The following publication Z. Deng, P. Xu, L. Xie, K. -S. Choi and S. Wang, "Transductive Joint-Knowledge-Transfer TSK FS for Recognition of Epileptic EEG Signals," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 8, pp. 1481-1494, Aug. 2018 is available at https://doi.org/10.1109/TNSRE.2018.2850308.

Transductive Joint-Knowledge-Transfer TSK FS for Recognition of Epileptic EEG Signals

Z. Deng, P. Xu, L. Xie, K.S. Choi, S Wang.

Abstract*—Intelligent recognition of electroencephalogram (EEG) signals is an important means to detect seizure. Traditional methods for recognizing epileptic EEG signals are usually based on two assumptions: 1) adequate training examples are available for model training, and 2) the training set and the test set are sampled from datasets with the same distribution. Since seizures occur sporadically, training examples of seizures could be limited. Besides, the training and test sets are usually not sampled from the same distribution for generic non-patient-specific recognition of EEG signals. Hence, the two assumptions in traditional recognition methods could hardly be satisfied in practice, which results in degradation of model performance. Transfer learning is a feasible approach to tackle this issue attributed to its ability to effectively learn the knowledge from the related scenes (source domains) for model training in the current scene (target domain). Among the existing transfer learning methods for epileptic EEG recognition, transductive transfer learning fuzzy systems (TTL-FSs) exhibit distinctive advantages – the interpretability that is important for medical diagnosis and the transfer learning ability that is absent from traditional fuzzy systems. Nevertheless, the transfer learning ability of TTL-FSs is restricted to a certain extent since the discrepancy in marginal distribution between the training data and test data is only considered. In this study, the enhanced transductive transfer learning Takagi-Sugeno-Kang fuzzy system construction method (ETTL-TSK-FS) is proposed to overcome the challenge by introducing two novel transfer learning mechanisms: (1) joint-knowledge is adopted to reduce the discrepancy between the two domains, and (2) an iterative transfer learning procedure is introduced to enhance transfer learning ability. Extensive experiments have been carried out to evaluate the effectiveness of the proposed method in recognizing epileptic EEG signals in the Bonn and CHB-MIT EEG datasets. The results show that the method is superior to or at least competitive with existing state-of-art methods under the scenario of transfer learning.

***Index Terms*—Joint-Knowledge Transfer, EEG, epilepsy detection, feature extraction, TSK fuzzy logic system (FLS), transfer learning.**

I. Introduction

Epilepsy is a cerebral functional disease caused by sudden abnormal brain neurons discharge [1]. It is one of the most common brain diseases. Intelligent recognition of electroencephalogram (EEG) signals is an important means of seizure detection since the signals contain a large amount of physiological and pathological information of the brain. However, this is not trivial due to the stochastic and non-stationary nature of EEG signals among others. In recent years, data-driven modeling techniques have attracted considerable attention for epilepsy detection and recognition because of their strong learning abilities [1].

In fact, many intelligent recognition methods have been adopted for the recognition of epileptic EEG signals, such as Support Vector Machine (SVM) [2], Linear Discriminant Analysis (LDA) [3], Neural Networks [4], Fuzzy System (FS) [5, 6], and so on. The purpose of these intelligent methods is to learn a prediction model using the existing data and then predict or classify the new data based on the trained model. However, when applied to epileptic EEG signals recognition, these methods often encounter the following challenges. First, since seizures occur sporadically and data labeling of the EEG signals is very time and cost consuming [7], adequate labeled data of EEG signals is not always available for clinical applications. Second, the majority of the traditional machine learning techniques used for training intelligent models are based on the assumption that the training and test data have the same distribution, which cannot be satisfied in generic non-patient-specific applications of EEG signal recognition. Although transfer learning techniques have been used to tackle these challenges, the transfer learning ability of the existing models is still insufficient. The issues will be addressed in this paper.

Transfer learning [8-11] is a feasible solution to the challenges discussed above for its ability to leverage the data or knowledge gained from the related scenes (commonly called the source domains in transfer learning) for the learning of the model in the current scene (commonly called the target domain in transfer learning) [11]. Transfer learning can be divided into two main categories according to the characteristics of the target domain data, namely, inductive transfer learning and transductive transfer learning [12]. Inductive transfer learning assumes that the datasets sampled from the source domain and the target domain have different distributions, and that many labeled data exist in the source domain while a small number of labeled data are available from the target domain [12]. Transductive transfer learning further assumes that labeled data are abundantly available from the source domain but no labeled

This work was supported in part by the Jiangsu Province Outstanding Youth Fund under Grant No. BK20140001, the National Key Research Program of China under Grant No. 2016YFB0800803, the NSFC under Grant No. 61772239 and the Hong Kong Research Grants Council under Grant No. PolyU 152040/16E (Corresponding author: Zhaohong Deng.)

Z.H. Deng, L.X. Xie, P. Xu and S.T. Wang are with the School of Digital Media, Jiangnan University, Wuxi 214122, China (e-mail: dengzhaohong@jiangnan.edu.cn; lixiaoxie@jiangnan.edu.cn; pengxu@jiangnan.edu.cn; wxwangst@aliyun.com)

K.S. Choi is with the Centre for Smart Health, School of Nursing, Hong Kong Polytechnic University, Hong Kong (e-mail: kschoi@ieee.org)

data from the target domain. With transfer learning, the performance of the model in the target domain can usually be improved by leveraging the knowledge from the source domain [8]. Compared with inductive transfer learning, the requirements of transductive transfer learning on the target domain data are less constrained, which is very suitable for epileptic EEG signals recognition when the labeled data of seizures in the target domain is scarce. In fact, some transductive transfer learning methods have been developed. For example, the large-margin-projected transductive SVM (LMPROJ), which is a classical transductive transfer learning method, has been developed and demonstrated effective recognition of epileptic EEG signals [13].

Among the existing intelligent models for epileptic EEG recognition, fuzzy systems exhibit the distinctive advantages of better interpretability and uncertainty modeling ability. In many applications, such as medical diagnosis, interpretability is crucial. At present, there have been some effective methods developed for epileptic EEG recognition based on fuzzy system and transfer learning techniques. In [14], two transductive transfer learning Takagi-Sugeno-Kang (TSK) FS modeling methods were proposed for the recognition of epileptic EEG signals. The transfer learning FS construction methods have shown strong adaptability to the discrepancy in distributions between the training dataset and the test dataset. However, the performance of these methods are still limited by the fact that they only consider the discrepancy of the marginal distributions between the source domain and the target domain in model training [14], so that a lot of useful knowledge, such as the labeling information in source domain, is not fully exploited for knowledge transfer. Thus, the capacity of knowledge transfer may be fully unleashed for epilepsy EEG recognition with the existing transfer fuzzy system modeling methods. To meet this challenge, an enhanced transductive transfer learning TSK fuzzy system (ETTL-TSK-FS) is investigated in this study. The proposed method improves the transfer learning abilities of TSK FS for epileptic EEG recognition from two aspects. First, a more comprehensive knowledge transfer mechanism called joint-knowledge transfer is introduced. Second, an iterative transfer learning mechanism is adopted to enhance the knowledge transfer ability by improving the labeling accuracy of the preliminarily labeled samples in the target domain.

The main contributions of this paper are as follows: (1) the novel TSK FS construction method ETTL-TSK-FS is proposed to handle the drifting in distributions between the training data and the test data, which can hinder epileptic EEG recognition; (2) two transfer learning strategies, i.e., joint-knowledge transfer and iterative knowledge transfer, are introduced to realize the proposed transfer learning method; and (3) a large amount of experiments are conducted to confirm the effectiveness of the proposed method.

The rest of the paper is organized as follows. Section II describes the traditional TSK FS and the maximum mean discrepancy, and other related work. Section III presents the enhanced transductive transfer learning TSK FS construction method. The experiments conducted to evaluate the performance of the proposed method and the results are discussed in section IV. Finally, conclusions and the potential future work are given in section V.

II. Related Work

The related works are reviewed in this section. Background of seizure detection and recognition is first introduced, followed by a review of some classical feature extraction methods of EEG signals. The fundamentals of the TSK fuzzy system and transductive transfer learning technique that the method proposed in the paper are based are also described briefly.

A. *Seizure detection*

Seizure detection generally refers to the use of intelligent models to recognize that a seizure is occurring through EEG signal analysis [15]. Here, intelligent models are mainly the classifiers that are used to identify the signals. In clinical applications, online analysis of EEG signals and real-time seizure detection is needed. The systems that realize real-time seizure detection without human intervention are often called automatic seizure detection and the corresponding models used in the systems are automated computational models [16, 17].

Seizure detection can be classified as seizure onset detection and seizure event detection [7]. The purpose of seizure onset detection is to recognize that a seizure has started with the shortest possible delay, but not necessarily with the highest possible accuracy [18]. In contrast, the purpose of seizure event detection is to recognize seizures with the greatest possible accuracy [19], but not necessarily with the shortest delay. In clinical applications, seizure event detection and seizure onset detection are both important for the requirements in different scenarios. The former is needed for applications requiring rapid response to a seizure, while the latter is necessary for applications requiring accurate identification of seizure activity over a period of time [2].

