rightarrow\sum_{i=1}^{N}(\lambda_i^++\lambda_i^-)=\frac{2}{\tau}, \quad (23b)$$

$$\frac{\partial L}{\partial \xi_i^+}=\frac{2}{\tau N}\xi_i^+-\lambda_i^+=0\Rightarrow\xi_i^+=\frac{N\tau}{2}\lambda_i^+, \quad (23c)$$

$$\frac{\partial L}{\partial \xi_i^-}=\frac{2}{\tau N}\xi_i^--\lambda_i^-=0\Rightarrow\xi_i^-=\frac{N\tau}{2}\lambda_i^-. \quad (23d)$$

Furthermore, we let

$$\lambda_i^+=2\lambda_i^{'+}/\tau, \quad (23e)$$

$$\lambda_i^-=2\lambda_i^{'-}/\tau, \quad (23f)$$

$$\boldsymbol{\psi}=(\mathbf{E}+2\lambda\boldsymbol{\Omega}^T)^{-1}. \quad (23g)$$

By substituting Eqs. (23a)-(23g) into Eq. (22), the dual problem in Eq. (21) can be represented as the quadratic programming problem below:

$$L(\lambda_i^{'+},\lambda_i^{'-})=\frac{2}{\tau^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\mathbf{x}_{gi}^T\left(\boldsymbol{\psi}^{\mathbf{T}}\boldsymbol{\psi}+\mathbf{2\lambda}\boldsymbol{\psi}^{\mathbf{T}}\boldsymbol{\Omega}\boldsymbol{\psi}-\mathbf{2}\boldsymbol{\psi}^{\mathbf{T}}\right)\mathbf{x}_{gj}\right)\cdot(\lambda_i^{'+}-\lambda_i^{'-})(\lambda_j^{'+}-\lambda_j^{'-})$$
$$+\left(\sum_{i=1}^{N}\left(\frac{2}{\tau}\mathrm{y}_i\right)\lambda_i^{'+}-\sum_{i=1}^{N}\left(\frac{2}{\tau}\mathrm{y}_i\right)\lambda_i^{'-}\right)-\left(\sum_{i=1}^{N}\frac{N}{\tau}\left(\lambda_i^{'+}\right)^2+\sum_{i=1}^{N}\frac{N}{\tau}\left(\lambda_i^{'-}\right)^2\right). \quad (24)$$
$$s.t.\ \sum_{i=1}^{N}(\lambda_i^{'+}+\lambda_i^{'-})=1,\lambda_i'^+,\lambda_i'^-\geq 0 \quad \forall i.$$

Furthermore, let

$$\tilde{\boldsymbol{\alpha}}=(\tilde{\alpha}_1,\ldots,\tilde{\alpha}_{2N})^T=((\lambda_i'^+)^T,(\lambda_i'^-)^T)^T, \quad (25a)$$

$$\mathbf{K}=[\tilde{k}_{ij}]_{N\times N}\ \ \tilde{k}_{ij}=\frac{2}{\tau^2}\mathbf{x}_{gi}^T\left(2\boldsymbol{\psi}^T-\boldsymbol{\psi}^T\boldsymbol{\psi}-2\lambda\boldsymbol{\psi}^T\boldsymbol{\Omega}\boldsymbol{\psi}\right)\mathbf{x}_{gj}+\frac{N}{\tau}\delta_{ij}, \quad (25b)$$

$$\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases}, \quad (25c)$$

$$\mathbf{H}=\begin{pmatrix}\mathbf{K} & -\mathbf{K}\\ -\mathbf{K} & \mathbf{K}\end{pmatrix}, \quad (25c)$$

$$\boldsymbol{\beta}=(\frac{2}{\tau}y^T,-\frac{2}{\tau}\mathrm{y}^T)^T,\ \ y=(y_1,y_2,\ldots,y_N)^T. \quad (25d)$$

Eq. (24) can be simplified into the compact form below:

$$\max_{\tilde{\alpha}} \ -\tilde{\boldsymbol{\alpha}}^T\mathbf{H}\tilde{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\alpha}}^{\mathbf{T}}\boldsymbol{\beta}$$
$$s.t. \ \ \tilde{\boldsymbol{\alpha}}^T\mathbf{1} = 1, \qquad \alpha_i \ge 0 \quad \forall i. \tag{26}$$

The ultimate optimal solution to Eq. (21) can be obtained by solving Eq. (24) or (26). According to the dual theory, the optimal solution to the primal optimization problem in Eq. (21) is given by

$$\mathbf{p}_g = (\mathbf{E}+2\lambda\boldsymbol{\Omega})^{-1}\sum_{i=1}^{N}(\lambda_i^+ - \lambda_i^-)\mathbf{x}_{gi} = (\mathbf{E}+2\lambda\boldsymbol{\Omega})^{-1}\cdot\frac{2}{\tau}\cdot\sum_{i=1}^{N}(\lambda'^+_i - \lambda'^-_i)\mathbf{x}_{gi}, \tag{27a}$$

$$\xi_i^+ = \frac{N\tau}{2}\lambda_i^+ = N\lambda_i'^+, \tag{27b}$$

$$\xi_i^- = \frac{N\tau}{2}\lambda_i^- = N\lambda_i'^-, \tag{27c}$$

$$\begin{aligned}\varepsilon &= \sum_{i=1}^{N}(\lambda_i^+ - \lambda_i^-)y_i - N\sum_{i=1}^{N}((\lambda_i^+)^2 + (\lambda_i^-)^2) - \frac{2}{\tau}\sum_{i=1}^{N}\sum_{j=1}^{N}(\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-)\cdot\mathbf{x}_{gi}^T\boldsymbol{\psi}^T\mathbf{x}_{gj} \\ &= \frac{2}{\tau}\sum_{i=1}^{N}(\lambda'^+_i - \lambda'^-_i)y_i - \frac{4N}{\tau^2}\sum_{i=1}^{N}((\lambda'^+_i)^2 + (\lambda'^-_i)^2) - \frac{8}{\tau^3}\sum_{i=1}^{N}\sum_{j=1}^{N}(\lambda'^+_i - \lambda'^-_i)(\lambda'^+_j - \lambda'^-_j)\cdot\mathbf{x}_{gi}^T\boldsymbol{\psi}^T\mathbf{x}_{gj},\end{aligned} \tag{27d}$$

where λ'^+_i and λ'^-_i are the solution to the dual problem in Eq. (26). Eq. (27a) is just the optimal solution to the consequent parameters of the trained TSK FLS. In particular, Eq. (26) is in a form that can be conveniently extended to scalable algorithms for large-scale datasets using the minimal enclosing approximation technique [20, 22]. However, this is not the focus of the present study and is not investigated further here.

C. TSK-TL-FLS (BC)

Next, the TSK-TL-FLS (BC) method proposed for binary classification is discussed here. For a training dataset $D_s = \{\mathbf{x}_i, y_i \mid \mathbf{x}_i \in R^d, y_i \in \{1, -1\}, i = 1, \cdots, N\}$ and a test dataset $D_t = \{\mathbf{z}_i \mid \mathbf{z}_i \in R^d, i = 1, \cdots, N\}$, by applying the transductive transfer learning mechanism for constructing a TSK FLS, the objective function is given by

$$\min_{\mathbf{p}_g,\,\xi_i} \frac{1}{2}\mathbf{p}_g^T\mathbf{p}_g + C\sum_{i=1}^{N}\xi_i \ + \lambda\mathrm{PMMD}^2(\tilde{D}_s, \tilde{D}_t)^2$$
$$s.t. \quad y_i\,\mathbf{p}_g^T\mathbf{x}_{gi} \ge 1 - \xi_i, \ \ \forall i. \tag{28}$$

where ξ_i are the slack variables; $\mathbf{x}_{gi}$ represents the new data in the mapping feature space mapped from $\mathbf{x}_i$ with the fuzzy rules; and $\mathbf{p}_g$ is the vector of the combined consequent parameters defined in Eq. (13) and used as the projective vector here. Then, Eq. (28) can be constructed as follows based on Eq. (20.c):

$$\begin{aligned}&\min_{\mathbf{p}_g,\,\xi_i} \frac{1}{2}\mathbf{p}_g^T\mathbf{p}_g + C\sum_{i=1}^{N}\xi_i + \lambda\mathbf{p}_g^T\boldsymbol{\Omega}\mathbf{p}_g \\ &s.t. \quad y_i\,\mathbf{p}_g^T\mathbf{x}_{gi} \geq 1-\xi_i\ .\end{aligned} \tag{29}$$

