Multiview Fuzzy Logic System With the Cooperation Between Visible and Hidden Views

Te Zhang, Zhaohong Deng ^(D), Senior Member, IEEE, Dongrui Wu ^(D), Senior Member, IEEE, and Shitong Wang ^(D)

Abstract—Multiview datasets are frequently encountered in learning tasks, such as web data mining and multimedia information analysis. Given a multiview dataset, traditional learning algorithms usually decompose it into several single-view datasets, from each of which a single-view model is learned. In contrast, a multiview learning algorithm can achieve better performance by cooperative learning on the multiview data. However, existing multiview approaches mainly focus on the views that are visible and ignore the hidden information behind the visible views, which usually contains some intrinsic information of the multiview data, or vice versa. To address this problem, this paper proposes a multiview fuzzy logic system which utilizes both the hidden information shared by the multiple visible views and the information of each visible view. Extensive experiments were conducted to validate its effectiveness.

Index Terms—Classification, cooperation between visible and hidden views, hidden space, multiview learning, Takagi–Sugeno– Kang fuzzy logic systems (TSK FLS).

I. INTRODUCTION

W ULTIVIEW learning is a machine learning paradigm that focuses on the data represented by different feature sets. The main reason for the arising of multiview learning is that more and more datasets can be represented by different attribute sets, i.e., they are collected with multiple views in practical scenes. For example, during food fermentation process, the

Manuscript received October 24, 2017; revised February 11, 2018; accepted September 13, 2018. Date of publication September 19, 2018; date of current version June 3, 2019. This work was supported in part by the National Key Research Program of China under Grant 2016YFB0800803, in part by the National Natural Science Foundation of China under Grant 61772239, in part by the Jiangsu Province Outstanding Youth Fund under Grant BK20140001, in part by the National First-Class Discipline Program of Light Industry Technology and Engineering under Grant LITE2018-02, and in part by the Basic Research Program of Jiangnan University Key Project in Social Sciences JUSRP1810ZD. (*Corresponding author: Zhaohong Deng.*)

T. Zhang, Z. Deng, and S. Wang are with the School of Digital Media, Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China (e-mail: 511865360@qq.com; dengzhaohong@jiangnan.edu.cn; wxwangst@aliyun.com).

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

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. The material provides parameters of the fuzzy rules of the FLSs generated by the proposed TSK-FLS-CVH method and the membership functions, as well as the potential linguistic explanation of each fuzzy subset in the antecedent of each fuzzy rule of the TSK FLS obtained. This material is 214 KB in size.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2018.2871005

food fermentation index can be recorded from two aspects: fermentation conditions and chemical index [3]. Another example is content based web image retrieval, where each image can be described by visible pixels and the associated text simultaneously [1]. Multiview learning has shown very promising results in such applications [2]–[7].

Multiview learning usually aims to learn one model by integrating the information of multiple views to improve the generalization performance. A naive solution for multiview learning concatenates all views into one single view and applies traditional single-view learning algorithms directly. However, this can easily result in overfitting when the training set is small, and also the specific statistical property of each view is lost. A noteworthy merit of multiview learning is that the performance on a single view dataset may still be improved by using artificially generated multiple views.

Many multiview learning approaches have been proposed in the literature. Representative ones are reviewed below.

- 1) Multiview classification, which applies multiview learning to classification problems. A frequently used technique is coregularization, which adds regularization terms to the objective function to make data from multiple views consistent. For example, an approach combining kernel canonical component analysis and support vector machine (SVM) was proposed in [7], a multiview transductive SVM was proposed in [5] by introducing global constraint variables, and a multiview Laplacian SVM was proposed in [6] by integrating manifold regularization and multiview regularization with a traditional SVM. Ensemble learning has also been extended to multiview settings [37], [43], and shown good performance on the highdimensional data and poem data. A multiview perceptron using a deep model for learning face identity and view representation was proposed in [38].
- 2) Multiview clustering, which is unsupervised. A collaborative clustering algorithm based on the classical fuzzy c-means (FCM) approach was proposed in [9], which used collaborative partition by controlling the fuzzy partitions among different views. Two different collaborative multiview clustering approaches were proposed in [2] using more sophisticated view corporation mechanisms. Two new multiview clustering algorithms based on kernel *k*-means and spectral clustering were proposed in [10]. An expectation-maximization-based collaborative multiview clustering method, Co-EM, was proposed in [11]. A canonical correlation analysis based multiview cluster

1063-6706 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information. ing technique was proposed in [12]. A novel tensor-based framework for integrating heterogeneous multiview data in the context of the spectral clustering was proposed in [39]. A new robust large-scale multiview clustering method that integrates multiple representations of large scale data was proposed in [40]. A multiview learning model, which integrates all features and learns the weight for every feature with respect to each cluster individually, was proposed in [41]. Most of the above-mentioned methods are based on the feature transformation [39]– [41], which transforms features form different views to a common feature space.

3) Multiview regression: A multiview regression approach using canonical correlation analysis was proposed in [13]. A multiview low-rank model that imposes low-rank constraints on the multiview regression model was proposed in [35]. A multiview regression model combining feature selection, low-rank, and subspace learning was proposed in [36].

Fuzzy sets and fuzzy logic systems (FLSs) have also been successfully used in multiview learning. For multiview clustering, a weighed fuzzy multiview collaborative clustering algorithm based FCM was proposed in [14] by introducing a penalty factor, and a minimax FCM multiview clustering algorithm was proposed in [15] by introducing minimax optimization into the classical FCM. For multiview classification, the large margin learning mechanism was introduced into the objective function of the classical Takagi–Sugeno–Kang (TSK) FLS, and then a collaborative learning based two-view fuzzy classification system was proposed in [16]. A multiview FLS for epilepsy EEG signal recognition was proposed in [17], which achieved better performance in epilepsy detection than many other methods.

Although existing multilearning techniques have shown promising performance in different applications, there is still room to improve due to the following fact. Most multiview data come from different domains or different feature extraction methods. Thus, there exists otherness among different views. Meanwhile, the features come from different views describe the same objects, so there also exists internal relation among different views. Thus, multiview learning usually follows two principles: complementarity principle and consistency principle. The complementarity principle aims to make full use of the otherness in different views to describe the data in a comprehensive way and enhance the model generalization ability. The consistency principle aims to use the internal relation between different views to maximize the consistency and enhance the model performance. However, most existing methods only focus on the information from the visible views, i.e., the otherness among different views, or only use the internal relation among them. In practice, not only the otherness but also the internal relation among different views are important for multiview modeling tasks. How to effectively integrate the complementarity principle with consistency principle in multiview modeling is a challenging task.

A novel multiview fuzzy classification system, i.e., TSK FLS with cooperation between visible and hidden views (TSK-FLS-CVH), is proposed in this paper. First, a TSK FLS is used as the base model to construct a multiview fuzzy classification system due to its flexibility and simplicity [18]. Then, nonnegative matrix factorization (NMF) is used to learn the hidden space that is shared among the visible views, and the corresponding data are obtained in this hidden view. The NMF based hidden view learning can effectively extract the consistency of different visible views. Next, cooperative learning between the visible and hidden views is implemented to construct the multiview TSK FLS. By cooperative learning, the visible views and the shared hidden view can benefit each other according to the complementarity principle. Compared with existing multiview FLSs and other related multiview approaches, the proposed method can effectively use the shared hidden information to implement the cooperation between visible and hidden views, i.e., it can effectively use the otherness and consistency in multiview data, which greatly enhances the generalization performance of the trained model.

The main contributions of this paper can be highlighted as follows.

- 1) An approach to extract shared hidden information among the visible views is proposed using nonnegative matrix factorization.
- A multiview FLS with cooperation between the visible and hidden views is proposed.
- 3) The proposed TSK-FLS-CVH is validated using extensive experiments.

The rest of this paper is organized as follows. Section II briefly reviews the concepts and principles of classical TSK FLSs and multiview learning. Section III proposes a strategy of learning the shared hidden space from multiple visible views, and describes the details of the multiview FLS, TSK-FLS-CVH. Section IV presents the experimental results. Finally, Section V draws conclusions and points out some future research directions.

II. BACKGROUND KNOWLEDGE

The basic concepts and principles of classical TSK FLSs and multiview learning are briefly reviewed in this section.

A. TSK FLS

There are three main categories of FLSs: TSK model [18], Mamdani model [19], and generalized fuzzy model [20]. Among them, the TSK model is most popular due to its simplicity and flexibility, and it is used in this paper.

