This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

Seizure Classification from EEG Signals using Transfer Learning, Semi-Supervised Learning and TSK Fuzzy System

Yizhang Jiang, *Member, IEEE*, Dongrui Wu, *Senior Member, IEEE*, Zhaohong Deng, *Senior Member, IEEE*, Pengjiang Qian, *Member, IEEE*, Jun Wang, *Member, IEEE*, Guanjin Wang, Fu-Lai Chung, *Member, IEEE*, Kup-Sze Choi, *Member, IEEE*, Shitong Wang

***Abstract*—Recognition of epileptic seizures from offline EEG signals is very important in clinical diagnosis of epilepsy. Compared with manual labeling of EEG signals by doctors, machine learning approaches can be faster and more consistent. However, the classification accuracy is usually not satisfactory for two main reasons: the distributions of the data used for training and testing may be different, and the amount of training data may not be enough. Additionally, most machine learning approaches generate black-box models that are difficult to interpret. In this paper, we integrate transductive transfer learning, semi-supervised learning and TSK fuzzy system to tackle these three problems. More specifically, we use transfer learning to reduce the discrepancy in data distribution between the training and testing data, employ semi-supervised learning to use the unlabeled testing data to remedy the shortage of training data, and adopt TSK fuzzy system to increase model interpretability. Two learning algorithms are proposed to train the system. Our experimental results show that the proposed approaches can achieve better performance than many state-of-the-art seizure classification algorithms.**

***Index Terms*—EEG recognition, seizure classification, transductive transfer learning, semi-supervised learning, TSK fuzzy system**

I. Introduction

Epilepsy is a common neurological disorder in which clusters of nerve cells in the brain function abnormally and cause seizures. Electroencephalogram (EEG) signals are commonly used to determine the presence and the type of epilepsy [1,2]. As it is very time-consuming for doctors to label the EEG signals manually, researchers have studied machine learning approaches to automatically detect seizures from EEG signals [3-11]. Many different methods, including decision tree [9], naïve Bayes [7,9], support vector machine [6,11], nearest-mean [7] and linear discriminant analysis [5,8,10], have been applied.

Compared with manual labeling, detection by machine learning methods is faster and more consistent. However, the detection accuracy could be an issue. Yang et al. [12] pointed out that the main reason for low detection accuracy is that most machine learning methods are developed based on the assumption that the distributions of the training and testing data are identical or similar, which may not be true in practice. For example, a classifier may be trained using data from a repository of existing subjects and then applied to a new subject. To cope with this issue, transfer learning (TL) [13-18,40,42], particularly, large margin projection (LMPROJ) [18], has been used to reduce the data distribution mismatch between the training and testing data. The results demonstrate that TL is very suitable for this problem.

This paper also utilizes LMPROJ TL to improve seizure classification performance. It further considers two more problems: 1) how to increase model interpretability which is very important in medical diagnostics, and 2) how to make use of the information contained in the unlabeled testing data to improve classification performance. To deal with the first problem, we adopt the Takagi-Sugeno-Kang (TSK) fuzzy system [19-22], which is intrinsically interpretable, as our base classifier. For the second problem, in addition to LMPROJ, we use a semi-supervised learning (SSL) approach to take advantage of the unlabeled testing data. We also propose two learning algorithms for this TL-SSL-TSK based model and demonstrate its outstanding performance.

The rest of this paper is organized as follows. Section II briefly reviews the typical feature extraction and machine learning methods that are used for epileptic EEG recognition. Section III introduces the TSK fuzzy system model and the learning algorithm. Section IV introduces the details of the proposed TL-SSL-TSK model and the two learning algorithms developed for the model. Section V compares the performance of the proposed algorithms with that of existing methods on six real EEG datasets. Section VI concludes the findings of the

This work was supported in part by the Hong Kong Polytechnic University under Grant G-UA3W, the Hong Kong Research Grants Council under grant PolyU 152040/16E, the National Natural Science Foundation of China under Grant 61702225, the Fundamental Research Funds for the Central Universities under Grant JUSRP11737, the Jiangsu Province Outstanding Youth Fund under Grant BK20140001, and by the Natural Science Foundation of Jiangsu Province under Grant BK20160187. (*Corresponding author*: *Zhaohong Deng*)

Y.Z. Jiang, Z.H. Deng, P.J. Qian, J. Wang and S.T. Wang and are with the School of Digital Media, Jiangnan University, Wuxi 214122, China (e-mail: 7121606003@vip.jiangnan.edu.cn; dzh666828@ yahoo.com.cn; qianpjiang@126.com; wangjun_sytu@hotmail.com ; wxwangst@yahoo.com.cn)

D. Wu is with DataNova, Clifton Park, NY 12065 USA (e-mail: drwu09@gmail.com).

G.J. Wang and K. S. Choi are with the Centre for Smart Health, School of Nursing, Hong Kong Polytechnic University, Hong Kong (e-mail: brendaguanguan@hotmail.com; kschoi@ieee.org)

F. L. Chung is with the Department of Computing, Hong Kong Polytechnic University, Hong Kong (e-mail: cskchung@comp.polyu.edu.hk)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

study.

II. Related Work

This section provides a brief review of typical methods used for EEG feature extraction and classification.

A. Feature Extraction Methods

Signal processing methods are used to extract features from the original raw EEG signals that are more concise and powerful to improve the classification performance [12,23-29,43] and to reduce the computational cost [4,7,8-10,12,23-29]. In general, there are three types of features: 1) time-domain features [23,25], e.g. principal component features of the raw EEG signals, 2) frequency-domain features [3,24], e.g. Fourier transform features, and 3) time-frequency features [3,25-27], e.g. wavelets. All these three types of features will be considered in this paper.

B. Machine Learning Approaches

Machine learning methods can potentially detect seizures from EEG signals more quickly and consistently than manual labeling by doctors. If well trained, the methods can also achieve very high accuracy. Many classical machine leaning methods have been studied for this purpose, such as decision tree [9], naïve Bayes [7,9], support vector machine [6,11], nearest-mean [7] and linear discriminant analysis [5,8,10].

Traditional machine learning methods usually assume that the distribution of the training and testing data are consistent, but they are indeed different in many real-world situations. As a result, the testing performance could be much worse than the training performance. TL [13] is a well-known approach for handling data distribution discrepancy. For example, Yang et al. [12] have used LMPROJ [18], a transductive TL algorithm, for seizure detection using EEG signals and achieved outstanding performance.

Another problem with many existing machine learning approaches is that black-box models can only be obtained, which lack the interpretability that is very important in medical diagnostics. In this paper, intrinsically interpretable TSK fuzzy system is used as our base model.

III. TSK Fuzzy System Model

The Mamdani model [20] and the TSK model [21,22,38] are two popular fuzzy system models. The latter is adopted in this study due to its simplicity and flexibility. This section introduces the TSK fuzzy system model and the corresponding training algorithm. The integration of the model with TL and SSL for achieving better classification performance will be discussed in the next section.

A. Concept and Principle

TSK fuzzy system is a rule based system. A typical rule can be described as

TSK fuzzy rule R^k :

$$\text{IF } x_1 \text{ is } A_1^k \ \wedge x_2 \text{ is } A_2^k \ \wedge\cdots\wedge x_d \text{ is } A_d^k \tag{1}$$

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

where K is the number of fuzzy rules, d is the number of inputs, A_i^k is a fuzzy set in the ith input domain for the kth rule, and $\wedge$ is a fuzzy conjunction operator. Each rule is premised on the input vector $\mathbf{x}=[x_1,x_2,\cdots,x_d]^T$ which is mapped to a singleton $f^k(\mathbf{x})$, a linear function of the input. The output of the TSK fuzzy system is computed as

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

where

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

and

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

in which $\mu_{A_i^k}(x_i)$ is the membership grade of x_i on A_i^k [17, 22].

Gaussian membership function is used in this paper:

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

where

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

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

in which N is the total number of dataset. u_{jk} denotes the fuzzy membership and can be obtained by Fuzzy C-means (FCM) clustering or the likes [30,39]. h is a scaling parameter which can be set manually, or optimized by some learning strategies such as cross-validation.

Denote

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

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

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

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

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

(2.a) can then be expressed as [17,22]:

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

B. Learning Algorithm for TSK Fuzzy Model

Given a training dataset of EEG signals (source domain) $D_S = \{\mathbf{x}_i, \mathbf{y}_i \mid \mathbf{x}_i \in R^d, \mathbf{y}_i \in R^C, i = 1,\cdots,N_S\}$, the least squares method can be used to optimize the consequent parameters $\mathbf{p}_g$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

The objective function is:

$$\min_{\mathbf{p}_{\mathrm{g}}} \quad J_{TSK}(\mathbf{p}_{\mathrm{g}}) = \frac{1}{2}\sum_{j=1}^{C}\sum_{i=1}^{N_{\mathrm{S}}}\left\|\mathbf{p}_{\mathrm{g},j}^{T}\mathbf{x}_{\mathrm{g}i} - y_{ij}\right\|^{2} + \frac{\lambda_1}{2}\sum_{j=1}^{C}\mathbf{p}_{\mathrm{g},j}^{T}\mathbf{p}_{\mathrm{g},j}\ , \quad (4)$$

where C is the number of classes, $\mathbf{p}_{\mathrm{g},j}$ is the consequent parameter vector of the jth class, $\mathbf{x}_i$ is the d-dimension input vector of the ith sample, $\mathbf{y}_i$ is the C-dimension label vector of the ith sample (y_{ij}=1 when the ith sample belongs to the jth class; otherwise, y_{ij}=0), and $\lambda_1 > 0$ is a regularization parameter, which controls the tradeoff between the complexity of the classifier and the tolerance of error. λ_1 can be set manually or determined by cross-validation [31].

The minimum of $J_{TSK}(\mathbf{p}_{\mathrm{g}})$ is obtained when the derivative of J_{TSK} w.r.t. each $\mathbf{p}_{\mathrm{g},j}$ is zero. The solution is:

$$\mathbf{p}_{\mathrm{g},j} = (\lambda_1\mathbf{I}_{((d+1)*K)\times((d+1)*K)} + \sum_{i=1}^{N_{\mathrm{S}}}\mathbf{x}_{\mathrm{g}i}\mathbf{x}_{\mathrm{g}i}^{T})^{-1}\left(\sum_{i=1}^{N_{\mathrm{S}}}\mathbf{x}_{\mathrm{g}i}y_{ij}\right). \quad (5)$$

The steps for training TSK fuzzy system is summarized in Algorithm 1.

Algorithm 1: *TSK fuzzy system model*

Initialization: Set the number of fuzzy rules K and the regularization parameter λ_1.

Stage 1: Construct dataset for linear regression

Step 2	Determine the antecedents of the TSK fuzzy system by clustering or other partition techniques to partition the dataset in the input space.
Step 3:	Construct the new dataset $\tilde{D} = \{\mathbf{x}_{gi}, \mathbf{y}_i\}$ using (3.a)-(3.c).

Stage 2: Obtain the decision function of the TSK fuzzy system model

Step 4:	Obtain the parameters of the TSK fuzzy system using (5) and construct the decision function (3.f).