The corresponding Lagrange function of Eq. (29) is

$$L(\mathrm{p}_g,\xi_i) = \frac{1}{2}\mathbf{p}_g^T\mathbf{p}_g + C\sum_{i=1}^{N}\xi_i + \lambda\mathbf{p}_g^T\boldsymbol{\Omega}\mathbf{p}_g - \lambda_i\sum_{i=1}^{N}(y_i\,\mathbf{p}_g\mathbf{x}_{gi} - 1 + \xi_i)\ . \tag{30}$$

where λ_i are the Lagrangian multipliers. According to optimization theory, the necessary conditions for an optimal solution are given by

$$\frac{\partial L}{\partial \mathbf{p}_g} = \mathbf{p}_g + 2\lambda(\mathbf{p}_g^T\boldsymbol{\Omega})^T - \sum_{i=1}^{N}\lambda_i\,\mathrm{y}_i\mathbf{x}_{gi} = 0 \Rightarrow \mathbf{p}_g = (\mathbf{E}+2\lambda\boldsymbol{\Omega}^T)^{-1}\sum_{i=1}^{N}\lambda_i\,\mathrm{y}_i\mathbf{x}_{gi} \tag{31a}$$

$$\frac{\partial L}{\partial \xi_i} = C - \sum_{i=1}^{N}\lambda_i = 0 \quad \Rightarrow \sum_{i=1}^{N}(\lambda_i) = C\ . \tag{31b}$$

Let

$$\boldsymbol{\psi} = (\mathbf{E}+2\lambda\boldsymbol{\Omega})^{-1}, \tag{31c}$$

By substituting Eqs. (31a)-(31c) into Eq. (30), the dual problem in Eq. (29) is given by

$$\begin{aligned}&\min_{\lambda_i} : L = \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\,\lambda_j\,y_i\,y_j\mathbf{x}_{gi}^T[\frac{1}{2}\boldsymbol{\psi}^T(\mathbf{E}+2\lambda\boldsymbol{\Omega})\boldsymbol{\psi} - \boldsymbol{\psi}^T]\mathbf{x}_{gi} + \sum_{i=1}^{N}\lambda_i \\ &s.t. \quad \sum_{i=1}^{N}\lambda_i = C, \quad 0 \leq \lambda_i \leq C.\end{aligned} \tag{32}$$

Furthermore, let

$$\tilde{\boldsymbol{\alpha}} = (\lambda_1,\ldots,\lambda_n)_{1\times N}^T, \tag{33a}$$

$$\boldsymbol{K} = [\tilde{k}_{ij}]_{N\times N}, \tag{33b}$$

$$\tilde{k}_{ij} = y_i\mathbf{x}_{\mathrm{gi}}^T\left(2\boldsymbol{\psi}^T - \boldsymbol{\psi}^T\boldsymbol{\psi} - 2\lambda\boldsymbol{\psi}^T\boldsymbol{\Omega}\boldsymbol{\psi}\right)\mathbf{x}_{\mathrm{gj}}y_j, \tag{33c}$$

$$\boldsymbol{\beta} = (1,\ldots,1)_{1\times N}^T\ . \tag{33d}$$

Eq. (32) can then be expressed in the following compact form

$$\begin{aligned}&\arg\max_{\tilde{\alpha}} \quad -\tilde{\boldsymbol{\alpha}}^T\mathbf{K}\boldsymbol{\alpha} + \tilde{\boldsymbol{\alpha}}^T\boldsymbol{\beta} \\ &s.t. \quad \tilde{\boldsymbol{\alpha}}^T\boldsymbol{1} = C, \quad 0 \leq \lambda_i \leq C \quad \forall i\ \cdot\end{aligned} \tag{34}$$

According to the dual theory, the solution to the optimization problem in Eq. (29) can be represented as

$$\mathbf{p}_g = (\mathbf{E} + 2\lambda\mathbf{\Omega}^T)^{-1}\sum_{i=1}^{N}\lambda_i^* \, y_i \mathbf{x}_{gi} \,, \tag{35}$$

where $\boldsymbol{\lambda} = (\lambda_i, \ldots, \lambda_i)^T$ is the optimal solution to the dual problem in Eq. (32) or (34). The solution in Eq. (35) is the optimal solution to the consequent parameters of the trained TSK FLS. Finally, the decision function of the TSK-TL-FLS (BC) for binary classification is given by

$$y = \begin{cases} 1 & f(\mathbf{x}) = \mathbf{p}_g^T \mathbf{x}_g > 0 \\ -1 & f(\mathbf{x}) = \mathbf{p}_g^T \mathbf{x}_g < 0 \end{cases}. \tag{36}$$

D. Algorithms of the TSK-TL-FLS (Reg) and TSK-TL-FLS (BC)

The algorithms of the proposed TSK-TL-FLS (Reg) and TSK-TL-FLS (BC) are presented in this subsection. In both methods, the algorithm consists of two stages: preprocessing and transfer learning. In the first stage, the antecedent parameters of the fuzzy rules are estimated using fuzzy clustering technique. The original input data are then mapped into the new feature space, and the new datasets in the mapping space are obtained. In the second stage, based on the data in the mapping space and the transfer learning mechanism, the consequent parameters are learned. Finally, the FLS is constructed with the parameters obtained in these two stages. The details of two algorithms are given in Table I. Based on the proposed methods for constructing a TSK FLS based on transfer learning and the existing methods for extracting features of EEG signals, we will present the TSK FLS based epileptic detection algorithm in the next subsection.

Table I Algorithms of the TSK-TL-FLS (Reg) and TSK-TL-FLS (BC)

Algorithm:	**TSK-TL-FLS (Reg)**	**TSK-TL-FLS (BC)**
Stage 1: Estimating the antecedent parameters of the TSK FLS and the construction of the data in the mapping new space.		
Step 1:	Set the number of fuzzy rules M and the parameter h in Eq. (7);	
Step 2:	Calculate the parameters of the antecedents, i.e. c_i^k and δ_i^k, by using fuzzy c-means clustering with Eqs. (6) and (7);	
Step 3:	Construct the training and test datasets in the mapping new space, i.e.	

	$D_{train}=\{\boldsymbol{x}_{gi},y_i\}, i=1,2,...,N$ and $D_{test}=\{\boldsymbol{z}_{gi}\}, i=1,2,...,M$ by using Eqs. (9)-(11).	
Stage 2: Transfer learning of the consequent parameter of the TSK FLS.		
Step 4:	Set the transfer learning parameters τ and λ in Eq. (19);	Set the transfer learning parameters C and λ in Eq. (28)
Step 5:	Use Eqs. (25a), (26), and (27a) to obtain the combined consequent parameters $\mathbf{p}_g$ based on the data in the mapping new space.	Use Eqs. (33a), (34) and (35) to obtain the combined consequent parameters $\mathbf{p}_g$ based on the data in the mapping new space.
Stage 3: Constructing the model of the final TSK FLS.		
Step 6:	Use Eqs. (12) and (13) to obtain the consequent parameters $\mathbf{p}_i$, $i=1,2,...,K$;	
Step 7:	Use the antecedent parameters and consequent parameters obtained in Step 2 and Step 6 to construct the final TSK FLS.	

E. Multi-class Classification of Epileptic EEG Signals

Although the TSK-TL-FLS (Reg) is developed for regression, it can easily be extended to perform multi-class classifications for detecting epileptic EEG signals. One common strategy is to use the regression function to approximate the class labels in the corresponding classification task. Once the model is trained, a future sample can be tested and the label nearest to the model output is taken as the label of the given test sample. In this paper, however, a more effective strategy enabling the regression method to be used for classification is introduced [36].

The idea of this strategy is to use a multiple output function for the task of classification. Given a classification dataset with *m* classes $\{\mathbf{x}_i, y_i\}$, $y_i \in \{1,...,m\}$, $i=1,...,N$, an *m*-output regression dataset $\{\mathbf{x}_i, \tilde{y}_i\}$ is constructed. If the original class label of the *i*th training sample in $\{\mathbf{x}_i, y_i\}$ is $y_i = p\,(1 \le p \le m)$, the corresponding output vector in the constructed *m*-output regression dataset $\{\mathbf{x}_i, \tilde{y}_i\}$ is defined as

$$\tilde{\mathbf{y}}_i = [0,\cdots,0,\overset{p}{1},0,\cdots,0]^T , \quad (37)$$

where only the *p*th element of $\tilde{y}_i$ is one, and the rest of the elements are set to zero.

With the *m*-output regression dataset, *m* single-output regression models can be trained. Once the *m* models are obtained, for a given test sample the output vector can be expressed as

$$\tilde{\mathbf{y}}_i^{\text{model}} = [y_{i,1}^{\text{model}},...,y_{i,m}^{\text{model}}]^T , \quad (38)$$

where $y_{i,l}^{\text{model}}$ denotes the lth output of the constructed model. Then, the predicted class label of the test sample is the index of the element with the highest value in the output vector. For example, if the value of $y_{i,l}^{\text{model}}$ is highest among all of the elements in vector $\tilde{\mathbf{y}}_i^{\text{model}}$, then the final predicted class label of the test sample will be l, which can be determined by

$$l = \arg\max_{1 \le k \le m}(y_{i,k}^{\text{model}}) . \tag{39}$$

Similarly, while the TSK-TL-FLS (BC) is developed for binary classification, it can also be extended for multi-class classifications using many strategies, such as the popular one-versus-one or the one-versus-rest strategy, to decompose a multi-class problem into multiple binary classification problems. The multiple binary classifiers are then used to predict the labels of the test data, and the final labels are determined by voting.