1) TSK Fuzzy Classification System: A TSK FLS employs the following rules:

$$R^{k}: \text{ IF } x_{1} \text{ is } A_{1}^{k} \wedge x_{2} \text{ is } A_{2}^{k} \wedge \ldots \wedge x_{d} \text{ is } A_{d}^{k}$$

THEN $f^{k}(x) = p_{0}^{k} + p_{1}^{k}x_{1} + \ldots + p_{d}^{k}x_{d}, k = 1, 2, \ldots, K$
(1)

where $\boldsymbol{x} = [x_1, x_2, \dots, x_d]^T$ is an input vector, A_i^k is a fuzzy set in the *i*th input domain for the *k*th rule, and \wedge is a fuzzy conjunction operator.

When the product *t*-norm and center-of-gravity defuzzification are used, the final output of a classical TSK FLS is calculated as

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

where

$$\mu^{k}\left(\boldsymbol{x}\right) = \prod_{i=1}^{d} \mu_{A_{i}^{k}}\left(x_{i}\right)$$
(2b)

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

in which $\mu_{A_i^k}(x_i)$ is the membership of x_i to fuzzy set A_i^k , $\mu^k(\boldsymbol{x})$ is the firing strength of the *k*th rule to input vector \boldsymbol{x} , and $\sum_{k'=1}^{K} \mu^{k'}(\boldsymbol{x})$ is a normalization term, which is the sum of all firing strengths of *K* fuzzy rules to \boldsymbol{x} . With the normalization term, the normalized firing strength of the *k*th rule to \boldsymbol{x} , i.e., $\tilde{\mu}^k(\boldsymbol{x})$ can be calculated. Gaussian membership functions are used in this paper

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

where the parameters can be estimated using different approaches, e.g., FCM clustering [1]

$$c_i^k = \sum_{j=1}^N u_{jk} x_{ji} / \sum_{j=1}^N u_{jk}$$
(2e)

$$\delta_{i}^{k} = h \cdot \sum_{j=1}^{N} u_{jk} \left(x_{ji} - c_{i}^{k} \right)^{2} / \sum_{j=1}^{N} u_{jk}$$
(2f)

in which u_{jk} denotes the membership of x_{ji} in the *k*th fuzzy partition obtained by FCM on the training dataset [23], and *h* is a scaling parameter that can be set manually or determined by cross-validation (CV).

When the antecedent parameters in the rules are fixed, a TSK FLS can be reexpressed as a linear model [3], [21], [24]. Let

$$\boldsymbol{x}_{\boldsymbol{e}} = \left(1, \boldsymbol{x}^T\right)^T \tag{3a}$$

$$\tilde{\boldsymbol{x}}^{k} = \tilde{\boldsymbol{\mu}}^{k}\left(\boldsymbol{x}\right)\boldsymbol{x}_{\boldsymbol{e}} \tag{3b}$$

$$\boldsymbol{x_g} = \left(\left(\tilde{\boldsymbol{x}}^1 \right)^T, \left(\tilde{\boldsymbol{x}}^2 \right)^T, \dots, \left(\tilde{\boldsymbol{x}}^K \right)^T \right)^T$$
(3c)

$$\boldsymbol{p^{k}} = \left(p_0^k, p_1^k, \dots, p_d^k\right)^T \tag{3d}$$

$$\boldsymbol{p}_{\boldsymbol{g}} = \left(\left(\boldsymbol{p}^{1} \right)^{T}, \left(\boldsymbol{p}^{2} \right)^{T}, \dots, \left(\boldsymbol{p}^{K} \right)^{T} \right)^{T}.$$
(3e)

Then, (2a) becomes

$$y^0 = \boldsymbol{p_g^T x_g} \tag{3f}$$

where $x_g \in R^{K(d+1)}$ is a data vector in the new feature space that is mapped from the data vector $x \in R^d$ in the original feature space, and p_g is the combination of the consequent parameters of all fuzzy rules, which can be optimized using the method proposed later in this paper.



Fig. 1. Classical multiview learning framework.

2) Classification Based on the TSK FLS Regression Model: TSK FLSs have been extensively used for regression tasks. A classification task can be transformed into a multioutput regression task, and then a TSK FLS can be applied. For example, a three-class classification problem can be transformed to a threeoutput regression problem by representing labels 1, 2, and 3 as $[1 \ 0 \ 0]^T$, $[0 \ 1 \ 0]^T$, and $[0 \ 0 \ 1]^T$, respectively. The predicted label of a test sample is the index of the maximum value in the output vector. For example, if the prediction output of a test sample is $[0.6 \ 0.3 \ 0.1]^T$, then the test sample is classified to Class 1.

B. Multiview Learning

Multiview learning trains a model by considering the correlation and cooperation among different views, as shown in Fig. 1. Generally the performance of a multiview learning model is better than each individual single-view model, as more information is exploited in the former.

FLSs are frequently used in building individual single-view models. For multiview data, a popular strategy is to build a single-view FLS for every view and then fuse their outputs, e.g., using averaging, as shown in Fig. 2.

The above-mentioned simple strategy provides a feasible way for using single-view models for multiview data, but it ignores the relationship among the multiple views, and hence may not be able to get the best learning performance. Research has been done to improve it. A novel two-view FLS that combines cooperation leaning and large margin criterion was proposed in [16], which demonstrated promising performance. In addition, a multiview FLS has also been proposed in [17] for epilepsy EEG recognition. In this paper, we propose further improvements to multiview FLSs, by exploiting the hidden information shared among different visible views and using the cooperation between the visible and hidden views.



Fig. 2. Multiview modeling using single-view TSK FLS.

III. MULTIVIEW FLS WITH COOPERATION BETWEEN VISIBLE AND HIDDEN VIEWS

This section proposes TSK-FLS-CVH (see Fig. 3), a novel multiview FLS with the cooperation between the visible and hidden views. It uses multiview cooperation to train multiple FLSs for different views. The cooperation not only exists among the visible views, but also exists between the visible and hidden views. In this way, it can simultaneously utilize the explicit information from the visible views and the shared hidden information among them.

The details of TSK-FLS-CVH are described next.

A. Hidden View Generation

Our hypothesis is that there exists a shared hidden space that can be used to generate different visible spaces by different mappings, as shown in Fig. 4. This section describes how to identify this hidden view.

Given a multiview dataset with N samples and K visible views, the feature subset associated with the kth visible view can be represented as $\{\tilde{x}_i^k\}_{i=1}^N \subset R^{d_k}$, and the corresponding matrix representation is $\bar{X}^k = [\tilde{x}_1^k, \dots, \tilde{x}_N^k]^T \in \mathbb{R}^{N \times d_k}$. Let $H = [h_1, h_2, \dots, h_N]^T \in \mathbb{R}^{N \times r}$ be the data representation in the shared hidden space, and $W^k \in \mathbb{R}^{r \times d_k}$ be the mapping between the hidden view and the kth visible view. They are then identified by solving the following minimization problem:

$$\min_{\left\{\boldsymbol{W}^{k}\right\}_{k=1}^{K},\boldsymbol{H}} \sum_{k=1}^{K} \left(\alpha^{k} l\left(\bar{\boldsymbol{X}}^{k}, \boldsymbol{W}^{k}, \boldsymbol{H}\right) + \beta^{k} Dist\left(\kappa_{\bar{\boldsymbol{X}}^{k}}, \kappa_{\boldsymbol{H}}\right) \right)$$
s.t.
$$\sum_{k=1}^{K} \alpha^{k} = 1$$
(4)

where $\alpha^k \geq 0$ and $\beta^k \geq 0$ represents the weight and the regularization coefficient of the *k*th visible view, respectively. The first term of (4), $l(\bar{\boldsymbol{X}}^k, \boldsymbol{W}^k, \boldsymbol{H})$ is the empirical loss in mapping from \boldsymbol{H} to $\bar{\boldsymbol{X}}^k$, which is defined below

$$l\left(ar{oldsymbol{X}}^k,oldsymbol{W}^k,oldsymbol{H}
ight) = \left\|ar{oldsymbol{X}}^k-oldsymbol{H}oldsymbol{W}^k
ight\|_F^2$$

s.t. $oldsymbol{W}^k,oldsymbol{H}\geq oldsymbol{0}.$



Fig. 3. Framework of TSK-FLS-CVH.



Fig. 4. Relationship between the hidden view data and the visible view data.