IV. Transductive TL and SSL

In this section, the TL-SSL-TSK model, which integrates TL, SSL with a TSK fuzzy system model, is proposed.

A. *Transductive TL*

As mentioned before, TL can be used to reduce the discrepancy in data distribution between the training data (source domain) and testing data (target domain). Maximum mean discrepancy (MMD) is used in this paper to measure the distribution distance between the two domains. By minimizing the MMD, the difference in data distribution between the source and the target domains can be reduced effectively, which makes the testing performance close to the training performance. The effectiveness of TL for EEG signal classification has been demonstrated in [12]. In this paper, the same technique is used to enhance the performance of the TSK fuzzy system.

Given a set of labeled training data $D_{\mathrm{S}} = \{\mathbf{x}_i, \mathbf{y}_i \mid \mathbf{x}_i \in R^d, \mathbf{y}_i \in R^C, i = 1, \cdots, N_{\mathrm{S}}\}$ in the source domain and a set of unlabeled testing data $D_{\mathrm{T}} = \{\mathbf{x}_i \mid \mathbf{x}_i \in R^d, i = 1, \cdots, N_{\mathrm{T}}\}$ in the target domain, the projected squared MMD distance between the source domain and target domain is defined as [12,18]:

$$\begin{aligned} d(P_{\mathrm{source}}, P_{\mathrm{target}}) &= \mathrm{PMMD}^2 = \sum_{j=1}^{C}\left\|\frac{1}{N_{\mathrm{S}}}\sum_{i=1}^{N_{\mathrm{S}}}\mathbf{p}_{g,j}^{T}\mathbf{x}_{gi,\mathrm{S}} - \frac{1}{N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{T}}}\mathbf{p}_{g,j}^{T}\mathbf{x}_{gi,\mathrm{T}}\right\|^{2} \\ &= \sum_{j=1}^{C}\Big(\frac{1}{N_{\mathrm{S}}^{2}}\sum_{i=1}^{N_{\mathrm{S}}}\sum_{j=1}^{N_{\mathrm{S}}}\mathbf{p}_{g,j}^{T}\mathbf{x}_{gi,\mathrm{S}}\mathbf{x}_{gj,\mathrm{S}}^{T}\mathbf{p}_{g,j} + \frac{1}{N_{\mathrm{T}}^{2}}\sum_{i=1}^{N_{\mathrm{T}}}\sum_{j=1}^{N_{\mathrm{T}}}\mathbf{p}_{g,j}^{T}\mathbf{x}_{gi,\mathrm{S}}\mathbf{x}_{gj,\mathrm{T}}^{T}\mathbf{p}_{g,j}, \\ &\quad - \frac{2}{N_{\mathrm{S}}N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{S}}}\sum_{j=1}^{N_{\mathrm{T}}}\mathbf{p}_{g,j}^{T}\mathbf{x}_{gi,\mathrm{S}}\mathbf{x}_{gj,\mathrm{T}}^{T}\mathbf{p}_{g,j}\Big) \end{aligned} \quad (6)$$

where $\mathbf{x}_{gi,\mathrm{S}}$ is the ith sample in the source domain (training dataset), $\mathbf{x}_{gi,\mathrm{T}}$ is the ith sample in the target domain (testing dataset), and $\mathbf{p}_{g}$ is an expected projection for the TSK-FS model. Let

$$\begin{aligned} \Omega = &\frac{1}{N_{\mathrm{S}}^{2}}\mathbf{x}_{g,\mathrm{S}}[\mathbf{1}]^{N_{\mathrm{S}}\times N_{\mathrm{S}}}\mathbf{x}_{g,\mathrm{S}}^{T} + \frac{1}{N_{\mathrm{T}}^{2}}\mathbf{x}_{g,\mathrm{T}}[\mathbf{1}]^{N_{\mathrm{T}}\times N_{\mathrm{T}}}\mathbf{x}_{g,\mathrm{T}}^{T}, \\ &- \frac{1}{N_{\mathrm{S}}N_{\mathrm{T}}}\mathbf{x}_{g,\mathrm{S}}[\mathbf{1}]^{N_{\mathrm{S}}\times N_{\mathrm{T}}}\mathbf{x}_{g,T}^{T} - \frac{1}{N_{\mathrm{S}}N_{\mathrm{T}}}\mathbf{x}_{g,\mathrm{T}}[\mathbf{1}]^{N_{\mathrm{T}}\times N_{\mathrm{S}}}\mathbf{x}_{g,\mathrm{S}}^{T} \end{aligned} \quad (7)$$

then (6) can be written as

$$d(P_{\mathrm{source}}, P_{\mathrm{target}}) = \sum_{j=1}^{C}\mathbf{p}_{g,j}^{T}\mathbf{\Omega}\mathbf{p}_{g,j}\ . \quad (8)$$

B. *SSL*

For offline classification, SSL can be adopted to further improve the classification performance by using the unlabeled testing data which also contain useful information. One approach to realize this idea is illustrated in Fig. 1. It is based on the assumption that the data within the same class are close to each other.

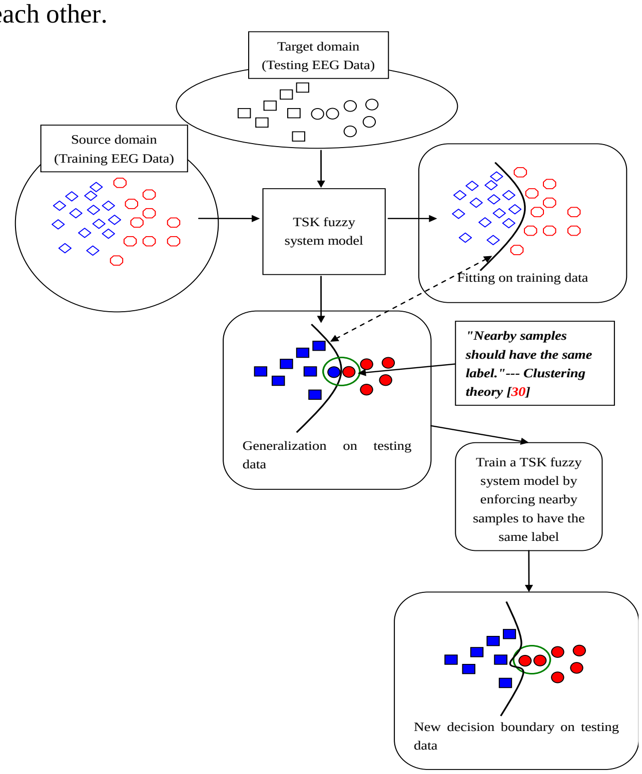

Fig. 1 Illustration of label clustering SSL.

The above heuristics is used to further improve the performance of our system. Specifically, a novel FCM-like [30] SSL approach is designed for label clustering:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

$$\begin{aligned} &\min \ J_{SSL}(\mathbf{U})=\sum_{j=1}^{C}\sum_{i=1}^{N_\mathrm{T}}\mu_{ij}^{m}\left\|\mathbf{p}_{\mathrm{g,S}}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{T}}-\boldsymbol{\theta}_{j}\right\|^{2} \\ &s.t. \quad \mu_{ij}\in[0,1] \ \text{and} \ \sum_{j=1}^{C}\mu_{ij}=1 \end{aligned} \tag{9}$$

where C is the total number of clusters, N_T is the total number of data of target domain (testing dataset), m is the fuzzy index in FCM, $\mathbf{x}_{gi,\mathrm{T}}$ is the ith sample in the target domain, $\mathbf{p}_{\mathrm{g,S}}$ is the expected projection of source domain for the TSK fuzzy system model, $\mathbf{U}=[\mu_{ij}]_{C\times N_\mathrm{T}}$ is the matrix of fuzzy partition with μ_{ij} denoting the label membership of the ith unlabeled sample of the target domain belonging to the jth cluster, and $\boldsymbol{\theta}_j=[0,\ldots,0,\underset{jth}{1},0,\ldots,0]^T$ is a known label vector of the jth cluster $\left(1\le j\le C\right)$.

C. Learning Algorithms of the TL-SSL-TSK Model

Two learning algorithms are proposed for the TL-SSL-TSK model, which result in two versions of the models, the simple version, namely the *S-TL-SSL-TSK* model, and the advanced version, the *A-TL-SSL-TSK* model.

1) Learning algorithm of S-TL-SSL-TSK

TL and SSL can be integrated with the TSK model using the simple learning objective function:

$$\begin{aligned} \min_{\mathbf{p}_\mathrm{g}} \ J_{S-TL-SSL-TSK} =& \frac{1}{2}\sum_{j=1}^{C}\sum_{i=1}^{N_\mathrm{S}}\left\|\mathbf{p}_{\mathrm{g},j}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{S}}-y_{ij,\mathrm{S}}\right\|^{2}+\frac{\lambda_1}{2}\sum_{j=1}^{C}\mathbf{p}_{\mathrm{g},j}^{T}\mathbf{p}_{\mathrm{g},j}+\lambda_2\sum_{j=1}^{C}\mathbf{p}_{\mathrm{g},j}^{T}\boldsymbol{\Omega}\mathbf{p}_{\mathrm{g},j} \\ &+\lambda_3\sum_{j=1}^{C}\sum_{i=1}^{N_T}\hat{\mu}_{ij}^{m}\left\|\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{T}}-\boldsymbol{\theta}_j\right\|^{2} \end{aligned} \tag{10}$$

where $\lambda_1>0$, $\lambda_2>0$, $\lambda_3>0$ are the regularization parameters. The first two terms in (10) are used to learn the regularized TSK model based only on the EEG data from the source domain. The third term is due to transductive TL, and the fourth term due to SSL. For simplicity, the label membership parameter $\hat{\mu}_{ij}$ is set to be a fixed knowledge transfer parameter which directly inherits from the source domain. Similar to classical FCM, $\hat{\mu}_{ij}$ is computed as:

$$\hat{\mu}_{ij}=\frac{\left(\dfrac{1}{\left\|\mathbf{p}_{\mathrm{g,S}}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{T}}-\boldsymbol{\theta}_j\right\|^{2}}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{C}\left(\dfrac{1}{\left\|\mathbf{p}_{\mathrm{g,S}}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{T}}-\boldsymbol{\theta}_k\right\|^{2}}\right)^{\frac{1}{m-1}}} \tag{11}$$

where the $\mathbf{x}_{\mathrm{g}i,\mathrm{T}}$ is the ith unlabeled sample in the target domain, $\mathbf{p}_{\mathrm{g,S}}$ is an expected projection of the source domain for the TSK-FS model and can be obtained using (5), $\boldsymbol{\theta}_j=[0,\ldots,0,\underset{jth}{1},0,\ldots,0]^T$ is a known label vector of the jth cluster.

