This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

EEG-Based Driver Drowsiness Estimation Using an Online Multi-View and Transfer TSK Fuzzy System

Yizhang Jiang, *Senior Member, IEEE*, Yuanpeng Zhang, *Member, IEEE*, Chuang Lin, Dongrui Wu, *Senior Member, IEEE*, and Chin-Teng Lin, *Fellow, IEEE*

***Abstract*— In the field of intelligent transportation, transfer learning (TL) is often used to recognize EEG-based drowsy driving for a new subject with few subject-specific calibration data. However, most of existing TL-based models are offline, non-transparent, and in which features are only represented from one view (usually only one algorithm is used to extract features). In this paper, we consider an online multi-view regression model with high interpretability. By taking the 1-order TSK fuzzy system as the basic regression component and injecting the nature of the multi-view settings into the existing transfer learning framework and enforcing the consistencies across different views, we propose an online multi-view & transfer TSK fuzzy system for driver drowsiness estimation. In this novel model, features in both the source domain and the target domain are represented from multi-view perspectives such that more pattern information can be utilized during model training. Also, comparing with offline training, the proposed online fuzzy system meets the practical requirements more competently. An experiment on a driving dataset demonstrates that the proposed fuzzy system has smaller drowsiness estimation errors and higher interpretability than introduced benchmarking models.**

***Index Terms*— Transfer learning, multi-view learning, TSK fuzzy systems, EEG.**

I. Introduction

TRAFFIC accidents are one of the serious social problems facing the world at present. They have been recognized as the first public hazard that threatens the safety of human life in the world today. At least 500,000 people die each year due to traffic accidents [48]. In the statistics of the causes of traffic accidents in various countries, drowsy driving all occupies a large proportion [1]. According to the statistics from the Ministry of Transport of China in 2001 that about 48% of highway traffic accidents were caused by drowsy driving [48]. In the United States, 4,121 people were killed in road traffic accidents caused by drowsy driving between 2011 and 2015 [45]. In addition, the National Sleep Foundation poll reported that 4% of drivers admitted that they had an accident or near-accident because they were dozing off or too tired to drive [46]. Some researchers estimated by modeling that 15%-33% fatal accidents might involve drowsy drivers [47].

As a result, how to recognize drivers' drowsiness levels by computational intelligent technologies and initiate interference measures so as to avoid accidents has become an active research topic in automotive safety engineering. As summarized in [2], there are two types of approaches for drowsiness level recognition. One is computer vision-based, which uses cameras with different functions to capture the characteristics of PERCLOS (percentage eye closure), the mouth state, the pupil size and/or nodding activities and then deduce a driver's drowsiness level. Although this type of approaches can work in an effective manner, it is susceptible to light. The other one is physiological signal monitoring-based, which uses, e.g., a brain-computer interference (BCI) system, to capture and analyze scalp EEG signals to estimate the response time, i.e., the time interval between the onset of lane deviation and the driver's first response, which is an indicator of a driver's drowsiness level. In this paper, our approach belongs to the second category.

Drowsiness recognition has been studied for serval years [49]–[51]. In [3], Khushaba *et al*. proposed an approach for driver drowsiness estimation (classification) by introducing an FWP-based (fuzzy wavelet packet) feature extraction algorithm. In [4], based on enhanced batch-mode active learning, Wu *et al*. proposed an EEG-based model for driver drowsiness estimation. Other relevant studies can be found in [39]–[42]. However, these are offline studies which do not meet the practical requirements that drowsiness level of a driver should be identified online, in real-time. Additionally, due to the differences between individuals, it is very challenging to design a drowsiness estimation model that fits all individuals. Some of subject-specific calibration data are usually needed to adjust the model for a new individual (an individual is also termed as a subject in the following sections). In order to improve the efficiency of the BCI system, we have to reduce the

Manuscript received September 23, 2019; revised December 18, 2019 and January 26, 2020; accepted February 10, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 81701793, Grant 61702225, Grant 61772241, and Grant 61873321, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20160187, in part by the 2018 Six Talent Peaks Project of Jiangsu Province under Grant XYDXX-127, in part by the Science and Technology Demonstration Project of Social Development of Wuxi under Grant WX18IVJN002, in part by the Youth Foundation of the Commission of Health and Family Planning of Wuxi under Grant Q201654, and in part by the Shenzhen Basic Research Grant under Grant JCYJ20170413152804728 and Grant JCYJ2018050718250885. The Associate Editor for this article was H. Gao. *(Corresponding author: Yuanpeng Zhang.)*

Yizhang Jiang is with the School of Digital Media, Jiangnan University, Wuxi 214122, China (e-mail: yzjiang@jiangnan.edu.cn).

Yuanpeng Zhang is with the Department of Medical Informatics, Nantong University, Nantong 226001, China (e-mail: maxbirdzhang@ntu.edu.cn).

Chuang Lin is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518000, China (e-mail: chuang.lin@siat.ac.cn).

Dongrui Wu is with the Key Laboratory of the Ministry of Education for Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: drwu@hust.edu.cn).

Chin-Teng Lin is with the Centre of Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: chin-teng.lin@uts.edu.au).

Digital Object Identifier 10.1109/TITS.2020.2973673

amount of such calibration data. To this end, transfer learning (TL) [5]–[7] can be adopted for driver drowsiness estimation since it is able to take full advantage of knowledge/data from other existing subjects. To our best knowledge, there have been only three studies on driver drowsiness estimation by using TL. Wei *et al*. [8] proposed a selective TL framework, where TL can be selectively turned on or off based on the power of session generalizability. Their experimental results showed that selective TL performs better than classical TL on driver drowsiness estimation. In [9], a transfer model termed as DAMF (domain adaptation with model fusion) is proposed by Wu *et al*. [9] for driver drowsiness estimation. In DAMF, very little knowledge is required such that its real-world applicability is significantly increased. Furthermore, in [10], Wu *et al*. also proposed a transfer regression model for driver drowsiness estimation based on weighting adaptation regularization. This model not only needs very little calibration data, but also can select a subset of useful source domains to reduce the computational cost.

In the above transfer models, subjects are described from only one perspective (view), e.g., the theta band power in [9] or the dBs in [4]. Wu *et al*. in [4] indicated that the theta band powers are highly relative to the dBs, and both of them have strong correlations with the drowsiness index. Therefore, if we can design a collaborative learning mechanism which can not only learn pattern information from each individual view, but also leverage the potential supplementary information across multiple views, then promising drowsiness estimation performance is anticipated. To this end, multi-view learning is often employed, and extensive prior studies [11]–[17] have demonstrated that leveraging the supplementary information across multiple views can indeed enhance the learning performance. For instance, in [11], the authors trained a within-view classifier from each individual view and then regularized the consistencies across different views. Moreover, the authors demonstrated that with the across-view consistency regularization, the Rademacher complexity of the function class can also be significantly reduced. The across-view consistency regularization was further studied in [13], in which it was integrated with multi-view semi-supervised learning, and helped the classifier achieve a substantial classification performance improvement. In [16], a similar idea was introduced to local learning, and a novel approach was designed to define the graph Laplacian. However, most existing multi-view learning models are designed for single-domain settings.

This paper proposes an online multi-view transfer Takagi-Sugeno-Kang (TSK) fuzzy system O-MV-T-TSK-FS for driver drowsiness estimation, which combines online learning, transfer learning and multi-view learning together. O-MV-T-TSK-FS is not a model which heuristically combines multi-view learning and transfer learning; instead, it injects the nature of the multi-view settings into the transfer learning framework and enforces the consistencies across different views, which restricts the capacity of the hypothesis output implicitly. Furthermore, it also emphasizes the data distribution difference between the source domain and the target domain, employing weighted labeled data to transfer knowledge from the source domain to set up an estimation model for the target domain. The main contributions of this study can be summarized as follows.

(1) An online multi-view & transfer learning framework is proposed. The framework has three merits: it can enforce the consistencies across different views; it emphasizes the data distribution difference between the source domain and the target domain; it can avoid the output being (nearly) a constant.

(2) The framework is applied to estimate drivers' drowsiness by EEG signals, where the theta band power and dBs extracted from EEG are considered as two views.

The remainder of this paper is organized as follows. Section II introduces the drivers' scalp EEG data and the classic TSK fuzzy system. Section III constructs an online multi-view & transfer learning framework and proposes our regression model. Section IV reports the experimental results. Section V draws conclusions.

II. Background

This section first briefly introduces the drivers' scalp EEG data including signal collection, preprocessing and feature extraction, then briefly presents the classic TSK fuzzy system which is considered as the basic regression component in this study.

A. Drivers' Scalp EEG Data

The original drivers' scalp EEG data are provided and authorized by the Institutional Review Board of the Taipei Veterans General Hospital. They are collected by the following procedures.

The Taipei Veterans General Hospital recruited 16 volunteers (subjects) with normal or corrected to normal vision from the community to participate in a sustained-attention driving experiment. Before the experiment began, all of the volunteers were informed to read and sign a consent form. The experimental installation consisted of a real vehicle that is mounted on a motion platform with 6 degrees of freedom immersed in a 360-degree virtual-reality (VR) scene. All experiments were conducted in the afternoon because at this time the circadian rhythm of sleep reached its peak, and the experimental results were recorded for about 60-90 minutes. In order to create drowsiness during driving, our VR devices simulated monotonous driving at a fixed speed of 100 km/h on a straight and empty highway. During the experiment, our virtual device randomly generated lane departure events every 5-10 seconds, and the drivers were instructed to adjust their vehicles as soon as possible to correct such simulation perturbations. All volunteers' driving performance and cognitive states were monitored via a vehicle trajectory record system and a surveillance video camera throughout the whole experiment. For each perturbation, the response time from each volunteer was recorded and finally converted to the drowsiness index indicating drowsiness driving. At the same time, volunteers' scalp EEG signals were also recorded via a 32-channel 500 Hz Neuroscan NuAmps Express system provided by Compumedics Ltd., VIC, Australia. The 32-channel contains 30-channel EEGs and 2-channel earlobes.

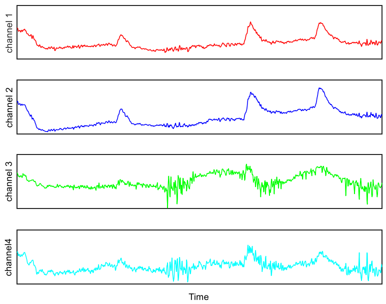


Fig. 1. EEG data after preprocessing.