(5)

The second term of (4), i.e., $\text{Dist}(\kappa_{\bar{\mathbf{X}}^k}, \kappa_{\mathbf{H}})$ is a regularization term. It is developed to guarantee that the data similarity $\kappa_{\mathbf{H}}$ described by the hidden view is close to the similarity $\kappa_{\bar{\mathbf{X}}^k}$ described by the original view. For this purpose, the manifold regularization is adopted in this paper. First, a nearest neighbor graph with N vertexes G^k for each visible view data is generated. Let $S^k \in \mathbb{R}^{N \times N}$ be the weight matrix of G^k . If \tilde{x}_i^k is the ε -nearest neighbor of \tilde{x}_j^k or \tilde{x}_j^k is the ε -nearest neighbor of \tilde{x}_i^k (the value of ε can be set manually or other strategies), then $S_{ij}^k = \kappa(\tilde{x}_i^k, \tilde{x}_j^k)$, otherwise, $S_{ij}^k = 0$. $\kappa(\tilde{x}_i^k, \tilde{x}_j^k)$ is a kernel function. If two points are neighbors in the original feature space, then they are also expected to be neighbors in the hidden space. Let $d_i^k = \sum_{j=1}^N S_{i,j}$, $D^k = \text{diag}(d_1^k, \ldots, d_N^k) \in \mathbb{R}^{N \times N}$, and then we define the regularization term $\text{Dist}(\kappa_{\bar{x}}^k, \kappa_H)$ in (4) as follows [27]:

Dist
$$\left(\kappa_{\bar{\boldsymbol{X}}^{k}}, \kappa_{\boldsymbol{H}}\right) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} ||\boldsymbol{h}_{\boldsymbol{i}} - \boldsymbol{h}_{\boldsymbol{j}}||^{2} S_{i,j}^{k}$$

= Trace $\left(\boldsymbol{H}^{T} \boldsymbol{L}^{k} \boldsymbol{H}\right)$ (6)

where $L^k = D^k - S^k$ is a Laplacian matrix.

Combining (5) and (6), (4) can be reexpressed as

$$\min_{\{\boldsymbol{W}^k\}_{k=1}^K, \boldsymbol{H}} \sum_{k=1}^K \left(\alpha^k \left\| \bar{\boldsymbol{X}}^k - \boldsymbol{H} \boldsymbol{W}^k \right\|_F^2 + \beta^k \operatorname{Trace} \left(\boldsymbol{H}^T \boldsymbol{L}^k \boldsymbol{H} \right) \right).$$
(7)

Next we explain how to solve for H and W^k from (7).

We first make all elements of \bar{X}^k nonnegative by normalization. Then, under the constraints $W^k \ge 0$ and $H \ge 0$, (7) can be solved by nonnegative matrix factorization (NMF). Existing solvers, such as coordinate descent [26], can be used directly. More specifically, (7) can be solved iteratively: optimize W^k by fixing H; and optimize H by fixing W^k .

1) Optimize W^k by fixing H.

When $H = H^{(t)}$ is fixed (where *t* is the current iteration number), the suboptimization problem of (7) that only involves W^k as the variable can be equivalently formulated as follows:

$$\min_{\boldsymbol{W}^{k}} \left\| \bar{\boldsymbol{X}}^{k} - \boldsymbol{H} \boldsymbol{W}^{k} \right\|_{F}^{2} \text{ s.t. } \boldsymbol{W}^{k} \ge \boldsymbol{0}$$
(8)

and \boldsymbol{W}^k can be updated as [26]

$$\left(\boldsymbol{W}^{k}\right)_{i,j}^{(t+1)} \leftarrow \left[\frac{\left(\left(\boldsymbol{\bar{X}}^{k}\right)^{T}\boldsymbol{H}\right)_{i,j}}{\left(\boldsymbol{H}^{T}\boldsymbol{H}\left(\boldsymbol{W}^{k}\right)^{(t)}\right)_{i,j}}\right] \left(\boldsymbol{W}^{k}\right)_{i,j}^{(t)} \quad (9)$$

2) Optimize H by fixing W^k .

When $W^k = (W^k)^{(t+1)}$ is fixed, the optimization problem that involves H becomes

$$\min_{H} \sum_{k=1}^{K} \alpha^{k} \left\| \bar{\boldsymbol{X}}^{k} - \boldsymbol{H} \boldsymbol{W}^{k} \right\|_{F}^{2} + \beta^{k} \operatorname{Trace} \left(\boldsymbol{H}^{T} \boldsymbol{L}^{k} \boldsymbol{H} \right)$$
s.t. $\boldsymbol{H} \geq 0$
(10)

and H can be updated as [26]

$$\boldsymbol{H}_{i,j}^{(t+1)} \leftarrow \left[\frac{\sum_{k=1}^{K} \alpha^{k} \left(\boldsymbol{\bar{X}}^{k} \left(\boldsymbol{W}^{k}\right)^{T}\right)_{i,j}}{\sum_{k=1}^{K} \left(\alpha^{k} \left(\boldsymbol{H}^{(t)} \boldsymbol{W}^{k} \left(\boldsymbol{W}^{k}\right)^{T}\right)_{i,j} + \beta^{k} \left(\boldsymbol{L}^{k} \boldsymbol{H}^{(t)}\right)_{i,j}\right)}\right] \boldsymbol{H}_{i,j}^{(t)}.$$
(11)

Clearly, if the initial $(\mathbf{W}^k)^{(0)}$ and $\mathbf{H}^{(0)}$ are nonnegative, then \mathbf{W}^k and \mathbf{H} are guaranteed to be nonnegative with the above updating rules.

B. Cooperative Learning Between Visible and Hidden Views

Given a multiview dataset with K visible views, we obtain a hidden view using the approach introduced in the previous Section III-A. Cooperative learning is used to make use of information in these K + 1 views.

Assume the TSK FLS corresponding to the *k*th visible view has L_k rules. Then, it can be represented as the following linear model:

$$f^{k}(\boldsymbol{x}^{k}) = (\boldsymbol{p}_{\boldsymbol{g}}^{k})^{T} \tilde{\boldsymbol{x}}_{g}^{k}$$
(12)

where $p_g^k, \tilde{x}_g^k \in R^{L_k(d_k+1)}, p_g^k$ is the vector consisting of all the consequent parameters of the FLS, \tilde{x}_g^k is the vector in the new feature space, which is mapped by fuzzy rules from the input vector \tilde{x}^k in the original feature space of View k.

Assume the TSK FLS corresponding to the hidden view has *J* rules. Then, it can be represented as

$$h(\tilde{\boldsymbol{h}}) = (\boldsymbol{p}_{\boldsymbol{g}}^{K+1})^T \tilde{\boldsymbol{h}}_g$$
(13)

where p_g^{K+1} , $\tilde{h}_g \in R^{J(r+1)}$, p_g^{K+1} is the vector consisting of all the consequent parameters, \tilde{h}_g is the vector in the new feature space, which is mapped by fuzzy rules from the input vector \tilde{h} in the original feature space of the hidden view.

Based on the above transformations, the following optimization problem is proposed for the multiview TSK FLS with the cooperation of the visible and hidden views:

$$\lim_{y,w} J = \sum_{k=1}^{K} w_k \sum_{i=1}^{N} \left\| \tilde{\boldsymbol{x}}_{gi}^k \boldsymbol{p}_{g}^k - y_i \right\|^2 \\
+ w_{K+1} \sum_{i=1}^{N} \left\| \tilde{\boldsymbol{h}}_{gi} \boldsymbol{p}_{g}^{K+1} - y_i \right\|^2 + \lambda_1 \sum_{k=1}^{K+1} w_k \ln w_k \\
+ \frac{1}{2} \lambda_2 \sum_{k=1}^{K+1} \left(\boldsymbol{p}_{g}^k \right)^T \boldsymbol{p}_{g}^k + \lambda_3 \sum_{k=1}^{K} \sum_{i=1}^{N} \left\| \tilde{\boldsymbol{x}}_{gi}^k \boldsymbol{p}_{g}^k - \bar{y}_i^k \right\|^2 \\
+ \lambda_3 \sum_{i=1}^{N} \left\| \tilde{\boldsymbol{h}}_{gi} \boldsymbol{p}_{g}^{K+1} - \bar{y}_i^{K+1} \right\|^2 \\$$
s.t. $\sum_{k=1}^{K+1} w_k = 1, w_k \ge 0$
(14)

where y_i is the label of sample *i*, p_g^k is the consequent parameter of the TSK FLS for the *k*th visible view, and w_k is its cor-

Authorized licensed use limited to: Huazhong University of Science and Technology. Downloaded on July 06,2020 at 15:29:44 UTC from IEEE Xplore. Restrictions apply.

 m_{p_g}

responding weight. $\sum_{k=1}^{K} \sum_{i=1}^{N} \|\tilde{\boldsymbol{x}}_{gi}^{k} \boldsymbol{p}_{\boldsymbol{g}}^{k} - \bar{y}_{i}^{k}\|^{2}$ and $\sum_{i=1}^{N} \|\tilde{\boldsymbol{h}}_{gi} p_{\boldsymbol{g}}^{k} - \bar{y}_{i}^{k+1}\|^{2}$ are the cooperation terms. $\bar{y}_{i}^{k} = \frac{1}{K} \sum_{j=1, j \neq k}^{K+1} \tilde{\boldsymbol{x}}_{gi}^{j} p_{\boldsymbol{g}}^{\boldsymbol{g}}$ (for convenience, here $\tilde{\boldsymbol{x}}_{gi}^{K+1} = \tilde{\boldsymbol{h}}_{gi}$). $\sum_{k=1}^{K+1} w_{k} \ln w_{k}$ is the negative Shannon entropy corresponding to the weights of different views (including the hidden view), and $\sum_{k=1}^{K+1} (\boldsymbol{p}_{\boldsymbol{g}}^{k})^{T} \boldsymbol{p}_{\boldsymbol{g}}^{k}$ is a regularization term.