Seizure prediction is another important clinical application in epilepsy management [20]. Seizure prediction algorithm is expected to forecast impending seizures before they start, we while seizure detection algorithm is expected to identify the presence of a seizure after it has begun. Although prediction algorithms are useful for clinical management of adult epilepsy, many intractable problems exists in the theory of seizure prediction [21]. Many methods adopted for seizure prediction are indeed similar with those used for seizure detection, where the prediction algorithms are based on detecting the patterns of EEG signals. Hence, seizure recognition is of higher importance for epilepsy diagnosis and treatment. It is also the basis of seizure detection. The recognition of epileptic EEG signals is reviewed in the following section.

B. *Recognition of Epileptic EEG Signals*

The basis of seizure detection is the recognition of epileptic EEG signals, or in other words, the classification of EEG signals. The focus of EEG signal recognition is to deal with the segments of epileptic EEG signals rather than the seizures that are concerned in seizure detection tasks. The recognition of epileptic EEG signals are studied from two aspects: feature extraction and classification. The common classification methods used for epileptic EEG recognition have already been reviewed in the second paragraph of Section I. The classical feature extraction methods for EEG signals are reviewed as follows, mainly on methods dealing with the segments of single-channel signals.

The major feature extraction methods can be divided into three categories, i.e. time domain analysis, frequency domain analysis and time-frequency analysis. In time domain analysis, features are extracted by analyzing the characteristics of the EEG waveforms. In clinic applications, mean, variance, amplitude and kurtosis and the other statistics of the waveforms can all be used [3, 22]. Owing to the non-stationary nature of EEG signals, entropy-based approaches are proposed to quantify the amount of regularity and unpredictability of fluctuations [4, 23]. In frequency-domain analysis, features in EEG signals are analyzed in the frequency domain. Power spectrum analysis of EEG signals is performed to transform changes in EEG signal amplitude into changes in EEG signal power to directly observe the variations in brain waves at different frequencies [24]. Fast Fourier Transforms (FFT) is a common method of this kind [25]. In time-frequency analysis, features are extracted from the time domain and frequency domain simultaneously to identify more effectively the non-stationary EEG signals. A representative method is wavelet transform which is widely used and shown to be advantageous [26, 27]. Wavelet transform is available in two forms: continuous wavelet transform (CWT) and discrete wavelet transform (DWT) forms. The wavelet packet decomposition (WPD) is a generalization of the DWT, where the time-frequency resolution can be adjusted as desired. In DWT, the calculation at each level is performed by passing in only the wavelet approximation coefficients from the previous level to give the approximate and detailed information at a higher level. However, in WPD, both the detailed and approximation coefficients are decomposed to create the full binary tree [2].

C. TSK FS

FSs are a type of important intelligent models with good interpretability and uncertainty modeling ability. TSK FS is one of the most widely used FS models [28] attributed to its strong learning abilities and flexibility. In theory, TSK FS can be regarded as a universal approximator [29], which can be expressed with fuzzy logic rules in different ways. The most popular one is as follows.

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

In (1), A_i^k is a fuzzy set; p_i^k is the real-valued parameters in the consequents and $f^k(x)(k=1,2,...,K)$ represents the output of the kth rule; K is the number of rules in the rule base. Given an input vector $\mathbf{x}=(x_1,\cdots,x_d)^{\mathbf{T}}$, when multiplication is employed as the conjunction and implication operator, addition as the combination operator, and the center of gravity as the defuzzification operator, the output $f(x)$ of the TSK FS can be expressed as the weighted average of $f^k(\mathbf{x})$, i.e.,

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

In (2), $\mu^k(\mathbf{x})$ is the fuzzy membership for the kth rule and $\tilde{\mu}^k(\mathbf{x})$ is the normalized form.

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

In (3), $\mu_{A_i^k}(x_i)$ is the membership degree of the *i*th dimension of the input vector $\mathbf{x}$ to the fuzzy subset A_i^k [30]. The Gaussian membership function in (4) is commonly used as the antecedent fuzzy membership function, and also implies (???) the fuzzy subset A_i^k indeed. The membership degree can be calculated by the membership function.

$$\mu_{A_i^k}(x_i) = \exp\left(\frac{-(x_i - c_i^k)^2}{2\delta_i^k}\right) \tag{4}$$

where δ_i^k and c_i^k are the kernel width and the center respectively. The parameters δ_i^k and c_i^k in the membership function are called the antecedent parameters which define the concrete forms of the fuzzy subsets A_i^k. There are several ways to generate the antecedent parameters for building FS. One effective and efficient method is to divide the input space by clustering techniques and then estimate the antecedent parameters according to the clustering result. For example, fuzzy-c-means (FCM) can be used to estimate the antecedent parameters. More details about the computation process of FCM can be found in [30].

Once the antecedent parameters are determined, the output of the TSK FS in (2) can be viewed as a linear regression model in a hidden-mapping feature space, which is expressed by

$$y = f(\mathbf{x}) = \mathbf{p}_g^{\mathrm{T}} \mathbf{x}_g, \tag{5}$$

where the corresponding data and parameters are obtained as follows,

$$\mathbf{x_g} = [(\tilde{\mathbf{x}}^1)^{\mathbf{T}}, (\tilde{\mathbf{x}}^2)^{\mathbf{T}}, \cdots, (\tilde{\mathbf{x}}^K)^{\mathbf{T}}]^{\mathbf{T}} \in R^{K(d+1)}, \tag{6a}$$

$$\tilde{\mathbf{x}}^K = \tilde{\mu}^k(\mathbf{x})\mathbf{x}_e, \tag{6b}$$

$$\mathbf{x}_e = (1, \mathbf{x}^{\mathbf{T}})^{\mathbf{T}}, \tag{6c}$$

$$\mathbf{p_g} = [(\mathbf{p}^1)^{\mathrm{T}}, (\mathbf{p}^2)^{\mathrm{T}}, \cdots, (\mathbf{p}^K)^{\mathrm{T}}]^{\mathbf{T}}, \tag{6d}$$

$$\mathbf{p}^k = (p_0^k, p_1^k, \cdots, p_d^k)^{\mathbf{T}}. \tag{6e}$$

In (6a), $\mathbf{x_g}$ represents the vector in the feature space mapped from the input data $\mathbf{x}$ in the original space according to the fuzzy rules; $\mathbf{p_g}$ is the combined vector of the consequent parameters of the fuzzy rules in the rule base of the TSK FS. The consequent parameters of the TSK FS, i.e., $\mathbf{p_g}$ in (5), can be obtained directly by optimization using the effective learning techniques developed for linear models.

D. Maximum Mean Discrepancy and Transductive Transfer Learning FS

The techniques of Maximum Mean Discrepancy (MMD) and Transductive Transfer Learning FS that are related to the work in this study are briefly described as follows.

1) MMD and projected MMD

The MMD and the projected MMD (PMMD) have shown to be effective techniques for transductive transfer learning [30, 31]. For the dataset $D_s = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$ and $D_t = \{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_M\}$ that are collected from the source and target domains with distributions *Ps* and *Pt* respectively, the discrepancy between these two distributions can be measured in terms of MMD by

$$\mathbf{d}(P_s, P_t)^2 = \mathbf{MMD}^2 = \left\| \frac{1}{N}\sum_{j=1}^{N}\phi(\mathbf{x}_j) - \frac{1}{M}\sum_{j=1}^{M}\phi(\mathbf{z}_j) \right\|^2 \quad , \qquad (7)$$

where $\phi(\mathbf{x}_j)$ is a mapping function, N and M represent the numbers of examples in the source and the target domains respectively.

When drifting in distributions between the two domains exists, the PMMD with the projective vector $\mathbf{p}$ is often adopted for model training to improve the effectiveness of the prediction model. Given the same datasets used in MMD, the PMMD of the two distributions is

$$\mathbf{d}(P_s, P_t, \mathbf{p})^2 = \mathbf{PMMD}^2 = \left\| \frac{1}{N}\sum_{i=1}^{N}\mathbf{p}^T\phi(\mathbf{x}_i) - \frac{1}{M}\sum_{i=1}^{M}\mathbf{p}^T\phi(\mathbf{z}_i) \right\|^2 . \qquad (8)$$

This measure has been used to estimate the difference between the distributions of the source and the target domains in transfer learning. For example, it was adopted to realize transfer learning for SVM and yield the LMPROJ [13, 31].

2) Transductive transfer learning TSK FS

The transductive transfer learning TSK FS is an integration of the transductive transfer learning technique and the TSK FS. It is proposed to develop transfer learning based intelligent models with good interpretability and uncertainty modeling ability. In transductive transfer learning scene, the training set D_s is usually taken as the source domain which contains labeled data pairs, i.e., $D_s = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$. Meanwhile, the test set D_t is taken as the target domain and only contains unlabeled data, i.e., $D_t = \{\mathbf{z}_1, \ldots, \mathbf{z}_M\}$. In TSK FS, with reference to (6a)-(6c), we can get the following mapping functions,

$$\mathbf{x}_i \to \phi(\mathbf{x}_i) = \mathbf{x}_{gi} \text{ and} \qquad (9)$$

$$\mathbf{z}_i \to \phi(\mathbf{z}_i) = \mathbf{z}_{gi} , \qquad (10)$$

where $\mathbf{x}_{gi}$, $\mathbf{z}_{gi}$ are the data in the new feature space, mapped from the data in the original feature space. The new feature space is constructed based on the fuzzy rules in the TSK FS rule base. Based on the mapping, the corresponding MMD can be estimated to measure of the distribution distance between the source domain and the target domain.