Based on Table I, the algorithms for a multi-class classification of the epileptic EEG signals in this study are presented in Table II.

Table II Algorithms for recognizing epileptic EEG signals

Algorithm:	**Recognizing epileptic EEG signals based on the TSK-TL-FLS (Reg)**	**Recognizing epileptic EEG signals based on the TSK-TL-FLS (BC)**
Step1:	Extract the effective features of EEG signals using the classic WPD or KPCA feature extraction techniques and obtain the data used for training the TSK FLS.	
Sep2:	Use the TSK-TL-FLS (Reg) algorithm to train m single-output TSK FLSs for an *m*-class classification based on the labeled data in the source domain and the unlabeled data in the target domain.	Use the TSK-TL-FLS (BC) algorithm and the one-versus-one strategy to obtain $m \cdot (m-1)/2$ binary classification TSK FLSs for an *m*-class classification based on the labeled data in the source domain and the unlabeled data in the target domain.
Step3:	Use Eq. (39) to determine the labels of the test data.	Determine the final labels of the test data by voting.

Remark 1: Regarding the effectiveness of the proposed epileptic EEG recognition algorithms, as the physiological EEG signals of healthy people differ from those of patients with epilepsy, detecting epilepsy by classifying the EEG signals is in principle a feasible approach. However, EEG signals are complicated in nature and the recording can be influenced by many factors, which present a great challenge to ensuring the accuracy of the classification algorithms. In this regard, transfer learning

is an important technique that can be used to construct robust intelligent models for classifying EEG signals. The algorithms proposed in the study are for tackling the issue of accuracy. They are expected to be a promising tool for classifying EEG signals due to their transfer learning abilities and the advantages inherited from fuzzy modeling.

Remark 2: Like other learning-based methods, e.g. LMPROJ, the two transductive learning based methods proposed in this study not only use the labeled training data in the source domain to train the TSK FLS, but also the unlabeled test data in the target domain. This approach is different from the learning procedure of conventional classifiers. It means that for a new test dataset, the transductive transfer learning based model must be trained again. While this characteristic could lead to a computational burden, the proposed transductive learning based methods are more adaptive than other methods and thus practical for many applications.

IV EXPERIMENTAL STUDY

A. Data and Setup

1) Data Source

The epileptic EEG data used in this study are publicly available on the web from the University of Bonn, Germany (http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html). The complete data archive contains five groups of data, groups A to E, each containing 100 single channel EEG segments with a duration of 23.6 seconds sampled at 173.6Hz. Groups A and B consist of segments acquired from a surface EEG recording performed on five healthy subjects using a standardized scheme of placing electrodes on the subjects when they were relaxed and in an awake state with their eyes open and closed, respectively. Groups C, D, and E are the data obtained from subjects with epilepsy during seizure-free intervals (C and D) and during seizure activity (E). Recordings in groups C and D were made from the hippocampus formation in the opposite hemisphere of the brain and the epileptogenic zone, respectively. The settings are described in Table III. A typical signal trace in each group is shown in Fig. 4 to provide an intuitive display of the visual differences in the signals of the five data groups. Further details about epileptic EEG signals can be found in [25].

Table III Settings for measuring epileptic EEG data

	Group	Setting
Healthy people	A	EEG signals obtained from healthy subjects with their eyes open.
	B	EEG signals obtained from healthy subjects with their eyes closed.
People with epilepsy	C	EEG signals acquired from the hippocampus formation in the opposite hemisphere of the brains of subjects with epilepsy during seizure-free intervals.
	D	EEG signals acquired from the epileptogenic zone of the brains of subjects with epilepsy during seizure-free intervals.
	E	EEG signals obtained from subjects with epilepsy during seizure activity.

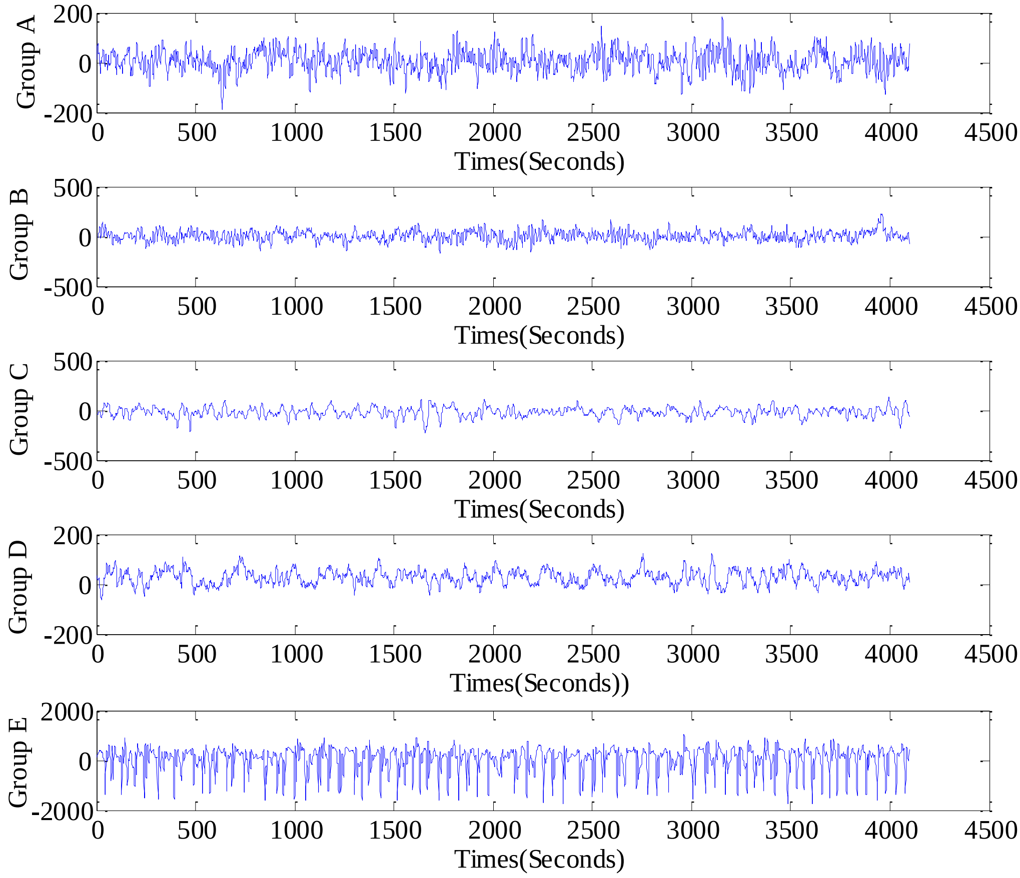

Fig. 4 Examples of traces of EEG signals

2) Experimental Datasets

Two types of datasets were constructed to compare the performances of different conventional classification algorithms and the proposed TSK-TL-FLS method: one consisting of training and test datasets drawn from the same distribution, and the other constructed with training and test datasets of different distributions. In total, eight datasets of these two types were constructed for a performance evaluation, as shown in Table IV. In each dataset, all of the training and test data were constructed with different proportions of the five groups of EEG signals. Taking dataset 7 as an example, the data in the training set were created using the data in groups A, C, and E, while the data in the test set were derived from groups B, C, and E.

The eight datasets include: (i) two datasets with data of identical distributions in the source and the target domain; (ii) six datasets with data of different distributions in the two domains. In particular, datasets 1-6 were constructed for binary classification and datasets 7-8 for multi-class classification. For binary classification, the task is to classify healthy and non-healthy (epileptic) subjects. For multi-class classification, the task is to identify three different groups of subjects, i.e., healthy subjects, subjects with epilepsy in the preictal state, and subjects with epilepsy in the ictal state.

Table IV Datasets constructed for the experiments

Data Distribution	Dataset Number	Training dataset	Test dataset
Identical distribution	1	BE-each 75 segments	BE-each 25 segments
	2	BDE-each 75 segments	BDE-each 25 segments
Different distribution	3	AE-each 25 segments	AC-each 25 segments
	4	AE-each 25 segments	AD-each 25 segments
	5	BE-each 25 segments	BC-each 25 segments
	6	BE-each 25 segments	BD-each 25 segments
	7	ACE-each 25 segments	BCE-each 25 segments
	8	ADE-each 25 segments	BDE-each 25 segments

3) Classification Methods

After extracting the features from the original EEG signals using WPD and KPCA respectively, ten classification methods, including seven conventional methods were employed to train and test all of the datasets. They were: LDA [26,27], DT [28], NB [29], NM [30-32], SVM[33], L2-TSK-FLS [20], a fuzzy system developed using fuzzy clustering and SVM called FS-FCSVM [54]; and three transductive transfer learning methods, i.e. LMPROJ[24], TSK-TL-FLS (Reg), and TSK-TL-FLS (BC). FS-FCSVM is an FLS construction method based on fuzzy clustering and SVM, where the structure of the FLS antecedents is generated by applying fuzzy clustering to the input data and the consequent parameters are trained using SVM to improve the system generalization performance. LMPROJ is a classic transductive transfer method, which has been used for recognizing epileptic EEG signals [55].