The roles of the terms in (14) include the following.

- 1) The first two terms, $\sum_{k=1}^{K} w_k \sum_{i=1}^{N} \|\tilde{\boldsymbol{x}}_{gi}^k \boldsymbol{p}_g^k y_i\|^2$ and $w_{k+1} \sum_{i=1}^{N} \|\tilde{\boldsymbol{h}}_{gi} \boldsymbol{p}_g^{K+1} y_i\|^2$, are used to train the FLSs for multiple visible views and the hidden view.
- 2) The cooperation terms, $\sum_{k=1}^{K} \sum_{i=1}^{N} \|\tilde{\boldsymbol{x}}_{gi}^{k} \boldsymbol{p}_{\boldsymbol{g}}^{k} \bar{y}_{i}^{k}\|^{2}$ and $\sum_{i=1}^{N} \|\boldsymbol{h}_{gi} \boldsymbol{p}_{\boldsymbol{g}}^{K+1} \bar{y}_{i}^{K+1}\|^{2}$, try to make the outputs from different views consensus, which may help generalization.
- 3) The Shannon entropy term is used to determine the optimal weights for different views. Let ∑_{k=1}^{K+1} w_k = 1, w_k ≥ 0. Then, the weights can be regarded as a probability mass distribution, whose Shannon entropy is -∑_{i=1}^{K+1} w_k ln w_k. Minimizing ∑_{i=1}^{K+1} w_k ln w_k will make w_k close to each other, reducing the risk that a certain view dominates the final output.
- The regularization parameters λ₁ > 0, λ₂ > 0, λ₃ > 0 are used to control the impact of the corresponding terms. They can be determined by CV.

C. Parameter Learning

The Lagrangian method is used for solving p_g^k and w_k in (14), and the updating rules are the following:

$$(\boldsymbol{p}_{\boldsymbol{g}}^{k})^{(t+1)} = \left[\lambda_{2} \boldsymbol{I}_{d} + \left(w_{k}^{(t)} + \lambda_{3} \right) \sum_{i=1}^{N} \left(\tilde{\boldsymbol{x}}_{gi}^{k} \right)^{\mathrm{T}} \tilde{\boldsymbol{x}}_{gi}^{k} \right]^{-1} \\ \left[w_{k}^{(t)} \sum_{i=1}^{N} \left(\tilde{\boldsymbol{x}}_{gi}^{k} \right)^{T} y_{i} + \lambda_{3} \sum_{i=1}^{K} \left(\tilde{\boldsymbol{x}}_{gi}^{k} \right)^{T} \left(\bar{y}_{i}^{k} \right)^{(t)} \right]$$
(15)
$$w_{k}^{(t+1)} = \frac{\exp\left(-\sum_{i=1}^{N} \left(\tilde{\boldsymbol{x}}_{gi}^{k} \left(\boldsymbol{p}_{\boldsymbol{g}}^{k} \right)^{(t+1)} - y_{i} \right) / \lambda_{1} \right)}{\sum_{k=1}^{K+1} \exp\left(-\sum_{i=1}^{N} \left(\tilde{\boldsymbol{x}}_{gi}^{k} \left(\boldsymbol{p}_{\boldsymbol{g}}^{k} \right)^{(t+1)} - y_{i} \right) / \lambda_{1} \right)}$$
(16)

For convenience, we have set $\tilde{x}_{gi}^{K+1} = \tilde{h}_{gi}$ in (16).

Given a test sample with K visible views, the output of the multiview FLS is

$$Y_{output} = \sum_{k=1}^{K+1} w_k f^k \left(\boldsymbol{x} \right)$$
$$= \left(\sum_{k=1}^{K} w_k \left(\tilde{\boldsymbol{x}}_{\boldsymbol{g}}^k \right)^T \boldsymbol{p}_{\boldsymbol{g}}^k \right) + w_{K+1} \tilde{\boldsymbol{h}}_{\boldsymbol{g}}^T \boldsymbol{p}_{\boldsymbol{g}}^{K+1} \qquad (17)$$

$$f^{k}(\boldsymbol{x}) = \begin{cases} \left(\tilde{\boldsymbol{x}}_{\boldsymbol{g}}^{k}\right)^{T} \boldsymbol{p}_{\boldsymbol{g}}^{k} & 1 \leq k \leq K \\ \tilde{\boldsymbol{h}}_{\boldsymbol{g}}^{T} \boldsymbol{p}_{\boldsymbol{g}}^{K+1} & k = K+1 \end{cases}$$
(18)

D. Generation of Hidden View Data of the Test Dataset

Given a test dataset with N_{te} sample and K views, the subset associated with the kth view can be represented as $\{\boldsymbol{x}_{i}^{k}\}_{i=1}^{N_{te}}$ $\subset R^{d_{k}}$, and the corresponding matrix form is $\bar{\boldsymbol{X}}_{te}^{k} = [\boldsymbol{x}_{1}^{k}, \ldots, \boldsymbol{x}_{N_{te}}^{k}]^{T} \in \mathbb{R}^{N_{te} \times d_{k}}$. When $\bar{\boldsymbol{X}}_{te}^{k}$ has already been normalized to be nonnegative, the hidden view data of the test data can be generated by optimizing the following objective function:

$$\min_{H} \sum_{k=1}^{K} \left(\alpha^{k} \left\| \bar{\boldsymbol{X}}_{te}^{k} - \boldsymbol{H} \boldsymbol{W}^{*k} \right\|_{F}^{2} + \beta^{k} \operatorname{Trace} \left(\boldsymbol{H}^{T} \boldsymbol{L}^{k} \boldsymbol{H} \right) \right)$$
s.t. $\boldsymbol{H} \geq 0$
(19)

where W^{*k} is the mapping matrix of view k and it is obtained based on the training data by solving (7). Since W^{*k} is fixed in (19), we only need to solve H with the following iterative rule:

$$\boldsymbol{H}_{i,j}^{(t+1)} \leftarrow \begin{bmatrix} \sum_{k=1}^{K} \alpha^{k} \left(\bar{\boldsymbol{X}}_{te}^{k} \left(\boldsymbol{W}^{*k} \right)^{T} \right)_{i,j} \\ \frac{\sum_{k=1}^{K} \left(\alpha^{k} \left(\boldsymbol{H}^{(t)} \boldsymbol{W}^{*k} \left(\boldsymbol{W}^{*k} \right)^{T} \right)_{i,j} + \beta^{k} \left(\boldsymbol{L}^{k} \boldsymbol{H}^{(t)} \right)_{i,j} \end{bmatrix}}{\left(\sum_{k=1}^{K} \left(\alpha^{k} \left(\boldsymbol{H}^{(t)} \boldsymbol{W}^{*k} \left(\boldsymbol{W}^{*k} \right)^{T} \right)_{i,j} + \beta^{k} \left(\boldsymbol{L}^{k} \boldsymbol{H}^{(t)} \right)_{i,j} \right)} \right] \boldsymbol{H}_{i,j}^{(t)}.$$

$$(20)$$

Here, (20) is obtained with the similar way to (11).

E. Theoretical Analysis for TSK-FLS-CVH

There exist two key problems in multiview learning: 1) how to make full use of the otherness of different views; and 2) how to fully exploit the consistency among different views. The proposed TSK-FLS-CVH answers the above-mentioned questions from the following two perspectives.