Furthermore, the PMMD can be estimated to build TSK FS with transfer learning abilities. In [14], a transfer learning TSK FS has been proposed with the PMMD defined as follows,

$$\mathbf{d}(P_s, P_t, \mathbf{p}_g)^2 = \mathbf{PMMD}^2 = \left\| \frac{1}{N}\sum_{i=1}^{N}\mathbf{p}_g^T\mathbf{x}_{gi} - \frac{1}{M}\sum_{i=1}^{M}\mathbf{p}_g^T\mathbf{z}_{gi} \right\|^2 \qquad (11)$$

where $\mathbf{p}_g$ is the projective vector, and $\mathbf{x}_{gi}$, $\mathbf{z}_{gi}$ are the new data in the new mapping space, which are defined by (6a)-(6c), (9) and (10).

By introducing the PMMD, two transductive transfer learning TSK FS training methods are proposed respectively for classification and regression [14]. The objective function for regression is as follows:

$$\begin{aligned} &\min_{\mathbf{p}_g} : \frac{1}{2}\mathbf{p}_g^T\mathbf{p}_g + C\sum_{i=1}^{N}[(\xi_i^+)^2 + (\xi_i^-)^2] + \frac{2}{\tau}\varepsilon + \lambda \cdot \mathbf{d}(P_s, P_t, \mathbf{p}_g)^2 \\ &s.t. \quad \begin{cases} y_i - \mathbf{p}_g^T\mathbf{x}_{gi} < \varepsilon + \xi_i^+ \\ \mathbf{p}_g^T\mathbf{x}_{gi} - y_i < \varepsilon + \xi_i^- \end{cases} \end{aligned} \qquad (12)$$

The TSK FS transfer learning methods developed based on binary classification and regression have been used to implement multi-class classification with different strategies for epileptic EEG signals recognition [14].

III. Enhanced Transductive transfer learning TSK FS

From probability theory, a joint distribution J can be factorized into the product of conditional distribution Q and marginal distribution $P : J = Q * P$, which means that the joint distribution J contains more information of the distribution than the conditional distribution Q or the marginal distribution P [32]. According to the discussion in section II-B, we know that the existing transductive learning TSK FS construction methods have been focusing on minimizing the discrepancy between the marginal distribution of the source domain P_s and that of the target domain P_t in the input space. However, the discrepancy between the conditional distribution of the source domain Q_s and that of the target domain Q_t is not considered for transfer learning in the TSK FS construction procedure. Since both the marginal and the conditional distributions contain useful information for transfer learning, it is anticipated that transductive transfer learning TSK FS construction method can be enhanced by the joint-knowledge. Hence, the use of joint-knowledge to enhance the performance of transfer learning is investigated in this study and the details are presented below.

A. Domain Adaptation with Joint-Knowledge-Transfer

Given the labeled data $D_s = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)\}$ in the source domain and the unlabeled data $D_t = \{\mathbf{x}_{N+1}, \mathbf{x}_{N+2}, \cdots, \mathbf{x}_{N+M}\}$ in the target domain of a transfer learning scene, the marginal distributions of the two domains $P_s(\mathbf{x})$ and $P_t(\mathbf{x})$ are different, i.e., $P_s(\mathbf{x}) \neq P_t(\mathbf{x})$, and so are the conditional distributions, i.e., $Q_s(y \mid \mathbf{x}) \neq Q_t(y \mid \mathbf{x})$. Hence, the joint distributions of the source and the target domain $Q_s(\mathbf{x}, y) = P_s(\mathbf{x})Q_s(y \mid \mathbf{x}) = P_s(y)Q_s(\mathbf{x} \mid y)$ and $Q_t(\mathbf{x}, y) = P_t(\mathbf{x})Q_t(y \mid \mathbf{x}) = P_t(y)Q_t(\mathbf{x} \mid y)$ are usually different. In order to train a classifier that can leverage the information in the source domain more sufficiently for the learning in the target domain, the knowledge of joint distribution should be considered [8, 9, 33].

The above analyses indicate that, for better adaptation between the two domains, it is necessary to minimize simultaneously the discrepancy in conditional distributions and marginal distributions between the two domains. However, it is difficult to obtain the conditional distributions $Q_s(y \mid \mathbf{x})$ and $Q_s(y \mid \mathbf{x})$ directly. A commonly used strategy is to replace $Q_s(y \mid \mathbf{x})$ and $Q_s(y \mid \mathbf{x})$ by $Q_s(\mathbf{x} \mid y)$ and $Q_t(\mathbf{x} \mid y)$ respectively [34, 35], so that the joint-knowledge, i.e., the margin and conditional distributions $P_s(\mathbf{x})$, $P_t(\mathbf{x})$, $Q_s(\mathbf{x} \mid y)$ and $Q_t(\mathbf{x} \mid y)$, can be made available for transfer learning.

While $Q_s(\mathbf{x} \mid y)$ can be estimated easily based on the labeled source data, it is necessary to preliminarily label the unlabeled data in the target domain. The $Q_t(\mathbf{x} \mid y)$ can be estimated by labeling the data in the target domain preliminarily with an existing classifier, e.g. a classical TSK FS or a transfer learning TSK FS, that has been trained beforehand [14].

Given a labeled dataset in the source domain and a prelabeled dataset in the target domain (i.e. the data are preliminarily labeled using an existing classifier), the PMMD of the conditional distributions $Q_t(\mathbf{x}|y)$ and $Q_s(\mathbf{x}|y)$ between the two domains is given by

$$D^{(c)}(Q_s,Q_t,\mathbf{p}_g)=\mathbf{PMMD}_{\mathrm{H}}^2(\mathbf{p}_g,D_s^{(c)},D_t^{(c)})$$
$$=\left\|\frac{1}{N^c}\sum_{x_i\in D_s^{(c)}}\mathbf{p}_g^T\mathbf{x}_{gi}-\frac{1}{M^c}\sum_{z_j\in D_t^{(c)}}\mathbf{p}_g^T\mathbf{z}_{gj}\right\|^2 \tag{13}$$

Here, $D_s^{(c)}=\{\mathbf{x}_i:\mathbf{x}_i\in D_s\wedge y(\mathbf{x}_i)=c\}$ represents a subset of examples belonging to the cth class in the source domain, $y(\mathbf{x}_i)$ is the true label of $\mathbf{x}_i$ and $N_s^{(c)}$ is the number of examples belonging to the cth class in the source domain. Similarly, $D_t^{(c)}=\{\mathbf{x}_j:\mathbf{x}_j\in D_t\wedge\hat{y}(\mathbf{x}_j)=c\}$ represents a subset of examples whose preliminary labels belong to the cth class in the target domain, $\hat{y}(\mathbf{x}_j)$ is the preliminary label of $\mathbf{x}_j$ and $N_t^{(c)}$ is the number of examples that have been classified as the cth class in the target domain by an existing classifier.

Based on (11) and (13), the following joint-knowledge term D_{JK} that integrates the marginal distribution and the conditional distribution is proposed for transfer learning in the TSK FS construction,

$$D_{JK}(J_s,J_t,\mathbf{p}_g)=\alpha\cdot D(P_s,P_t,\mathbf{p}_g)+(1-\alpha)\cdot\sum_{c=1}^{C}D^{(c)}(Q_s,Q_t,\mathbf{p}_g) \tag{14}$$

where $\alpha\in[0,1]$ is a parameter to balance the contribution of the two terms. $D_{JK}(J_s,J_t,\mathbf{p}_g)$ degenerates into $D(P_s,P_t,\mathbf{p}_g)$ in (11) when $\alpha=1$. In our experimental studies, $\alpha=0.5$.

B. Objective Function and Optimization

In order to enhance the transfer learning ability for TSK FS, an objective function based on ε-insensitive loss and joint-knowledge-transfer is proposed below.

$$\min_{\mathbf{p}_g,\xi_i^+,\xi_i^-,\varepsilon}\ \frac{1}{\tau N}\sum_{i=1}^{N}((\xi_i^+)^2+(\xi_i^-)^2)+\frac{1}{2}\mathbf{p}_g^T\mathbf{p}_g+\frac{2}{\tau}\varepsilon+\lambda D_{JK}(J_s,J_t,\mathbf{p}_g),$$
$$s.t\ \begin{cases}y_i-\mathbf{p}_g^T\mathbf{x}_{\rho i}<\varepsilon+\xi_i^+\\ \mathbf{p}_g^T\mathbf{x}_{\rho i}-y_i<\varepsilon+\xi_i^-\end{cases}\ \forall i \tag{15}$$

The first three terms in (15) are inherited from the ε-insensitive loss based TSK FS construction methods [31, 32], and the fourth term is the joint-knowledge-transfer term in (14). The main difference between (15) in this paper and (19) in [14] is that enhancement is made here by introducing the last term, which is the key innovation of our work.