Let $\hat{\boldsymbol{U}}=[\hat{\boldsymbol{U}}_1,\ldots,\hat{\boldsymbol{U}}_C]\in R^{1\times CN_\mathrm{T}}$ be the label membership values for all the unlabeled data where $\hat{\boldsymbol{U}}_j=[\hat{\mu}_{1j},\ldots,\hat{\mu}_{ij},\ldots,\hat{\mu}_{N_\mathrm{T}j}]\in R^{1\times N_\mathrm{T}}$ and $j=1\ldots C$, $\hat{\boldsymbol{U}}=diag(\hat{\boldsymbol{U}})\in R^{CN_\mathrm{T}\times CN_\mathrm{T}}$ be a diagonal matrix of $\hat{\boldsymbol{U}}$, $\mathbf{V}=[\underbrace{\mathbf{E},\ldots,\mathbf{E}}_{C}]\in R^{N_\mathrm{T}\times CN_\mathrm{T}}$ be a transformation matrix where $\mathbf{E}$ is an identity matrix of size N_T, and $\mathbf{Q}=[\mathbf{q}_1,...,\mathbf{q}_C]\in R^{C\times CN_\mathrm{T}}$ be a transformation matrix where all the values of the jth row of $\mathbf{q}_j\in R^{C\times N_\mathrm{T}}$ are 1 and the rest are 0. Then, (10) can be further expressed as:

$$\begin{aligned} \min_{\mathbf{p}_\mathrm{g}} \ J_{S-TL-SSL-TSK} =& \frac{1}{2}tr\left((\mathbf{p}_\mathrm{g}^{T}\mathbf{x}_{\mathrm{g,S}}-\mathbf{y}_\mathrm{S})(\mathbf{p}_\mathrm{g}^{T}\mathbf{x}_{\mathrm{g,S}}-\mathbf{y}_\mathrm{S})^{T}\right)+\frac{\lambda_1}{2}tr(\mathbf{p}_\mathrm{g}^{T}\mathbf{p}_\mathrm{g})+\lambda_2 tr\left(\mathbf{p}_\mathrm{g}^{T}\boldsymbol{\Omega}\mathbf{p}_\mathrm{g}\right) \\ &+\lambda_3 tr\left((\mathbf{p}_\mathrm{g}^{T}\mathbf{x}_{\mathrm{g,T}}\mathbf{V}-\mathbf{Q})\hat{\mathbf{U}}(\mathbf{p}_\mathrm{g}^{T}\mathbf{x}_{\mathrm{g,T}}\mathbf{V}-\mathbf{Q})^{T}\right) \end{aligned} \tag{12}$$

where

$$\boldsymbol{\Omega}=\frac{1}{{N_\mathrm{T}}^2}\mathbf{x}_{g,\mathrm{T}}[\mathbf{1}]^{N_\mathrm{T}\times N_\mathrm{T}}\mathbf{x}_{g,\mathrm{T}}^{T}+\frac{1}{{N_\mathrm{S}}^2}\mathbf{x}_{g,\mathrm{S}}[\mathbf{1}]^{N_\mathrm{S}\times N_\mathrm{S}}\mathbf{x}_{g,\mathrm{S}}^{T}-\frac{1}{N_\mathrm{S}N_\mathrm{T}}\mathbf{x}_{g,\mathrm{T}}[\mathbf{1}]^{N_\mathrm{T}\times N_\mathrm{S}}\mathbf{x}_{g,\mathrm{S}}^{T}-\frac{1}{N_\mathrm{S}N_\mathrm{T}}\mathbf{x}_{g,\mathrm{S}}[\mathbf{1}]^{N_\mathrm{S}\times N_\mathrm{T}}\mathbf{x}_{g,\mathrm{T}}^{T}.$$

The optimal consequent parameters $\mathbf{p}_g$ can be computed by setting the derivatives of $J_{S-TL-SSL-TSK}$ w.r.t. $\mathbf{p}_\mathrm{g}$ to zero, and the solution is

$$\mathbf{p}_\mathrm{g}=\begin{pmatrix}\mathbf{x}_{\mathrm{g,S}}\mathbf{x}_{\mathrm{g,S}}^{T}+\lambda_1\mathbf{I}_{((d+1)*K)\times((d+1)*K)} \\ +2\lambda_2\boldsymbol{\Omega}+2\lambda_3\mathbf{x}_{\mathrm{g,T}}\mathbf{V}\hat{\mathbf{U}}\mathbf{V}^{T}\mathbf{x}_{\mathrm{g,T}}^{T}\end{pmatrix}^{-1}(\mathbf{x}_{\mathrm{g,S}}\mathbf{y}^{T}+2\lambda_3\mathbf{x}_{\mathrm{g,T}}\mathbf{V}\ \hat{\mathbf{U}}\mathbf{Q}^{T}). \tag{13}$$

The algorithm for S-TL-SSL-TSK is given in Algorithm 2.

Algorithm 2: *Learning algorithm for S-TL-SSL-TSK.*

Initialization: Set the number of fuzzy rules K, the regularization parameters $\lambda_1,\lambda_2,\lambda_3$, and the fuzzy index m.

Stage 1: Construct dataset for linear regression

Step 1: Determine the antecedents of the TSK fuzzy system by clustering or other partition techniques to partition the dataset in the input space.

Step 2: Construct the new training dataset $\tilde{D}_\mathrm{S}=\left\{\mathbf{x}_{g,\mathrm{S}},\mathbf{y}_\mathrm{S}\right\}$ in the source domain and the new testing datasets $\tilde{D}_\mathrm{T}=\left\{\mathbf{x}_{g,\mathrm{T}}\right\}$ by using (3.a)-(3.c).

Stage 2: Obtain the knowledge transfer parameter

Step 3: Obtain the consequent parameters $\mathbf{p}_{\mathrm{g,S}}$ of the source domain using (5).

Step 4: Obtain the knowledge transfer parameter, i.e. label membership $\hat{\mu}_{ij}$, using (11) with the optimized $\mathbf{p}_{\mathrm{g,S}}$;

Stage 3: Generate the S-TL-SSL-TSK model

Step 5: Obtain the consequent parameters $\mathbf{p}_\mathrm{g}$ of the target domain using (13) and get the decision function (3.f) of the S-TL-SSL-TSK model.

Remark: Although the S-TL-SSL-TSK algorithm introduces

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

the transductive TL and SSL mechanism for fuzzy system training, the learning ability of this algorithm can still be further enhanced. Since the label membership parameter $\hat{\mu}_{ij}$ in S-TL-SSL-TSK is a fixed parameter, i.e. directly inheriting from the source domain in step 4 of Algorithm 2, the algorithm is weak in adapting to $\hat{\mu}_{ij}$. In the next subsection, a more adaptive algorithm will be proposed.

2) Learning algorithm of A-TL-SSL-TSK

To further enhance the abilities of TL and SSL in the proposed TL-SSL-TSK model, a more sophisticated label clustering mechanism is proposed to replace the fourth term in (10):

$$J_{label-clustering} = \sum_{i=1}^{N_{\mathrm{T}}}\sum_{j=1}^{C}\left(\eta\mu_{ij}^{m} + (1-\eta)\hat{\mu}_{ij}^{m}\right)\left\|\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{T}} - \boldsymbol{\theta}_{j}\right\|^{2}, \quad (14)$$

where $\eta\in[0,1]$ is a trade-off parameter controlling the degree of knowledge transfer between the source domain and the target domain. When $\eta\to 1$, the knowledge in the target domain, i.e., parameter μ_{ij}, is emphasized. In contrast, when $\eta\to 0$, the knowledge in the source domain, i.e., parameter $\hat{\mu}_{ij}$, is emphasized.

Substituting (14) into (10), the objective function of A-TL-SSL-TSK is expressed as:

$$\begin{aligned}\min_{\mathbf{p}_{\mathrm{g}}}\ J_{A-TL-SSL-TSK} &= \frac{1}{2}\sum_{j=1}^{C}\sum_{i=1}^{N_{\mathrm{S}}}\left\|\mathbf{p}_{\mathrm{g}j}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{S}} - y_{ij,\mathrm{S}}\right\|^{2} + \frac{\lambda_1}{2}\sum_{j=1}^{C}\mathbf{p}_{\mathrm{g}j}^{T}\mathbf{p}_{\mathrm{g}j} + \lambda_2\sum_{j=1}^{C}\mathbf{p}_{\mathrm{g}j}^{T}\boldsymbol{\Omega}\mathbf{p}_{\mathrm{g}j}\\ &+\lambda_3\sum_{i=1}^{N_{\mathrm{T}}}\sum_{j=1}^{C}\left(\eta\mu_{ij}^{m} + (1-\eta)\hat{\mu}_{ij}^{m}\right)\left\|\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{T}} - \boldsymbol{\theta}_{j}\right\|^{2}\\ s.t.\quad & \mu_{ij}\in[0,1] \text{ and } \sum_{j=1}^{C}\mu_{ij}=1\end{aligned} \quad (15)$$

Similar to (12), (15) can be expressed as:

$$\begin{aligned}\min_{\mathbf{p}_{\mathrm{g}}}\ J_{A-TL-SSL-TSK} &= \frac{1}{2}tr\left((\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g,S}} - \mathbf{y}_{\mathrm{S}})(\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g,S}} - \mathbf{y}_{\mathrm{S}})^{T}\right) + \frac{\lambda_1}{2}tr(\mathbf{p}_{\mathrm{g}}^{T}\mathbf{p}_{\mathrm{g}})\\ &+\lambda_2 tr\left(\mathbf{p}_{\mathrm{g}}^{T}\boldsymbol{\Omega}\mathbf{p}_{\mathrm{g}}\right) + \lambda_3 tr\left((\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g,T}}\mathbf{V} - \mathbf{Q})\left(\eta\breve{\mathbf{U}} + (1-\eta)\hat{\mathbf{U}}\right)(\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g,T}}\mathbf{V} - \mathbf{Q})^{T}\right)\\ s.t.\quad & \mu_{ij}\in[0,1] \text{ and } \sum_{j=1}^{C}\mu_{ij}=1\end{aligned} \quad (16)$$

The joint optimization of $\mathbf{p}_g$ and μ_{ij} in (16) makes it non-convex and a closed-form solution is not available. Thus, an iterative optimization method [15,32] is adopted in this paper, which contains the two steps below.

i) ***Step 1***: Compute $\mathbf{p}_g$:

When the parameter μ_{ij} is fixed, i.e., $\breve{\mathbf{U}}$ is fixed, the minimization of $J_{A-TL-SSL-TSK}$ can be computed by setting the derivatives of (16) w.r.t. $\mathbf{p}_{\mathrm{g}}$ to zero, and we obtain:

$$\mathbf{p}_{\mathrm{g}} = \begin{pmatrix}\mathbf{I}_{((d+1)*K)\times((d+1)*K)} + \lambda_1\mathbf{x}_{\mathrm{g,S}}\mathbf{x}_{\mathrm{g,S}}^{T} + 2\lambda_2\boldsymbol{\Omega}\\ +2\lambda_3\mathbf{x}_{\mathrm{g,T}}\mathbf{V}\left(\eta\mathbf{U} + (1-\eta)\hat{\mathbf{U}}\right)\mathbf{V}^{T}\mathbf{x}_{\mathrm{g,T}}^{T}\end{pmatrix}^{-1}\begin{pmatrix}\lambda_1\mathbf{x}_{\mathrm{g,S}}\mathbf{y}^{T}\\ +2\lambda_3\mathbf{x}_{\mathrm{g,T}}\mathbf{V}\ \left(\eta\mathbf{U} + (1-\eta)\hat{\mathbf{U}}\right)\mathbf{Q}^{T}\end{pmatrix} \quad (17)$$

ii) ***Step 2***: Compute μ_{ij}:

When the parameter $\mathbf{p}_g$ is fixed, by minimizing (17) using the Lagrangian method, the following update equation for the label fuzzy membership μ_{ij} is obtained.

$$\mu_{ij} = \frac{\left(\dfrac{1}{\left\|\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{T}} - \boldsymbol{\theta}_{j}\right\|^{2}}\right)^{\frac{1}{m-1}}}{\sum_{k=1}^{C}\left(\dfrac{1}{\left\|\mathbf{p}_{\mathrm{g}}^{T}\mathbf{x}_{\mathrm{g}i,\mathrm{T}} - \boldsymbol{\theta}_{k}\right\|^{2}}\right)^{\frac{1}{m-1}}} \quad (18)$$

The complete learning algorithm for A-TL-SSL-TSK is given in Algorithm 3.