- By introducing hidden view data, the proposed TSK-FLS-CVH can take advantage of the consistency information of different views. Because the hidden space is shared by different views, the hidden view data can reflect the internal relation between different views, i.e., the hidden view data contain the consistency information of the visible views. And to enhance the consistency among different views, the proposed method tries to make the outputs from different views consensus, which may also help generalization.
- 2) The proposed TSK-FLS-CVH exploits the otherness among different views by first training a TSK-FLS for each view data, including the hidden view data, respectively, and then assigning different weights to different views. It uses the maximum entropy in optimization to make the weights more sensible.

IV. EXPERIMENTS

Extensive experimental results are presented in this section to validate the performance of the proposed TSK-FLS-CVH. Five datasets were from the UCI machine learning repository and another is Epileptic EEG dataset [17]. The details are given in Table I. The performance index was the classification accuracy in five-fold cross validation.

1167

	Number Number Visible View 1				Visible View 2	
Dataset	of	of		Number		Number
	Samples	Classes	Description	0Î Faaturaa	Description	01 Eastures
	_		Lungs have designed based	Features	Speatnum view Speatnum values	Features
Forest Type	326	4	information of the data	9	and the difference values	18
SPECTF	267	2	Stress view: Single proton emission computed tomography image of heart in stress	22	Rest view: Single proton emission computed tomography image of heart in rest	22
Dermatology	366	6	Histopathological view: Histopathological information of a case	12	<i>Clinical view</i> : Clinical information of a case	22
Multiple Features	2000	10	<i>Fourier coefficients view</i> : Fourier coefficients of the character shapes	76	Zernike moments view: Zernike moments of the character shapes	47
Image Segmentation	2310	7	<i>Shape view</i> : Shape information of an image	9	<i>RGB view</i> : RGB information of an image	10
Epileptic EEG	500	2	DWT view: The features are extracted by DWT method as in [42].	20	 WPD view: The features are extracted by WPD method. After decomposition of the EEG signals using WPD, nodes (1, 0), (2, 0), (3, 0), (4, 0), (5, 0) and (5, 1) in the binary tree were used to calculate the corresponding energies. A 6-dimensional feature vector was then generated for each signal. 	6

 TABLE I

 MULTIVIEW DATASETS USED IN OUR EXPERIMENTS

 TABLE II

 CLASSIFICATION ACCURACIES OF THE TSK-FLS-CVH AND FIVE SINGLE-VIEW ALGORITHMS

Dataset	TSK-FLS-CVH	F-ELM	IQP	LESSLI	GENFIS2	L2-TSK-FS
Forest Type	0.9086 ± 0.0206	0.8700 ± 0.0200	0.8127 ± 0.0246	0.8108 ± 0.0232	0.8451 ± 0.0048	0.8509 ± 0.0359
SPECTF	0.8688 ± 0.0350	0.8203 ± 0.0306	0.7978 ± 0.0302	0.8240 ± 0.0391	0.7604 ± 0.0692	0.8164 ± 0.0250
Dermatology	0.9836 ± 0.0061	0.9016 ± 0.0264	0.8472 ± 0.0660	0.8715 ± 0.0669	0.9699 ± 0.0115	0.9781 ± 0.0184
Multiple Feature	0.8735 ± 0.0136	0.6230 ± 0.0418	0.4955 ± 0.0229	0.4580 ± 0.0265	0.8695 ± 0.0096	0.8405 ± 0.0130
Image Segmentation	0.8879 ± 0.0136	0.8675 ± 0.0412	0.7970 ± 0.0247	0.6970 ± 0.0192	0.8468 ± 0.0135	0.6476 ± 0.1037
Epileptic EEG	0.9700 ± 0.0084	0.9460 ± 0.0305	0.9680 ± 0.0110	$0.9640 \pm \ 0.0114$	0.9640 ± 0.0167	0.8960 ± 0.0182
Mean	0.9154	0.8381	0.7864	0.7709	0.8760	0.8382

A. Comparison With Single-View Algorithms

The TSK-FLS-CVH was first compared with five single-view FLS algorithms: F-ELM [28], IQP [29], LESSLI [29], GENFIS [30], and L2-TSK-FS [31]. For the latter, we constructed a single-view dataset by combining features from all visible views.

In order to maintain the interpretability and compactness of the FLSs, the number of fuzzy rules for each single-view FLS was determined using grid search in {10, 12, 14, 16, 18, 20}. The average classification accuracies and the standard deviations of the six algorithms are presented in Table II. The proposed TSK-FLS-CVH achieved the highest accuracies in all six datasets, suggesting that multiview learning is advantageous to the single-view learning.

B. Comparison With Multiview Algorithms

The TSK-FLS-CVH was next compared with two multiview algorithms, TwoV-TSK-FCS [16] and AMVMED [32]. The latter two are only available for binary classification. We used a 1-v-1 strategy to extend them to multiclass classification.

We also transformed the five single-view algorithms in Section IV-A to the corresponding multiview algorithms: a single-view classifier was trained separately for each visible view, and then their outputs were averaged. To distinguish from the original single-view algorithms, the transformed algorithms are denoted as F-ELM (MV), IQP (MV), LESSLI (MV), GEN-FIS2 (MV), and L2-TSK-FS (MV), respectively.

The experimental results on different multiview datasets are presented in Tables III–VIII.

The proposed TSK-FLS-CVH obtained the best classification accuracies in four out of the six multiview datasets. On some datasets, the classification accuracies of some multiview algorithms were worse than those obtained from a single view. For example, Table V shows that on the Dermatology dataset, the classification accuracy of IQP (MV) when using only the histopathological view was 0.8177, but its multiview learning accuracy was only 0.7405. This suggests that a simple average integration of multiple views may not always enhance the performance. However, TSK-FLS-CVH, which employs a more sophisticated strategy to integrate different views, can always achieve better performance than each individual view.

TABLE III Classification Accuracies (Mean \pm Std) of the Eight Multiview Algorithms on the Forest Type Dataset

Algorithm	Forest type						
Algorithm	Image Band View	Spectrum View	Hidden View	Integration of Multiple Views			
TSK-FLS-CVH	0.8872 ± 0.0139	0.8816 ± 0.0309	0.7686 ± 0.0182	0.9086 ± 0.0206			
TwoV-TSK-FCS	0.8604 ± 0.0106	0.6578 ± 0.0345	N/A	0.8604 ± 0.0106			
AMVMED	0.8853 ± 0.0064	0.8873 ± 0.0293	N/A	0.9026 ± 0.0243			
F-ELM (MV)	0.8834 ± 0.0104	0.8413 ± 0.0148	N/A	0.8872 ± 0.0139			
IQP (MV)	0.8108 ± 0.0284	0.7859 ± 0.0329	N/A	0.8203 ± 0.0262			
LESSLI (MV)	0.8050 ± 0.0357	0.7668 ± 0.0377	N/A	0.8146 ± 0.0310			
GENFIS2 (MV)	0.8489 ± 0.0327	0.8624 ± 0.0169	N/A	0.8643 ± 0.0227			
L2-TSK-FS (MV)	0.8298 ± 0.0239	0.6272 ± 0.0844	N/A	0.7783 ± 0.0483			

TABLE IV Classification Accuracies (Mean \pm Std) of the Eight Multiview Algorithms on the SPECTF Dataset

	SPECTF						
Algorithm	Stress View	Rest View	Hidden View	Integration of Multiple Views			
TSK-FLS-CVH	0.8015 ± 0.0172	0.8203 ± 0.0275	0.8462 ± 0.0416	0.8688 ± 0.0350			
TwoV-TSK-FCS	0.7940 ± 0.0021	0.6029 ± 0.0544	N/A	0.8277 ± 0.0277			
AMVMED	0.7940 ± 0.0021	0.8015 ± 0.0161	N/A	0.8015 ± 0.0161			
F-ELM (MV)	0.7753 ± 0.0290	0.8167 ± 0.0415	N/A	0.8092 ± 0.0460			
IQP (MV)	0.7867 ± 0.0398	0.7679 ± 0.0419	N/A	0.8017 ± 0.0397			
LESSLI (MV)	0.7904 ± 0.0325	0.7979 ± 0.0227	N/A	0.8166 ± 0.0355			
GENFIS2 (MV)	0.7978 ± 0.0153	0.7754 ± 0.0212	N/A	0.7940 ± 0.0021			
L2-TSK-FS (MV)	0.7940 ± 0.0021	0.8052 ± 0.0117	N/A	0.8090 ± 0.0242			

TABLE V Classification Accuracies (Mean \pm Std) of the Eight Multiview Algorithms on the Dermatology Dataset