According to the theory of reproducing kernel Hilbert space [35, 36], the projection vector $\mathbf{p}_g$ in the TSK FS can be expressed as

$$\mathbf{p}_g=\sum_{i=1}^{N+M}\beta_i\Phi(s)_i=\mathbf{\Phi(s)\beta}, \tag{16}$$

where $\mathbf{\Phi(s)}=[\phi(\mathbf{s}_1),\phi(\mathbf{s}_2),...,\phi(\mathbf{s}_{N+M})]=[\mathbf{x}_{g1},...,\mathbf{x}_{gN},\mathbf{z}_{g1},...,\mathbf{z}_{gM}]$. By substituting (16) into (14), $D_{JK}(J_s,J_t,\mathbf{p_g})$ can be expressed as (17). The detailed derivation procedure is similar as that in [14].

$$D_{JK}(J_s,J_t,\mathbf{p}_g)=\alpha\cdot D(P_\mathbf{s},P_\mathbf{t},\mathbf{p}_g)+(1-\alpha)\cdot\sum_{c=1}^{C}D^{(c)}(\mathbf{Q}_s^{(c)},\mathbf{Q}_t^{(c)},\mathbf{p}_g)$$
$$=\alpha\cdot\mathbf{\beta^T\Omega_0\beta}+(1-\alpha)\cdot\mathbf{\beta^T}(\sum_{c=1}^{C}\mathbf{\Omega}_c)\mathbf{\beta}$$
$$=\mathbf{\beta^T\tilde{\Omega}\beta}$$

(17)

In (17), $\tilde{\mathbf{\Omega}}=\alpha\cdot\mathbf{\Omega}_0+(1-\alpha)\cdot\sum_{c=1}^{C}\mathbf{\Omega}_c$, where $\mathbf{\Omega}_0$ and $\mathbf{\Omega}_c$ are matrices of size $(N+M)\times(N+M)$ and can be calculated respectively as follows.

$$\mathbf{\Omega}_0=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{\Phi(s)^T x}_{gi}\mathbf{x}_{gj}^T\mathbf{\Phi(s)}+\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbf{\Phi(s)^T z}_{gi}\mathbf{z}_{gj}^T\mathbf{\Phi(s)}-\frac{2}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}\mathbf{\Phi(s)^T x}_{gi}\mathbf{z}_{gj}^T\mathbf{\Phi(s)} \tag{18a}$$

$$\mathbf{\Omega}_c=\frac{1}{(N^c)^2}\sum_{\mathbf{x}_i\in D_s^{(c)}}\sum_{\mathbf{x}_j\in D_t^{(c)}}\mathbf{\Phi(s)^T x}_{gi}\mathbf{x}_{gj}^T\mathbf{\Phi(s)}+\frac{1}{(M^c)^2}\sum_{\mathbf{z}_i\in D_t^{(c)}}\sum_{\mathbf{z}_j\in D_t^{(c)}}\mathbf{\Phi(s)^T z}_{gi}\mathbf{z}_{gj}^T\mathbf{\Phi(s)}-\frac{2}{N^cM^c}\sum_{\mathbf{x}_i\in D_s^{(c)}}\sum_{\mathbf{z}_j\in D_t^{(c)}}\mathbf{\Phi(s)^T x}_{gi}\mathbf{z}_{gj}^T\mathbf{\Phi(s)} \tag{18b}$$

Specifically, when $\mathbf{\Omega}=(\tilde{\mathbf{\Omega}}+\tilde{\mathbf{\Omega}}^T)/2$, we have $\mathbf{\beta^T\tilde{\Omega}\beta}=\mathbf{\beta^T\Omega\beta}$ for (17), where $\mathbf{\Omega}$ is a symmetric matrix. Finally, based on (16)-(18b), the objective function in (15) can be expressed as:

$$\min_{\mathbf{\beta},\xi_i^+,\xi_i^-,\varepsilon}\ \frac{1}{\tau N}\sum_{i=1}^{N}((\xi_i^+)^2+(\xi_i^-)^2)+\frac{1}{2}\mathbf{\beta}^T\mathbf{\Phi(s)^T\Phi(s)\beta}+\frac{2}{\tau}\varepsilon+\lambda\mathbf{\beta}^T\mathbf{\Omega\beta}$$
$$\mathbf{s.t}\ \begin{cases}y_i-\mathbf{\beta}^T\mathbf{\Phi(s)^T}\boldsymbol{x}_{\rho i}<\varepsilon+\xi_i^+\\ \mathbf{\beta}^T\mathbf{\Phi(s)^T}\boldsymbol{x}_{\rho i}-y_i<\varepsilon+\xi_i^-\end{cases}\ \forall i \tag{19}$$

Based on optimization theory, the dual problem of (19) can be transformed into the following quadratic programing form.

$$\arg\max_{\tilde{\alpha}}-\tilde{\mathbf{\alpha}}^\mathbf{T}\mathbf{H}\tilde{\mathbf{\alpha}}+\tilde{\mathbf{\alpha}}^\mathbf{T}\mathbf{f}$$
$$\mathbf{s.t}\ \tilde{\mathbf{\alpha}}^\mathbf{T}1=1,\alpha_i\geq 0,\forall i \tag{20}$$

Here, $\tilde{\boldsymbol{\alpha}}$ is the Lagrangian multipliers, and the symbols in (20) are presented with more details as follows:

$$\tilde{\boldsymbol{\alpha}}=\left(\lambda_1'^+,\dots\lambda_N'^+,\lambda_1'^-,\cdots,\lambda_N'^-\right)^\mathrm{T} \tag{21a}$$

$$\boldsymbol{f}=\left(\frac{2}{\tau}\boldsymbol{y}^\mathrm{T},-\frac{2}{\tau}\boldsymbol{y}^\mathrm{T}\right)^\mathrm{T},\ \boldsymbol{y}=\left(y_1,y_2,\cdots,y_N\right)^\mathrm{T} \tag{21b}$$

$$\mathbf{H}=\begin{bmatrix}\mathbf{K} & \mathbf{-K}\\ \mathbf{-K} & \mathbf{K}\end{bmatrix} \tag{21c}$$

$$\mathbf{K}=\left[\tilde{k}_{ij}\right]_{N\times N},\ \tilde{k}_{ij}=\frac{2}{\tau^2}\boldsymbol{x}_{\rho i}^\mathbf{T}\mathbf{\Phi}(\mathbf{s})\mathbf{\Psi}\mathbf{\Phi}(\mathbf{s})^\mathbf{T}\boldsymbol{x}_{\rho i}+\frac{N}{\tau}\delta_{ij} \tag{21d}$$

The detailed derivation of (20) from (19) is provided in appendix A (*please see Part 6 in the Supplementary Materials*).

According to the duality theory and the result in appendix A, the relationship of the optimal solutions between the primal problem (19) and the dual problem (20) can be expressed as follows.

$$\mathbf{\beta}=\frac{2}{\tau}\mathbf{\Psi}\sum_{i=1}^{N}\left(\lambda_i'^+-\lambda_i'^-\right)\mathbf{\Phi}(\mathbf{s})^\mathbf{T}\mathbf{x}_{gi} \tag{22}$$

Based on (16) and (22), the optimal solution of $\mathbf{p}_g$, i.e., the optimal consequent parameters of the trained TSK FS, can be obtained.

C. Iterative Transfer Learning

As shown in (16), the labeled data in both the source domain and the target domain must be available in order to realize the adaptation of conditional distribution. Since there are no labeled data in the target domain for transductive transfer learning scene, preliminarily labeling of the data in the target domain is needed. This requires a certain classifier that has been trained with other methods for the target domain. Thus, an issue here is how to improve the accuracy of the preliminary labels in the target domain in order to leverage the conditional distribution appropriately for knowledge transfer. For this challenge, an iterative transfer learning strategy as shown in Fig.1 is proposed.

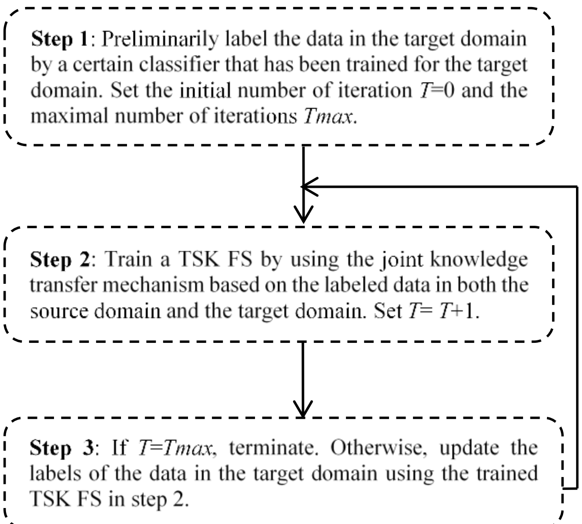

Fig.1 Iterative transfer learning

D. Description of Algorithm

Based on the abovementioned descriptions and analyses, the algorithm of the proposed enhanced transductive transfer learning method for TSK FS construction, ETTL-TSK-FS, is presented in Table I (*please see Part 1 in the Supplementary Materials*).

E. Multi-classification Strategy

The TSK FS model concerned in the paper is commonly used for regression, but it is also suitable for classification. For the purpose of classification, a widely used strategy is to transform the multi-class classification problem to a multi-output regression problem. The procedure is as follows. Given a dataset having m classes: $\{\mathbf{x}_i, y_i\}, \quad y_i \in \{1,2,\cdots,m\}, i=1,2,\cdots,N$, we can construct a regression dataset with multiple outputs. If the label of the ith example $(\mathbf{x}_i, y_i)$ is $y_i = p \ (1 \le p \le m)$, the corresponding label vector in the regression data with m outputs is $\tilde{\mathbf{y}}_i = [0,\cdots,1_p,0,\cdots,0]^T$, i.e. the pth element of the label vector is 1 and the other elements are 0. A regression model with m outputs can be divided into m regression models each with a single output. Once the m regression models are obtained, for a given test example, the output vector of the trained model can be expressed as $\tilde{\mathbf{y}}_i^{\mathbf{model}} = [\tilde{y}_{i1}^{model},\cdots,\tilde{y}_{im}^{model}]^T$. Then, the predicted label of a test example is the subscript of the largest element in the output vector [10]. For instance, if $\tilde{y}_{il}^{\mathbf{model}}$ is the largest element in the vector $\tilde{\mathbf{y}}_i^{\mathbf{model}} (i=1,...,m)$, the final label of the test example is l.