Algorithm 3: *Learning algorithm for the A-TL-SSL-TSK model.*

Initialization: Set the number of fuzzy rules K, regularization parameters $\lambda_1,\lambda_2,\lambda_3$, transfer parameter η, convergence threshold ε, fuzzy index m, maximum number of iterations T. Initialize the fuzzy partition $\mu_{ij}^{(s)}$ and set the iterative index $l=1$

Stage 1: Construct dataset for linear regression

Step 1:	Determine the antecedents of TSK fuzzy system by clustering or other partition techniques to partition the dataset in the input space.
Step 2:	Construct the new training dataset $\tilde{D}_S = \{\mathbf{x}_{g,\mathrm{S}}, \mathbf{y}_{\mathrm{S}}\}$ of source domain and the new testing datasets $\tilde{D}_t = \{\mathbf{x}_{g,\mathrm{T}}\}$ by using (3.a)-(3.c).

Stage 2: Obtain the knowledge transfer parameter

Step 3:	Obtain the consequent parameters $\mathbf{p}_{\mathrm{g,S}}$ of the source domain by (5).
Step 4:	Obtain the knowledge transfer parameter, i.e., label membership $\hat{\mu}_{ij}$, using (11) and the optimized $\mathbf{p}_{\mathrm{g,S}}$.

Stage 3: Generate the A-TL-SSL-TSK model

Step 5:	Compute the consequent parameter $\mathbf{p}_g^{(l+1)}$ using (17) with $\mu_{ij}^{(l)}$
Step 6:	Compute the label fuzzy membership $\mu_{ij}^{(l+1)}$ using (18) with $\mathbf{p}_g^{(l+1)}$;
Step 7:	If $\left\|U^{(l+1)} - U^{(l)}\right\| < \varepsilon$ or the number of iterations $l > T$, terminate and output the consequent parameter $\mathbf{p}_g^{(l+1)}$ of the A-TL-SSL-TSK model; otherwise, set $l = l+1$ and go to step 5.
Step 8:	Obtain the optimized consequent parameters $\mathbf{p}_g$ of the target domain by following steps 5 to 7 and get the decision function (3.f) of the A-TL-SSL-TSK model.

V. Experimental Study

In this section, the two proposed models, i.e., S-TL-SSL-TSK and A-TL-SSL-TSK, are evaluated by performing two-class classification between EEG signals of healthy subjects and that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

of epileptic subjects captured during seizure. In addition, they are compared with six classical non-TL-based methods and four TL-based or SSL-based methods, as shown in Table I.

To ensure fair comparison, the same EEG data and feature extraction methods are used for all the algorithms, as in our previous work [12]. The details of the experimental settings and the scenarios of TL are described below.

A. Experimental Settings

In our experiments, all the algorithms were implemented using MATLAB on a computer with Intel Core i5-3317U 1.70 GHz CPU and 16GB RAM. The experimental settings are summarized in Table I. Note that TSVM and S4VM are SSL-based methods, where the unlabeled EEG data in the target domain are used for learning. LMPROJ and GTL2 are TL-based methods.

TABLE I. EXPERIMENTAL SETTINGS

	Non-TL-based methods	TL-based or SSL-based methods
Methods for comparison	1. LDA [5,8,10] 2. DT [9] 3. NB [7,9] 4. NM [7] 5. SVM [6] **6. TSK**	1. TSVM [33] 2. S4VM [34] 3. LMPROJ [18] 4. GTL2 [35] **5. S-TL-SSL-TSK** **6. A-TL-SSL-TSK**
EEG data	The EEG data used in this study are the same as those in [12], and can be downloaded from http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3 [41]. It contains five groups of data (Groups A to E) and each group contains 100 single channel EEG segments of 23.6s duration. The sampling rate of all the datasets was 173.6 Hz. Table II gives a detailed description of the five groups. Fig. 2 shows some typical EEG signals in each group.	
Feature extraction methods	1. Wavelet packet decomposition (WPD) 2. Short time Fourier transform (STFT) 3. Kernel principal component analysis (KPCA)	
Performance evaluation measures	1. Accuracy: Number of correctly predicted testing data divided by the number of the total testing data. 2. Friedman test combined with Holm's post hoc test [36,37].	
Method-specific settings	1. For LDA, DT, NB, NM, SVM and LMPROJ methods, we adopted the same experimental setting in [12]. 2. For TSVM, S4VM and GTL2 methods, we adopted the same experimental setting in [33], [34] and [35] respectively. 3. For TSK, S-TL-SSL-TSK and A-TL-SSL-TSK, the number of fuzzy rules was selected from {5,10,15,20,25,30}, the regularization parameters λ_1, λ_2 and λ_3 were selected from $\{10^{-3}, 10^{-2}, \ldots, 10^{2}, 10^{3}\}$, the transfer balance parameter η was selected from $\{0, 0.1, 0.2, 0.3, \ldots, 0.9, 1\}$, and the fuzzy index m was selected from $\{1.1, 1.5, 2, 2.5\}$. Five-fold cross-validation on the training data was applied for all the methods.	

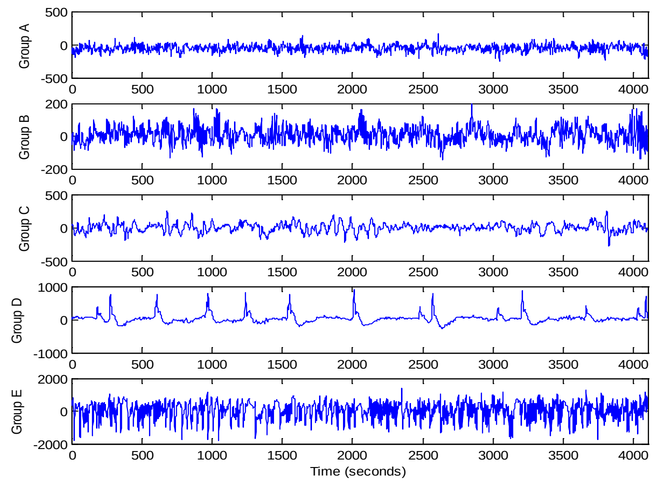

Fig. 2 Typical EEG signals in Groups A-E.

TABLE II DESCRIPTION OF THE EEG DATA

Subjects	Groups	Size of groups	Descriptions of datasets
Healthy	A	100	EEG signals measured from healthy people with eyes open
	B	100	EEG signals measured from healthy people with eyes closed
Epileptic	C	100	EEG signals obtained in the hippocampal formation of the opposite hemisphere of the brain during seizure free intervals
	D	100	EEG signals obtained from within the epileptogenic zone during seizure free intervals
	E	100	EEG signals measured during seizure

B. TL Scenarios for Epileptic EEG Recognition

The two TL scenarios used in our previous work [12] were adopted in this study. In the first scenario, the data distributions of the source and the target domain were the same, as shown in the first part of Table III. In the second scenario, the data distributions were different, as shown in the second part of Table III. More details of these two scenarios can be found in Section 4.1.2 of [12].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

TABLE III DETAILS OF THE TL SCENARIOS FOR EPILEPTIC EEG RECOGNITION

Scenario	**Datasets**	**Source domain (training dataset)**	**Target domain (testing dataset)**	**Number of samples in the source domain**	**Number of samples in the target domain**
Scenario 1: Same distribution	D1	Groups A and E	Groups A and E	150	50
	D2	Groups A, B and E	Groups A, B and E	150	50
Scenario 2: Different distribution	D3	Groups A and E	Groups A and C	50	50
	D4	Groups A and E	Groups A and D	50	50
	D5	Groups A, C and E	Groups B, C and E	75	75
	D6	Groups A, D and E	Groups B, D and E	75	75

TABLE IV PERFORMANCE COMPARISON OF THE CLASSIFIERS USING WPD FEATURES.

Part A: Non-TL methods						
Datasets	LDA	DT	NB	NM	SVM	TSK
D1	0.922±0.012	0.887±0.01	0.851±0.006	0.798±0.008	0.918±0.004	0.914±0.025
D2	0.927±0.087	0.936±0.008	0.922±0.012	0.833±0.02	0.92±0.001	0.928±0.009
D3	0.544±0.007	0.825±0.02	0.527±0.007	0.425±0.004	0.824±0.06	0.826±0.056
D4	0.561±0.009	0.808±0.02	0.573±0.008	0.469±0.01	0.816±0.044	0.812±0.068
D5	0.772±0.006	0.802±0.018	0.559±0.005	0.515±0.008	0.807±0.03	0.812±0.054
D6	0.739±0.01	0.778±0.07	0.622±0.026	0.535±0.009	0.81±0.06	0.824±0.038
Part B: TL methods						
Datasets	TSVM	S4VM	LMPROJ	GTL2	S-TL-SSL-TSK	A-TL-SSL-TSK
D1	0.924±0.009	**0.967±0.038**	0.937±0.004	0.931±0.022	0.950±0.010	0.958±0.011
D2	0.929±0.007	**0.965±0.018**	0.936±0.001	0.945±0.036	0.957±0.009	0.961±0.006
D3	0.843±0.034	0.856±0.033	0.941±0.011	0.929±0.049	0.953±0.007	**0.957±0.005**
D4	0.837±0.025	0.860±0.025	0.945±0.008	0.917±0.067	0.960±0.012	**0.966±0.008**
D5	0.821±0.039	0.829±0.056	0.925±0.008	0.939±0.048	0.948±0.017	**0.956±0.016**
D6	0.845±0.041	0.868±0.060	0.945±0.005	0.922±0.051	0.966±0.006	**0.968±0.007**

TABLE V PERFORMANCE COMPARISON OF THE CLASSIFIERS USING STFT FEATURES.