B. Signal Preprocessing and Feature Extraction

Generally speaking, the raw EEG data cannot be directly used for model training since they often contain high-frequency muscle artifacts, line-noise contaminations, and DC drift. Therefore, they should be preprocessed before being used. The open source EEGLAB toolbox [18] is often employed for raw EEG data preprocessing. In our experiments, we used a 1-50 Hz bandpass filter to filter out line-noise contaminations, DC drift, and high-frequency muscle artifacts contained in the raw EEG data. Then, we down-sampled them from 500 Hz to 250 Hz and re-referenced to averaged earlobes. Fig.1 shows the EEG data (first 4 channels from 1 to 1000 seconds) of one subject after preprocessing.

For the EEG data after preprocessing, we still need feature extraction algorithms to extract different kinds of features to represent EEG data from different views. Generally, there can be three types of features: the time-domain features [21], [23], the frequency-domain features [19], [20], and the time-frequency features [19], [21], [22]. Once the features are obtained, we can then build a regression model for drowsiness estimation.

C. TSK Fuzzy System

Among the existing regression models, most of them work in a black-box manner, e.g., the support vector regression model [24]. In this study, to avoid the proposed model working in a black-box manner, we select a TSK fuzzy system as the basic regression component due to its good balance between high interpretability and promising approximation ability. Here, we first introduce the classic one-order TSK fuzzy system (1-TSK-FS), then in the next section, we further extend it to online multi-view & transfer learning.

In 1-TSK-FS, the kth fuzzy rule in the feature space can be expressed as

$$\begin{aligned}&\text{If } x_{i1} \text{ is } A_1^k \wedge x_{i2} \text{ is } A_2^k \wedge \ldots \wedge x_{id} \text{ is } A_d^k,\\ &\text{then } f^k(\mathbf{x}_i) = p_0^k + p_1^k x_{i1} + \ldots + p_d^k x_{id}, \quad k = 1, 2, \ldots, K.\end{aligned} \tag{1}$$

where A_j^k is a fuzzy set subscribed by the input feature x_{ij} for the kth fuzzy rule, $\wedge$ is a operator for fuzzy conjunction and K is the number of fuzzy rules. Each fuzzy rule is premised on the feature space ($\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T \in R^d$) and maps the fuzzy sets in the feature space into a varying singleton represented by $f^k(\mathbf{x}_i)$. After inference and defuzzification, the output of 1-TSK-FS can be formulated as

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

where

$$\mu^k(\mathbf{x}_i) = \prod_{j=1}^{d} \mu_{A_j^k}(x_{ij}). \tag{3}$$

When the Gaussian function being adopted as the fuzzy membership function, $\mu_{A_j^k}(x_{ij})$ is formulated as

$$\mu_{A_j^k}(x_{ij}) = \exp\left(\frac{-(x_{ij} - c_j^k)^2}{2(\delta_j^k)^2}\right), \tag{4}$$

where c_j^k and δ_j^k denote the kernel center and kernel width, respectively.

From the above equations, we see that c_j^k and δ_j^k in the antecedents and $\mathbf{p}^k = [p_0^k, p_1^k, \ldots, p_d^k]^T$ in the consequents are two sets of parameters needed to learn in the training procedure of 1-TSK-FS. In generally, parameters in antecedent and parameters in consequent are learned independently. Parameters in the antecedents are often obtained by clustering. For example, if fuzzy c-means (FCM) clutering [25], [26] is employed, c_j^k and δ_j^k can be estimated by

$$c_j^k = \sum_{i=1}^{N} \mu_{ik} x_{ij} \sum_{i=1}^{N} \mu_{ik}, \tag{5}$$

$$\delta_j^k = h \sum_{i=1}^{N} u_{ik} (x_{ij} - c_j^k)^2 \sum_{i=1}^{N} u_{ik}, \tag{6}$$

where μ_{ik} represents the fuzzy membership degree of $\mathbf{x}_i$ belonging to cluster k. The parameter h is a regularized constant that is often set to 0.5 empirically or determined by cross-validation. For consequent learning, suppose parameters in the antecedent are determined, let

$$\mathbf{x}_e = (1, (\mathbf{x}_i)^T)^T, \tag{7.a}$$
$$\tilde{\mathbf{x}}_i^k = \tilde{\mu}^k(\mathbf{x}_i)\mathbf{x}_e, \tag{7.b}$$
$$\mathbf{x}_{gi} = ((\tilde{\mathbf{x}}_i^1)^T, (\tilde{\mathbf{x}}_i^2)^T, \ldots, (\tilde{\mathbf{x}}_i^K)^T)^T, \tag{7.c}$$
$$\mathbf{p}^k = (p_0^k, p_1^k, \ldots, p_d^k)^T, \tag{7.d}$$
$$\mathbf{p}_g = ((\mathbf{p}^1)^T, (\mathbf{p}^2)^T, \ldots, (\mathbf{p}^K)^T)^T, \tag{7.e}$$

then the decision result of 1-TSK-FS can be rewritten as

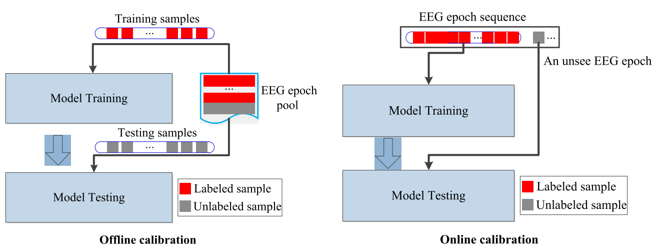
$$y^o(\mathbf{x}_i) = \mathbf{p}_g^T \mathbf{x}_{gi}. \tag{8}$$

Observing from (8), we see that solving $\mathbf{p}_g$ is obviously a linear regression problem [27]–[30]. Therefore, different criteria can generate different solutions. According to [31], $\mathbf{p}_g$ can be solved from the following objective function,

$$J_{1-TSK-FS}(\mathbf{p}_g) = \frac{1}{2}(\mathbf{p}_{g,c})^T \mathbf{p}_{g,c} + \frac{\eta}{2} \sum_{i=1}^{N} \left\| (\mathbf{p}_g)^T \mathbf{x}_{gi} - y_i \right\|^2, \tag{9}$$

Algorithm 1 The 1-TSK-FS Algorithm

Input: Training set $\mathbf{X} = \{\mathbf{x}_i, y_i\}_{i=1}^{N}$ with N labeled examples;
Number of fuzzy rules K;
Parameters η, h.
Output: The 1-TSK-FS model.
Use the FCM (the number of clusters is set to K) clustering algorithm to obtain the kernel centers and widthes for the fuzzy memebership fucntions by (5) and (6);
Use (7.a)~(7.c) to map the input vector $\mathbf{x}_i$ in the training set;
Compute $\mathbf{p}_g$ by (10).
Return $\mathbf{p}_g$ and hence $f(\mathbf{x}) = \mathbf{p}_g^T \mathbf{x}_g$ for prediction.



(a) Offline calibration (b) Online calibration

Fig. 2. Two kinds of calibration in BCI.

where $\frac{1}{2}(\mathbf{p}_g)^T \mathbf{p}_g$ is a regularization term that can improve the generalization ability of 1-TSK-FS for regression. $\mathbf{x}_{gi}$ denotes the input vector mapped from $\mathbf{x}_i$ through (7.c). $\sum_{i=1}^{N} \left\| (\mathbf{p}_g)^T \mathbf{x}_g - y_i \right\|^2$ is the error term, and $\eta > 0$ is used to control the the tolerance of errors and the complexity of 1-TSK-FS. By setting $\partial J_{1-TSK-FS}(\mathbf{p}_g)/\partial \mathbf{p}_g$ to 0, the optimal $\mathbf{p}_g$ can be analytically solved as

$$\mathbf{p}_g = \left(\mathbf{I}_{k(d+1)\times k(d+1)} + \sum_{i=1}^{N} \mathbf{x}_{gi} (\mathbf{x}_{gi})^T \right)^{-1} \times \left(\eta \sum_{i=1}^{N} \mathbf{x}_{gi} y_i \right). \tag{10}$$

The pseudocode for the complete 1-TSK-FS algorithm is described in Algorithm 1.

III. O-MV-T-TSK-FS for EEG-Based Driver Drowsiness Estimation

In [9], two scenarios of calibration in BCIs were defined, i.e., offline calibration and online calibration. Next, we briefly review the differences between them.

Fig.2 illustrates the two calibration scenarios. In Fig.2(a), i.e., offline calibration, a pool of unlabeled EEG epochs has been obtained. Experts are employed to label part of them manually. Then, the labeled EEG epochs are used to train a machine learning model to predict the labels of unlabeled epochs in the pool. Fig.2(b) illustrates the online calibration, where a regression model is trained by the labeled EEG epochs which are obtained on-the-fly.

Obviously, there are two main differences between offline calibration and online calibration. First, with regards to offline calibration, we can make use of labeled EEG epochs associating with unlabeled EEG epochs (semi-supervised learning) to train a machine learning model. However, as for online calibration, there are no unlabeled EEG epochs being used. Second, as for offline calibration, we can retrieval any EEG epoch in the pool for its label. However, as for online calibration, the sequence of the EEG epochs is often determined in advance and the experts have little control on which epochs to see next.

In this study, we only consider online calibration since that it meets the practical requirements of driver drowsiness estimation. Next, we first introduce the problem statement of this study, then present the online multi-view & transfer learning framework. Lastly, we design the corresponding objective function, give its optimization strategy, and describe the detailed algorithm steps.

A. Notations and Problem Statement for Driver Drowsiness Estimation

Suppose a domain in multi-view & transfer learning can be defined as $D = \{\chi^{(m)}, P(\mathbf{x}^{(m)})\}_{m=1}^{M}$, where $\chi^{(m)}$ and $P(\mathbf{x}^{(m)})$ represent a feature space and a marginal probability distribution of the m-th view, respectively, M is the number of views, and $\mathbf{x}^{(m)} \in \chi^{(m)}$. Two domains D_s and D_t derived from D are considered as being different if $\chi_t^{(m)} \neq \chi_s^{(m)}$, and/or $P_t(\mathbf{x}^{(m)}) \neq P_s(\mathbf{x}^{(m)})$ in each view.