4) Parameter Setting

For the methods based on fuzzy rules, the number of fuzzy rules K was set to 40 in the experiments. For the datasets with data of identical distributions in the target domain and source domain, the classic five-fold cross-validation strategy was used to determine the optimal hyperparameters of the different algorithms in the experiments. For the datasets with data of different distributions in the two domains, since the classic cross-validation strategy could not be directly applied, a strategy resembling cross validation was adopted instead, as illustrated in Fig. 5 with dataset 7 as an example. The data in groups ACE and BCE were divided and combined to give eight different datasets for training and testing respectively, eventually resulting in 64 combinations of datasets to evaluate each of the algorithms in the experiments.

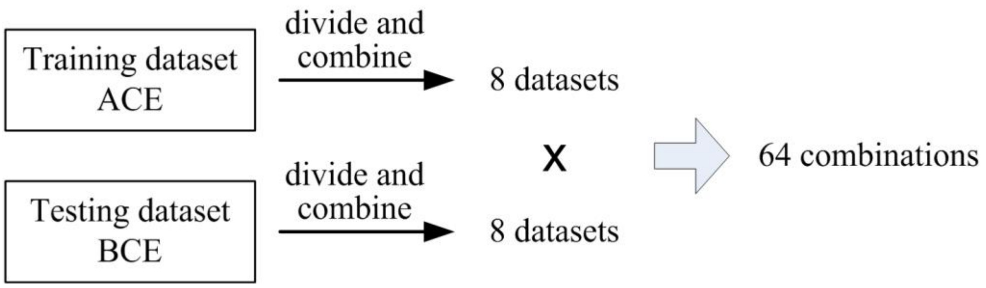

Fig. 5 The strategy resembling cross validation applied to dataset 7.

5) Evaluation Index

To measure the performance, the mean and standard deviation (SD) of the classification accuracy achieved with each algorithm were calculated. The p-values of the t-tests were also calculated based on the classification results that were obtained.

B. Results and Analyses

The experimental results are shown in Table V and Table VI, and also in Fig. 6 and Fig. 7. From these results the following observations can be made.

1) The classification results of the conventional classifiers exhibited poor performance when the data distribution between the training and test datasets differed. Meanwhile, promising results were observed for the transfer learning based models, i.e. LMPROJ, the proposed TSK-TL-FLS (Reg) and TSK-TL-FLS (BC). For

example, the classification accuracies of these transfer learning based models were all greater than 93%.

2) Among the three transfer learning based methods, the TSK-TL-FLS (Reg) and TSK-TL-FLS (BC) outperformed LMPROJ, which demonstrates that the method based on fuzzy rules has distinctive advantages for classification. Furthermore, the results to be presented in the next subsection will also show that the methods based on fuzzy rules are advantageous in terms of interpretability, which is a feature not available from LMPROJ.

3) Of the two proposed methods, the TSK-TL-FLS (BC), which is based on a binary classification model, is found to offer more advantages than the TSK-TL-FLS (Reg), which is based on a regression model. In particular, the classification accuracies of the TSK-TL-FLS (BC) was over 96%. This finding indicates that a specific design is necessary for classification tasks based on the TSK-FLS.

4) As shown in Table IV, for datasets 1 and 2 where the training and test data have identical data distributions, transfer learning is not necessary since the classic intelligent methods are already very effective for these scenes. Nevertheless, even if transfer learning is not critical for these scenes, the proposed methods based on transfer learning are also effective. As shown in the experimental results, when compared with the classic intelligent models, the proposed TSK FLSs based on transfer learning demonstrated competitive performance on these datasets

Overall, it can be concluded that the performance of the proposed TSK-TL-FLS (Reg) and TSK-TL-FLS (BC) in classifying EEG signals for detecting epilepsy is better than that of the classic non-transfer-learning methods and the transfer learning based LMPROJ method. This finding shows that transfer learning technology can effectively improve the recognition accuracy of the TSK FLS for detecting epilepsy using EEG.

While it has been proven that the introduction of the transductive transfer learning mechanism improves the accuracy of epileptic EEG signal detection, one inherent problem with the proposed methods is also due to the introduction of such a mechanism. In transductive transfer learning, both the training and test dataset are required in the learning procedure for training models. It is always necessary to re-train the classifier every time a different test dataset is used to optimize the parameters of the classifiers in the proposed TSK-TL-FLS (Reg) and TSK-TL-FLS (BC).

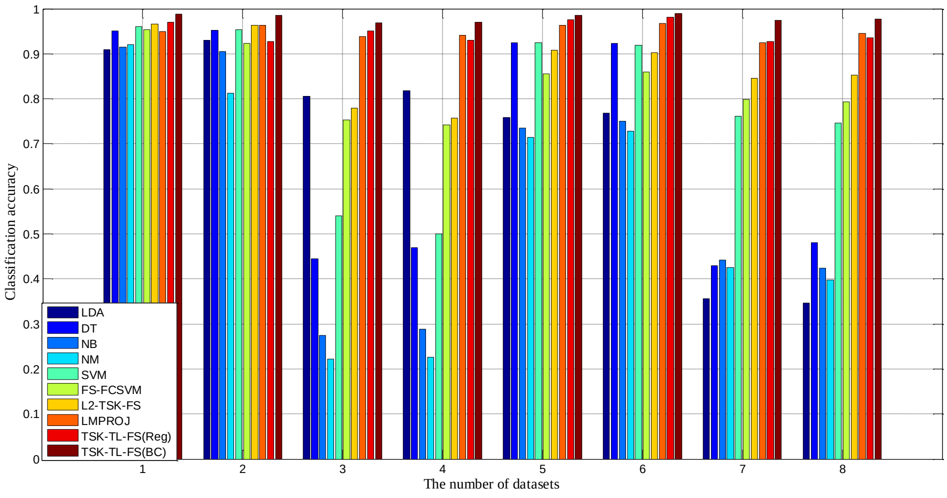

Fig. 6 Comparing the performances of different methods of classification with WPD used for feature extraction.

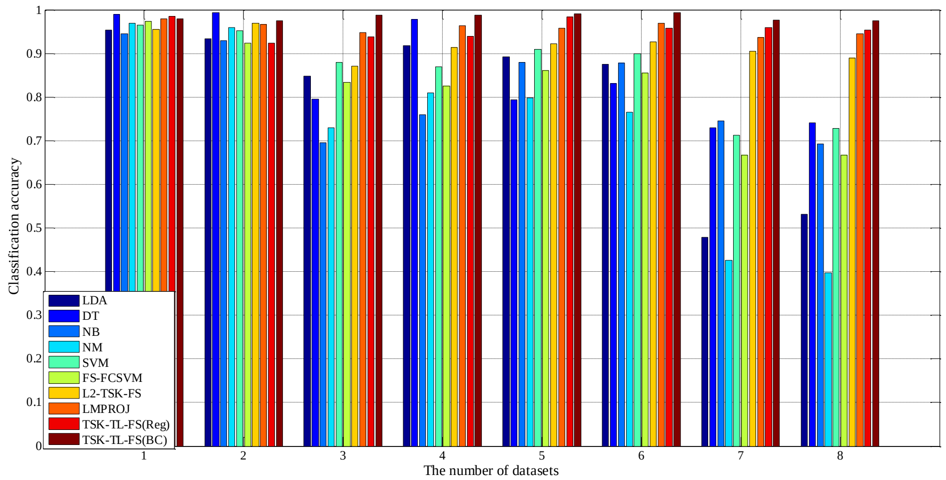

Fig. 7 Comparing the performances of different methods of classification with KPCA used for feature extraction.

Table V Comparing the accuracy of the ten classifiers (mean±SD) with WPD used for feature extraction.