Datasets	Part A: Non-TL methods					
	LDA	DT	NB	NM	SVM	TSK
D1	0.983±0.036	0.994±0.003	0.947±0.005	0.990±0.019	0.988±0.003	0.988±0.014
D2	0.981±0	0.951±0.02	0.933±0.017	0.890±0.036	0.994±0.002	0.992±0.011
D3	0.536±0.03	0.62±0.09	0.53±0.022	0.495±0.003	0.51±0.015	0.564±0.066
D4	0.565±0.015	0.675±0.05	0.632±0.003	0.537±0.007	0.562±0.031	0.586±0.057
D5	0.622±0.02	0.476±0.04	0.417±0.043	0.483±0.02	0.864±0.02	0.873±0.036
D6	0.601±0.03	0.46±0.02	0.375±0.022	0.484±0.018	0.874±0.054	0.876±0.073
Datasets	Part B: TL methods					
	TSVM	S4VM	LMPROJ	GTL2	S-TL-SSL-TSK	A-TL-SSL-TSK
D1	0.992±0.010	**0.996±0.008**	0.977±0.001	0.992±0.026	0.994±0.013	0.994±0.009
D2	**0.998±0.004**	0.994±0.009	0.994±0.001	0.993±0.057	0.994±0.006	0.996±0.006
D3	0.526±0.036	0.860±0.036	0.945±0.008	0.910±0.027	0.948±0.007	**0.963±0.009**
D4	0.598±0.072	0.872±0.027	0.938±0.01	0.940±0.018	**0.957±0.005**	**0.957±0.003**
D5	0.876±0.034	0.884±0.037	0.952±0.008	0.931±0.049	0.960±0.018	**0.967±0.012**
D6	0.891±0.075	0.905±0.039	0.949±0.01	0.924±0.039	0.953±0.025	**0.974±0.036**

TABLE VI PERFORMANCE COMPARISON OF THE CLASSIFIERS USING KPCA FEATURES.

Part A: Non-TL methods						
Datasets	LDA	DT	NB	NM	SVM	TSK
D1	0.882±0.058	0.932±0.017	0.82±0.075	0.747±0.064	0.92±0.02	0.936±0.038
D2	0.827±0.065	0.86±0.026	0.682±0.094	0.649±0.041	0.91±0.011	0.924±0.026
D3	0.741±0.09	0.795±0.07	0.628±0.05	0.673±0.11	0.79±0.02	0.752±0.033
D4	0.765±0.09	0.944±0.06	0.625±0.13	0.721±0.12	0.82±0.04	0.792±0.028
D5	0.46±0.061	0.739±0.043	0.491±0.3	0.447±0.06	0.710±0.02	0.766±0.051
D6	0.53±0.035	0.711±0.02	0.641±0.10	0.487±0.04	0.721±0.02	0.777±0.012
Part B: TL methods						
Datasets	TSVM	S4VM	LMPROJ	GTL2	S-TL-SSL-TSK	A-TL-SSL-TSK
D1	0.942±0.035	**0.984±0.021**	0.958±0.01	0.943±0.023	**0.984±0.015**	**0.984±0.012**
D2	0.924±0.022	**0.971±0.023**	0.948±0.004	0.906±0.026	0.958±0.014	0.966±0.013
D3	0.862±0.048	0.908±0.021	0.949±0.05	0.930±0.030	0.952±0.027	**0.957±0.029**
D4	0.894±0.059	0.918±0.016	0.963±0.009	0.940±0.039	0.978±0.022	**0.984±0.012**
D5	0.773±0.072	0.842±0.076	0.948±0.015	0.956±0.030	0.962±0.047	**0.978±0.018**
D6	0.764±0.067	0.891±0.043	0.95±0.008	0.946±0.021	0.958±0.036	**0.965±0.025**

TABLE VII. RESULTS OF FRIEDMAN TEST ON ALL THE DATASETS IN TERMS OF AVERAGE PERFORMANCE ($\alpha = 0.05$)

Part A: Friedman Test for S-TL-TSK-FS				**Part B: Friedman Test for A-TL-TSK-FS**			
Algorithms	**Friedman Rank**	***p*-value**	**Hypothesis**	**Algorithms**	**Friedman Rank**	***p*-value**	**Hypothesis**
LDA	8.6667	0	**Rejected**	LDA	8.6667	0	**Rejected**
DT	6.9444			DT	6.9444		
NB	9.7222			NB	9.7222		
NM	10.3889			NM	10.3889		
SVM	7.3889			SVM	7.4167		
TSK	6.5556			TSK	6.5556		
TSVM	5.2222			TSVM	5.2222		
S4VM	3.2222			S4VM	3.25		
LMPROJ	3			LMPROJ	3.0278		
GTL2	3.4722			GTL2	3.4722		
S-TL-SSL-TSK	1.4167			A-TL-SSL-TSK	1.3333		

TABLE VIII. HOLM'S POST HOC COMPARISON OVER THE RESULTS OF FRIEDMAN TEST ON ALL THE DATASETS IN TERMS OF AVERAGE PERFORMANCE ($\alpha = 0.05$)

Part A: Holm's Post Hoc Comparison for S-TT-TSK-FS					
i	**Algorithms**	$z = (R_0 - R_i) / SE$	p	$Holm = \alpha / i$	**Hypothesis**
10	NM	8.11568	0	0.005	**Rejected**
9	NB	7.512658	0	0.005556	**Rejected**
8	LDA	6.557872	0	0.00625	**Rejected**
7	SVM	5.402078	0	0.007143	**Rejected**
6	DT	5.000063	0.000001	0.008333	**Rejected**
5	TSK	4.6483	0.000003	0.01	**Rejected**
4	TSVM	3.442255	0.000577	0.0125	**Rejected**
3	GTL2	1.85932	0.062982	0.016667	**Not rejected**
2	S4VM	1.633186	0.10243	0.025	**Not rejected**
1	LMPROJ	1.432179	0.152093	0.05	**Not rejected**
Part B: Holm's Post Hoc Comparison for A-TT-TSK-FS					
i	**Algorithms**	$z = (R_0 - R_i) / SE$	p	$Holm = \alpha / i$	**Hypothesis**
10	NM	8.191058	0	0.005	**Rejected**
9	NB	7.588036	0	0.005556	**Rejected**
8	LDA	6.63325	0	0.00625	**Rejected**
7	SVM	5.502582	0	0.007143	**Rejected**
6	DT	5.075441	0	0.008333	**Rejected**
5	TSK	4.723678	0.000002	0.01	**Rejected**
4	TSVM	3.517632	0.000435	0.0125	**Rejected**
3	GTL2	1.934698	0.053027	0.016667	**Not rejected**
2	S4VM	1.73369	0.082973	0.025	**Not rejected**
1	LMPROJ	1.532683	0.125354	0.05	**Not rejected**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

C. Experimental results and discussions

1) Recognition performance analysis

The following arrangements were made to achieve comprehensive comparison. Three different feature extraction methods were used, i.e., WPD, STFT, and KPCA. For each method, we compared S-TL-SSL-TSK and A-TL-SSL-TSK with six non-TL-based methods and four sets of features. The classification results are shown in Tables IV-VI and the findings below are drawn.

(1) Among the six non-TL-based methods (LDA, DT, NB, NM, SVM and TSK), TSK showed the best overall performance. The results indicate that TSK is an effective approach for epileptic EEG recognition. In addition, it has better interpretability than other black-box methods like SVM.

(2) For Datasets D1 and D2, where the source and the target domains had the same data distribution, most non-TL-based methods achieved high accuracy. However, for Datasets D3- D6, where the source and the target domains had different data distributions, their performance deteriorated significantly. This is because the data distributions of the training and testing datasets were inappropriately assumed to be the same, and hence they were not able to adapt to the new data of different distribution in the target domain.

(3) Compared with the six classical non-TL-based methods, the six TL-based or SSL-based methods achieved higher accuracy consistently. These results indicate that TL and SSL can indeed be used to improve the classification performance of epileptic EEG recognition.

(4) Among the six TL-based or SSL-based methods, S-TL-SSL-TSK and A-TL-SSL-TSK showed the best overall recognition performance, especially for Datasets D3-D6. This is because TSVM and S4VM only used SSL, and LMPROJ and GTL2 only used TL, whereas the proposed S-TL-SSL-TSK and A-TL-SSL-TSK used both, i.e., S-TL-SSL-TSK and A-TL-SSL-TSK explicitly reduced the discrepancy in data distribution between the source and the target domains, and made use of the information contained in the unlabeled EEG samples of the target domain at the same time.

(5) A-TL-SSL-TSK almost always outperformed S-TL-SSL-TSK. This is expected, since A-TL-SSL-TSK employed a more flexible TL mechanism and was tuned using more parameters. However, the computational cost of A-TL-SSL-TSK is higher. So, the choice between S-TL-SSL-TSK and A-TL-SSL-TSK is a tradeoff between computational cost and classification accuracy.

To evaluate whether the performance difference among the algorithms were statistically significant, Friedman test [36,37] and the Holm post hoc test [36,37] were performed. Friedman test was used to compute the average ranks of the different methods, and to evaluate whether statistically significant difference existed among them. The null hypothesis was that there was no statistically significant difference. If the *p*-value was smaller than 0.05, the null hypothesis was rejected. The Holm post hoc test was also performed to verify if there was statistically significant difference between the control approach, i.e., the one achieving the best Friedman rank, and the other approaches.

We first compared S-TL-SSL-TSK with all the other methods except A-TL-SSL-TSK, and then A-TL-SSL-TSK with all the other methods except S-TL-SSL-TSK. The results are summarized in Tables VII and VIII. The results of the Friedman test indicate that there was statistically significant difference in accuracy between the two proposed methods and the other ten methods. The Holm's post hoc test shows that S-TL-SSL-TSK and A-TL-TSK-FS significantly outperformed all the six non-TL-based methods and TSVM, but not so for GTL2, S4VM and LMPROJ. Although the increase in performance of our proposed methods over GTL2, S4VM and LMPROJ was not statistically significant, they improved the interpretability which is important for medical diagnostics.

In summary, the experimental results demonstrate that S-TL-SSL-TSK and A-TL-SSL-TSK are suitable for epileptic EEG recognition: their classification performance is comparable to or ever better than many state-of-the-art classification algorithms, and they have better interpretability.

2) Model interpretability analysis

The interpretability of A-TL-SSL-TSK is analyzed here to demonstrate the advantage of the proposed methods. A model constructed by A-TL-SSL-TSK using the Dataset D1 is shown in Table IX, which result in five fuzzy rules.

In Table IX, the parameters involved in the fuzzy sets of five fuzzy rules are given. Fig. 3 presents the corresponding membership functions (MFs) of each fuzzy set obtained for all the fuzzy rules, where each MF corresponds to a fuzzy linguistic description, such as "the energy of a band of EEG signal is Low (or *A little low*, *Medium*, *A little high*, *High*). The given linguistic description is only a possible explanation for the IF-Part of fuzzy rule, since different medical experts may use different linguistic descriptions for the same rule.