Suppose a task in multi-view & transfer learning can be defined as $T = \{\Upsilon, P(y|\mathbf{x}^{(m)})\}_{m=1}^{M}$, where Υ and $P(y|\mathbf{x}^{(m)})$ represent an output feature space and a conditional probability distribution of the m-th view, respectively, and $y \in \Upsilon$. Two tasks T_s and T_t derived from T are considered as being different if $\Upsilon_t \neq \Upsilon_s$, or $P_t(y|\mathbf{x}^{(m)}) \neq P_s(y|\mathbf{x}^{(m)})$ in each view.

With the s-th source domain D_s containing N_s samples represented by $\{\mathbf{x}_i^{(m)}, y_i\}_{i=1}^{N_s}$ and a target domain D_t containing C calibration data (labeled samples) represented by $\{\mathbf{x}_j^{(m)}, y_j\}_{j=1}^{C}$, where $1 \leq m \leq M$, multi-view & transfer learning expects to learn a regression function $f(\mathbf{x}) : \mathbf{x} \mapsto y$ with the low expected error on the target domain D_t, under the assumption that $\chi_t^{(m)} = \chi_s^{(m)}$, $\Upsilon_t = \Upsilon_s$, $P_t(\mathbf{x}^{(m)}) = P_s(\mathbf{x}^{(m)})$ and $P_t(y|\mathbf{x}^{(m)}) = P_s(y|\mathbf{x}^{(m)})$.

When we estimate driver drowsiness, EEG epochs from a new subject are consdered as being in D_t, while EEG epochs from the s-th existing subject are considered as being in the s-th source domain. In D_t and D_s, although the features in the corresponding views are extracted in the same way, i.e., $\chi_t^{(m)} = \chi_s^{(m)}$, $m = 1, 2, .., M$, generally speaking, the corresponding marginal probability distribution and conditional probability distribution of each view are different, i.e., $P_t(\mathbf{x}^{(m)}) \neq P_s(\mathbf{x}^{(m)})$ and $P_t(y|\mathbf{x}^{(m)}) \neq P_s(y|\mathbf{x}^{(m)})$ since different subjects have similar but non-identical drowsy neural responses. Therefore, samples from D_s cannot accurately represent samples in D_t, and must be integrated with some labeled samples from D_t to induce the regression function for D_t.

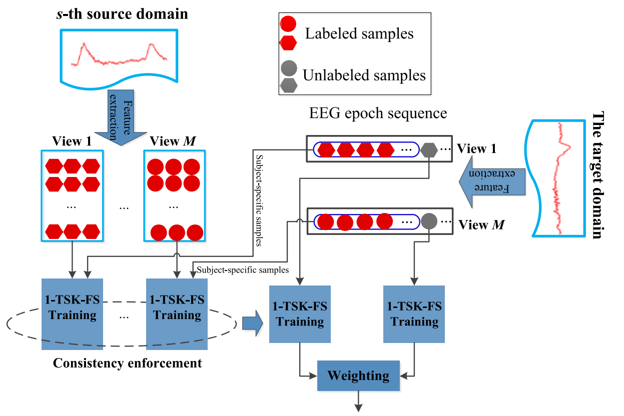


Fig. 3. Framework of online multi-view & transfer learning.

B. Framework of Multi-View & Transfer Learning

In order to combine multi-view learning and transfer learning, hence enforce the consistencies across different views and use the data from the source domains in the target domain, we design a new multi-view & transfer learning framework which takes 1-TSK-FS as the basic component, as shown in Fig.3.

In order to make use of the data from the s-th source domain in the target domain, based on the Bayesian principle that

$$f(\mathbf{x}^{(m)}) = P(y|\mathbf{x}^{(m)}) = \frac{P(\mathbf{x}^{(m)}, y)}{P(\mathbf{x}^{(m)})} = \frac{P(\mathbf{x}^{(m)}|y)P(y)}{P(\mathbf{x}^{(m)})}, \quad (11)$$

the dissimilarity between the marginal probability distributions of each view in the two domains $d(P_t(\mathbf{x}^{(m)}), P_s(\mathbf{x}^{(m)}))$ and the dissimilarity between the conditional probability distributions of each view in two domains $d(P_t(\mathbf{x}^{(m)}|y), P_s(\mathbf{x}^{(m)}|y))$ should be minimized to guarantee that $P_t(\mathbf{x}^{(m)})$ is similar to $P_s(\mathbf{x}^{(m)})$, and $P_t(\mathbf{x}^{(m)}|y)$ is similar to $P_s(\mathbf{x}^{(m)}|y)$.

Additionally, for multi-view learning, it is expected that the distribution of y under $\mathbf{x}^{(m)}$, i.e., the output of $\mathbf{x}^{(m)}$ in the m-th view is infinitely close to those in other views. Therefore, in order to enforce the consistencies across different views, we should minimize $\upsilon(P_s(y|\mathbf{x}^{(m)}), \frac{1}{M-1}\sum_{l=1,l\neq m}^{M} \tilde{P}_s(y|\mathbf{x}^{(m)}))$ in the source domain and $\upsilon(P_t(y|\mathbf{x}^{(m)}), \frac{1}{M-1}\sum_{l=1,l\neq m}^{M} \tilde{P}_t(y|\mathbf{x}^{(m)}))$ in the target domain, respectively. Here, $\frac{1}{M-1}\sum_{l=1,l\neq m}^{M} \tilde{P}_s(y|\mathbf{x}^{(m)})$ is the mean of the prior distribution of y under $\mathbf{x}^{(m)}$ of all views except the m-th view in the source domain. Homoplastically, $\frac{1}{M-1}\sum_{l=1,l\neq m}^{M} \tilde{P}_t(y|\mathbf{x}^{(m)})$ is the mean of the prior distribution of y under $\mathbf{x}^{(m)}$ of all views except the m-th view in the target domain, and υ a sum of squared errors (SSE) function.

Therefore, the multi-view & transfer learning framework can be formulated as

$$\begin{aligned} f =& \arg\min_{f} \sum_{m=1}^{M} \ell(\Upsilon_s, P_s(y|\mathbf{x}^{(m)})) + \sigma_t \sum_{m=1}^{M} \ell(\Upsilon_t, P_t(y|\mathbf{x}^{(m)})) \\ &+ \lambda_1 \left[\begin{aligned} &\sum_{m=1}^{M} \upsilon(P_s(y|\mathbf{x}^{(m)}), \frac{1}{M-1} \sum_{l=1,l\neq m}^{M} \tilde{P}_s(y|\mathbf{x}^{(m)})) \\ &+ \sigma_t \sum_{m=1}^{M} \upsilon(P_t(y|\mathbf{x}^{(m)}), \frac{1}{M-1} \sum_{l=1,l\neq m}^{M} \tilde{P}_t(y|\mathbf{x}^{(m)})) \end{aligned} \right] \\ &+ \lambda_2 \sum_{m=1}^{M} \Big[d(P_t(\mathbf{x}^{(m)}), P_s(\mathbf{x}^{(m)})) \\ &+ d(P_t(\mathbf{x}^{(m)}|y), P_s(\mathbf{x}^{(m)}|y)) \Big] \\ &- \lambda_3 \sum_{m=1}^{M} \tilde{r}^2(\Upsilon, P(y|\mathbf{x}^{(m)})), \end{aligned} \quad (12)$$

where λ_1, λ_2, and λ_3 are three positive regularization parameters, and σ_t is a positive overall weighting parameter used to control the contribution of calibration data in the target domain.

In (12), the first two terms are used to minimize the sum of squared errors in the source domain and the target domain, respectively. Here, ℓ is an SSE function (such as the hinge loss [32]). The third term is used to enforce the consistencies across different views in the two domains. The fourth term is used to minimize the dissimilarity between the marginal probability distributions of each view and the dissimilarity between the conditional probability distributions of each view in the two domains. The last term in (12) is used to maximize the approximate sample Pearson's correlation coefficient between outputs and input vectors in each view, which can avoid the prediction being (nearly) a constant.

Next, each term in (12) will be explained and computed in detail.

1) Sum of Squared Error Minimization: With regards to the problem stated in Section III.A, the first two terms in (12) can be re-written as

$$\sum_{m=1}^{M}\sum_{i=1}^{N_s} \left\| y_i - f(\mathbf{x}_i^{(m)}) \right\|^2 + \sigma_t \sum_{m=1}^{M}\sum_{j=1}^{C} \left\| y_j - f(\mathbf{x}_j^{(m)}) \right\|^2 \quad (13)$$

In (13), $f(\cdot)$ is the regression function which can be expressed as $f(\mathbf{x}_i^{(m)}) = (\mathbf{p}_g^{(m)})^T \mathbf{x}_{gi}^{(m)}$ in the basic regression component 1-TSK-FS, where $\mathbf{x}_{gi}^{(m)}$ is the input vector mapped from $\mathbf{x}_i^{(m)}$ in the m-th view through (7.c). $\mathbf{p}_g^{(m)}$ is the regression parameter, i.e., the consequent parameter which needs to be found.