IV. Experimental Studies

A. Datasets

1) Bonn dataset

The Bonn dataset is a single-channel EEG signals dataset from the University of Bonn, Germany [37]. It contains five groups of data, namely, A to E, in the EEG database, where each group consists of one hundred examples. The EEG signals were recorded under the different conditions with five patients and five healthy volunteers. The detailed descriptions of five groups are briefly described in Table II. The first sample of each group is shown in Fig. 2 (*please see Part 2 in the Supplementary Materials*).

The data were recorded using a 128-channel amplifier system with 12 bit resolution and digitized at a sampling rate of 173.61Hz. There were 4096 sampling points in each sample during 23.6s. These epochs were cut out from continuous multichannel EEG recordings after visual inspection for artifacts due to muscle activity or eye movement.

2) CHB-MIT dataset

CHB-MIT dataset is a multi-channel EEG signals dataset collected at the Children's Hospital Boston. It consisted of EEG recordings from pediatric subjects with intractable seizures. The recordings were grouped into 24 cases collected from 23 subjects. Each case contained dozens of hours of continuous EEG signals from a single subject.

All signals were sampled at 256 Hz with 16 bit resolution. Most of recordings contained 23 channels (24 or 26 in a few cases). There were 198 seizures in the set of 24 cases and the beginning and ending time of all seizures were given.

In the experimental studies, 24 cases and 23 channels were selected. The recordings with less than 23 channels were removed and the selected 23 channels were the channels contained in the patient 1 (???). The details about the EEG recordings adopted in the experiments are described in Table III, where ID represents the number of patients, T_{min} and T_{max} represents for the minimum and maximum duration of the seizures(???).

Table II: The five groups of data in the Bonn dataset

	Group	Dataset description
Healthy people	A	EEG signals obtained when healthy volunteer kept their eyes open.
	B	EEF signals obtained when healthy volunteer kept their eyes closed.
Patients	C	EEG signals of patients in the hippocampus of the brain during periodic lulls.
	D	EEG signals of patients in the epileptogenic areas of the brain during periodic lulls.
	E	EEG signals measured during the onset of epileptic seizure.

B. Feature Extraction and Datasets Construction

1) Feature extraction method for segments of EEG signals

To leverage the information in both the time and frequency domains simultaneously, WPD, which has been extensively used for feature extraction from EEG signals [38, 39], was adopted in this study. In the experiments, Daubechies order-4 (db4) and level 5 were used to decompose the original EEG signals. The sub-band energies of the wavelet coefficients are adopted as the features [40]. In consideration of the different sampling rate of the Bonn and CHB-MIT datasets, different sub-bands were selected to calculate the energies. For the Bonn dataset, nodes (1, 0), (2, 0), (3, 0), (4, 0), (5, 0) and (5, 1) of the wavelet packet tree were selected. For the CHB-MIT dataset, nodes (1, 1), (2, 1), (3, 1), (4, 1), (5,1) and (5, 0) of the wavelet packet tree were selected. A 6-dimensional feature vector was generated for each segment of the EEG signals.

2) Datasets construction for Bonn dataset

In the Bonn dataset, there were 500 samples in total with 100 samples for each group, each sample corresponded to an epoch with 4096 sampling points. They were used to extract features in the study. After WPD feature extraction, each sample contained 6 new features and were used to train a classifier. The labels of the samples were the corresponding categories that the EEG signals belonged to. The features extracted from all the samples in five groups are shown in Fig.3 (*please see part 3 in the Supplementary Materials*). Based on the data described in Table II, ten datasets were constructed for evaluating the performance of the proposed ETTL-TSK-FS method. The details of the ten datasets are given in Table IV. For datasets 1 and 2, the training data in the source domain and the test data in the target domain had the same distribution. For the remaining datasets, i.e., datasets 3-10, the distributions of the training data and test data were different. Furthermore, the samples in the training data and test data for all the datasets were different.

3) Datasets construction for CHB-MIT datasets

The EEG signals in the CHB-MIT dataset are continuous. They must be segmented for feature extraction. The duration of non-seizure signals for each patient is long, while the time interval of seizure onset is short. That is, the training data for the non-seizure interval is enough but the training data for seizure onset is scare. Here, the moving-window technique was used to segment the EEG signals into epochs. The non-overlapping moving-window with 10 seconds (2560 samples) was used to extract features from non-seizure signals and the 1-second overlapping moving-windows with 10 seconds was used to extract features from seizure onset signals. 23 channels were selected for all the signals. Each epoch of each channel had 6 features and the features of all the channels were concatenated to produce a sample with 138 (6*23) dimensions.

Table III: The data adopted from the CHB-MIT dataset

ID-Gender-Age	Number of seizures (T_{min}-T_{max})	Seizure time (number of segments)	Non-seizure time (number of segments)
1-F-11	7 (27-101)	442 (433)	145546 (14552)
2-M-11	3 (9-82)	172 (163)	126787 (12677)
3-F-14	7 (47-69)	402 (393)	136404 (13636)
4-M-22	4 (49-116)	378 (369)	561456 (56137)
5-F-7	5 (96-120)	558 (549)	139852 (13984)
6-F-1.5	10 (12-20)	153 (144)	240093 (24006)
7-F-14.5	3 (86-143)	325 (316)	241063 (24105)
8-M-3.5	5 (134-264)	919 (910)	71104 (7108)
9-F-10	4 (62-79)	276 (267)	244062 (24402)
10-M-3	7 (35-89)	447 (438)	179637 (17961)
11-F-12	3 (22-752)	806 (797)	124451 (12442)
12-F-3	27 (13-847)	1810 (1801)	72666 (7261)
13-F-3	10 (17-70)	440 (431)	39160 (3913)
14-F-9	8 (14-41)	169 (160)	93431 (9339)
15-F-16	20 (31-205)	1992 (1983)	138442 (13838)
16-F-7	8 (6-14)	69 (60)	61131 (6111)
17-F-12	3 (88-115)	293 (284)	68131 (6812)
18-F-18	6 (30-68)	317 (308)	124357 (12434)
19-F-19	3 (77-81)	236 (227)	103910 (10390)
20-F-6	8 (29-49)	294 (285)	99072 (9903)
21-F-13	4 (12-81)	199 (190)	117990 (11796)
22-F-9	3 (58-74)	204 (195)	111407 (11139)
23-F-6	7 (20-113)	424 (415)	95186 (9516)
24-NR-NR	16 (16-70)	511 (502)	76156 (7611)

Table IV: The ten datasets constructed for performance comparison

Distribution	No. of dataset	Training set (source domain)	Test set (target domain)	Number of classes	Descriptions of class labels
Identical	1	BE-each 75	BE-each 25	2	**Binary classification** Class 1: EEG signals obtained from the healthy volunteer with eye open or closed. Class 2: EEG signals obtained from patients during periodic lulls or onset of seizure. **Three class classification** Class 1: EEG signals obtained from the healthy volunteer keeping eye open or closed. Class 2: EEG signals obtained from patients during periodic lulls. Class 3: EEG signals obtained from patients during the onset of seizure.
	2	BDE-each 75	BDE-each 25	3	
Different	3	AE-each 50	AC-each 50	2	
	4	AE-each 50	AD-each 50	2	
	5	BE-each 50	BC-each 50	2	
	6	BE-each 50	BD-each 50	2	
	7	ACE-each 50	BCE-each 50	3	
	8	ADE-each 50	BDE-each 50	3	
	9	ACE-each 50	ADE-each 50	3	
	10	BCE-each 50	BDE-each 50	3	

C. Experimental Settings

The purpose of the proposed method is to improve the performance of epileptic EEG signal recognition in the scene of transductive transfer learning. After feature extraction and dataset construction, two sets of data were made available for the experiments and the settings are described as follows.

1) Experimental settings on Bonn dataset

For the Bonn dataset, accuracy was selected as the performance metric to demonstrate the performance of the algorithms more intuitively. The definition of accuracy was the proportion of the number of correctly classified epochs to the total number of epochs.

Six existing modeling methods were adopted for the performance comparison with the proposed ETTL-TSK-FS on the Bonn dataset. On one hand, the proposed method were compared with three classical intelligent methods without transfer learning abilities, i.e., support vector machine (SVM) [41], TSK fuzzy system (TSK FS) [28] and radial basis function neural networks (RBF-NN) [42]. Radial basis function was selected as the kernel function of SVM. For these three non-transfer learning methods, the training and test datasets listed in Table IV were fed into the models directly without additional settings as in the traditional training strategies. On the other hand, the proposed method was also compared with three classical transfer learning methods, i.e., TSVM [43], LMPROJ [13] and TSK-TL-FS [14]. TSVM is an SVM with transductive inference ability; LMPROJ is an SVM with transductive transfer learning abilities based on MMD; and TSK-TL-FS is a TSK FS construction method with transductive transfer learning abilities using MMD [14]. The strategies of training and testing for all these transfer learning methods were same as that of the proposed method.

For all these methods, the hyper-parameters are determined using five-fold cross-validation strategy on given search grids. The hyper-parameters and the corresponding search grids are listed in Table V (*please see part 3 in the Supplementary Materials*). The means and the standard deviations of the classification accuracy of 10 runs of the experiment were recorded for performance evaluation.