To provide further explanation, take the rules in the second row of Fig. 3 as an example. According to the antecedent parameters (centers $\mathbf{c}$ and variance $\boldsymbol{\delta}$) of Band 1 in Fig. 3, i.e., (**3.58**, 1.87) for 1st fuzzy rule, (**4.27**, 2.87) for 2nd fuzzy rule, (**5.11**, 2.06) for 3rd fuzzy rule, (**1.66**, 2.33) for 4th fuzzy rule, and (**8.34**, 2.52) for 5th fuzzy rule, five MFs can be generated to represent this feature (Band 1). In addition, these five MFs can be linguistically expressed as "*A little low*", "*Medium*", "*A little high*", "*Low*", and "*High*" in ascending order of the values of the centers. Similarly, the other features can also be divided into these five classes. Finally, with the linguistic expression of the *IF-Part* of the fuzzy rule and the corresponding linear function of the *THEN-Part*, the five fuzzy rules that are generated based on the WPD features can be described linguistically as follows:

The 1st fuzzy rule:

IF the energy of the EEG signal in the frequency band 1 is ***A little low****,*
and the energy of the EEG signal in the frequency band 2 is ***A little high****,*
and the energy of the EEG signal in the frequency band 3 is ***A little high****,*
and the energy of the EEG signal in the frequency band 4 is ***A little high****,*
and the energy of the EEG signal in the frequency band 5 is ***A little low****,*
and the energy of the EEG signal in the frequency band 6 is ***A little low****,*
THEN this rule gives the decision values of the two outputs with the following formula:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

$$f^1(\mathbf{x}) = \begin{bmatrix} 0.2714 + 0.4287x_1 - 0.5325x_2 + 0.1676x_3 - 0.1119x_4 + 0.0872x_5 + 0.0031x_6, \\ -0.2616 - 0.4189x_1 + 0.5427x_2 - 0.1576x_3 + 0.1219x_4 - 0.0772x_5 - 0.0025x_6 \end{bmatrix}.$$

The 2nd fuzzy rule:

*IF the energy of the EEG signal in the frequency band 1 is **Medium**,*
*and the energy of the EEG signal in the frequency band 2 is **Medium**,*
*and the energy of the EEG signal in the frequency band 3 is **Medium**,*
*and the energy of the EEG signal in the frequency band 4 is **A little low**,*
*and the energy of the EEG signal in the frequency band 5 is **Medium**,*
*and the energy of the EEG signal in the frequency band 6 is **A little high**,*
THEN this rule gives the decision values of the two outputs with the following formula:

$$f^2(\mathbf{x}) = \begin{bmatrix} 0.1024 + 0.2909x_1 - 0.2746x_2 - 0.0503x_3 + 0.1071x_4 - 0.0213x_5 + 0.0015x_6, \\ -0.0928 - 0.2813x_1 + 0.2849x_2 + 0.0603x_3 - 0.0971x_4 + 0.0313x_5 - 0.0009x_6 \end{bmatrix}.$$

The 3rd fuzzy rule:

*IF the energy of the EEG signal in the frequency band 1 is **A little high**,*
*and the energy of the EEG signal in the frequency band 2 is **A little low,***
*and the energy of the EEG signal in the frequency band 3 is **Low**,*
*and the energy of the EEG signal in the frequency band 4 is **Low**,*
*and the energy of the EEG signal in the frequency band 5 is **High**,*
*and the energy of the EEG signal in the frequency band 6 is **High**,*
THEN this rule gives the decision values of the two outputs with the following formula:

$$f^3(\mathbf{x}) = \begin{bmatrix} 0.0569 - 0.0160x_1 - 0.0416x_2 - 0.0115x_3 - 0.0264x_4 + 0.0326x_5 - 6.04e-5x_6, \\ -0.0499 + 0.0227x_1 + 0.0547x_2 + 0.0215x_3 + 0.0363x_4 - 0.0226x_5 + 0.0006x_6 \end{bmatrix}.$$

The 4th fuzzy rule:

*IF the energy of the EEG signal in the frequency band 1 is **Low**,*
*and the energy of the EEG signal in the frequency band 2 is **High**,*
*and the energy of the EEG signal in the frequency band 3 is **High**,*
*and the energy of the EEG signal in the frequency band 4 is **Medium**,*
*and the energy of the EEG signal in the frequency band 5 is **Low**,*
*and the energy of the EEG signal in the frequency band 6 is **Low**,*
THEN this rule gives the decision values of the two outputs with the following formula:

$$f^4(\mathbf{x}) = \begin{bmatrix} -0.0159 + 0.2942x_1 - 0.2838x_2 - 0.0550x_3 + 0.0895x_4 + 0.0046x_5 + 0.0003x_6, \\ 0.0257 - 0.2844x_1 + 0.2940x_2 + 0.0648x_3 - 0.0793x_4 + 0.0052x_5 + 0.0002x_6 \end{bmatrix}.$$

The 5th fuzzy rule:

*IF the energy of the EEG signal in the frequency band 1 is **High**,*
*and the energy of the EEG signal in the frequency band 2 is **Low**,*
*and the energy of the EEG signal in the frequency band 3 is **A little low,***
*and the energy of the EEG signal in the frequency band 4 is **High**,*
*and the energy of the EEG signal in the frequency band 5 is **A little high,***
*and the energy of the EEG signal in the frequency band 6 is **Medium**,*
THEN this rule gives the decision values of the two outputs with the following formula:

$$f^5(\mathbf{x}) = \begin{bmatrix} -0.0195 - 0.0178x_1 + 0.0161x_2 + 0.0021x_3 - 0.0106x_4 + 0.0097x_5 - 0.0002x_6, \\ 0.0266 + 0.0249x_1 - 0.0032x_2 + 0.0078x_3 + 0.0205x_4 + 0.0003x_5 + 0.0007x_6 \end{bmatrix}.$$

In a similar way, the fuzzy systems that are learned based on the STFT features and the KPCA features can be interpreted accordingly.

An example is given in Fig. 4 to further explain the usage and the importance of the rules generated by the proposed method. In Fig. 4, the features of the original EEG signals of a patient, extracted by WPD, are used for diagnosis based on the trained A-TL-SSL-TSK fuzzy system. A vector is used to encode the output of the system, with [1,0] indicating the control (i.e., healthy people) and [0,1] indicating the epileptic patient. Using Eq.(3.f), the proposed A-TL-SSL-TSK fuzzy system yields Y=[-0.015,1.015]. According to the "winner takes all" criterion, we further obtain the final output Y=[0,1], which indicates an epileptic patient. It can be seen from the figure that the absolute values of the components in Fuzzy rule 1 (f1=[-0.015,1.015]) is much closer to [0,1] than those in other fuzzy rules, which implies that Fuzzy rule 1 in Fig.4 takes a predominant role in the whole identification process and thus the final decision is primarily determined by this rule.

The above analysis illustrates that the A-TL-SSL-TSK fuzzy system model is an interpretative model for identifying epileptic patient using the fuzzy rules generated.

TABLE IX. AN A-TL-SSL-TSK MODEL WITH FIVE RULES TRAINED USING THE DATASET D1 WITH WPD FEATURES

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 $f^k(\mathbf{x}) = p_0^k + p_1^k x_1 + \cdots + p_d^k x_d$.		
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	$\mathbf{c}^1$ = [3.58, 11.96, 15.17, 19.77, 24.16, 25.32], $\boldsymbol{\delta}^1$ = [1.87, 5.04, 1.59, 0.69, 1.33, 2.38]	$\mathbf{p}_1$ = [0.2714, 0.4287, -0.5325, 0.1676, -0.1119, 0.0872, 0.0031; -0.2616, -0.4189, 0.5427, -0.1576, 0.1219, -0.0772, -0.0025]
2	$\mathbf{c}^2$ = [4.27, 8.73, 12.80, 19.46, 26.17, 28.53], $\boldsymbol{\delta}^2$ = [2.87, 6.02, 2.41, 1.74, 2.11, 4.41]	$\mathbf{p}_2$ = [0.1024, 0.2909, -0.2746, -0.0503, 0.1071, -0.0213, 0.0015; -0.0928, -0.2813, 0.2849, 0.0603, -0.0971, 0.0313, -0.0009]
3	$\mathbf{c}^3$ = [5.11, 2.24, 7.35, 19.13, 31.80, 34.35], $\boldsymbol{\delta}^3$ = [2.06, 4.77, 5.31, 5.87, 5.46, 9.19]	$\mathbf{p}_3$ = [0.0569, -0.0160, -0.0416, -0.0115, -0.0264, 0.0326, -6.04e-05; -0.0499, 0.0227, 0.0547, 0.0215, 0.0363, -0.0226, 0.0006]
4	$\mathbf{c}^4$ = [1.66, 16.62, 17.90, 19.66, 21.78, 22.35], $\boldsymbol{\delta}^4$ = [2.33, 9.10, 2.86, 0.51, 2.31, 3.72]	$\mathbf{p}_4$ = [-0.0159, 0.2942, -0.2838, -0.0550, 0.0895, 0.0046, 0.0003; 0.0257, -0.2844, 0.2940, 0.0648, -0.0793, 0.0052, 0.0002]
5	$\mathbf{c}^5$ = [8.34, 2.12, 10.47, 22.95, 27.69, 28.40], $\boldsymbol{\delta}^5$ = [2.52, 4.96, 3.50, 3.26, 2.52, 3.75]	$\mathbf{p}_5$ = [-0.0195, -0.0178, 0.0161, 0.0021, -0.0106, 0.0097, -0.0002; 0.0266, 0.0249, -0.0032, 0.0078, 0.0205, 0.0003, 0.0007]

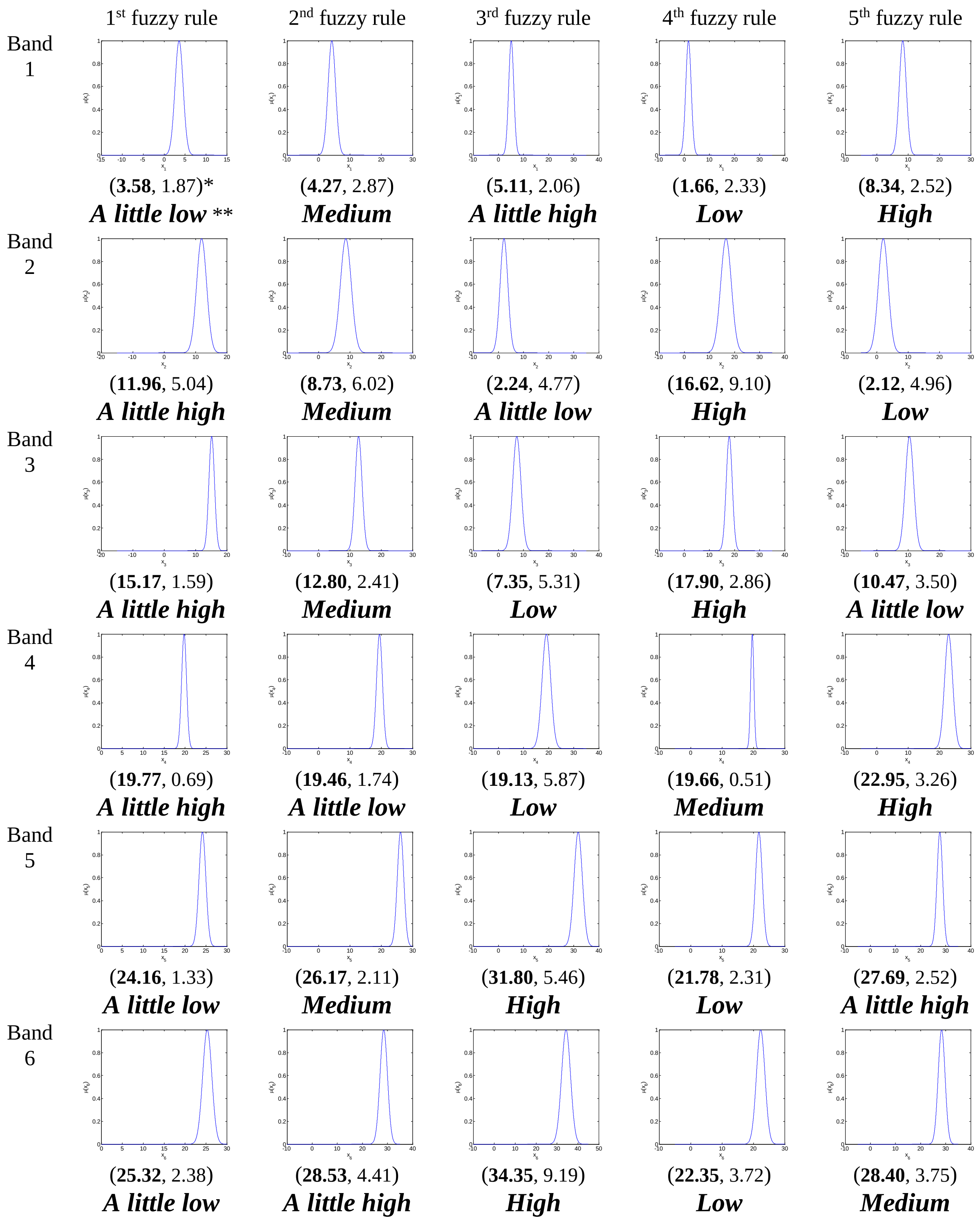

Fig.3 The membership functions and the possible linguistic explanation of each fuzzy subset in the antecedent of the fuzzy rules of the TSK fuzzy system. The system is obtained by the A-TL-SSL-TSK based on the WPD features.