For easy minimization, let

$$\mathbf{X}^{(m)} = [\mathbf{x}_{g1}^{(m)}, \ldots, \mathbf{x}_{gN_s}^{(m)}, \ldots, \mathbf{x}_{g(N_s+C)}^{(m)}]^T \quad (14)$$

and

$$\mathbf{y} = [y_1, \ldots, y_{N_s}, \ldots, y_{N_s+C}]^T, \quad (15)$$

where the first N_s elements in $\mathbf{X}^{(m)}$ and $\mathbf{y}$ denote the column inputs and the corresponding outputs of the m-th view in the source domain, the next C elements denote the column inputs and the corresponding outputs of the m-th view in the target domain. Additionally, we define a $N_s + N_c$ diagonal matrix $\mathbf{\Theta}$ as

$$\mathbf{\Theta}(i,i)=\begin{cases}1 & 1\le i\le N_s\\ \sigma_t & N_s+1\le i\le N_s+C.\end{cases}\tag{16}$$

As a result, (13) can be simplified into

$$\begin{aligned}&\sum_{m=1}^{M}\left[(\sum_{i=1}^{N_s}\left\|y_i-(\mathbf{p}_g^{(m)})^T\mathbf{x}_{gi}^{(m)}\right\|^2)\right.\\ &\qquad\left.+\sigma_t(\sum_{j=1}^{N_c}\left\|y_j-(\mathbf{p}_g^{(m)})^T\mathbf{x}_{gj}^{(m)}\right\|^2)\right]\\ &==\sum_{m=1}^{M}\left(\sum_{i=1}^{N_s+C}\mathbf{\Theta}(i,i)\left\|y_i-(\mathbf{p}_g^{(m)})^T\mathbf{x}_{gi}^{(m)}\right\|^2\right)\\ &=\sum_{m=1}^{M}\left[(\mathbf{y}^T-(\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T)\mathbf{\Theta}(\mathbf{y}-(\mathbf{X}^{(m)})\mathbf{p}_g^{(m)})\right]\end{aligned}\tag{17}$$

We find that σ_t plays an important role to the regression accuracy. In our model, the setting of σ_t obeys the following guideline,

$$\sigma_t=\max(2,\delta\cdot C/N_s).\tag{18}$$

where δ is a positive adjustable parameter. This guideline is deduced from the following three heuristics.

1) When the size of the target domain (C) is small, all samples in the target domain should be assigned a large weight so as to avoid being overwhelmed by samples in the source domain.

2) As the size of the target domain increases, the weight assigned to samples in the target domain should decrease gradually so as to avoid samples in the source domain being not overwhelmed by samples in the target domain.

3) Samples in the target domain should always be assigned larger weights than samples in the source domain due to that our proposed regression model will be eventually used for the target domain prediction.

2) Multi-View Collaborative Learning: For an unseen sample in a domain, it is expected that its prediction result in each view should be consistent as far as possible. To this end, the third term in (12) can be further expanded as,

$$\begin{aligned}&\sum_{m=1}^{M}\sum_{i=1}^{N_s}\left\|f(\mathbf{x}_i^{(m)})-\frac{1}{M-1}\sum_{l=1,l\neq m}^{M}\tilde{f}(\mathbf{x}_i^{(l)})\right\|^2\\ &\quad+\sigma_t\sum_{m=1}^{M}\sum_{j=1}^{C}\left\|f(\mathbf{x}_j^{(m)})-\frac{1}{M-1}\sum_{l=1,l\neq m}^{M}\tilde{f}(\mathbf{x}_j^{(l)})\right\|^2,\end{aligned}\tag{19}$$

which measures the sum of output squared errors between the current view and the remnant views both in the source domain and the target domain, respectively. Here, please note that $\tilde{f}(\mathbf{x}_i^{(l)})$ can be computed by $\tilde{\mathbf{p}}_g^T\mathbf{x}_{gi}^{(l)}$ based on (8), where $\tilde{\mathbf{p}}_g^T$ is considered as the *prior knowledge* which can be obtained before the beginning of multi-view learning in l-th view. With the diagonal matrix $\mathbf{\Theta}$ defined in (16), (19) can be simplified into

$$\sum_{m=1}^{M}\left[\begin{aligned}&((\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T-\frac{1}{M-1}\sum_{l=1,l\neq m}^{M}(\tilde{\mathbf{p}}_g^{(l)})^T(\mathbf{X}^{(l)})^T)\\ &\times\mathbf{\Theta}(\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}-\frac{1}{M-1}\sum_{l=1,l\neq m}^{M}\mathbf{X}^{(l)}\tilde{\mathbf{p}}_g^{(l)})\end{aligned}\right].\tag{20}$$

By minimizing (20), multi-view collaborative learning can be achieved such that the consistent prediction results of all views are able to be expected.

3) Probability Distribution Adaptation-Based Transfer Learning: With the maximum mean discrepancy (MMD) used in [33], [34], $d(P_t(\mathbf{x}^{(m)}),P_s(\mathbf{x}^{(m)})$ in the fourth term in (12) can be calculated by

$$\begin{aligned}&d(P_t(\mathbf{x}^{(m)}),P_s(\mathbf{x}^{(m)})\\ &\quad=[\frac{1}{N_s}\sum_{i=1}^{N_s}f(\mathbf{x}_i^{(m)})-\frac{1}{C}\sum_{j=1}^{N_c}f(\mathbf{x}_j^{(m)})]^2.\\ &\quad=(\mathbf{p}_g^{(m)})^T\mathbf{X}^{(m)}\mathbf{\Phi}X^{(m)}(\mathbf{p}_g^{(m)}),\end{aligned}\tag{21}$$

where $\mathbf{\Phi}$ is the MMD matrix defined as

$$\mathbf{\Phi}=\begin{cases}1/N_s^2, & 1\le i\le N_s,1\le j\le N_s\\ 1/C^2, & N_s+1\le i,j\le N_s+C\\ -1/N_sC & \text{otherwise.}\end{cases}\tag{22}$$

In [10], Wu *et al.*, transform the conditional probability distribution adaptation for classification tasks into regression tasks based on fuzzy set theory [43]. As a result, based on [9], the dissimilarity $d(P_t(\mathbf{x}^{(m)}|y),P_s(\mathbf{x}^{(m)}|y))$ can be solved by

$$\begin{aligned}&d(P_t(\mathbf{x}^{(m)}|y),P_s(\mathbf{x}^{(m)}|y))\\ &\quad=\sum_{c=1}^{3}[\sum_{\mathbf{x}_i^{(m)}\in D_s}\bar{\mu}_{ic.s}^{(m)}f(\mathbf{x}_i^{(m)})-\sum_{\mathbf{x}_j^{(m)}\in D_t}\bar{\mu}_{ic.t}^{(m)}f(\mathbf{x}_j^{(m)})]^2\\ &\quad=\sum_{c=1}^{3}(\mathbf{p}_g^{(m)})^T\mathbf{X}^{(m)}M_c^{(m)}\mathbf{X}^{(m)}(\mathbf{p}_g^{(m)})\\ &\quad=(\mathbf{p}_g^{(m)})^T\mathbf{X}^{(m)}\mathbf{\Omega}^{(m)}\mathbf{X}^{(m)}(\mathbf{p}_g^{(m)}),\end{aligned}\tag{23}$$

where $\mathbf{\Omega}^{(m)}=\sum_{c=1}^{3}\mathbf{\Omega}_c^{(m)}$ in which $\mathbf{\Omega}_c^{(m)}$ is the MMD matrix defined as

$$\mathbf{\Omega}_c^{(m)}(i,j)=\begin{cases}\bar{\mu}_{ic.s}^{(m)}\bar{\mu}_{jc.s}^{(m)} & \mathbf{x}_i^{(m)},\mathbf{x}_j^{(m)}\in D_{s,c}\\ \bar{\mu}_{ic.t}^{(m)}\bar{\mu}_{jc.t}^{(m)} & \mathbf{x}_i^{(m)},\mathbf{x}_j^{(m)}\in D_{t,c}\\ -\bar{\mu}_{ic.s}^{(m)}\bar{\mu}_{jc.t}^{(m)} & \mathbf{x}_i^{(m)}\in D_{s,c},\mathbf{x}_j^{(m)}\in D_{t,c}\\ -\bar{\mu}_{ic.t}^{(m)}\bar{\mu}_{jc.s}^{(m)} & \mathbf{x}_i^{(m)}\in D_{t,c},\mathbf{x}_j^{(m)}\in D_{s,c}\\ 0 & \text{otherwise}\end{cases}\tag{24}$$

In (24), $D_{s,c}$ and $D_{t,c}$ are derived from D_s and D_t based on the fuzzy theory, where $c=1,2,3$. Here, c represents one of three classes corresponding to three triangular fuzzy sets, Small^s, Medium^s, and Large^s we defined. $\bar{\mu}_{ic.s}^{(m)}=\mu_{ic.s}^{(m)}/\sum_{i=1}^{N_s}\mu_{ic.s}^{(m)}$ where $\mu_{ic.s}^{(m)}$ represents the fuzzy membership degree of the output of $\mathbf{x}_i^{(m)}$ from the s-th source domain in Class c.

4) Maximize the Pearson Correlation Coefficient: Based on [44], the classic Pearson correlation coefficient (PCC) $r(\Upsilon, P(y|\mathbf{x}^{(m)}))$ of each view is defined as the following form,

$$r(\Upsilon, P(y|\mathbf{x}^{(m)})) = \frac{\mathbf{y}\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}}{\|\mathbf{y}\| \cdot \left\|\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}\right\|} = \frac{\mathbf{y}\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}}{\sqrt{\mathbf{y}^T\mathbf{y}} \cdot \sqrt{(\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}}}, \tag{25}$$

and hence,

$$r^2(\Upsilon, P(y|\mathbf{x}^{(m)})) = \frac{(\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T\mathbf{y}\mathbf{y}^T\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}}{(\mathbf{y}^T\mathbf{y}) \cdot (\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}}. \tag{26}$$

In (26), since the classic $r^2(\Upsilon, P(y|\mathbf{x}^{(m)}))$ has $\mathbf{p}_g^{(m)}$ in the denominator, it is almost impossible for us to find a closed-form solution to maximize it. Fortunately, we observe that $r^2(\Upsilon, P(y|\mathbf{x}^{(m)}))$ increases as the numerator increase, and decreases as $(\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}$ increases. Therefore, it is expected to maximize $\tilde{r}^2(\Upsilon, P(y|\mathbf{x}^{(m)}))$ instead of maximizing $r^2(\Upsilon, P(y|\mathbf{x}^{(m)}))$ to obtain a closed-form solution,

$$\begin{aligned}\tilde{r}^2(\Upsilon, P(y|\mathbf{x}^{(m)})) &= \frac{(\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T\mathbf{y}\mathbf{y}^T\mathbf{X}^{(m)}\mathbf{p}_g^{(m)} - (\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}}{\mathbf{y}^T\mathbf{y}} \\ &= \frac{(\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T(\mathbf{y}\mathbf{y}^T - \mathbf{I})\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}}{\mathbf{y}^T\mathbf{y}},\end{aligned} \tag{27}$$

where $\mathbf{I}$ is a $(N_s + C)$ by $(N_s + C)$ identity matrix. We see that $\tilde{r}^2(\Upsilon, P(y|\mathbf{x}^{(m)}))$ in (27) follows the same property $r^2(\Upsilon, P(y|\mathbf{x}^{(m)}))$ has.