2) *Experimental settings on CHB-MIT dataset*

As in most literatures that are based on the CHB-MIT dataset, the classification performance metrics for the seizure and non-seizure segments adopted here are sensitivity, specificity and accuracy, which are defined as follows.

$$Sen = TP/TP + FN \tag{23a}$$

$$Spe = TN/TN + FP \tag{23b}$$

$$Acc = (TN + TP)/(TP + TN + FP + FN) \tag{23c}$$

where TP (true positive) is the number of segments correctly detected as seizure, FN (false negative) is the number of segments incorrectly detected as non-seizure, TN (true negative) is the number of segments correctly detected as non-seizure, and FP (false positive) is the number of segments incorrectly detected as seizure.

For patient-specific recognition of EEG signals in the scene of transductive transfer learning, the data in the target domain can be selected from a specific patient, say patient A, and the data in the source domain can be selected from the non-seizure data of the same patient A and from the seizure data of other patients. Two thousand randomly selected non-seizure segments of patient A and one thousand seizure segments of other patients were combined as the data of the source domain, where the seizure segments from other patients were treated as auxiliary data to realize transfer learning. In addition, another two thousand non-seizure segments and all the seizure segments from patient A were selected as the data in the target domain. When five-fold cross-validation strategy was adopted for evaluation, all the segments with labels in the source domain and four-fold of the segments without labels in target domain were combined as training data, one-fold of the segments in target domain were the test data.

D. *Experimental results on Bonn dataset*

1) *Comparing non-transfer and transfer learning methods*

The experimental results obtained with the ten epileptic EEG datasets are shown in Table VI, where the means and standard deviations (inside brackets) of the classification accuracies of the 10 runs are reported. The following findings can be drawn from the results.

(1) For datasets 1 and 2, since the distribution of the data in the source domain and the target domain are identical, the classification performance of transfer learning methods and non-transfer methods are comparable. For datasets 3 to 10, the transfer learning methods LMPROJ, TSK-TL-FS and ETTL-TSK-FS obviously outperform the non-transfer learning methods in classification performance.

(2) The results on datasets 1 and 2 show that the proposed ETTL-TSK-FS method is also suitable for classical modeling scenes where the training data and the test data have the same distribution. In this situation, although the classification accuracies of ETTL-TSK-FS are not significantly better than that of the classical non-transfer-learning methods, transfer learning based method is still recommended since the distributions of the training and the test data are usually not known to users in advance.

(3) Among the transfer learning methods, the proposed ETTL-TSK-FS has demonstrated better performance than the others. For example, the classification accuracies of ETTL-TSK-FS on all the datasets are higher than 95%.

(4) Comparing with TSVM and LMPOJ, the proposed ETTL-TSK-FS not only has better classification accuracies but also better interpretability as it inherits the characteristics of fuzzy logic rules based methods.

(5) Comparing with the TSK-TL-FS, the proposed ETTL-TSK-FS inherits the good interpretability of TSK-TK-FS. It also demonstrates better transfer learning ability, attributed to the improved transfer learning mechanism proposed that considers both marginal and conditional distributions of the data simultaneously.

An important innovation of the proposed ETTL-TSK-FS is the iterative transfer learning ability. This is illustrated in Fig. 4 with datasets 1, 4, 6 and 9, and the classification accuracies of 10 iterations are recorded (*please see part 4 in the Supplementary Materials*). We can see that when the number of iteration increases, there is an improving trend of classification accuracies, until the performance stabilizes at a certain level. The results also show empirically that ten iterations are usually enough for epileptic EEG recognition, which offers a practical advantage that the iterative transfer learning process is time efficient.

2) *Statistical analysis*

To further evaluate the proposed ETTL-TSK-FS, the nonparametric Friedman test [55] was used to evaluate whether the difference in classification performance among the methods concerned in the study is of statistical significance. In the test, the significance level α was set to 0.05 as usual. If the *p*-value of the Friedman test was smaller than α, the null hypothesis that the classification performance of all the methods was the same was rejected and there existed significant difference in their performance. Furthermore, if the null hypothesis was rejected, the post-doc test [55] was conducted to analyze the difference between the best method and the other six methods. The results of the Friedman test and post-hoc test are given in Table VII (*please see part 5 in the Supplementary Materials*) and VIII respectively.

Table VI: Classification accuracies of the seven methods on the ten sets of data derived from the Bonn dataset

Datasets	**Non-transfer methods**				**Transfer methods**		
	SVM	TSK-FS	RBF-NN	TSVM	LMPROJ	TSK-TL-FS	ETTL-TSK-FS
1	0.9200 (0.0016)	**0.9800** (0.0190)*	0.9700 (0.0058)	0.9267 (0.0346)	0.9480 (0.0235)	0.9700 (0.0134)	0.9701 (0.0230)
2	0.9200 (0.0046)	0.953 (0.0031)	0.9401 (0.0064)	0.9500 (0.0593)	0.9370 (0.0258)	0.9270 (0.0215)	**0.9700** (0.0120)
3	0.9400 (0.0120)	0.9102 (0.0240)	0.9304 (0.0146)	0.8903 (0.0715)	0.9380 (0.0717)	0.9510 (0.0128)	**0.9601** (0.0300)
4	0.9510 (0.0220)	0.9201 (0.0340)	0.9500 (0.0303)	0.9200 (0.0789)	0.9590 (0.0232)	0.9300 (0.0235)	**0.9750** (0.0093)
5	0.9500 (0.1006)	0.9400 (0.0187)	0.9301 (0.0290)	0.8901 (0.0994)	0.9380 (0.1740)	0.9760 (0.0146)	**0.9850** (0.0045)
6	0.9000 (0.0105)	0.9300 (0.0104)	0.9501 (0.0106)	0.9200 (0.0919)	0.9590 (0.0354)	0.9810 (0.1039)	**0.9850** (0.0435)
7	0.8830 (0.0209)	0.8500 (0.1230)	0.8304 (0.1322)	0.7344 (0.0752)	0.9470 (0.0532)	0.9280 (0.0415)	**0.9500** (0.0176)
8	0.8800 (0.0100)	0.8500 (0.0140)	0.8011 (0.0250)	0.7501 (0.0572)	0.9380 (0.0355)	0.9350 (0.0340)	**0.9500** (0.0375)
9	0.8709 (0.0008)	0.8601 (0.0040)	0.8110 (0.0300)	0.7625 (0.0588)	0.9288 (0.0253)	0.9453 (0.0410)	**0.9567** (0.0333)
10	0.8950 (0.0306)	0.8702 (0.0106)	0.8020 (0.0200)	0.7805 (0.0430)	0.9301 (0.0357)	0.9472 (0.0520)	**0.9533** (0.0165)
Average	0.9109 (0.0213)	0.9064 (0.0261)	0.8915 (0.0303)	0.8525 (0.0669)	0.9423 (0.0503)	0.9525 (0.0358)	**0.9655** **(0.0227)**

Table VIII: Post-hoc test between the proposed ETTL-TSK-FS and the other methods

i	Algorithms	$z=(R_0-R_i)/SE$	p	$\text{Holm}=\alpha/i,\ \alpha=0.05$	Null hypothesis
6	TSVM	5.27900	0	0.008333	Rejected
5	RBF-NN	3.881619	0.000104	0.01	Rejected
4	SVM	3.415825	0.000636	0.0125	Rejected
3	TSK-FS	3.208805	0.001333	0.016667	Rejected
2	LMPROJ	2.277216	0.022773	0.025	Rejected
1	TSK-TL-FS	1.500893	0.133383	0.05	Not rejected

The result of Friedman test in Table VII shows that the classification performance of the seven methods is significantly different and that the ranking of the proposed ETTL-TSK-FS is highest (with the lowest ranking value), which indicates its superiority to all the other methods. The post-hoc test was then conducted to compare the ETTL-TSK-FS with the other six methods. The results shown in Table VIII indicate that the performance of the proposed ETTL-TSK-FS clearly exceeds that of TSVM, RBF-NN, SVM, TSK FS and LMPROJ. Meanwhile, it can be seen from Tables VII and VIII that the performance of the ETTL-TSK-FS is better than that of the TSK-TL-FS to some extent although the improvement is not statistically significant.

3) Further comparison

In this subsection, the proposed method was further evaluated by comparing with methods that had been developed for epileptic EEG recognition on the Bonn dataset. Four methods reported in the literature were adopted, namely, *FFT+DT*, *WT+SLFN*, *WT+ANFIS* and *ApEn+SLFN* [4, 6, 44, 45]. They are only described briefly here. The detailed procedure and the experiments settings of these methods can be found in [4, 6, 44, 45].

FFT+DT [44]: This method is based on frequency domain analysis and consists of two stages: feature extraction using the FFT-based Welch method and classification with decision tree. 129 features were obtained by the FFT based Welch method to train a decision tree (DT) for classification.

WT+SLFN [45]: It is based on time-frequency domain analysis using DWT. The db4 wavelet coefficients were used to decompose the original EEG signals and the statistics of wavelet coefficients were calculated for the frequency bands A5 and D3-D5. The statistics used were mean, average power, standard deviation and the ratio of the absolute mean values of the adjacent sub-bands. A total of 16 features were used to train the single hidden layer feedforward neural networks (SLFN) with 10 hidden neurons and sigmoid activation functions.

WT+ANFIS [6]: The feature extraction in this method adopted was similar to that in [45]. Each signal was first divided into 16 segments by a rectangular window composed of 256 sampling points. Then, db2 wavelet coefficients were used to decompose each segmented signal into D1-D4 and A4 sub-bands. Four statistics were extracted from each segmented signal, i.e., maximum, minimum, mean and standard deviation of the wavelet coefficients. Datasets of 20 dimensions and with 8000 vectors (1600 vectors from each class) were used to train the ANFIS classifier, with three fuzzy partitions for each dimension.