Data sets	Index	LDA	DT	NB	NM	SVM	FS-FCSVM	L2-TSK-FS	LMPROJ	TSK-TL-FLS(Reg	TSK-TL-FLS(BC)
1	Mean	0.91	0.951	0.915	0.92	0.961	0.954	0.966	0.95	0.97	**0.988**
	SD	0.01	0.08	0.019	0.016	0.003	0.008	0.02	0.004	0.01	0.006
	p1*	0.005$^{(+)}$	0.115$^{(+)}$	0.01$^{(+)}$	0.003$^{(+)}$	0.19$^{(+)}$	0.09$^{(-)}$	0.5$^{(-)}$	0.073$^{(+)}$		
	p2	0.002$^{(+)}$	1e-03$^{(+)}$	0.002$^{(+)}$	0.007$^{(+)}$	0.008$^{(+)}$	0.016$^{(+)}$	0.02$^{(+)}$	5e-03$^{(+)}$		
2	Mean	0.93	0.952	0.905	0.813	0.954	0.923	0.963	0.963	0.927	**0.986**
	SD	0.013	0.006	0.01	0.02	0.004	0.002	0.023	0.005	0.005	0.006
	p1	0.62$^{(-)}$	0.008$^{(-)}$	0.03$^{(+)}$	8e-03$^{(+)}$	0.009$^{(-)}$	0.63$^{(-)}$	0.003$^{(+)}$	0.003$^{(-)}$		
	p2	6.6e-03$^{(+)}$	0.004$^{(+)}$	0.001$^{(+)}$	4e-03$^{(+)}$	0.006$^{(+)}$	0.004$^{(+)}$	0.008$^{(+)}$	0.016$^{(+)}$		
3	Mean	0.805	0.444	0.275	0.222	0.54	0.753	0.779	0.938	0.951	**0.969**
	SD	0.042	0.029	0.027	0.03	0.005	0.017	0.015	0.003	0.004	0.005
	p1	0.005$^{(+)}$	3.7e-05$^{(+)}$	1.9e-05$^{(+)}$	2.6e-05$^{(+)}$	5.9e-07$^{(+)}$	1.5e-06$^{(+)}$	2.2e-05$^{(+)}$	0.007$^{(+)}$		
	p2	0.003$^{(+)}$	3e-05$^{(+)}$	1.1e-05$^{(+)}$	2.0e-05$^{(+)}$	2.3e-07$^{(+)}$	2.7e-06$^{(+)}$	2.2e-05$^{(+)}$	0.001$^{(+)}$		
4	Mean	0.818	0.469	0.289	0.226	0.5	0.742	0.757	0.941	0.93	**0.97**
	SD	0.038	0.018	0.023	0.017	0	0.008	0.013	0.008	0.004	0.008
	p1	0.008$^{(+)}$	1.3e-05$^{(+)}$	1.6e-05$^{(+)}$	2.6e-06$^{(+)}$	2.4e-07$^{(+)}$	5e-07$^{(+)}$	1.5e-04$^{(+)}$	0.14$^{(+)}$		
	p2	0.005$^{(+)}$	3.5e-05$^{(+)}$	4.0e-06$^{(+)}$	1.0e-05$^{(+)}$	1.4e-06$^{(+)}$	2.8e-07$^{(+)}$	8e-05$^{(+)}$	0.002$^{(+)}$		
5	Mean	0.759	0.924	0.735	0.714	0.924	0.856	0.908	0.963	0.976	**0.985**
	SD	0.009	0.009	0.009	0.018	0.01	0.007	0.014	0.005	0.005	0.002
	p1	4.5e-05$^{(+)}$	5.0e-04$^{(+)}$	3.8e-05$^{(+)}$	1.3e-04$^{(+)}$	5.1e-04$^{(+)}$	8.6e-05$^{(+)}$	8.2e-04$^{(+)}$	0.01$^{(+)}$		
	p2	3.3e-05$^{(+)}$	2.9e-04$^{(+)}$	1.7e-05$^{(+)}$	8.5e-05$^{(+)}$	5.8e-04$^{(+)}$	9.4e-05$^{(+)}$	4.4e-05$^{(+)}$	0.004$^{(+)}$		
6	Mean	0.768	0.923	0.751	0.728	0.919	0.859	0.902	0.968	0.981	**0.989**
	SD	0.018	0.009	0.008	0.007	0.006	0.012	0.017	0.005	0.005	0.003
	p1	1.6e-04$^{(+)}$	0.002$^{(+)}$	8.0e-06$^{(+)}$	1.8e-06$^{(+)}$	0.001$^{(+)}$	9.4e-05$^{(+)}$	5.5e-05$^{(+)}$	0.002$^{(+)}$		
	p2	1.8e-04$^{(+)}$	9.8e-04$^{(+)}$	2.0e-05$^{(+)}$	8.4e-06$^{(+)}$	3.4e-04$^{(+)}$	6.6e-05$^{(+)}$	2.2e-04$^{(+)}$	0.007$^{(+)}$		
7	Mean	0.356	0.429	0.442	0.426	0.762	0.799	0.846	0.925	0.928	**0.974**
	SD	0.04	0.08	0.02	0.015	0.019	0.015	0.008	0.007	0.006	0.006
	p1	1e-03$^{(+)}$	0.001$^{(+)}$	9e-06$^{(+)}$	3.6e-06$^{(+)}$	6e-03$^{(+)}$	0.001$^{(+)}$	0.003$^{(+)}$	0.28$^{(+)}$		
	p2	1.5e-03$^{(+)}$	8e-03$^{(+)}$	2e-05$^{(+)}$	8e-06$^{(+)}$	9e-05$^{(+)}$	5.1e-04$^{(+)}$	4.7e-04$^{(+)}$	0.004$^{(+)}$		
8	Mean	0.346	0.481	0.424	0.397	0.746	0.793	0.853	0.945	0.935	**0.977**
	SD	0.036	0.11	0.016	0.02	0.03	0.013	0.022	0.01	0.007	0.003
	p1	9e-05$^{(+)}$	0.004$^{(+)}$	2e-05$^{(+)}$	2e-05$^{(+)}$	0.001$^{(+)}$	0.001$^{(+)}$	0.003$^{(+)}$	0.09$^{(-)}$		
	p2	4e-05$^{(+)}$	0.003$^{(+)}$	6e-06$^{(+)}$	6e-06$^{(+)}$	0.001$^{(+)}$	4.7e-04$^{(+)}$	4e-04$^{(+)}$	0.008$^{(+)}$		

Note: The superscripts (+) and (-) denote whether the proposed TSK-TL-FLS method is better or worse than the method under comparison based on t-test results. The smaller the p-value, the more significant the mean difference. The result is considered statistically significant with $p<0.05$. The p-values p1 and p2 are computed by comparing the TSK-TL-FLS (Reg) and TSK-TL-FLS (BC) respectively with the other methods.

Table VI Comparing the accuracy of the ten classifiers (mean±SD) with KPCA used for feature extraction.