C. Solution to O-MV-T-TSK-FS

By substituting (17), (20), (21), (23) and (27) into (12), we can obtain the objective function of O-MV-T-TSK-FS,

$$\begin{aligned}J(\mathbf{p}_g^{(m)}) &= \arg\min_{\mathbf{p}_g^{(m)}} \\ &\times \left\{\begin{aligned}&\sum_{m=1}^{M}\left[(\mathbf{y}^T - (\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T)\mathbf{\Theta}(\mathbf{y} - (\mathbf{X}^{(m)})\mathbf{p}_g^{(m)})\right] \\ &+ \lambda_1\sum_{m=1}^{M}\left[\begin{aligned}&((\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T \\ &- \frac{1}{M-1}\sum_{l=1,l\neq m}^{M}(\tilde{\mathbf{p}}_g^{(l)})^T(\mathbf{X}^{(l)})^T) \\ &\times \mathbf{\Theta}(\mathbf{X}^{(m)}\mathbf{p}_g^{(m)} - \frac{1}{M-1}\sum_{l=1,l\neq m}^{M}\mathbf{X}^{(l)}\tilde{\mathbf{p}}_g^{(l)})\end{aligned}\right]. \\ &+ \lambda_2\sum_{m=1}^{M}\left[\begin{aligned}&(\mathbf{p}_g^{(m)})^T\mathbf{X}^{(m)}\mathbf{\Phi}\mathbf{X}^{(m)}(\mathbf{p}_g^{(m)}) \\ &+ (\mathbf{p}_g^{(m)})^T\mathbf{X}^{(m)}\mathbf{\Omega}^{(m)}\mathbf{X}^{(m)}(\mathbf{p}_g^{(m)})\end{aligned}\right] \\ &- \lambda_3\sum_{m=1}^{M}\frac{(\mathbf{p}_g^{(m)})^T(\mathbf{X}^{(m)})^T(\mathbf{y}\mathbf{y}^T - \mathbf{I})\mathbf{X}^{(m)}\mathbf{p}_g^{(m)}}{\mathbf{y}^T\mathbf{y}}\end{aligned}\right\}\end{aligned} \tag{28}$$

By setting $\partial J(\mathbf{p}_g^{(m)})/\partial\mathbf{p}_g^{(m)} = 0$, the solution in terms of $\mathbf{p}_g^{(m)}$ to O-MV-T-TSK-FS appears, i.e.,

$$\begin{aligned}\mathbf{p}_g^{(m)} &= \left[(\mathbf{X}^{(m)})^T\left(2\mathbf{\Theta} + \lambda_2\mathbf{\Phi} + \lambda_2\mathbf{\Omega}^{(m)} + \lambda_3\frac{\mathbf{I} - \mathbf{y}\mathbf{y}^T}{\mathbf{y}^T\mathbf{y}}\right)\mathbf{X}^{(m)}\right]^{-1} \\ &\times \left(\mathbf{X}^{(m)}\mathbf{\Theta}\boldsymbol{y} + \mathbf{X}^{(m)}\mathbf{\Theta}\frac{1}{M-1}\sum_{l=1,l\neq m}^{M}\mathbf{X}^{(l)}\tilde{\mathbf{p}}_g^{(l)}\right)\end{aligned} \tag{29}$$

We use $\mathbf{p}_g^{(m)}$ to estimate all training samples in m-th view and record the corresponding RMSE as $\rho^{(m)}$. Then a global regression function is constructed as a weighted average of the functions for all views, i.e.,

$$f(\mathbf{x}) = \sum_{m=1}^{M}(1/\rho^{(m)})(\mathbf{p}_g^{(m)})^T\mathbf{x}_g^{(m)} / \sum_{m=1}^{M}(1/\rho^{(m)}). \tag{30}$$

D. Algorithm of O-MV-T-TSK-FS

Firstly, O-MV-T-TSK-FS is separately trained on each source domain so as to generate S models. Secondly, each model is taken as a basic component to construct the final regression model by weighting strategy. The weight assigned to each basic component is set to the reciprocal of the training regression accuracy of the corresponding basic component. The obtained final regression model is used to predict the unlabeled EEG epochs. The pseudocode of O-MV-T-TSK-FS is given in Algorithm 2.

IV. Experimental Results

In this section, extensive experimental results about drowsiness estimation are reported to demonstrate the performance and interpretability of the proposed online multi-view & transfer fuzzy system O-MV-T-TSK-FS.

A. Setup

The experiment setup basically followed our previous work in [10]. The data collected from the 16 individuals had different sizes since the lane-departure events were presented very 5-10s randomly. In the following experiments, we used data from 15 subjects and discarded one subject whose data were recorded incorrectly. Although all data were collected offline and the corresponding drowsiness indices (labels) of all EEG epochs were known, we used the following steps (shown in Fig.4) to simulate an incremental online calibration scenario to evaluate the proposed fuzzy system. For simplicity, we do not illustrate the multi-view learning in Fig.4.

1) We labeled all EEG epochs from 14 subjects and took each subject as a source domain.

2) We iteratively and sequentially generated C labeled samples from the 15-th subject as the labeled subject-specific samples and gradually added them to the target domain on-the-fly. To be more specific, for fair studies, a random integer m_0 belonging to $[1, C]$ was first generated to denote the starting position. Then, we gradually added labeled epochs $m_0, m_0 + 1, \ldots, m_0+C-1$ which were taken as subject-specific samples

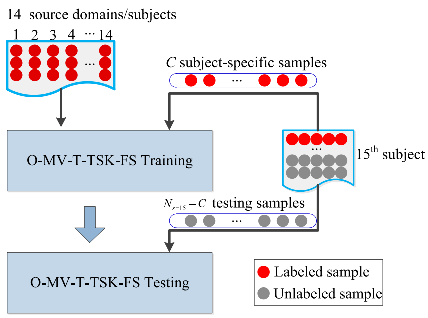


Fig. 4. Online calibration scenario.

Algorithm 2 The O-MV-T-TSK-FS Algorithm

Input: S source domains, and the s-th ($s = 1, 2, \ldots, S$) source domain contains N_s samples $\{\mathbf{x}_i, y_i\}_{i=1}^{N_s}$, and each sample $\mathbf{x}_i$ can be represented from M views, i.e., $\{\mathbf{x}_i^{(m)}\}_{m=1}^{M}$;
C target domain samples and each sample can also be represented from M views.
Parameters λ_1, λ_2, λ_3 in (12) and in δ (17).

Output: The regression function in O-MV-T-TSK-FS.

For $s = 1$ to S
  Use **Algorithm 1** to compute the prior knowledge $\tilde{\mathbf{p}}_g^{(m)}$ of each view.
  For $m = 1$ to M
    Construct $\mathbf{X}^{(m)}$ in (14), $\mathbf{y}$ in (15), $\boldsymbol{\Theta}$ in (16), $\boldsymbol{\Phi}$ in (22), $\boldsymbol{\Omega}^{(m)}$ in (24);
    Compute $\mathbf{p}_g^{(m)}$ by (29);
  End For
  Use (30) to estimate the outputs denoted as $f_s(\mathbf{x})$ for the $N_s + C$ samples from S source domains and the target domain, and record RMSE as ϖ_s;
  Assign $\omega_s = 1/\varpi_s$ as the weight to the s-th regression model;
End

Return $f(\mathbf{x}) = \dfrac{\sum_{s=1}^{S} \omega_s f_s(\mathbf{x})}{\sum_{s=1}^{S} \omega_s}$ for prediction.

to the training set for O-MV-T-TSK-FS training in the first iteration, and hence we obtained different trained regression models. Finally, O-MV-T-TSK-FS testing on the remaining $N_{s=15} - C$ epochs was carried out. During the second iteration, subject-specific samples $m_0 + C, m_0 + C + 1, \ldots, m_0 + 2C - 1$ were added to the training set to train different regression models, then their performance was verified on the remaining $N_{s=15} - 2C$ epochs. The iterations did not stop until the iteration number reached its maximum. In order to make the performance being statistically meaningful, we repeated the process 10 times, each time with a random starting position m_0.

To make sure each subject had a chance to become the "15-th" subject, the whole process shown in Fig.4 was repeated 15 times.

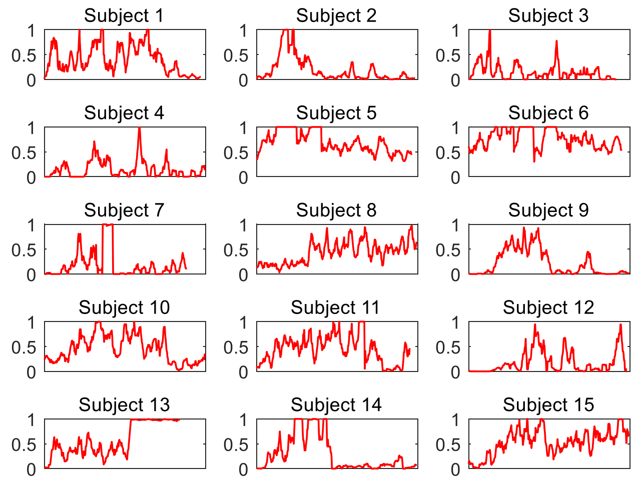


Fig. 5. Drowsiness index for each subject.

B. Performance Evaluation

In our experiments, we adopted the RMSE between the true drowsiness indices and the corresponding estimated values to evaluate all regression models, including the baselines.

The drowsiness index y for performance evaluation in our experiments can be defined as a function with regards to the response time to an event:

$$y = \max\left\{0, \frac{1 - e^{-(\tau - \tau_0)}}{1 + e^{-(\tau - \tau_0)}}\right\}, \tag{31}$$

where according to [8], τ_0 was set to 1. We employ a 90-s square moving-average window to smooth the drowsiness indices and hence reduce the variations. The smoothing did not reduce the sensitivity of the drowsiness index since the cycle lengths of drowsiness fluctuations are longer than 4 min [35]. Fig.5 illustrates the smoothed drowsiness fluctuations for the 15 subjects. From Fig.5, we observe that there are some drowsiness indices in each subject at or close to 1, which indicates drowsy driving.