ApEn+SLFN [4]: This method is based on approximate entropy (ApEn). Every EEG segment of 4096 sampling points for each class was divided into four equal epochs to obtain 2000 epochs, each containing 1024 sampling points. For each

EEG epoch, the value of ApEn was calculated. Finally, SLFN with 10 hidden neurons and sigmoid activation functions was trained to be a classifier with the extracted new features.

For fair comparison, the same datasets that were generated from the University of Bonn's epileptic EEG database as described above were adopted for performance comparison. The results are reported in Table IX. The results show that for the scenes where the source and target domains have the same distribution (e.g. datasets 1 and 2), the performance of the proposed methods are comparable with that of the existing methods, but when drifting in data distributions between the two domains exists (datasets 3-10), the proposed methods are obviously outstanding.

Table IX: Further comparison with four other algorithms on the Bonn dataset

	FFT+DT [44]	WT+SLFN [45]	WT+ANFIS [6]	ApEn+SLFN [4]	ETTL-TSK-FS
1	0.9937 (0.0035)	**0.9820** (0.0063)	0.9654 (0.0054)	0.8890 (0.0032)	0.9701 (0.0230)
2	0.9503 (0.0044)	0.9347 (0.0042)	0.8704 (0.0074)	0.6160 (0.0191)	**0.9700** (0.0120)
3	0.6222 (0.0131)	0.6060 (0.0184)	0.5731 (0.0141)	0.8317 (0.0024)	**0.9601** (0.0300)
4	0.6796 (0.0084)	0.6770 (0.0160)	0.7844 (0.0226)	0.8763 (0.0060)	**0.9750** (0.0093)
5	0.8134 (0.0173)	0.6480 (0.0162)	0.5151 (0.0044)	0.7680 (0.0052)	**0.9850** (0.0045)
6	0.9270 (0.0054)	0.9360 (0.0295)	0.8335 (0.0090)	0.6403 (0.0121)	**0.9850** (0.0435)
7	0.8879 (0.0106)	0.8993 (0.0139)	0.7935 (0.0083)	0.6345 (0.0207)	**0.9500** (0.0176)
8	0.8879 (0.0106)	0.8993 (0.0139)	0.7935 (0.0083)	0.6345 (0.0207)	**0.9500** (0.0176)
9	0.8879 (0.0106)	0.8993 (0.0139)	0.7935 (0.0083)	0.6345 (0.0207)	**0.9500** (0.0176)
10	0.8879 (0.0106)	0.8993 (0.0139)	0.7935 (0.0083)	0.6345 (0.0207)	**0.9500** (0.0176)
Average	0.8391 (0.0090)	0.8119 (0.0142)	0.7622 (0.0101)	0.7508 (0.0098)	**0.9500** (0.0375)

E. Experimental results on CHB-MIT dataset

For the classification of signals segments, the three performance metrics – sensitivity, specificity and accuracy – were recorded for each patient. The classification results were listed in Table X. The first column represents the target patients whose data are used for classification. The second column represents the patients whose data are used as the auxiliary seizure segments for the target patients.

The results of some recent studies are summarized in Table XI. It is difficult to directly compare the proposed method with the existing methods due to the different experimental settings adopted by these methods. The main difference among them are the number of selected channels, the number of selected patients, the number of selected seizures, the methods for feature extraction and the methods for constructing the training and test data. While various performance metrics were reported in the literatures of these methods, the common ones, i.e. sensitivity, specificity and accuracy, are given listed in Table XI where the NR represents that the values are not reported in the corresponding method.

(conclusions from the results, To be added late):

Table X: Classification results on CHB-MIT dataset

Patient	Auxiliary	Sensitivity	Specificity	Accuracy
1	3,4,7	100.00	98.75	98.95
2	1,3,5	100.00	98.00	98.15
3	1,5,7	94.12	98.39	98.00
4				
5	1,3,7	94.17	98.00	97.22
6				
7	3,5,9	96.67	96.50	96.52
8				
9	7,11	97.92	100	99.81
10				
11	7,9,13	96.88	95.68	97.86
12				
13				
14				
15				
16				
17				
18				
19	17,18,20,23	93.48	97.73	99.10
20				
21	17,18,19,20	97.37	89.75	90.41
22	17,20,21,23	100	99.50	99.54
23	17,20,22,24	95.18	94.25	94.41
24	18,20,23	88.12	97.75	95.81
Average				

Table XI: The experimental settings and performance results of existing methods on CHB-MIT dataset

No.	Authors	Features Extraction Methods	Patients	Channels	Seizures	Sen-Spe-Acc (%)
1	Rafiuddin et.al. [46]	Two wavelet based features and two statistical features	23	23	195	NR -NR-80.16
2	Khan et.al. [47]	Relative energy and normalized coefficient of variation	5	NR	65	83.6-100-91.8
3	Kiranyaz et.al. [48]	Features for time, frequency, time-frequency, non-linear and MFCC	21	18	NR	89.0-94.7-NR
4	Zabihi et.al. [49]	Seven features extracted from intersection sequence	23	23	161	89.1-94.8-94.6
5	Samiee et.al. [50]	Multivariate textural features from gray level co-occurrence matrix	23	23	153	70.1-97.7-NR
6	Abhijit et.al. [51]	Three features extracted using multivariate extension of EWT	23	5	157	97.9-99.6-99.4
7	Proposed	6*23 energy features extracted from the wavelet coefficients	24	23	176	

F. Discussions

It has been validated experimentally on the Bonn dataset and the CHB-MIT dataset that the proposed ETTL-TSK-FS exhibits better classification performance that various existing algorithms. This results also shows that joint-knowledge can effectively reduce the discrepancy between the source and target domains, and that the proposed iterative transfer learning procedure can enhance the transfer learning ability. Although the training procedure could be quite time consuming, it can be implemented offline. Once a model has been trained, for a future test example, the testing time will be very small as the previously trained model can be used directly. Therefore, it is appropriate for online applications.

As mentioned previously, the two focuses of seizure recognition are on the feature extraction methods and the classification methods. The paper mainly concerns the improvement of classifiers in the scene of transfer learning. Recent development in this area is further discussed here. It can be seen from the literatures that multilayer perceptron neural networks have been widely used, with various configurations for seizure recognition in different methods [52]. Similar with our work that is based on fuzzy modeling techniques, interval type-2 fuzzy inference has been introduced into support vector machine classifier for the classification of three seizures phases [53], which achieves superior learning performance due to its ability of uncertainty fuzzy modeling. Deep learning methods based on stacked sparse autoencoder [54] and recurrent neural networks [55] are explored for seizure recognition. Promising performance is achieved as a result of their outstanding feature learning ability. For the challenge due to the scarcity of labeled seizure data and the voluminous non-seizure data, weighted extreme learning machine is proposed to tackle the class imbalance problem in seizure classification [56], while the problem of lacking labeled seizure training data can be alleviated by the proposed ETTL-TSK-FS by enhancing the transfer learning ability.

V. Conclusion

The enhanced transductive transfer learning TSK FS construction method ETTL-TSK-FS is proposed in this study in order to increase the effectiveness of epileptic EEG signal recognition. Compared with the existing transfer learning TSK FS construction methods, the ETTL-TSK-FS has sufficiently considered the adaptation of both the conditional and the marginal distributions simultaneously, which constitutes the joint-knowledge for transfer learning. Meanwhile, the iterative joint-knowledge-transfer strategy is also introduced to enhance transfer learning ability. The results of the extensive experimental studies show that the ETTL-TSK-FS is a promising approach for epileptic EEG signal recognition in which drifting in data distributions between the source domain and the target domain exists.

Despite the encouraging results, further investigation is need to deal with certain issues . For example, negative transfer may occur in the proposed ETTL-TSK-FS if the joint-knowledge is not properly constructed in the transfer learning procedure. To tackle this problem, cross-validation strategy is adopted in our study to determine the appropriate values of the transfer learning parameters. Nevertheless, this strategy is very time consuming and it is necessary to investigate the theory that governs the setting of the transfer learning parameters. Research into effective approaches to avoid negative transfer is an important future work of the study.

References

[1] A. Talevi, C. L. Bellera, E. A. Castro, and L. E. Bruno-Blanch, "A successful virtual screening application: prediction of anticonvulsant activity in MES test of widely used pharmaceutical and food preservatives methylparaben and propylparaben," *Journal of Computer-Aided Molecular Design*, vol. 21, no. 9, pp. 527-538, 2007.

[2] Y. Zhang, Y. Zhang, J. Wang, and X. Zheng, "Comparison of classification methods on EEG signals based on wavelet packet decomposition," *Neural Computing & Applications*, vol. 26, no. 5, pp. 1217-1225, 2015.

[3] B. R. Greene, S. Faul, W. P. Marnane, G. Lightbody, I. Korotchikova, and G. B. Boylan, "A comparison of quantitative EEG features for neonatal seizure detection," *Clinical Neurophysiology*, vol. 119, no. 6, pp. 1248-1261, 2008.

[4] Q. Yuan, W. Zhou, S. Li, and D. Cai, "Epileptic EEG classification based on extreme learning machine and nonlinear features," *Epilepsy Research*, vol. 96, no. 1–2, pp. 29-38, 2011.