A 1	Dermatology						
Algorithm	Clinical View	Histopathological View	Hidden View	Integration of Multiple Views			
TSK-FLS-CVH	0.8660 ± 0.0396	0.9454 ± 0.0194	0.9563 ± 0.0062	0.9836 ± 0.0061			
TwoV-TSK-FCS	0.4507 ± 0.0509	0.7979 ± 0.0305	N/A	0.8715 ± 0.0538			
AMVMED	0.8032 ± 0.0539	0.9344 ± 0.1103	N/A	0.9699 ± 0.0179			
F-ELM (MV)	0.7649 ± 0.0435	0.8906 ± 0.0680	N/A	0.9841 ± 0.0179			
IQP (MV)	0.6612 ± 0.0571	0.8177 ± 0.0557	N/A	0.7405 ± 0.0393			
LESSLI (MV)	0.6448 ± 0.0291	0.8034 ± 0.0635	N/A	0.7760 ± 0.0108			
GENFIS2 (MV)	0.8496 ± 0.0392	0.9399 ± 0.0155	N/A	0.9672 ± 0.0076			
L2-TSK-FS (MV)	0.7895 ± 0.0582	0.9426 ± 0.0225	N/A	0.9699 ± 0.0225			

TABLE VI Classification Accuracies (Mean \pm Std) of the Eight Multiview Algorithms on the Multiple Features Dataset

A 1 * 41	Multiple Features						
Algorithm	Fourier Coefficients View	Zemike View	Hidden View	Integration of Multiple Views			
TSK-FLS-CVH	0.8180 ± 0.0226	0.8215 ± 0.0060	0.7475 ± 0.0095	0.8735 ± 0.0191			
TwoV-TSK-FCS	0.5765 ± 0.0201	0.6370 ± 0.0310	N/A	0.7225 ± 0.0225			
AMVMED	0.4300 ± 0.0419	0.7995 ± 0.0114	N/A	0.8660 ± 0.0096			
F-ELM (MV)	0.5525 ± 0.0326	0.6340 ± 0.0552	N/A	0.7375 ± 0.0326			
IQP (MV)	0.3690 ± 0.0194	0.4050 ± 0.0232	N/A	0.4500 ± 0.0208			
LESSLI (MV)	0.3120 ± 0.0259	0.3615 ± 0.0464	N/A	0.3490 ± 0.0460			
GENFIS2 (MV)	0.7825 ± 0.0170	0.8060 ± 0.0080	N/A	0.8430 ± 0.0149			
L2-TSK-FS (MV)	0.7180 ± 0.0251	0.7035 ± 0.0251	N/A	0.7990 ± 0.0154			

In particular, although the performance of the proposed method is better than many existing related methods on the binary classification dataset SPECTF, its advantage seems not obvious as that in some multiclass datasets. It may be due to the fact that multiclass classification is usually more complicated than binary classification. In this situation, the proposed method will be more adaptive than the other related methods due to its comprehensive cooperative learning abilities from both the visible views and the shared hidden view simultaneously.

C. Statistical Analysis

1) *Friedman test combined with post-Holm test:* Friedman test [33], [34] was used to check if there was statistically significant difference among the 13 algorithms. The null

A 1	Image Segmentation						
Algorithm	Shape View	RGB View	Hidden View	Integration of Multiple Views			
TSK-FLS-CVH	0.8015 ± 0.0172	0.8203 ± 0.0275	0.8462 ± 0.0416	0.8879 ± 0.0136			
TwoV-TSK-FCS	0.3870 ± 0.0179	0.7009 ± 0.0213	N/A	0.7009 ± 0.0213			
AMVMED	0.7009 ± 0.0213	0.7740 ± 0.0104	N/A	0.9364 ± 0.0129			
F-ELM (MV)	0.4701 ± 0.0865	0.8134 ± 0.0148	N/A	0.8771 ± 0.0146			
IQP (MV)	0.4320 ± 0.0073	0.6065 ± 0.0199	N/A	0.5455 ± 0.0136			
LESSLI (MV)	0.3939 ± 0.0363	0.6273 ± 0.0513	N/A	0.4662 ± 0.0529			
GENFIS2 (MV)	0.4853 ± 0.0182	0.7342 ± 0.0210	N/A	0.8429 ± 0.0172			
L2-TSK-FS (MV)	0.3693 ± 0.0238	0.6745 ± 0.0200	N/A	0.6710 ± 0.0766			

TABLE VIII Classification Accuracies (Mean \pm Std) of the Eight Multiview Algorithms on the Epileptic EEG Dataset

Algorithm	Epileptic EEG dataset.						
Algorium	WAV View	WPD View	Hidden View	Integration of Multiple Views			
TSK-FLS-CVH	0.9360 ± 0.0207	0.9480 ± 0.0130	0.7320 ± 0.0130	0.9700 ± 0.0084			
TwoV-TSK-FCS	0.6620 ± 0.0740	0.9080 ± 0.0179	N/A	0.9240 ± 0.0182			
AMVMED	0.6520 ± 0.0327	0.8640 ± 0.0297	N/A	0.9500 ± 0.0381			
F-ELM (MV)	0.7660 ± 0.1062	0.9080 ± 0.0164	N/A	0.9540 ± 0.0114			
IQP (MV)	0.9600 ± 0.0187	0.9480 ± 0.0130	N/A	0.9700 ± 0.0158			
LESSLI (MV)	0.9540 ± 0.0241	0.9440 ± 0.0195	N/A	0.9700 ± 0.0148			
GENFIS2 (MV)	0.9260 ± 0.0114	0.9300 ± 0.0158	N/A	0.9440 ± 0.0130			
L2-TSK-FS (MV)	0.6760 ± 0.0564	0.8800 ± 0.0354	N/A	0.9180 ± 0.0259			

 TABLE IX

 Ranking of the 13 Algorithms Based on Friedman Test

Algorithm	Ranking
TSK-FLS-CVH	1.333
AMVMED	4.6667
F-ELM(MV)	5.6667
GENFIS2	6.25
F-ELM	6.3333
GENFIS2(MV)	7.1667
L2-TSK-FS	7.3333
TwoV-TSK-FCS	7.4167
LESSLI	8.3333
L2-TSK-FS(MV)	8.8333
IQP	9
LESSLI(MV)	9.1667
IQP(MV)	9.5

hypothesis, which says that the classification performances of all methods were the same, was rejected (p = 0.0023), i.e., the classification performances of the 13 algorithms had statistically significant difference. Table IX shows the ranking of the 13 algorithms. A lower ranking indicates a better performance. Clearly, TSK-FLS-CVH had the best performance.

A post-Holm test was then used to compare the best method, i.e., TSK-FLS-CVH with the other 12 algorithms. The *p*-value and the statistical magnitude are presented in Table X. The null hypothesis, i.e., there does not exist statistically significant difference between two algorithms, is rejected if p < 0.05. Table X shows that there was statistically significant difference between

TSK-FLS-CVH and L2-TSK-FS, L2-TSK-FS (MV), IQP, IQP (MV), LESSLI, LESSLI (MV), GENFIS2, GENFIS2 (MV), F-ELM, and TwoV-TSK-FCS, respectively. Although the null hypothesis was not rejected when compared with F-ELM (MV) and AMVMED, Tables II–VIII show that TSK-FLS-CVH still outperformed them slightly.

2) *t-test:* t-test based on the classification accuracies of fivefold cross validation on different dataset were used to further compare the proposed method with others. The results are given in Table XI. The null hypothesis, i.e., there does not exist statistically significant difference between two algorithms, is rejected if p < 0.05, and the "Rejected numbers" index is the number of datasets on which the hypothesis is rejected. Table XI shows that the proposed method significantly outperformed GENFIS2, GENFIS2(MV), L2-TSK-FS, L2-TSK-FS(MV), IQP, IQP(MV), LESSLI, LESSLI(MV), and TwoV-TSK-FCS on most datasets (no less than half of the number of datasets). Although the null hypothesis was not rejected on most datasets when compared with F-ELM, F-ELM(MV), and AMVMEND, the proposed method still outperformed them slightly, as presented in Tables II-VIII.