Data sets	Index	LDA	DT	NB	NM	SVM	FS-FCSVM	L2-TSK-FLS	LMPROJ	TSK-TL-FLS(Reg	TSK-TL-FLS(BC)
1	Mean	0.954	**0.99**	0.945	0.97	0.965	0.973	0.955	0.98	0.985	0.98
	SD	0.008	0.012	0.02	0.012	0.004	0.01	0.017	0	0.006	0.006
	p1*	0.018$^{(+)}$	0.5$^{(-)}$	0.047$^{(+)}$	0.1$^{(+)}$	0.002$^{(+)}$	0.11$^{(-)}$	0.016$^{(+)}$	0.18$^{(+)}$		
	p2	2e-03$^{(+)}$	0.22$^{(-)}$	0.016$^{(+)}$	0.2$^{(+)}$	0.05$^{(+)}$	0.41$^{(-)}$	0.012$^{(+)}$	1$^{(-)}$		
2	Mean	0.933	**0.994**	0.93	0.96	0.952	0.924	0.97	0.967	0.923	0.975
	SD	0.011	0.008	0.003	0.012	0.004	0.011	0.009	0.003	0.009	0.005
	p1	0.18$^{(-)}$	0.003$^{(-)}$	0.29$^{(-)}$	0.035$^{(-)}$	0.015$^{(-)}$	0.88$^{(-)}$	0.002$^{(+)}$	5e-03$^{(-)}$		
	p2	0.015$^{(+)}$	0.001$^{(-)}$	0.015$^{(+)}$	0.03$^{(+)}$	0.01$^{(+)}$	0.004$^{(+)}$	0.42$^{(-)}$	0.2$^{(+)}$		
3	Mean	0.848	0.795	0.696	0.73	0.88	0.834	0.871	0.948	0.938	**0.988**
	SD	0.011	0.007	0.011	0.032	0.012	0.022	0.017	0.005	0.01	0.005
	p1	0.003$^{(+)}$	6.2e-05$^{(+)}$	8.8e-05$^{(+)}$	5.5e-04$^{(+)}$	1.8e-04$^{(+)}$	0.003$^{(+)}$	0.006$^{(+)}$	0.09$^{(-)}$		
	p2	2.7e-04$^{(+)}$	6.5e-05$^{(+)}$	2.5e-05$^{(+)}$	7.4e-04$^{(+)}$	6.8e-04$^{(+)}$	8.8e-04$^{(+)}$	8.3e-04	0.002$^{(+)}$		
4	Mean	0.918	0.978	0.76	0.81	0.87	0.825	0.914	0.963	0.94	**0.988**
	SD	0.007	0.006	0.042	0.03	0.017	0.014	0.011	0.005	0	0.006
	p1	0.006$^{(+)}$	0.001$^{(-)}$	0.003$^{(+)}$	0.003$^{(+)}$	0.004$^{(+)}$	0.001$^{(+)}$	0.049$^{(+)}$	0.003$^{(-)}$		
	p2	6.0e-04$^{(+)}$	0.13$^{(+)}$	0.001$^{(+)}$	0.001$^{(+)}$	4.4e-04$^{(+)}$	1.8e-04$^{(+)}$	0.001$^{(+)}$	0.035$^{(+)}$		
5	Mean	0.893	0.794	0.88	0.798	0.91	0.861	0.922	0.958	0.983	**0.991**
	SD	0.016	0.03	0.013	0.033	0.008	0.02	0.016	0.02	0.01	0.002
	p1	0.005$^{(+)}$	0.002$^{(+)}$	4.2e-04$^{(+)}$	9.9e-04$^{(+)}$	1.6e-04$^{(+)}$	0.001$^{(+)}$	0.005$^{(+)}$	0.03$^{(+)}$		
	p2	0.002$^{(+)}$	9.6e-04$^{(+)}$	4.1e-04$^{(+)}$	0.001$^{(+)}$	1.4e-04$^{(+)}$	9e-04$^{(+)}$	0.004$^{(+)}$	0.035$^{(+)}$		
6	Mean	0.875	0.831	0.878	0.766	0.90	0.855	0.927	0.97	0.958	**0.993**
	SD	0.015	0.033	0.017	0.016	0.008	0.01	0.012	0.008	0.01	0.003
	p1	0.003$^{(+)}$	0.009$^{(+)}$	0.002$^{(+)}$	1.6e-04$^{(+)}$	0.008$^{(+)}$	3.8e-04$^{(+)}$	0.005$^{(+)}$	0.08$^{(+)}$		
	p2	6.3e-04$^{(+)}$	0.002$^{(+)}$	8.3e-04$^{(+)}$	1.5e-04$^{(+)}$	8.4e-05$^{(+)}$	1.5e-04$^{(+)}$	7.4e-04$^{(+)}$	0.014$^{(+)}$		
7	Mean	0.478	0.729	0.746	0.426	0.712	0.667	0.905	0.936	0.959	**0.976**
	SD	0.06	0.04	0.03	0.06	0.02	0	0.013	0.006	0.01	0.006
	p1	5e-03$^{(+)}$	0.001$^{(+)}$	4e-03$^{(+)}$	4e-03$^{(+)}$	5e-03$^{(+)}$	1e-06$^{(+)}$	0.02$^{(+)}$	0.03$^{(+)}$		
	p2	4e-03$^{(+)}$	0.001$^{(+)}$	8e-03$^{(+)}$	3e03$^{(+)}$	1e-03$^{(+)}$	3.9e-07$^{(+)}$	0.01$^{(+)}$	0.004$^{(+)}$		
8	Mean	0.532	0.741	0.692	0.397	0.728	0.667	0.89	0.945	0.954	**0.975**
	SD	0.04	0.02	0.08	0.035	0.01	0	0.03	0.009	0.01	0.005
	p1	2e-03$^{(+)}$	2e-03$^{(+)}$	0.007$^{(+)}$	2e-03$^{(+)}$	4e-05$^{(+)}$	1e-06$^{(+)}$	0.019$^{(+)}$	0.03$^{(+)}$		
	p2	1e-03$^{(+)}$	2e-03$^{(+)}$	0.005$^{(+)}$	1e-03$^{(+)}$	7e-06$^{(+)}$	1.8e-06$^{(+)}$	0.01$^{(+)}$	0.008$^{(+)}$		

Note: The superscripts (+) and (-) denote whether the proposed TSK-TL-FLS method is better or worse than the method under comparison based on t-test results. The smaller the p-value, the more significant the mean difference. The result is considered statistically significant with $p<0.05$. The p-values p1 and p2 are computed by comparing the TSK-TL-FLS (Reg) and TSK-TL-FLS (BC) respectively with the other methods.

C. Model Analysis

Compared with other intelligent methods, the most distinctive characteristic of the FLS is its high interpretability. In this subsection, this characteristic is demonstrated by analyzing a model. In Table VII, a TSK FLS model with eight fuzzy rules was obtained with dataset 3 using the proposed TSK-TL-FLS (Reg) method. The table illustrates a fuzzy rule base for fuzzy inference corresponding to the features extracted from the EEG signals of group A using WPD. The fuzzy inference rules can be easily explained and described linguistically. Fig. 8 shows the membership function of each fuzzy subset of the first fuzzy rule that was obtained. Each membership function corresponds to a fuzzy subset that can be interpreted linguistically with expert knowledge. For example, the first fuzzy rule for recognizing epileptic EEG signals can be described as follows, with the energy of the EEG signals in different frequency bands expressed in terms of percentage.

If the energy of the EEG signal in frequency band 1 is about 3.8743%, and
if the energy of the EEG signal in frequency band 2 is about 3.4374%, and
if the energy of the EEG signal in frequency band 3 is about 6.8249%, and
if the energy of the EEG signal in frequency band 4 is about 19.6789%, and
if the energy of the EEG signal in frequency band 5 is about 31.9831%, and
if the energy of the EEG signal in frequency band 6 is about 32.2114%,
then this rule gives the decision value computed using the following formula:

$$f_1(\mathbf{x}) = 0.334 + 0.0335x_1 - 0.0340x_2 - 0.0021x_3 + 0.0092x_4 - 0.0069x_5 - 0.0016x_6 \quad (39)$$

It can be seen that the FLS obtained using the proposed TSK-TL-FLS (Reg) training method has better interpretability than the conventional non-rules-based methods. Meanwhile, the FLS also demonstrates better adaptability as validated by the higher classification accuracy reported in section IV-B.

Table VII A TSK FS with eight rules trained by the TSK-TL-FLS

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 $A_2^k(c_2^k,\delta_2^k)$ $\wedge\cdots\wedge x_d$ is $A_d^k(c_d^k,\delta_d^k)$, Then $f_k(\mathbf{x}) = p_{k0} + p_{k1}x_1 + \cdots + p_{kd}x_d$.		
No. of rules	**Antecedent parameters** (Gaussian membership function parameters)	**Consequent parameters** (linear function parameters)
k	$\mathbf{c}^k = (c_1^k,\cdots,c_d^k)^T$, $\boldsymbol{\delta}^k = (\delta_1^k,\cdots,\delta_d^k)^T$	$\mathbf{p}_k = (p_{k0}, p_{k1},\cdots, p_{kd})^T$
1	$\mathbf{c}^1 =$ [3.8743, 3.4374, 6.82, 1.967, 31.9831, 32.2114] $\boldsymbol{\delta}^1 =$ [0.1465, 0.6364, 0.6959, 0.1075, 0.62, 0.8945]	$\mathbf{p}_1 =$ [0.334, 0.0335, -0.0340,-0.0021, 0.0092, -0.0069, -0.0016]
2	$\mathbf{c}^2 =$ [3.842, 6.994, 1067, 19.1524, 28.816, 30.524] $\boldsymbol{\delta}^2 =$ [0.16, 0.405, 0.357, 0.127, 0.345, 0.4914]	$\mathbf{p}_2 =$ [0.0384, 0.0382, -0.0375, 6.283e-4, 9.41e-4, -0.0018, 0.0037]
3	$\mathbf{c}^3 =$ [5.11, 11.023, 16.072, 20.855, 23.158, 23.782] $\boldsymbol{\delta}^3 =$ [0.1473, 0.4177,0.352, 0.0622, 0.3136, 0.4059]	$\mathbf{p}_3 =$ [-0.0062, 0.0013, -0.0053, -0.0030, 0.0012, 0.0036, -0.020]
4	$\mathbf{c}^4 =$ [2.56, 16.277, 18.83, 20.3494, 20.9364, 21.049] $\boldsymbol{\delta}^4 =$ [0.14, .0.789, 0.4912, 0.043, 0.4065, 0.5457]	$\mathbf{p}_4 =$ [-0.0018, 0.0067, -0.0021, 0.0055, 0.0021, - 0.0050, 1.467e-5]
5	$\mathbf{c}^5 =$ [3.807, 14.015, 17.795, 20.466, 21.83, 22.0924] $\boldsymbol{\delta}^5 =$ [0.117, 0.5406, 0.4076, 0.0466, 0.3423, 0.4572]	$\mathbf{p}_5 =$ [0.0412, 0.0413, -0.405, -.0027, 0.202, -0.0181,-0.0265]
6	$\mathbf{c}^6 =$ [7.658, 3.75, 10.7778, 21.953, 27.57, 28.2898] $\boldsymbol{\delta}^6 =$ [0.3175, 0.6987, 0.357, 0.1278, 0.2837, 0.3641]	$\mathbf{p}_6 =$ [0.0547, 0.0635, -0.0646, 5.695e-4, 0.0094, -0.0071, 0.0024]
7	$\mathbf{c}^7 =$ [4.56, 7.997, 12.37, 21.3412, 26.4607, 27.2702] $\boldsymbol{\delta}^7 =$ [0.1658, 0.4143, 0.315, 0.0994, 0.2724, 0.3615]	$\mathbf{p}_7 =$ [0.0229, 0.302, -0.027, 1.2514e-4, 0.0014, -0.0011, -0.0171]
8	$\mathbf{c}^8 =$ [3.1556, 5.836, 8.92, 19.36, 30.3, 32.46] $\boldsymbol{\delta}^8 =$ [0.1657, 0.4249, 0.4572, 0.1168, 0.4306, 0.65]	$\mathbf{p}_8 =$ [0.0258, 0.0298, -.0.0281, -0.0026, 0.0128, -0.0103, 4.7e-5]