[5] J. M. Mendel, "Uncertain rule-based fuzzy logic systems: introduction and new directions," *Ptr Upper Saddle River Nj*, vol. 133, no. 02, 2001.

[6] I. Güler and E. D. Ubeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of Neuroscience Methods*, vol. 148, no. 2, p. 113, 2005.

[7] A. H. Shoeb and J. V. Guttag, "Application of Machine Learning To Epileptic Seizure Detection," in *International Conference on Machine Learning*, 2010, pp. 975-982.

[8] W. Dai, Q. Yang, G. R. Xue, and Y. Yu, "Boosting for transfer learning," *International Conference on Machine Learning*, pp. 193-200, 2007.

[9] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain Adaptation via Transfer Component Analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199-210, 2011.

[10] Z. Deng, K. Choi, Y. Jiang, and S. Wang, "Generalized Hidden-Mapping Ridge Regression, Knowledge-Leveraged Inductive Transfer Learning for Neural Networks, Fuzzy Systems and Kernel Methods," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 44, no. 12, pp. 2585-2599, 2014.

[11] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge & Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.

[12] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in *international conference on machine learning*, 2007, pp. 759-766.

[13] B. Quanz and J. Huan, "Large margin transductive transfer learning," in *ACM Conference on Information and Knowledge Management*, 2009, pp. 1327-1336.

[14] C. Yang, Z. Deng, K. S. Choi, and S. Wang, "Takagi–Sugeno–Kang Transfer Learning Fuzzy Logic System for the Adaptive Recognition of Epileptic Electroencephalogram Signals," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 5, pp. 1079-1094, 2016.

[15] Y. U. Khan and J. Gotman, "Wavelet based automatic seizure detection in intracerebral electroencephalogram," *Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology*, vol. 114, no. 5, pp. 898-908, 2003.

[16] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalography & Clinical Neurophysiology*, vol. 54, no. 5, pp. 530-540, 1982.

[17] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Computational Intelligence & Neuroscience*, vol. 2007, no. 4, p. 80510, 2007.

[18] A. H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," *Massachusetts Institute of Technology*, 2009.

[19] S. M. Akareddy and P. K. Kulkarni, "EEG signal classification for Epilepsy Seizure Detection using Improved Approximate Entropy," *International Journal of Public Health Science*, vol. 2, no. 1, 2013.

[20] J. C. Sackellares, D. S. Shiau, K. M. Kelly, and S. P. Nair, *Seizure Prediction in Epilepsy: From Basic Mechanisms to Clinical Applications*. 2008, pp. 237-248.

[21] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain A Journal of Neurology,* vol. 130, no. Pt 2, p. 314, 2007.

[22] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalography & Clinical Neurophysiology,* vol. 29, no. 3, pp. 306-310, 1970.

[23] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients," *Biomedical Engineering, IEEE Transactions on,* vol. 50, no. 5, pp. 603-615, 2003.

[24] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *Epilepsia,* vol. 52, no. 10, pp. 1761-1770, 2011.

[25] E. A. Vivaldi and A. Bassi, "Frequency domain analysis of sleep EEG for visualization and automated state detection," in *International Conference of the IEEE Engineering in Medicine & Biology Society*, 2006, p. 3740.

[26] H. Adeli and Z. N. Zhou, "Analysis of EEG records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods,* vol. 123, no. 1, p. 69, 2003.

[27] N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew, "Classification of EEG signals using the wavelet transform," *Signal Processing,* vol. 59, no. 1, pp. 61-72, 1997.

[28] J. M. Won, S. Y. Park, and J. S. Lee, "Parameter conditions for monotonic Takagi–Sugeno–Kang fuzzy system," *Fuzzy Sets & Systems,* vol. 132, no. 2, pp. 135-146, 2002.

[29] 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 & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society,* vol. 40, no. 3, pp. 798-818, 2010.

[30] Z. Deng, Y. Jiang, H. Ishibuchi, K. S. Choi, and S. Wang, "Enhanced Knowledge-Leverage-Based TSK Fuzzy System Modeling for Inductive Transfer Learning," *Acm Transactions on Intelligent Systems & Technology,* vol. 8, no. 1, p. 11, 2016.

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

[32] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation Regularization: A General Framework for Transfer Learning," *IEEE Transactions on Knowledge & Data Engineering,* vol. 26, no. 5, pp. 1076-1089, 2014.

[33] Wu, Pengcheng, Dietterich, and G. Thomas, "Improving SVM accuracy by training on auxiliary data sources," *Icml,* 2004.

[34] Z. Deng, Y. Jiang, K. S. Choi, F. L. Chung, and S. Wang, "Knowledge-Leverage-Based TSK Fuzzy System Modeling," *IEEE Transactions on Neural Networks & Learning Systems,* vol. 24, no. 8, p. 1200, 2013.

[35] Q. Sun *et al.*, "A Two-Stage Weighting Framework for Multi-Source Domain Adaptation," *Advances in Neural Information Processing Systems,* pp. 505-513, 2011.

[36] K. Zhang, V. W. Zheng, Q. Wang, J. T. Kwok, Q. Yang, and I. Marsic, "Covariate shift in Hilbert space: A solution via surrogate kernels," in *International Conference on Machine Learning*, 2013, pp. 388-395.

[37] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and 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," *Physical Review E Statistical Nonlinear & Soft Matter Physics,* vol. 64, no. 6 Pt 1, p. 061907, 2001.

[38] Y. Zhang, B. Liu, X. Ji, and D. Huang, "Classification of EEG Signals Based on Autoregressive Model and Wavelet Packet Decomposition," *Neural Processing Letters,* vol. 45, no. 2, pp. 1-14, 2017.

[39] D. Hu, W. Li, and X. Chen, "Feature extraction of motor imagery EEG signals based on wavelet packet decomposition," in *Ieee/icme International Conference on Complex Medical Engineering*, 2011, pp. 694-697.

[40] W. Tu and Q. Wei, "Classification of Three-Class Motor Imagery EEG Data by Combining Wavelet Packet Decomposition and Common Spatial Pattern," in *International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2009, pp. 188-191.

[41] R. Panda, P. S. Khobragade, P. D. Jambhule, S. N. Jengthe, P. R. Pal, and T. K. Gandhi, "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure diction," in *International Conference on Systems in Medicine and Biology*, 2010, pp. 405-408.

[42] S. K. Satapathy, S. Dehuri, and A. K. Jagadev, "EEG signal classification using PSO trained RBF neural network for epilepsy identification," *Informatics in Medicine Unlocked,* vol. 6, pp. 1-11, 2017.

[43] T. Joachims, "Transductive Inference for Text Classification using Support Vector Machines," in *international conference on machine learning*, 1999, pp. 200-209.

[44] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Applied Mathematics & Computation,* vol. 187, no. 2, pp. 1017-1026, 2007.

[45] M. Li, W. Chen, and T. Zhang, "Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble," *Biomedical Signal Processing & Control,* vol. 31, pp. 357-365, 2017.

[46] N. Rafiuddin, Y. Uzzaman Khan, and O. Farooq, "Feature extraction and classification of EEG for automatic seizure detection," in *International Conference on Multimedia, Signal Processing and Communication Technologies*, 2011, pp. 184-187.

[47] Y. U. Khan, N. Rafiuddin, and O. Farooq, "Automated seizure detection in scalp EEG using multiple wavelet scales," in *IEEE International Conference on Signal Processing, Computing and Control*, 2012, pp. 1-5.

[48] S. Kiranyaz, T. Ince, M. Zabihi, and D. Ince, "Automated patient-specific classification of long-term Electroencephalography," *Journal of Biomedical Informatics,* vol. 49, no. 6, pp. 16-31, 2014.

[49] M. Zabihi, S. Kiranyaz, R. A. Bahrami, A. Katsaggelos, M. Gabbouj, and T. Ince, "Analysis of High-Dimensional Phase Space via Poincaré Section for Patient-Specific Seizure Detection," *IEEE Transactions on Neural Systems & Rehabilitation Engineering,* vol. 24, no. 3, pp. 386-398, 2015.

[50] K. Samiee, S. Kiranyaz, and M. Gabbouj, "Long-term epileptic EEG classification via 2D mapping and textural features," *Expert Systems with Applications An International Journal,* vol. 42, no. 20, pp. 7175-7185, 2015.

[51] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient specific EEG seizure detection using empirical wavelet transform," *IEEE Transactions on Biomedical Engineering,* vol. PP, no. 99, pp. 1-1, 2017.

[52] S. Raghu and N. Sriraam, "Optimal configuration of multilayer perceptron neural network classifier for recognition of intracranial epileptic seizures," *Expert Systems with Applications,* vol. 89, 2017.

[53] U. Ekong *et al.*, "Classification of epilepsy seizure phase using interval type-2 fuzzy support vector machines," *Neurocomputing,* vol. 199, no. C, pp. 66-76, 2016.

[54] Q. Lin *et al.*, "Classification of Epileptic EEG Signals with Stacked Sparse Autoencoder Based on Deep Learning," 2016.

[55] L. S. Vidyaratne and K. M. Iftekharuddin, "Real-Time Epileptic Seizure Detection Using EEG," *IEEE Transactions on Neural Systems & Rehabilitation Engineering,* vol. 25, no. 11, pp. 2146-2156, 2017.

[56] Q. Yuan *et al.*, "Epileptic seizure detection based on imbalanced classification and wavelet packet transform," *Seizure,* vol. 50, p. 99, 2017.