C. Multi-View Scenario

In [4] and [9], Wu *et al.* respectively extracted the theta band powers and principal component features from EEG signals for driver drowsiness estimation since both kinds of features have strong correlations with the drowsiness index. Here, in our experiments, we united the two kinds of features to train O-MV-T-TSK-FS synergistically.

To be specific, we first epoched 30-s EEG signals right before each sample point, and employed Welch's method [36] to calculate the average power spectral density (PSD) in the theta band (4-7.5 Hz) for each channel. We considered the 30 theta band powers as one view in our experiments. Then, the 30 theta band powers were converted to dBs, which were considered as the second view. During the conversion, some channels whose maximum dBs were larger than 20 were removed in order to filter bad channel readings and noise. The dBs of each channel were then normalized to mean zero and standard deviation one.

TABLE I
PARAMETER SETTINGS OF ALL BECHMARKING MODELS

Models	Search ranges
1-TSK-FS	The number of fuzzy rules $K \in \{5,10,15,...,30\}$; The regularization parameter $\eta_{\mathbf{p}_g} \in \{10^{-3},10^{-2},...,10^{3}\}$.
MV-TSK-FS	The number of fuzzy rules $K \in \{5,10,15,...,30\}$; The regularization parameters $\lambda^{\mathbf{p}_g} \in \{10^{-3},10^{-2},...,10^{3}\}$, $\lambda^{co} \in \{10^{-3},10^{-2},...,10^{3}\}$ and $\lambda^{w} \in \{10^{-3},10^{-2},...,10^{3}\}$.
O-MV-T-TSK-FS	The number of fuzzy rules $K \in \{5,10,15,...,30\}$; The regularization parameters $\lambda_1 \in \{1,2,...,50\}$, $\lambda_2 \in \{1,2,...,50\}$, $\lambda_3 \in \{0.01,0.02,...,2\}$; The parameter δ in (18) $\delta \in \{0.1,0.2,...,1\}$;

D. Baseline Models

Four baseline regression models were introduced for comparative studies to highlight the characteristics of the proposed fuzzy system O-MV-T-TSK-FS.

1) Baseline 1 (BL1), which was a 1-order TSK fuzzy system trained on 14 existing subjects, and then tested on the new subject. That's, BL1 was taken as a subject-independent regression model which ignored subject-specific samples from the new subject completely.

2) Baseline 2 (BL2), which was also a 1-order TSK fuzzy system but trained only on subject-specific samples from the new subject and then tested on the remaining samples from the new subject. That's, BL2 did not depend on any existing subjects.

3) DAMF [9], which trained 14 1-order TSK fuzzy systems as basic component (ridge regression models in the original reference [9]) by making use of data from each existing subject associating with data from the new subject, respectively, and then the final DAMF is constructed as a weighted average of all basic components. The weights assigned to the basic components are set as the same as the proposed model.

4) MV-TSK-FS [37], which combined data from all 14 existing subjects, trained a multi-view TSK fuzzy system, and applied it to the new subject.

All parameters in the baseline models were determined by 5-fold cross-validation. Searching ranges of all parameters are given in Table I.

E. Performance Discussion

Since BL1, BL2 and DAMF are three single-view regression models, we trained them on each view, respectively, and then report the average performance on two views. Fig.6 shows the average performance in terms of RMSE of all regression models across 15 subjects, and Fig.7 the RMSE of each subject. Overall, O-MV-T-TSK-FS achieved the best performance among all models. In more details,

1) Among all baseline models, BL2 performed the worst, because it cannot learn enough pattern information by using a small size of subject-specific samples. Moreover, we should note that, when there are no subject-specific samples at all, BL2 can no longer be trained, whereas other models can be trained by using data from other subjects.

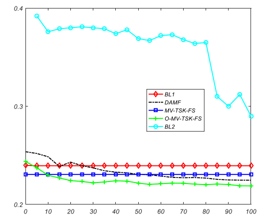


Fig. 6. Average performance in terms of RMSE of all regression models across 15 subjects.

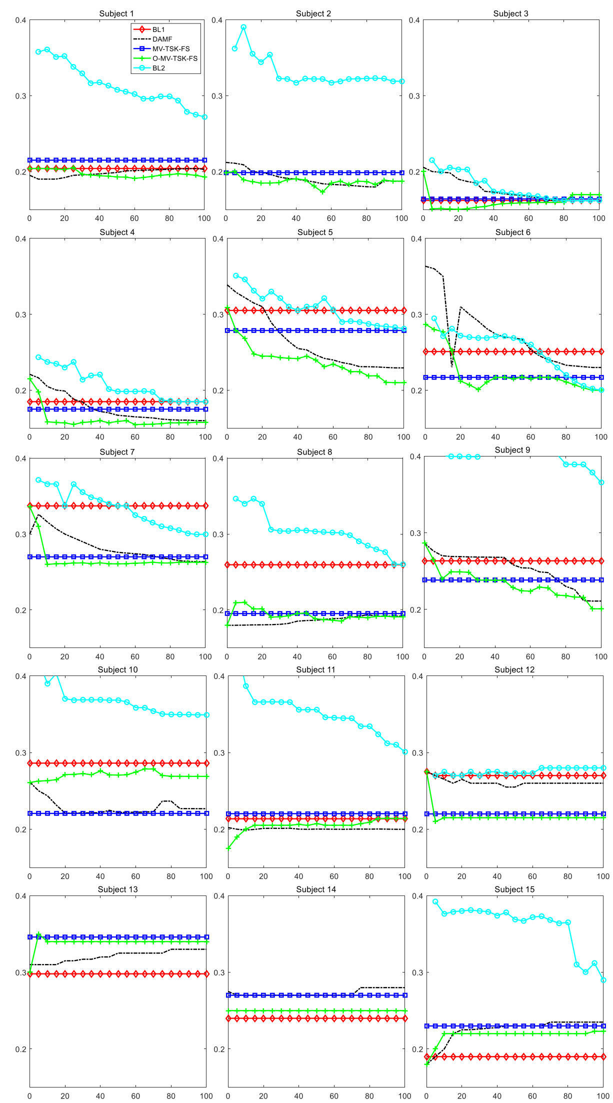


Fig. 7. RMSE of each subject.

2) Since MV-TSK-FS and BL1 built their models without subject-specific samples, their performances were not dependent on C. As for DAMF and O-MV-TSK-FS, it is intuitive that when C increased, the RMSE decreased.

3) MV-TSK-FS outperformed BL1, and O-MV-TSK-FS always had smaller RMSE than DAMF. This indicates that there existed correlations between the theta band powers and the principal component features, and hence multi-view learning across the two views is desirable, which can capture

TABLE II

FUZZY RULES TRAINED ON THE THETA BAND POWERS

Fuzzy rules-based O-MV-T-TSK-FS			
Fuzzy rules: If x_{i1} is $A_1^k \wedge x_{i2}$ is $A_2^k \wedge \ldots \wedge x_{id}$ is A_d^k, then $f^k(\mathbf{x}_i) = p_0^k + p_1^k x_{i1} + \ldots + p_d^k x_{id}, k = 1,2,\ldots,K$			
View	Number of fuzzy ruels	Antecedent parameters $\mathbf{c}^k = [c_1^k, c_2^k, \ldots, c_d^k]^T, \boldsymbol{\delta}^k = [\delta_1^k, \delta_2^k, \ldots, \delta_d^k]^T$	Consequent parameters $\mathbf{p}^k = [p_0^k, p_1^k, \ldots, p_d^k]^T$
The theta band powers	1	$\mathbf{c}^1 = [16.13, 15.96, 19.88, 17.01, -12.01, 36.32, \ldots]$, $\boldsymbol{\delta}^1 = [0.12, 0.12, 0.15, 0.12, 0.14, 0.18, \ldots]$	$\mathbf{p}^1 = [2.0034, -1.9821, 0.4215, 1.1124, 0.0023, 2.3501, \ldots]^T$
	2	$\mathbf{c}^2 = [-13.90, 26.75, 32.41, 2.18, 37.76, 16.31, \ldots]$, $\boldsymbol{\delta}^2 = [0.42, 0.58, 0.11, 0.55, 0.15, 0.17, \ldots]$	$\mathbf{p}^2 = [3.8724, 1.8250, 1.6781, -3.8721, 0.6511, 0.0034, \ldots]^T$
	3	$\mathbf{c}^3 = [11.94, 21.55, 2.78, 32.29, -22.51, -34.44, \ldots]$, $\boldsymbol{\delta}^3 = [0.52, 0.32, 0.36, 0.18, 0.55, 0.41, \ldots]$	$\mathbf{p}^3 = [0.0037, -0.8924, 2.6781, -1.8714, 0.9081, -0.1146, \ldots]^T$
	4	$\mathbf{c}^4 = [31.89, 20.14, -12.64, 37.11, 28.11, 16.46, \ldots]$, $\boldsymbol{\delta}^4 = [0.40, 0.18, 0.32, 0.21, 0.15, 0.17, \ldots]$	$\mathbf{p}^4 = [2.4671, 3.7621, 0.4462, -2.0342, 1.0891, -0.9910, \ldots]^T$
	5	$\mathbf{c}^5 = [15.91, 2.45, 27.46, 12.20, 24.41, 15.71, \ldots]$, $\boldsymbol{\delta}^5 = [0.44, 0.23, 0.43, 0.18, 0.25, 0.31, \ldots]$	$\mathbf{p}^5 = [1.7698, 2.9109, 0.3618, -1.9809, 1.2092, -0.8567, \ldots]^T$

more pattern information than learning separately on each view.

4) When $C = 0$, O-MV-TSK-FS and DAMF performed worse than MV-TSK-FS and BL1. However, when C increased, O-MV-TSK-FS and DAMF became superior in RMSEs, and the performance of O-MV-TSK-FS surpassed MV-TSK-FS and BL1 quickly. This indicates that the differences among the subjects were significant. Thus, it is not advisable to train a subject-independent regression model.

5) The performance of O-MV-TSK-FS gradually stabilized when the second batch of subject-specific samples was added. This suggests that O-MV-TSK-FS is very suitable for practical application because it only needs very few subject-specific samples during its training process.