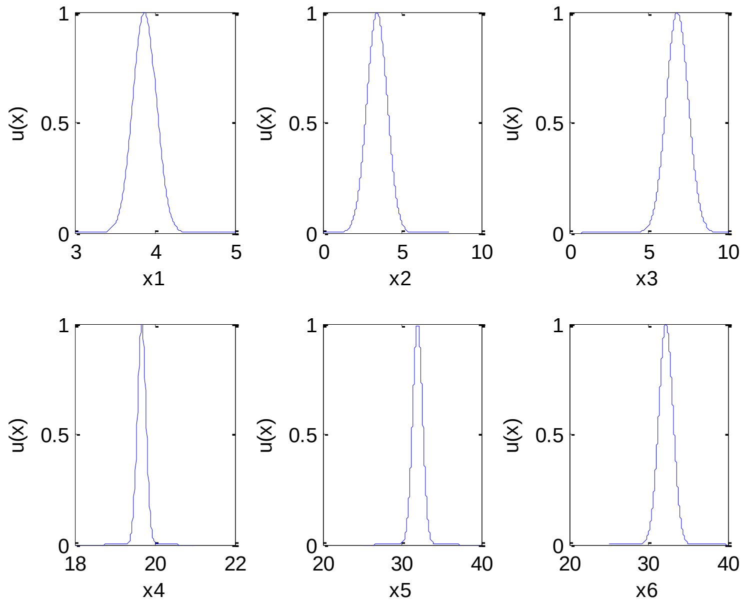

Fig. 8 The membership functions of each fuzzy subset in the antecedent of the first fuzzy rule trained by the TSK-TL-FLS (Reg) on dataset 3.

Since the TSK FLS is a kind of classic fuzzy-rules-based model, methods based on the TSK FLS are expected to have better interpretability than other classic intelligent models such as neural networks and kernel methods. However, some other rule-based intelligent models, e.g. the classic Mamdani-Larsen-type FLS (ML FLS) or decision tree, may demonstrate better interpretability than the TSK FLS. Therefore, introducing a transfer learning mechanism to the construction of these models is a meaningful future endeavor to pursue for the development of more interpretable and adaptive methods of intelligently recognizing epileptic EEG signals.

V CONCLUSION

The aim of the study was to tackle a key issue in conventional intelligent methods – i.e., for the recognition of epileptic EEG signals it is not practical to assume that the distribution of data in the training and test datasets is identical. Thus, methods for constructing the TSK FLS to detect epileptic EEG signals were proposed based on transductive transfer learning. Two TSK FLS training methods, the TSK-TL-FLS (Reg) and TSK-TL-FLS (BC), were developed based on a regression model and a binary classification model, respectively. Experimental studies have demonstrated that the proposed approach has distinct advantages in the detection of epileptic EEG signals, outperforming various classic intelligent recognition methods. However, the transfer learning mechanism adopted in the present study is admittedly relatively simple. Research will be conducted to introduce a more effective transfer learning mechanism for training FLS models in order to further enhance the accuracy of detecting epilepsy with EEG signals. While in this study transfer learning was only applied to the training of the parameters of the consequents in the TSK FLS, it is expected to be applicable to the training of the parameters of the antecedents as well. For example, it is beneficial to take into account the difference in the distribution of data between the training and testing domains when choosing the membership functions of the fuzzy sets. However, this is not a trivial work and deserves dedicated effort. This issue will be addressed in a future study.

REFERENCES

[1] A. Talevi, M. S. Cravero, and E. A. Castro, "Discovery of anticonvulsant activity of abietic acid through application of linear discriminant analysis," *Bioorganic & Medicinal Chemistry Letters*, vol. 17, pp. 1684-1690, 2007.

[2] R. Kohavi, "Scaling up the accuracy of naive-Bayes classifiers: A decision-tree hybrid," Portland: KDD-96, 1996.

[3] R. A. Fisher, "The use of multiple measures in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179-188, 1936.

[4] I. Goker, I. Osman, S. Ozekes, M. B. Baslo, et al., "Classification of Juvenile Myoclonic Epilepsy Data Acquired through Scanning Electromyography with Machine Learning Algorithms," *Journal of Medical Systems*, vol. 36, no. 5, pp. 2705-2711, 2012.

[5] J. R. Quinlan, "C4.5: programs for machine learning," San Mateo: Morgan Kaufmann, 1993.

[6] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, Jan. 1967.

[7] G. N. Fatma, and E. D. Ubeyli, "Multiclass Support Vector Machines for EEG-Signals Classification," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 2, pp. 117-126, Mar. 2007.

[8] I. Guler and E. D. Ubeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of Neuroscience Method*s, vol. 148, no. 2, pp. 113-121, 2005.

[9] A. Aarabi, R. Fazel-Rezal, and Y. Aghakhani, "A fuzzy rule-based system for epileptic seizure detection in intracranial EEG," *Clinical Neurophysiology*, vol. 120, no. 9, pp. 1648-1657, 2009.

[10] R. Harikumar and B. S. Narayanan, "Fuzzy techniques for classification of epilepsy risk level from EEG signals," in 2003 *Conference on Convergent Technologies for the Asia-Pacific Region*, 1, pp. 209-213.

[11] I. Guler and E. D. Ubeyli, "Application of adaptive neuro-fuzzy inference system for detection of electrocardiographic changes in patients with partial epilepsy using feature extraction," *Expert Systems with Applications*, vol. 27, no. 3, pp. 323-330, 2004.

[12] B. Litt, R. Esteller, and J. Echauz, "Epileptic seizures may begin hours in advance of clinical onset: a report of five patients," *Neuron*, vol. 30, pp. 51-64, 2011.

[13] J. L. Kenemans, P. C. M. Molenaar, M. N. Verbaten, and J. L. Slangen, "Removal of the Ocular Artifact from the EEG: A Comparison of Time and Frequency Domain Methods with Simulated and Real Data," *Psychophysiology*, vol. 28, no. 1, pp. 114-121,1991.

[14] G. C. Y Fong, N. G. Manyee, M. S. Jose, et al., "Childhood Absence Epilepsy with Tonic-Clonic Seizures and Electroencephalogram 3–4-Hz Spike and Multispike–Slow Wave Complexes: Linkage to Chromosome 8q24," *AJHG*, vol. 63, no. 4, pp. 1117-1129, 1998.

[15] L. D. Aubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. S, pp. 961-1005, 1990.

[16] R. Coifman and M. V. Wickerhauser, "Entropy based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 313-318, 1992.

[17] C. F. Juang, "Zero order TSK-type fuzzy system learning using a two-phase swarm intelligence algorithm," *Fuzzy Sets and System*, vol. 159, no. 21, pp. 2910-2926, 2008.

[18] L. O. Hall and D. B. Goldgof. "Convergence of the single-pass and online fuzzy c-means algorithms". *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 4, pp. 792-794, Aug. 2011.

[19] L. Zhu, F. L. Chung, and S. T. Wang, "Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions," *IEEE Transactions on SMC-B*, vol. 39, no. 3, pp. 578-591, Jun. 2009.

[20] Z. H. Deng, K. S. Choi, F. L. Chung, and S. T. Wang, "Scalable TSK fuzzy modeling for very large dataset sousing minima-enclosing-ball approximation," *IEEE Transactions on* Fuzzy *Systems*, vol. 19, no. 2, pp. 210-226, Apr. 2011.

[21] J. Laski, "TSK-fuzzy modeling based on E-insensitive learning," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 2, pp. 181-193, Apr. 2005.

[22] I. W. Tsang, J. T. Kwok, and J. M. Urara, "Generalized core vector machines," *IEEE Transaction on Neural Networks*, vol. 17, no. 5, pp. 1126-1140, Sept. 2006.

[23] R. E. Fan, P. H. Chan, and C. J. Lin, "Working set selection using second order information for training support vector machines," *Journal of Machine Learning Research*, vol. 6, pp. 1889-1918, 2005.