* The antecedent parameter (c_1^1 , δ_1^1) of Band 1 (first dimension of the data) of the first fuzzy rule.

** A possible explanation for the fuzzy set obtained.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

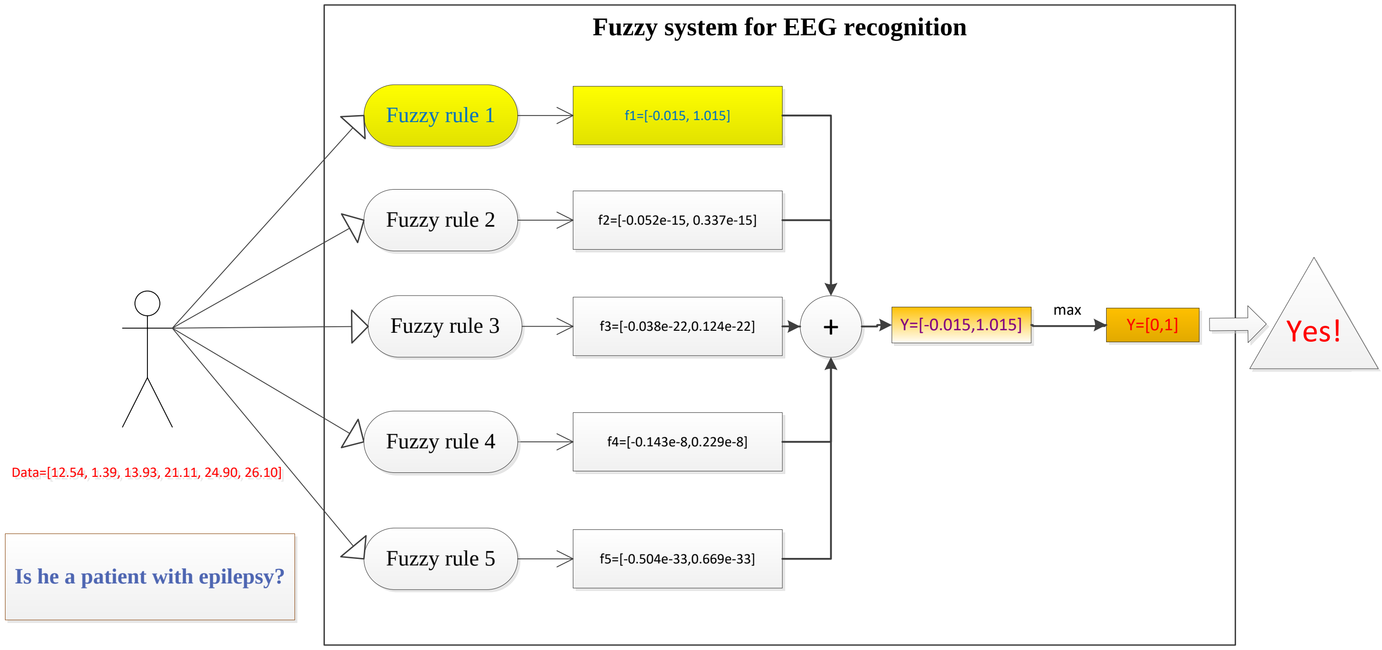

Fig. 4 An example showing the identification of epileptic patient using the fuzzy rules generated and the fuzzy system, where '+' denotes the combination operation and 'max' denotes the operation that sets the maximal element in Y to 1 and the others to 0.

TABLE X. COMPUTATIONAL COMPLEXITY OF THE TL-SSL-TSK MODEL TRAINED BY THE TWO PROPOSED ALGORITHMS ON DATASET D1

Feature extraction Method	Algorithms	Index (Second)		
		Average training time	Average testing time	Real time* (One sample)
WPD	S-TL-SSL-TSK	1.85	1.08e-05	(0.0852+1.56e-07)
	A-TL-SSL-TSK	22.19	1.32e-05	(0.0822+1.69e-07)
STFT	S-TL-SSL-TSK	1.63	1.56e-05	(0.1289+1.56e-07)
	A-TL-SSL-TSK	18.49	1.20e-05	(0.1316+1.44e-07)
KPCA	S-TL-SSL-TSK	2.03	1.09e-05	(0.0057+2.05e-07)
	A-TL-SSL-TSK	21.97	1.02e-05	(0.0053+1.81e-07)

* The first and the second component inside the bracket are the running time of feature extraction and classification respectively.

3) Computational complexity

In this subsection, the computational complexity of the fuzzy systems trained by the proposed S-TL-SSL-TSK and the A-TL-SSL-TSK algorithms were compared. The running time of the two algorithms on dataset D1 is reported in Table X and the following observations are made.

1) The average training time of the proposed A-TL-SSL-TSK was nearly 10 times longer than that of the S-TL-SSL-TSK since the former contained more model parameters than the latter.

2) The average testing time of both algorithms was comparable since they adopted the same the decision function, i.e. Eq. (3.f).

3) The real-time performance can be analyzed by investigating the EEG signal classification process. When raw EEG data were obtained, they were first processed by a feature extraction method (e.g. WPD, STFT or KPCA), and the extracted feature were then processed by a classifier (e.g. S-TL-SSL-TSK or A-TL-SSL-TSK). Thus, the computation time for real-time applications contains two parts, i.e., the running time of the feature extraction method and the running time of the classifier. As shown in Table X, the testing time of A-TL-SSL-TSK and S-TL-SSL-TSK method was comparable, and the difference was due to the variation in running time of different feature extraction methods. The results show that the algorithms (S-TL-SSL-TSK or A-TL-SSL-TSK) developed based on KPCA were much faster than those developed based on WPD or STFT.

It can be concluded from the above analyses that when real-time performance is concerned, if rapid model training is desired, the S-TL-SSL-TSK method is preferable. However, if high recognition accuracy is needed, the A-TL-SSL-TSK is a better choice.

In addition, the running time of S-TL-SSL-TSK and A-TL-SSL-TSK are both acceptable for real-time applications since the main computational cost is feature extraction, as shown in Table X. Furthermore, the KPCA-based S-TL-SSL-TSK or A-TL-SSL-TSK is recommended for applications demanding real-time performance.

VI. CONCLUSIONS

This paper proposes the TL-SSL-TSK model which integrates TL, SSL and TSK fuzzy system model to increase the robustness, accuracy and interpretability of the classifier for EEG signal classification. It also proposes two learning algorithms, S-TL-SSL-TSK and A-TL-SSL-TSK, to train the model. Experimental results show that the proposed approaches

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2017.2748388, IEEE Transactions on Neural Systems and Rehabilitation Engineering

can achieve better performance than many state-of-the-art classification algorithms. In addition, according to our experiments, 50 labeled data for source domain and 50 unlabeled data for target domain are usually adequate for the proposed methods to produce satisfactory results, which is also clinically practical. For example, the average accuracy of the two proposed methods are higher than 95% in most cases. Future research will be conducted to reduce the computational cost of the algorithms, and to extend them to other relevant application domains, including brain-computer interface.

REFERENCES

[1] L. D. Iasemidis, D. S. Shiau, W. Chaovalitwongse, J. C. Sackellares, P. N. Pardalos, J. C. Principe, P. R. Carney, A. Prasad, B. Veeramani, and K. Tsakalis, “Adaptive epileptic seizure prediction system,” *IEEE Trans. Biomed. Eng.*, vol. 50, no. 5, pp. 616–627, 2003.

[2] S. Benbadis and W. Hauser, "An estimate of the prevalence of psychogenic nonepileptic seizures," *Seizure*, vol. 9, pp. 280-281, 2000.

[3] V. Srinivasan, C. Eswaran and N. Sriraam, “Artificial neural network based epileptic detection using time-domain and frequency-domain features,” *J. Med. Syst.*, vol. 29, no. 6, pp. 647–660, 2005.

[4] A. Talevi A, M. S. Cravero, E. A. Castro, L. E. Bruno-Blanch, “Discovery of anticonvulsant activity of abietic acid through application of linear discriminant analysis,” *Bioorganic & Medicinal Chemistry Letters*, vol.17, no.6, pp.684-90, 2007.

[5] A. Dorai, K. Ponnambalam, "Automated epileptic seizure onset detection," *in Proc. the International Conference on Autonomous and Intelligent Systems*, pp. 1-4, 2010.

[6] C. Cortes, V. Vapnik, "Support vector networks," *Machine Learning*, vol.20, no.3, pp.273- 95,1995.

[7] Z. Iscan, Z. Dokur, T. Demiralp, “Classification of electroencephalogram signals with combined time and frequency features,” *Expert Systems with Applications*, vol.38, no.8, pp.10499-505, 2011.

[8] A. Subasi, M. Ismail Gursoy, “EEG signal classification using PCA, ICA, LDA and support vector machines,” *Expert Systems with Applications*, vol.3, no.12, pp.8659-66, 2010.

[9] A. T. Tzallas, M. G. Tsipouras, D. I. Fotiadis, "Epileptic seizure detection in eegs using time-frequency analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol.13, no.5, pp.703-10, 2009.

[10] P. Peng, B. L. Lu, "Immune clonal algorithm based feature selection for epileptic EEG signal classification," *in Proc. the 11th International Conference on Information Science, Signal Processing and their Applications*, pp. 848-53, 2012.

[11] I. Güler, and EDÜ Beyli, "Multiclass support vector machines for EEG-signals classification," *IEEE Trans. Information Technology in Biomedicine*, vol.11, no.2, pp.117-126, 2007.