In summary, O-MV-TSK-FS outperformed MV-TSK-FS, BL1 and BL2, which did not use TL, and also DAMF, which did not use multi-view learning. In other words, TL associating with multi-view learning are indeed beneficial for driver drowsiness estimation.

F. Interpretability Analysis

In our experiments, we know that the proposed fuzzy system contains two TSK fuzzy systems for two views, i.e., the theta band powers and the dBs, respectively. Here, we take the first view, i.e., the theta band powers as an example and analysis the interpretability of O-MV-T-TSK-FS on the first source domain. Table II gives the antecedent and consequent parameters learned in each fuzzy rule. Please note that since there are 30 features contained in the theta band powers, we only list the parameters for the first 6 features, i.e., C3, C4, CP3, CP4, CPZ and CZ. Fig.8 shows the fuzzy membership functions of all fuzzy sets trained in each fuzzy rule. Each fuzzy membership function in Fig.8 can be described from a fuzzy linguistic perspective. For example, "the theta band power C4 is "a litter high" in which "a litter high" can be replaced by "a litter low", "medium", or "high". This linguistic description of a fuzzy membership function may be different when being observed by different medical experts since that for a fuzzy rule, different experts have different interpretations.

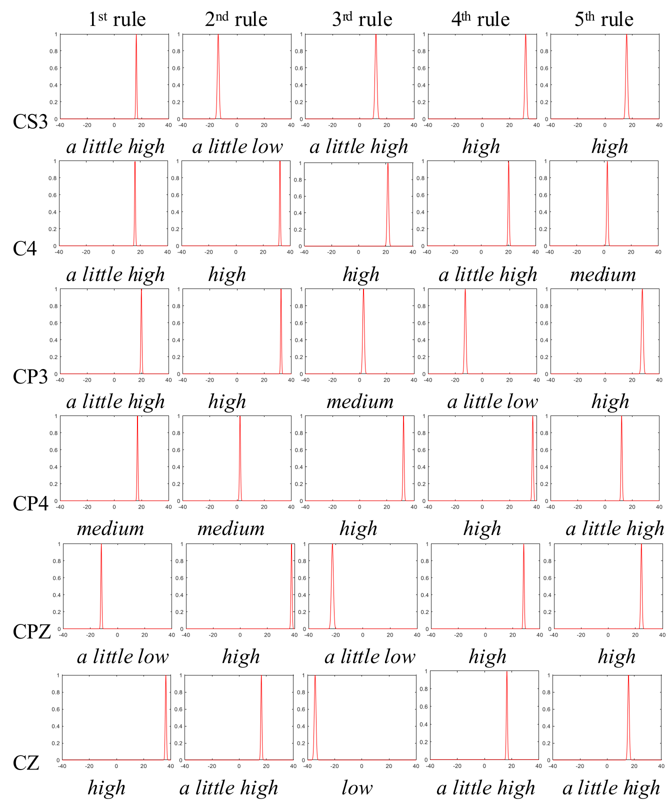


Fig. 8. Linguistic explanation of each fuzzy subset in the theta band powers.

Now, let us take the first row in Fig.8 as an example to further discuss the interpretability of the fuzzy rule. The antecedent parameters (c_1^k, δ_1^k) of the feature C3 in each fuzzy rule are (16.13, 0.12), (-13.90, 0.42), (11.94, 0.52), (31.89, 0.40), and (15.91, 0.44), respectively. Based on the antecedent parameters, the feature C3 can be presented by the corresponding 5 fuzzy membership functions which can be linguistically described by using "a little high", "a little low", "medium", "low", or "high" according to the centers of the functions. Finally, with the linguistic descriptions in the antecedent and the corresponding linear function in the consequent, respectively, we can obtain the 5 fuzzy rules

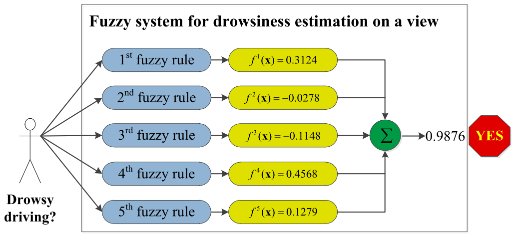


Fig. 9. Example of drowsiness estimation on a view.

for the view of the theta band powers. The first fuzzy rule is expressed as follows, and the remaining 4 rules can be described in a similar way.

The 1^{st}fuzzy rule:

If the theta band power C3 is "a litter high",
the theta band power C4 is "a litter high",
the theta band power CP3 is "a litter high",
the theta band power CP4 is "medium",
the theta band power CPZ is "a litter low",
the theta band power CZ is "high",

Then the regression (decision) value of this rule can be formulated as:

$$f^1(\mathbf{x}) = 2.0034 - 1.9821x_1 + 0.4215x_2 + 1.1124x_3 + 0.0023x_4 + 2.3501x_5 \ldots$$

With the trained fuzzy rules in the theta band powers, for an epoch in Fig.9, its regression value (prediction result) is 0.9876 which indicates drowsy driving. The fuzzy rules trained on the dBs view can be formulated in a similar way.

V. Conclusion

We recently proposed an offline transfer regression model for driver drowsiness estimation from EEG signals. It only needs few subject-specific calibration data to adjust itself for a new subject. In this paper, we have extended it to an online and multi-view setting. In the new model, features from EEG data both in the source domain and the target domain are represented from multiple perspectives and multi-view settings are injected into our previous TL framework to enforce the consistencies across different views. Moreover, the online training can make the proposed model competent for practical requirements.

Although the performance of the prosed model is promising, more future studies are also expected. For example, samples in the target domain are all assigned the same weight σ_t in the current objective function, which may not accurately capture the sample distribution. Maybe a mechanism that assigns a weight to each sample in the target domain can further improve the performance. Also, it is a hard task to fix 4 parameters simultaneously. In our future works, we will try our best to solve the above-mentioned limitations.

References

[1] (2011). Traffic Safety Facts Crash Stats: Drowsy Driving. U.S. Department of Transportation, National Highway Traffic Safety Administration, Washington, DC, USA. [Online]. Available: http://wwwnrd.nhtsa.dot.gov/pubs/811449.pdf

[2] B. Reddy, Y. Kim, S. Yun, C. Seo, and J. Jang, "Real-time driver drowsiness detection for embedded system using model compression of deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 438–445.

[3] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011.

[4] D. Wu, V. J. Lawhern, S. Gordon, B. J. Lance, and C. Lin, "Offline EEG-based driver drowsiness estimation using enhanced batch-mode active learning (EBMAL) for regression," in *Proc. IEEE Int. Conf. Syst., Man, (SMC)*, Budapest, Hungary, Oct. 2016, pp. 730–736.

[5] Y. Jiang *et al.*, "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2270–2284, Dec. 2017.

[6] Z. Deng, P. Xu, L. Xie, K.-S. Choi, and S. Wang, "Transductive joint-knowledge-transfer TSK FS for recognition of epileptic EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 8, pp. 1481–1494, Aug. 2018.

[7] 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 Trans. Fuzzy Syst.*, vol. 24, no. 5, pp. 1079–1094, Oct. 2016.

[8] C.-S. Wei *et al.*, "Selective transfer learning for EEG-based drowsiness detection," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Hong Kong, Oct. 2015, pp. 3229–3232.

[9] D. Wu, C.-H. Chuang, and C.-T. Lin, "Online driver's drowsiness estimation using domain adaptation with model fusion," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Xi'an, China, Sep. 2015, pp. 904–910.

[10] D. Wu, V. J. Lawhern, and B. J. Lance, "Reducing BCI calibration effort in RSVP tasks using online weighted adaptation regularization with source domain selection," in *Proc. Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Xi'an, China, 2015, pp. 567–573.

[11] J. Farquhar, D. Hardoon, H. Meng, J. Shawe-Taylor, and S. Szedmak, "Two view learning: SVM-2K, theory and practice," in *Proc. NIPS*, vol. 18, 2006, p. 355.

[12] T. Joachims and N. Cristianini, "Composite kernels for hypertext categorisation," in *Proc. ICML*, 2001, pp. 250–257.

[13] G. Li, S. C. H. Hoi, and K. Chang, "Two-view transductive support vector machines," in *Proc. SDM*, 2010, pp. 235–244.

[14] D. Rosenberg, V. Sindhwani, P. Bartlett, and P. Niyogi, "A kernel for semi-supervised learning with multi-view point cloud regularization," *IEEE Signal Process. Mag.*, to be published.

[15] V. Sindhwani and P. Niyogi, "A co-regularized approach to semi-supervised learning with multiple views," in *Proc. ICML Workshop Learn. Multiple Views*, 2005, pp. 74–79.

[16] D. Zhang, F. Wang, C. Zhang, and T. Li, "Multi-view local learning," in *Proc. AAAI*, 2008, pp. 752–757.

[17] T. Zhang, A. Popescul, and B. Dom, "Linear prediction models with graph regularization for Web-page categorization," in *Proc. SIGKDD*, 2006, pp. 821–826.

[18] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[19] V. Srinivasan, C. Eswaran, and A. 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, Dec. 2005.

[20] E. A. Vivaldi and A. Bassi, "Frequency domain analysis of sleep eeg for visualization and automated state detection," in *Proc. 28th Annu Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 3740–3743.

[21] G. C. Y. Fong *et al.*, "Childhood absence epilepsy with tonic-clonic seizures and electroencephalogram 3–4-Hz spike and multispike–slow wave complexes: Linkage to chromosome 8q24," *Amer. J. Hum. Genet.*, vol. 63, no. 4, pp. 1117–1129, Oct. 1998.

[22] Z. Zhang, H. Kawabata, and Z. Q. Liu, "EEG analysis using fast wavelet transform," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Nov. 2000, pp. 2959–2964.

[23] B. Litt *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] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2009.

[25] Z. Deng, Y. Jiang, F.-L. Chung, H. Ishibuchi, K.-S. Choi, and S. Wang, "Transfer prototype-based fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 5, pp. 1210–1232, Oct. 2016.

[26] K. Kiliç, Ö. Uncu, and I. B. Türksen, "Comparison of different strategies of utilizing fuzzy clustering in structure identification," *Inf. Sci.*, vol. 177, no. 23, pp. 5153–5162, Dec. 2007.