[24] B. Quaint, J. Human, "Large margin transductive transfer learning," *Proceedings of the 18th*

ACM Conference on Information and Knowledge Management, USA: ACM, 2009, pp. 1327-1336.

[25] R. G. Andrzejak, K Lehnertz, F Mormann, et al., “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” *Physical Review E*, vol. 64, no. 6, pp. 1907-1914, 2001.

[26] K. Patel, C. Chern-Pin, S. Fau, and C. J. Bleakley, “Low power real-time seizure detection for ambulatory EEG,” *Proceedings of the 3rd International Conference on Pervasive Computing Technologies for Healthcare*, NJ: IEEE Press, 2009, pp. 1-7.

[27] S. Mihandoost, M. C. Amirani, and B. Z. Varghahan, “Seizure detection using wavelet transform and a new statistical feature,” *Proceedings of 5th International Conference on Application of Information and Communication Technologies*. Piscataway, NJ: IEEE Press, 2001, pp. 1-5.

[28] K. Polat and S. Güneş, “Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform,” *Applied Mathematics and Computation*, vol. 187, no. 12, pp. 1017-1026, 2007.

[29] P. Valenti, E. Cazamajou, M. Scarpettini, A. Aizemberg, et al., “Automatic detection of interictal spikes using data mining models,” *Journal of Neuroscience Methods*, vol. 150, no. 1, pp. 105-110, 2006.

[30] F. Fukunaga, *Introduction to statistical pattern recognition*, Boston: Academic Press, 1990.

[31] D. Shin and S. Kim, “Nearest mean classification via one-class SVM,” 2009 *Proceedings of the International Joint Conference on Computational Sciences and Optimization*. Piscataway: IEEE Press, 2009, pp. 593-596.

[32] P. Wahlberg and G. Salomonsson, “Feature extraction and clustering of EEG epileptic spikes,” *Computers and Biomedical Research*, vol. 29, no. 5, pp. 382-394, 1996.

[33] M. Shen, J. Chen, and C. Lin, “Modeling of nonlinear medical signal based on local support vector machine,” 2009 *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, I2MTC, Piscataway, NJ: IEEE Press, 2009, pp. 675-679.

[34] Y. J. Xiong, R Zhang, C Zhang, and X. L. Yu, “A novel estimation method of fatigue using EEG based on KPCA-SVM and complexity parameters,” *Applied Mechanics and Materials*, vol. 373, pp. 965-969, 2013.

[35] A. R. Teixeiraa, A. M. Tomea, K Stadlthannerb, and E. W. Lang, “KPCA denoising and the pre-image problem revisited,” *Digital Signal Processing*, vol. 18, no. 4, pp. 568-580, 2008.

[36] Z. H. Deng, K. S. Choi, Y. Z. Jiang, and S. T. Wang, “Generalized Hidden-Mapping Ridge Regression, Knowledge-Leveraged Inductive Transfer Learning for Neural Networks, Fuzzy Systems and Kernel Methods,” *IEEE Transactions on Cybernetics*, Digital Object Identifier: 10.1109/TCYB.2014.2311014, 2014.

[37] S. S. Durbha and R. L. King, “Evaluating transfer learning approaches for image information mining applications,” 2011 *Proceedings of the Geoscience and Remote Sensing Symposium Conference*, IGARSS, NJ: IEEE Press, 2011, pp. 1457-1460.

[38] S. Yuan, Z. L. Zhen, L Wei, and B Wei, “Extending Semi-supervised Learning Methods for Inductive Transfer Learning,” 2009 *IEEE International Conference on Data Mining*, ICDM, NJ: IEEE Press, 2009, pp. 483-492.

[39] M. Tsuchiya, Y. Yamauchi, H Fujiyoshi, and T. H. Yamashita, “Transfer Learning for Efficient Learning in Object Detection,” *2013 2nd IAPR Asian Conference on Pattern Recognition*, ACPR, NJ: IEEE Press, 2013, pp. 69-73.

[40] A Bamdadian, C Guan, K. K. Ang, and J. X. Xu, “Improving session-to-session transfer performance of motor imagery-based BCI using Adaptive Extreme Learning Machine,” *2013 35th IEEE International Conference on EMBS*, Japan, 2013, NJ: IEEE Press, pp. 2188-2191.

[41] M. J. Mendel, *Uncertain rule-based fuzzy logic system: introduction and new directions*, Prentice-Hall, Upper-Saddle River, NJ, 2001.

[42] M. Biglarbegian, W. W. Melek, and J. M. Mendel, “On the stability of interval type-2 TSK fuzzy logic control systems,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2010, 40(3), pp. 798-818.

[43] J. M. Mendel, R. I. John, and F. Liu, “Interval type-2 fuzzy logic systems made simple,” *IEEE Transactions on Fuzzy Systems*, 2006, 14(6), pp. 808-821.

[44] Q. Liang and J. M. Mendel, “An introduction to type-2 TSK fuzzy logic systems,” *Proceedings of FUZZ-IEEE'99*, 1999, vol. 3, pp. 1534-1539.

[45] M. Almaraashi, R. John, S. Coupland, and A. Hopgood, “Time series forecasting using a TSK fuzzy system tuned with simulated annealing,” *Proceedings of FUZZ-IEEE'2010*, 2010, pp. 1-6.

[46] G. M. Mendez and O. Castillo, “Interval type-2 TSK fuzzy logic systems using hybrid learning algorithm,” *Proceedings of FUZZ-IEEE'2005*, 2005, pp. 230-235.

[47] J. R. Castro, O. Castillo, P. Melin, and A. Rodríguez-Díaz, “A hybrid learning algorithm for a class of interval type-2 fuzzy neural networks,” *Information Sciences*, 2009, vol. 179, no. 13, pp. 2175-2193.

[48] O. Castillo and P. Melin, “Optimization of type-2 fuzzy systems based on bio-inspired methods: a concise review,” Information Sciences, 2012, vol. 205, pp. 1-19.

[49] P. Melin and O. Castillo, “A review on the applications of type-2 fuzzy logic in classification and pattern recognition,” *Expert Systems with Applications*, 2013, vol. 40, no. 13, pp. 5413-5423.

[50] G. M. Mendez, “Interval type-1 non-singleton type-2 TSK fuzzy logic systems using the hybrid training method RLS-BP,” *In Analysis and Design of Intelligent Systems Using Soft Computing Techniques*, 2007, pp. 36-44, Springer Berlin Heidelberg.

[51] G. M. Mendez and L. A. Leduc, “First-order interval type-1 non-singleton type-2 TSK fuzzy logic systems,” *In MICAI 2006: Advances in Artificial Intelligence*, 2006, pp. 81-89, Springer Berlin Heidelberg.

[52] Z. H. Deng, Y. Z. Jiang, K. S. Choi, F. L. Chung, and S. T. Wang, “Knowledge-leverage-based TSK fuzzy system modeling,” *IEEE Transactions on Neural Networks and Learning Systems*, 2013, vol. 24, no. 8, pp. 1200-1212.

[53] Z. H. Deng, Y. Z. Jiang, L. B. Cao and S. T. Wang, “Knowledge-leverage based TSK fuzzy system with improved knowledge transfer,” *Proceedings of FUZZ-IEEE'2014*, 2014, pp. 178-185.

[54] C. F. Juang, S. F. Chiu, and S. J. Shiu, “Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation,” *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2007, vol. 37, no. 6, pp. 1077-1087.

[55] C. J. Yang, Z. H. Deng, K. S. Choi, Y. Z. Jiang, and S. T. Wang, “Transductive domain adaptive learning for epileptic electroencephalogram recognition,” *Artificial intelligence in medicine*, 2014, vol. 62, no. 3, pp. 165-177.

[56] P. J. Kindermans, M. Tangermann, K. R. Müller K R, and B. Schrauwen, “Integrating dynamic stopping, transfer learning and language models in an adaptive zero-training ERP

speller,” *Journal of neural engineering*, 2014, vol. 11, no. 3, 035005.

[57] S. Dalhoumi, G. Derosiere, G. Dray, J. Montmain, and S. Perrey, “Graph-Based Transfer Learning for Managing Brain Signals Variability in NIRS-Based BCIs,” *in Information Processing and Management of Uncertainty in Knowledge-Based Systems*, Springer International Publishing, 2014, pp. 294-303.

[58] I. Iturrate, L. Montesano, and J. Minguez, “Robot Reinforcement Learning using EEG-based reward signals,” *in IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 4822-4829.

[59] D. Wu, B. Lance, and T. D. Parsons, “Collaborative filtering for brain-computer interaction using transfer learning and active class selection,” *PloS one*, 2013, vol. 8, no. 2, e56624.

[60] S. Makeig, C. Kothe, T.Mullen, N. Bigdely-Shamlo, Z. Zhang, and K. Kreutz-Delgado, “Evolving signal processing for brain – computer interfaces,” *Proceedings of the IEEE*, 2012, no. 100 (Special Centennial Issue), pp. 1567-1584.