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

[13] S.J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowledge Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[14] D. Wu, B. Lance, and T.D. Parsons, "Collaborative Filtering for Brain-Computer Interaction Using Transfer Learning and Active Class Selection," *PLoS ONE*, 2013.

[15] S.J. Pan, I.W. Tsang, J.T. Kwok, and Q. Yang, “Domain adaptation via transfer component analysis,” *IEEE Trans. Neural Networks*, vol. 22, no.2, pp.199-210, 2011.

[16] Z.H. Deng, Y.Z. Jiang, F.L. Chung, H. Ishibuchi, S.T. Wang, "Knowledge-Leverage Based Fuzzy System and Its Modeling." *IEEE Trans. Fuzzy Systems*, vol. 21, no. 4, pp. 597-609, 2013.

[17] Z.H. Deng, Y.Z. Jiang, K.S. Choi, F.L. Chung, S.T. Wang, "Knowledge-Leverage-Based TSK Fuzzy System Modeling," *IEEE Trans. Neural Networks and Learning Systems*, vol. 24, no. 8, pp. 1200-1212, 2013.

[18] B. Quanz, J. Huan, “Large margin transductive transfer learning,” *in Proc. the 18th ACM conference on Information and knowledge management*, pp. 1327-1336, 2009.

[19] M.F. Azeem, M. Hanmandlu, and N. Ahmad, "Generalization of adaptive neural-fuzzy inference systems," *IEEE Transactions on Neural Networks*, vol.11, no.6, pp.1332-1346, 2000.

[20] E.H. Mamdani, Application of fuzzy logic to approximate reasoning using linguistic synthesis, *IEEE Transactions on Computer*, vol.26, no.12, pp.1182-1191, 1977.

[21] T. Takagi and M. Sugeno, Fuzzy identification of systems and its application to modeling and control, *IEEE Transactions on Systems Man and Cybernetics*, vol.15, no.1, pp.116-132, 1985.

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

[23] B. Litt, R. Esteller, J. Echauz, M. D'Alessandro, R. Shor, T. Henry, et al, "Epileptic seizures may begin hours in advance of clinical onset: A Report of Five Patients," *Neuron*, vol.30, no.1, pp.51-64, 2001.

[24] E. A. Vivaldi, A. Bassi A, "Frequency domain analysis of sleep eeg for visualization and automated state detection," *in Proc. the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3740-43, 2006.

[25] G. C. Y. Fong, P. U. Shah, M. N. Gee, J. M. Serratosa, I. P. Castroviejo, S. Khan, et al, " Childhood absence epilepsy with tonic-clonic seizures and electroencephalogram 3–4-Hz spike and multispike–slow wave complexes: Linkage to chromosome 8q24," *The American Journal of Human Genetics*, vol.63, no.4, pp.1117-29, 1998.

[26] S. Blanco, S. Kochen, O. A. Rosso, P. Salgado, "Applying time-frequency analysis to seizure EEG activity," *IEEE Engineering in Medicine and Biology Magazine*, vol.16, no.1, pp.64-71, 1997.

[27] Z. Zhang, H. Kawabata, and Z. Q. Liu, "EEG analysis using fast wavelet transform," *in Proc. the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2959-64, 2000.

[28] J. Meng, L. Yao, X. Sheng, D. Zhang and X. Zhu, "Simultaneously optimizing spatial spectral features based on mutual information for EEG classification," *IEEE Trans. Biomedical Engineering*, vol.62, no.1, pp.227-240, 2015.

[29] J. J. Daly and J. E. Huggins, "Brain-Computer Interface: Current and Emerging Rehabilitation Applications," Archives of physical medicine and rehabilitation, vol.96, no.3, pp.1-7, 2015.

[30] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York, 1981.

[31] K. Ito, R. Nakano, “Optimizing support vector regression hyperparameters based on cross-validation,” *Proceedings of the International Joint Conference on Neural Networks*, pp. 2077-2082, 2003.

[32] Y Jiang, F L Chung, H Ishibuchi,et al, "Multitask TSK Fuzzy System Modeling by Mining Intertask Common Hidden Structure", *IEEE Transactions on Cybernetics*, vol.45, no.3,pp.548 - 561, 2015.

[33] O. Chapelle, V. Sindhwani, and S. Keerthi, “Optimization techniques for semi-supervised support vector machines,” *J. Mach. Learn. Res.* ,vol.9, pp. 203–233, Jun. 2008.

[34] Y.-F. Li and Z.-H. Zhou, “Towards making unlabeled data never hurt,” *in Proc. 28th Int. Conf. Mach. Learn.*, pp. 1081–1088, 2011.

[35] M. Long, J. Wang, G. Ding, D. Shen and Q. Yang "Transfer learning with graph co-regularization", *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp.1805 -1818, 2014.

[36] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol.7, pp. 1–30, 2006.

[37] S. García, F. Herrera, "An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons," *Journal of Machine Learning Research*, vol.9, pp.2677–2694, 2008.

[38] Y. Jiang, Z. Deng, F.-L. Chung, S. Wang, "Multi-task TSK fuzzy system modeling using inter-task correlation information," *Information Sciences*, vol.298, pp.512-533, 2015.

[39] Y. Jiang, F.-L. Chung, S. Wang, Z. Deng, J. Wang, P. Qian, "Collaborative Fuzzy Clustering From Multiple Weighted Views", *IEEE Transactions on Cybernetics*, vol.45, no.4, pp.688-701, 2015.

[40] D. Wu, B. Lance, V. Lawhern, "Transfer learning and active transfer learning for reducing calibration data in single-trial classification of visually-evoked potentials." *In Proceedings of 2014 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2801-2807, 2014.

[41] R. G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David , C. E. Elger, "Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, 64:, 061907, 2011

[42] N. Dhulekar, S. Nambirajan, B. Oztan B, et al. "Seizure Prediction by Graph Mining, Transfer Learning, and Transformation Learning," *International Conference on Machine Learning and Data Mining*. pp.32-52,2015.

[43] Y. Zhang, G. Zhou, J. Jin, M. Wang, X. Wang, A. Cichocki,. L1-regularized multiway canonical correlation analysis for SSVEP-based BCI, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no.6, pp. 887-896, 2013.

Yizhang Jiang (M'2012) received the Ph.D. degree in School of Digital Media from Jiangnan University, Wuxi, China, in 2016. He was also a research assistant in the computing department of the Hong Kong Polytechnic University for almost two years. He has published several papers in international journals including IEEE Trans. Fuzzy Systems, IEEE Trans. Neural Networks and Learning Systems, IEEE Trans. Cybernetics, and ACM Trans. Intelligent Systems and Technology. He is the author or coauthor of more than 20 research papers in international/national journals. He has served as the reviewer of several international conferences and journals, including ICDM, TKDE, TFS, TNNLS, TCYB, TSMCA, TII, and Neurocomputing. His research interests include pattern recognition, intelligent computation and their applications.

Dongrui Wu (S'05–M'09-SM'14) received a B.E. degree in Automatic Control from the University of Science and Technology of China, Hefei, China, in 2003, an M.Eng. degree in Electrical Engineering from the National University of Singapore, Singapore, in 2005, and a Ph.D. degree in Electrical Engineering from the University of Southern California, Los Angeles, in 2009. His research interests include affective computing, brain-computer interface, computational intelligence, intelligent control, machine learning, and optimization. He has more than 80 publications, including a book Perceptual Computing (with J. M. Mendel, New York: Wiley-IEEE, 2010).

Dr. Wu received the 2005 IEEE International Conference on Fuzzy Systems Best Student Paper Award, the 2012 IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award, the 2014 IEEE Transactions on Fuzzy Systems Outstanding Paper Award, the 2014 NAFIPS Early Career Award, and 10 Above and Beyond Awards from GE Global Research. He is currently an Associate Editor of the IEEE Transactions on Fuzzy Systems, the IEEE Transactions on Human-Machine Systems, and the IEEE Computational Intelligence Magazine.

Zhaohong Deng (M'2012-SM'2014) received the B.S. degree in physics from Fuyang Normal College, Fuyang, China, in 2002, and the Ph.D. degree in information technology and engineering from Jiangnan University, Wuxi, China, in 2008. He is currently an associate professor in the School of Digital Media, Jiangnan University. He has visited University of California-Davis and the Hong Kong Polytechnic University for more than two years. His current research interests include neuro-fuzzy systems, pattern recognition, and their applications. He is the author or coauthor of more than 80 research papers in international/national journals. He has served as an associate editor of several international Journals, such as Neurocomputing and PLOS ONE.

Pengjiang Qian (M'2012) received the Ph.D. degree from Jiangnan University, Wuxi, China, in 2011.He is an Associate Professor with the School of Digital Media, Jiangnan University, and Case Western Reserve University, Cleveland, OH, USA, as a Research Scholar in medical image processing. His current research interests include data mining, pattern recognition, bioinformatics and their applications, such as analysis and processing for medical imaging, intelligent traffic dispatching, and advanced business intelligence in logistics. He has authored/co-authored over 30 papers in international/national journals and conferences.

Jun Wang (M'2014) received his Ph.D. degree in pattern recognition and intelligence systems from the School of Computer Science and Technology in Nanjing University of Science and Technology, Nanjing (NUST), China, in 2011. From January to June in 2010, he was a Research Assistant in Department of Computing, Hong Kong Polytechnic University. He is currently an Associate Professor with the School of Digital Media, Jiangnan University, China, and a visiting scholar in the Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, USA. He has published more than 40 articles in international/national journals. His research interests include machine learning, fuzzy clustering and medical image classification.

Guanjin Wang is currently pursuing the Ph.D. degree from the School of Nursing, Hong Kong Polytechnic University, Hong Kong. Her current research interests include transfer learning, multi-task learning and health informatics.

Fu-Lai Chung (M'95) received the B.Sc. degree from the University of Manitoba, Winnipeg, MB, Canada, in 1987, and the M.Phil. and Ph.D. degrees from the Chinese University of Hong Kong, Hong Kong, in 1991 and 1995, respectively. In 1994, he joined the Department of Computing, Hong Kong Polytechnic University, where he is currently an Associate Professor. He has authored or coauthored over 80 journal papers published in the areas of soft computing, data mining, machine intelligence, and multimedia. His current research interests include transfer learning, social network analysis and mining, kernel learning, dimensionality reduction, and big data learning.

Kup-Sze Choi (M’97) received the Ph.D. degree in computer science and engineering from the Chinese University of Hong Kong. He is currently an Associate Professor at the School of Nursing, Hong Kong Polytechnic University, as well as the Director of the Centre for Smart Health and the PolyU-Henry G. Leong Mobile Integrative Health Centre. His research interests include virtual reality and artificial intelligence, and their applications in medicine and healthcare.

Shitong Wang received the M.S. degree in Computer Science from Nanjing University of Aeronautics and Astronautics, China, in 1987. He visited London University and Bristol University in U.K., Hiroshima International University and Osaka Prefecture University in Japan, Hong Kong University of Science and Technology, Hong Kong Polytechnic University, as a Research Scientist, for over six years. Currently, he is a Full Professor in the School of Digital Media, Jiangnan University, China. His research interests include artificial intelligence, neuro-fuzzy systems, pattern recognition, and image processing. He has published about 100 papers in international/national journals and has authored seven books.