[27] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-15, no. 1, pp. 116–132, Jan./Feb. 1985.

[28] Z. Deng, K.-S. Choi, F.-L. Chung, and S. Wang, "Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 2, pp. 210–226, Apr. 2011.

[29] Y. Jiang, F.-L. Chung, H. Ishibuchi, Z. Deng, and S. Wang, "Multi-task TSK fuzzy system modeling by mining intertask common hidden structure," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 534–547, Mar. 2015.

[30] Y. Jiang, Z. Deng, F.-L. Chung, and S. Wang, "Multi-task TSK fuzzy system modeling using inter-task correlation information," *Inf. Sci.*, vol. 298, pp. 512–533, Mar. 2015.

[31] Z. Deng, K.-S. 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 Trans. Cybern.*, vol. 44, no. 12, pp. 2585–2599, Dec. 2014.

[32] B. Scholkopf and A. Smola, *Learning With Kernels*. Cambridge, MA, USA: MIT Press, 2002.

[33] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation regularization: A general framework for transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1076–1089, May 2014.

[34] B. Quanz and J. Huan, "Large margin transductive transfer learning," in *Proc. 18th ACM Conf. Inf. Knowl. Manage.*, Hong Kong, Nov. 2009, pp. 1327–1336.

[35] S. Makeig and M. Inlow, "Lapses in alertness: Coherence of fluctuations in performance and EEG spectrum," *Electroencephalogr. Clin. Neurophysiol.*, vol. 86, pp. 23–35, Feb. 1993.

[36] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, no. 2, pp. 70–73, Jun. 1967.

[37] Y. Jiang *et al.*, "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 3–20, Feb. 2017.

[38] R. Zhang, Y. Li, and X. Li, "Topology inference with network tomography based on t-test," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 921–924, Jun. 2014.

[39] T. Igasaki, K. Nagasawa, N. Murayama, and Z. Hu, "Drowsiness estimation under driving environment by heart rate variability and/or breathing rate variability with logistic regression analysis," in *Proc. 8th Int. Conf. Biomed. Eng. Informat. (BMEI)*, Shenyang, China, 2015, pp. 189–193.

[40] M. Tsujikawal, Y. Onishil, Y. Kiuchil, T. Ogatsul, A. Nishino, and S. Hashimoto, "Drowsiness estimation from low-frame-rate facial videos using eyelid variability features," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Honolulu, HI, USA, Jul. 2018, pp. 5203–5206.

[41] T. Nakamura, A. Maejima, and S. Morishima, "Driver drowsiness estimation from facial expression features computer vision feature investigation using a CG model," in *Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP)*, Lisbon, Portugal, 2014, pp. 207–214.

[42] O. Rigane, K. Abbes, C. Abdelmoula, and M. Masmoudi, "A fuzzy based method for driver drowsiness detection," in *Proc. IEEE/ACS 14th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Hammamet, Tunisia, Oct. 2017, pp. 143–147.

[43] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.

[44] R. W. Walpole, R. H. Myers, A. Myers, and K. Ye, *Probability & Statistics for Engineers and Scientists*, 8th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2007.

[45] National Highway Traffic Safety Administration. (Oct. 15, 2017). *Traffic Safety Facts Crash Stats: Drowsy Driving 2015 [EB/OL]*. Accessed: Jun. 29, 2018. [Online]. Available: https://crashstats.nhtsa.gov/Api/Public/ViewPublication/812446

[46] National Sleep Foundation. *Facts and Stats*. Accessed: Jan. 13, 2015. [Online]. Available: http://drowsydriving.org/about/facts-and-stats/

[47] B. C. Tefft, "Asleep at the wheel: The prevalence and impact of drowsy driving," AAA Found. Traffic Saf., Washington, DC, USA, Tech. Rep. 202-638-5944, 2010. [Online]. Available: https://www.aaafoundation.org/sites/default/files/2010DrowsyDrivingReport1.pdf

[48] Sohu. *People Died in Each Year Due to Traffic Accidents. [EB/OL]*. Accessed: Feb. 25, 2018. [Online]. Available: http://www.sohu.com/a/223922497_116921

[49] S. Wang, Y. Zhang, C. Wu, F. Darvas, and W. A. Chaovalitwongse, "Online prediction of driver distraction based on brain activity patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 136–150, Feb. 2015.

[50] H. S. Kim, D. Yoon, H. S. Shin, and C. H. Park, "Predicting the EEG level of a driver based on driving information," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1215–1225, Apr. 2019.

[51] T. Teng, L. Bi, and Y. Liu, "EEG-based detection of driver emergency braking intention for brain-controlled vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 6, pp. 1766–1773, Jun. 2018.

Yizhang Jiang (Senior Member, IEEE) received the Ph.D. degree from Jiangnan University in 2015. He has also been a Research Assistant at the Department of Computing, The Hong Kong Polytechnic University, for two years. His research interests include pattern recognition, intelligent computation, and their applications. He is the author/coauthor of more than 40 research papers in international/national journals, including the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, and *Information Sciences*. He has served as a reviewer or co-reviewer for several international conferences and journals, such as the ICDM, TKDE, TFS, TNNLS, *Pattern Recognition*, *Neurocomputing*, and *Neural Computing & Applications*. He has also served as a lead guest editor or guest editor for several international journals, such as the *Journal of Ambient Intelligence and Humanized Computing* and the *Journal of Cloud Computing*.

Yuanpeng Zhang (Member, IEEE) received the Ph.D. degree from the Information Engineering School of Computer Application Technology, Jiangnan University, in 2018. He is currently an Associate Professor with the Department of Medical Informatics, Nantong University. He is also a Post-Doctoral Fellow at the Department of Health Information Technology, The Hong Kong Polytechnic University. His main research interests include pattern recognition and data mining. He has published about 20 articles in international/national journals, including the IEEE TRANSACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, and the *ACM Transactions on Multimedia Computing Communications and Applications*.

Chuang Lin received the M.Sc. and Ph.D. degrees in signal processing from the Harbin Institute of Technology, Harbin, China, in 2004 and 2008, respectively. He was a Research Scientist of the European Research Council Project DEMOVE with the Department of Neurorehabilitation Engineering, University Medical Center Göttingen, Georg-August University, Germany, within the Bernstein Center for Computational Neuroscience and the Bernstein Focus Neurotechnology Göttingen. He is currently an Associate Professor with the Chinese Academy of Sciences, Shenzhen Institutes of Advanced Technology, Shenzhen, China. He has presided and participated in many grants, including the National Natural Science Foundation of China under Grants such as the "Hundred Talents Program" of the Chinese Academy of Sciences, and the Shenzhen Governmental Basic Research Grant. He has coauthored more than 20 articles in peer-reviewed journals and over 30 conference papers/abstracts, and over 10 patents. His research interests include biomedical signal processing, neural rehabilitation engineering, pattern recognition, and machine learning.

Dongrui Wu (Senior Member, IEEE) received the B.E. degree in automatic control from the University of Science and Technology of China in 2003, the M.E. degree in electrical engineering from the National University of Singapore in 2005, and the Ph.D. degree in electrical engineering from the University of Southern California in 2009. He is currently a Professor with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, and the Deputy Director of the Key Laboratory of Image Processing and Intelligent Control, Ministry of Education. His research interests include affective computing, brain–computer interface, computational intelligence, and machine learning. He has more than 120 publications, including a book *Perceptual Computing* (Wiley-IEEE Press, 2010) and 40 IEEE Transactions papers. He received the IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award in 2012, the IEEE Transactions on Fuzzy Systems Outstanding Paper Award in 2014, the NAFIPS Early Career Award in 2014, the IEEE Systems, Man and Cybernetics (SMC) Society Early Career Award in 2017, and the IEEE SMC Society Best Associate Editor Award in 2018. He was also the Finalist of another three Best Paper Awards. He was an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS (2011–2018), the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS (2014), the *IEEE Computational Intelligence Magazine* (2017), and the IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING (2019).

Chin-Teng Lin (Fellow, IEEE) received the B.S. degree in control engineering from the National Chiao-Tung University (NCTU), Hsinchu, Taiwan, in 1986, and the master's and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively. He is currently a Distinguished Professor with the Faculty of Engineering and Information Technology, University of Technology Sydney. He also holds an Honorary Chair Professorship of Electrical and Computer Engineering, NCTU, the International Faculty of University of California at San-Diego, La Jolla, CA, USA, and an Honorary Professorship of the University of Nottingham, Nottingham, U.K. He is the coauthor of *Neural Fuzzy Systems* (Prentice-Hall, 1996), and the author of *Neural Fuzzy Control Systems with Structure and Parameter Learning* (World Scientific, 1994). He has published more than 220 journal articles (Total Citation: 18 236, H-index: 57, i10-index: 295) and holds 97 patents in the areas of computational intelligence, fuzzy neural networks, natural cognition, brain–computer interface, intelligent system, multimedia information processing, machine learning, robotics, and intelligent sensing and control, including approximately 108 IEEE journal articles. He was elevated to an IEEE Fellow for his contributions to biologically inspired information systems in 2005 and was elevated to an International Fuzzy Systems Association Fellow in 2012. He received the Merit National Science Council Research Fellow Award, Taiwan, in 2009, the Outstanding Electrical and Computer Engineer, Purdue University, in 2011, the Outstanding Achievement Award by Asia–Pacific Neural Network Assembly in 2013, and the IEEE Fuzzy Systems Pioneer Award in 2016. He has also served on the Board of Governors of the IEEE Systems, Man, Cybernetics Society from 2003 to 2005, the IEEE Circuits and Systems (CAS) Society from 2005 to 2008, and the IEEE Computational Intelligence Society (CIS) from 2008 to 2010. He was the Chair of the IEEE Taipei Section from 2009 to 2010. He was the Program Chair of the IEEE International Conference on Systems, Man, and Cybernetics in 2005, and the General Chair of the 2011 IEEE International Conference on Fuzzy Systems. He has served as the Deputy Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II from 2006 to 2008, and as the Editor-in-Chief for the IEEE TRANSACTIONS ON FUZZY SYSTEMS from 2011 to 2016. He was a Distinguished Lecturer of the IEEE CAS Society from 2003 to 2005 and the CIS Society from 2015 to 2017.