D. Effect of the Hidden View

The effect of the hidden view is further studied in this section. First, to test whether the hidden view can indeed enhance the multiview learning performance, we compared the classification accuracies of TSK-FLS-CVH with and without the hidden view. The results are presented in Table XII. Observe that the hidden view helped increase the classification accuracies on most datasets. GENFIS2/TSK-FLS-CVH

L2-TSK-FS/ TSK-FLS-CVH

F-ELM/ TSK-FLS-CVH

AMVMED/ TSK-FLS-CVH

Algorithm	$z = (R_0 - R_i) / SE$	p	$Holm = \alpha / i,$	Null Hypothesis
			$\alpha = 0.05$	
TwoV-TSK-FCS/ TSK-FLS-CVH	3.632122	0.000281	0.004167	Rejected
GENFIS2(MV)/ TSK-FLS-CVH	3.483872	0.000494	0.004545	Rejected
L2-TSK-FS(MV)/ TSK-FLS-CVH	3.409747	0.00065	0.005	Rejected
IQP(MV)/ TSK-FLS-CVH	3.335622	0.000851	0.005556	Rejected
LESSLI(MV)/ TSK-FLS-CVH	3.113247	0.00185	0.00625	Rejected
IQP/ TSK-FLS-CVH	2.70556	0.006819	0.007143	Rejected
F-ELM(MV)/ TSK-FLS-CVH	2.668498	0.007619	0.008333	Rejected
LESSLI TSK-FLS-CVH	2.594373	0.009476	0.01	Rejected

0.026165

0.028765

0.053949

0.138208

0.0125

0.016667

0.025

0.05

TABLE X THE RESULTS OF POST HOLM TEST

TABLE XI THE RESULTS (p Value) OF t-TEST

2.223748

2.186685

1.927248

1.482499

Algorithm		Dataset					Rejected Numbers
	Forest Type	SPECTF	Dermatology	Multiple Feature	Image Segmentation	Epileptic EEG	
TwoV-TSK-FCS/ TSK-FLS-CVH	0.0183	0.2506	0.0016	0.0001	0.0031	0.0026	5
L2-TSK-FS(MV)/ TSK-FLS-CVH	0.0005	0.1008	0.2829	0.0067	0.0002	0.0032	4
IQP(MV)/ TSK-FLS-CVH	0.0003	0.0911	0	0	0	1	4
LESSLI(MV)/ TSK-FLS-CVH	0.0043	0.1699	0	0	0	1	4
IQP/TSK-FLS-CVH	0.0001	0.0536	0.0021	0	0.0012	0.8974	4
LESSLI/ TSK-FLS-CVH	0.0003	0.3804	0.0055	0	0	0.6395	4
GENFIS2(MV)/ TSK-FLS-CVH	0.0116	0.0127	0.3972	0.8363	0.0386	0.7485	3
GENFIS2/ TSK-FLS-CVH	0.0116	0.0127	0.3972	0.8363	0.0386	0.7485	3
L2-TSK-FS/ TSK-FLS-CVH	0.0032	0.1508	0.7637	0.2097	0.0025	0.0002	3
F-ELM(MV) / TSK-FLS-CVH	0.0994	0.1402	0.0815	0.0006	0.3744	0.2543	2
F-ELM/ TSK-FLS-CVH	0.0766	1	0.0002	0	0.5077	0.1143	2
AMVMED/ TSK-FLS-CVH	0.691	0.0847	0.3356	0.717	0.007	0.2914	1

TABLE XII CLASSIFICATION ACCURACY OF TSK-FLS-CVH WITH AND WITHOUT THE HIDDEN VIEW

Dataset	TSK-FLS-CVH (without the hidden view)	TSK-FLS-CVH (with the hidden view)
Forest Type	0.8949 ± 0.0132	0.9086 ± 0.0206
SPECTF	0.8054 ± 0.0196	0.8688 ± 0.0350
Dermatology	0.9754 ± 0.0114	0.9836 ± 0.0061
Multiple Features	0.8480 ± 0.0135	0.8735 ± 0.0136
Image Segmentation	0.9078 ± 0.0094	0.8879 ± 0.0136
EEG Singals	0.9560 ± 0.0089	0.9700 ± 0.0084
Average	0.8979	0.9154

The dimensionality of the hidden view is an adjustable parameter. Its effect is also studied. Let the minimum dimensionality of all visible views be r. We extracted 9 hidden view datasets with dimensionalities $[0.1 * r], [0.2 * r], \dots, [0.9 * r]$ ([] denotes the round operation), respectively. The regularization parameters $\lambda_1, \lambda_2, \lambda_3$ in TSK-FLS-CVH were optimized without using the hidden view, and then hidden views with different dimensionalities were introduced. Experimental results on Forest Type, SPECT, and Dermatology are shown in Fig. 5, which indicate that the dimensionality of the hidden view can slightly

influence the classification accuracy. How to optimally set this parameter is still an open problem. In our experiment, hidden views with different dimensionalities were first trained, and the one with the best classification accuracy was then used in the subsequent cooperative learning between the visible and hidden views.

E. Example of the Generated Multiview FLSs

In this section, an example of the generated multiview TSK FLS by the proposed TSK-FLS-CVH based on the forest type

Rejected

Rejected

Not rejected

Not rejected



Fig. 5. Influence of the dimensionality of the hidden view on the classification accuracy. (a) Forest type. (b) SPECTF. (c) Dermatology.

dataset is shown. In order to readily display the results, we set the number of fuzzy rules of FLSs in visible and hidden views to three and four, respectively. Due to page limit, more results are presented in the supplementary materials. In Tables S1 and S2 of the supplementary materials, the parameters of the fuzzy rules of the FLSs generated in the image band view and the shared hidden view are listed, respectively. Meanwhile, the membership functions of the fuzzy subsets in fuzzy rules and the corresponding linguistic descriptions are shown for two views in Figs. S1 and S2, respectively. Based on these results, the first fuzzy rule in the two FLSs of two views can be linguistically described as follows.

The first fuzzy rule of the TSK FLS generated in the image band view is:

If the band 1 of ASTER image is low, and the band 2 of ASTER image is medium, and the band 3 of ASTER image is medium,

and the band 4 of ASTER image is low,

and the band 5 of ASTER image is medium,

and the band 6 of ASTER image is medium,

- and the band 7 of ASTER image is low,
- and the band 8 of ASTER image is medium,
- and the band 9 of ASTER image is low,

then this rule gives decision value of the first output with the following formula:

$$f^{2}(\boldsymbol{x}) = \begin{bmatrix} 0.0066 + 0.0372x_{1} - 0.0620x_{2} \\ -0.0263x_{3} - 0.0034x_{4} - 0.0215x_{5} + \\ 0.0143x_{6} - 0.0345x_{7} - 0.0550x_{8} + 0.0588x_{9} \end{bmatrix}.$$

The first fuzzy rule of the TSK FLS generated in the shared hidden view is:

If the shared hidden feature 1 is high,

and the shared hidden feature 2 is low,

and the shared hidden feature 3 is High,

- and the shared hidden feature 4 is a little low,
- and the shared hidden feature 5 is high,
- and the shared hidden feature 6 is low,

and the shared hidden feature 7 is a Little High,

then this rule gives decision value of the first output with the following formula:

$$f^{1}(\boldsymbol{x}) = \begin{bmatrix} -0.2350 - 0.2496x_{1} - 0.2348x_{2} - 0.2229x_{3} \\ -0.2293x_{4} - 0.2660x_{5} - 0.2241x_{6} - 0.4652x_{7} \end{bmatrix}$$

Finally, the decision of the multiview FLS for the first output can be given as follows

$$f\left(oldsymbol{x}
ight) = w_{1}f(oldsymbol{x})_{ ext{image - band - view}} + w_{2}f(oldsymbol{x})_{ ext{spectrum - view}}$$

 $w_3 f(oldsymbol{x})_{ ext{shared - hidden - view}}.$

V. CONCLUSION

In this paper, a multiview FLS with the cooperation between visible and hidden views has been proposed. It uses not only the information from multiple visible views, but also the information of the shared hidden view among multiple visible views, to enhance the learning performance. Experimental results demonstrated that the proposed approach outperformed many traditional single-view approaches and some state-of-theart multiview approaches.

Our future research includes how to efficiently optimize the hyper-parameters, and how to optimally determine the dimensionality of the hidden view.

REFERENCES

- A. Blum and T. Mitchell, "Combining labeled and unlabeled data with cotraining," in Proc. 11th Conf. Comput. Learn. Theory, 1998, pp. 92–100.
- [2] G. Cleuziou, M. Exbrayat, L. Martin, and J. H. Sublemontier, "COFKM: A centralized method for multiple-view clustering," in *Proc. 9th IEEE Int. Conf. Data Mining*, 2009, pp. 752–757.
- [3] Z. Deng, Y. Jiang, K. S. Choi, F. L. Chung, and S. Wang, "Knowledge leverage-based TSK FLS modeling," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 8, pp. 1200–1212, Aug. 2013.
- [4] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [5] G. Li, K. Chang, and S. C. H. Hoi, "Multi-view semi-supervised learning with consensus," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 11, pp. 2040– 2051, Nov. 2012.
- [6] S. Sun, "Multi-view Laplacian support vector machines," in Proc. Int. Conf. Adv. Data Mining Appl., 2011, pp. 209–222.
- [7] J. D. R. Farquhar, D. R. Hardoon, H. Meng, J. Shawe-Taylor, and S. Szedmak, "Two view learning: SVM-2K, theory and practice," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2005, pp. 355–362.
- [8] Q. Chen and S. Sun, "Hierarchical multi-view Fisher discriminant analysis," in *Proc. Int. Conf. Neural Inf. Process.*, 2009, vol. 5864, pp. 289–298.
- [9] W. Pedrycz, "Collaborative fuzzy clustering," *Pattern Recognit. Lett.*, vol. 23, no. 14, pp. 1675–1686, 2002.
- [10] G. Tzortzis and A. Likas, "Kernel-based weighted multi-view clustering," in Proc. IEEE Int. Conf. Data Mining, 2012, pp. 675–684.
- [11] Y. Yamanishi, J. P. Vert, and M. Kanehisa, "Protein network inference from multiple genomic data: a supervised approach," in *Proc. 12th Int. Conf. Intell. Syst. Molecular Biol./Third Eur. Conf. Comput. Biol.*, Glasgow, Uk, Aug. 2004, pp. 363–370.
- [12] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, 2009, pp. 129–136.

- [13] S. M. Kakade and D. P. Foster, "Multi-view regression via canonical correlation analysis," in *Proc. Conf. Learn. Theory*, Spring-Verlag, 2007, pp. 82–96.
- [14] Y. Jiang, F. L. Chung, S. Wang, Z. Deng, J. Wang, and P. Qian, "Collaborative fuzzy clustering from multiple weighted views," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 688–701, Apr. 2015.
- [15] Y. Wang and L. Chen, "Multi-view fuzzy clustering with minimax optimization for effective clustering of data from multiple sources," *Expert Syst. Appl.*, vol. 72, pp. 457–466, 2017.
- [16] Y. Jiang, Z. Deng, F. L. Chung, and S. Wang, "Realizing two-view TSK fuzzy classification system by using collaborative learning," *IEEE Trans. Syst., Man Cybern.–Syst.*, vol. 47, no. 1, pp. 145–160, Jan. 2017.
- [17] Y. Jiang *et al.*, "Recognition of epileptic EEG signals using a novel multiview TSK FLS," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 3–20, Feb. 2017.
- [18] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *Readings Fuzzy Sets Intell. Syst.*, vol. 15, no. 1, pp. 387–403, 1993.
- [19] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," in *Proc. Int. Symp. Multiple Valued Logic*, 1976, pp. 196–202.
- [20] M. F. Azeem, M. Hanmandlu, and N. Ahmad, "Generalization of adaptive neuro-fuzzy inference systems," *IEEE Trans. Neural Netw.*, vol. 11, no. 6, pp. 1332–1346, Nov. 2000.
- [21] Z. H. Deng, J. B. Zhang, Y. Z. Jiang, Y. Z. Shi, and S. T. Wang, "Fuzzy subspace clustering based zero-order l2-norm TSK FLS," *J. Electron. Inf. Technol.*, vol. 37, no. 9, pp. 2082–2088, 2015.
- [22] Y. Jiang, F. L. Chung, and S. Wang, "Enhanced fuzzy partitions vs data randomness in FCM," J. Intell. Fuzzy Syst. vol. 27, no. 4, pp. 1639–1648, 2014.
- [23] J. Bezdek and C. James, "Pattern recognition with fuzzy objective function algorithms," *Adv. Appl. Pattern Recognit.*, vol. 22, no. 1171, pp. 203–239, 1981.
- [24] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man Cybern.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.
- [25] L. Yang, L. Jing, B. Liu, and J. Yu, "Common latent space identification for heterogeneous co-transfer clustering," *Neurocomputing*, vol. 269, pp. 395– 406, 2015.
- [26] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization (NMF)," *Nature*, vol. 401, no. 6755, pp. 788–788, 1999.
- [27] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Int. Conf. Neural Inf. Process. Syst.: Natural Synthetic*, 2001, pp. 585–591.
- [28] S. Y. Wong, K. S. Yap, H. J. Yap, S. C. Tan, and S. W. Chang, "On equivalence of FIS and ELM for interpretable rule-based knowledge representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1417–1430, Jul. 2015.
- [29] J. M. Leski, "TSK-fuzzy modeling based onε;-insensitive learning," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 2, pp. 181–193, 2005.
- [30] S. L. Chiu, "Fuzzy model identification based on cluster estimation," J. Intell. Fuzzy Syst., vol. 2, pp. 267–278, 1994.
- [31] 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.
- [32] G. Chao and S. Sun, "Alternative multi-view maximum entropy discrimination," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 7, pp. 1445– 1456, Jul. 2016.
- [33] T. Joachims, "Transductive inference for text classification using support vector machines," in *Proc. 16th Int. Conf. Mach. Learn.*, 1999, pp. 200– 209.
- [34] J. H. Friedman and J. H. Friedman, "On bias, variance, 0/1—loss, and the curse-of-dimensionality," *Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 55–77, 1997.
- [35] S. Zheng, X. Cai, C. Ding, F. Nie, and H. Huang, "A closed form solution to multi-view low-rank regression," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1973–1979.
- [36] R. Hu *et al.*, "Low-rank feature selection for multi-view regression", *Multimedia Tools Appl.*, vol. 76, no. 16, pp. 17479–17495, Aug. 2017.
- [37] Z. Xu and S. Sun, "An algorithm on multi-view adaboost," *Lecture Notes Comput. Sci.*, vol. 6443, pp. 355–362, 2010
- [38] Z. Zhu and P. Luo, "Multi-view perceptron: a deep model for learning face identity and view representations," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 217–225.

- [39] X. Liu, S. Ji, W. Glanzel, and B. D. Moor, "Multi-view partitioning via tensor methods," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no.5, pp. 1056– 1069, May 2013.
- [40] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2011, pp. 2750–2756.
- [41] H. Wang, F. Nie, and H. Huang, "Multi-view clustering and feature learning via structured sparsity," in *Proc. Int. Conf. Mach. Learn.*, 2013. pp. 352–360.
- [42] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1084— 1093, 2007.
- [43] V. Kumar and S. Minz, "Multi-view ensemble learning for poem data classification using sentiwordnet" in *Proc. 2nd Int. Conf. Adv. Comput.*, *Netw. Inform.*, 2014, pp. 57–66.



Te Zhang received the M.S. degree in software engineering from the School of Digital Media, Jiangnan University, Wuxi, China, in 2018.

He is currently a Research Assistant with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. His current research interests include machine learning, fuzzy systems, data mining, and intelligent computation.

Zhaohong Deng (M'12–SM'14) received the B.S. degree in physics from Fuyang Normal College, Fuyang, China, in 2002 and the Ph.D. degree in information technology and engineering from Jiangnan University, Wuxi, China, in 2008.

He is currently a Professor with the School of Digital Media, Jiangnan University. He has visited University of California-Davis and the Hong Kong Polytechnic University for more than two years. He is the author or coauthor of more than 100 research papers in international/national journals. His current

research interests include uncertainty modeling, neuro-fuzzy systems, pattern recognition, and their applications.

He has served as an Associate Editor or Guest Editor of several international Journals, such as IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTA-TIONAL INTELLIGENCE, Neurocomputing and PLOS ONE.



Dongrui Wu (S'05–M'09–SM'14) received the B.E. degree in automatic control from the University of Science and Technology of China, Hefei, China, in 2003, the M.E. degree in electrical engineering from the National University of Singapore, Singapore, in 2005, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2009.

He is currently a Professor with the School of Automation, Huazhong University of Science and Technology, Wuhan, China. His research interests include

affective computing, brain-computer interface, computational intelligence, and machine learning. He has more than 100 publications, including a book entitled Perceptual Computing (Wiley-IEEE, 2010).

Dr. Wu received the IEEE CIS 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, and the IEEE SMC Society Early Career Award in 2017. He is an Associate Editor of the IEEE TRANS-ACTIONS ON FUZZY SYSTEMS, the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS, and the IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE.



Shitong Wang received the M.S. degree in computer science from the Nanjing University of Aeronautics and Astronautics, China, in 1987.

His research interests include artificial intelligence, neuro-fuzzy systems, pattern recognition, and image processing. He has authored and coauthored more than 100 papers in international/national journals and seven books.

Zhaoho degree Fuyang, formatic Universi He i