Received June 16, 2019, accepted June 25, 2019, date of publication June 28, 2019, date of current version July 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2925660

Entropy-Based Fuzzy Twin Bounded Support Vector Machine for Binary Classification

SUGEN CHEN[1,2], JUNFENG CAO[3], ZHONG HUANG[4], AND CHUANSHENG SHEN[1,2]

[1]School of Mathematics and Computational Science, Anqing Normal University, Anqing 246133, China
[2]Key Laboratory of Modeling, Simulation and Control of Complex Ecosystem in Dabie Mountains of Anhui Higher Education Institutes, Anqing Normal University, Anqing 246133, China
[3]School of Science, Jiangnan University, Wuxi 214122, China
[4]School of Physics and Electrical Engineering, Anqing Normal University, Anqing 246133, China

Corresponding author: Sugen Chen (chensugen@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702012, in part by the University Outstanding Young Talent Support Project of Anhui Province of China under Grant gxyq2017026, in part by the Natural Science Foundation of Anhui Province of China under Grant 1908085MF195, in part by the University Natural Science Research Project of Anhui Province of China under Grant KJ2017A361 and Grant KJ2017A368, and in part by the Program for Innovative Research Team in Anqing Normal University.

ABSTRACT Twin support vector machine (TWSVM) is a new machine learning method, as opposed to solving a single quadratic programming problem in support vector machine (SVM), which generates two nonparallel hyperplanes by solving two smaller size quadratic programming problems. However, the TWSVM obtains the final classifier by giving the same importance to all training samples which may be important for classification performance. In order to address this problem, in this paper, we propose a novel entropy-based fuzzy twin bounded support vector machine (EFTBSVM) for binary classification problems. By considering the fuzzy membership value for each sample and assigning it based on the entropy value, the samples with higher class certainty are assigned to relatively larger fuzzy membership. In addition, the proposed EFTBSVM not only maintains the superior characteristics of the TWSVM but also exploits the structural risk minimization principle by introducing a regularization term. The experimental results achieved on synthetic datasets and benchmark datasets illustrate the effectiveness of the proposed method.

INDEX TERMS Pattern classification, twin support vector machine, information entropy, fuzzy membership.

I. INTRODUCTION

Traditional support vector machine (SVM) is one of the most popular machine learning approaches, which is based on VC dimensional theory and structural risk minimization principle [1], [2]. Thus, it has already been extensively used in a variety of practical applications such as speaker identification [3], face recognition [4], bioinformatics [5], intrusion detection [6] and text categorization [7]. However, the computational cost of SVM is very high, i.e. $O(m^3)$, where m is the number of training samples. Followed by the success of the basic SVM, some improvements have been proposed. For instance, Fung and Mangasrian [8] proposed a proximal support vector machine (PSVM) for binary classification. In the spirit of PSVM, Mangasrian and Wild [9] proposed a generalized eigenvalue multisurface proximal SVM (GEPSVM), which aims at seeking a pair of nonparallel proximal hyperplanes such that each hyperplane is closer to one class and as far as possible from the other. Following GEPSVM, Jayadeva *et al.* [10] proposed a novel nonparallel hyperplane classifier for binary classification problems, called twin support vector machine (TWSVM). The key innovation of TWSVM aims at constructing two nonparallel hyperplanes, which needs to solve two small size quadratic programming problems (QPPs) rather than a single large size QPP. It makes training speed of TWSVM four times faster than SVM. From then on, various TWSVM methods have been widely investigated [11]–[23], such as least squares twin SVM (LSTSVM) [11], twin bounded SVM (TBSVM) [12], twin parametric-margin SVM (TPMSVM) [13], coordinate descent margin based twin SVM (CDMTSVM) [14], robust twin SVM (RTSVM) [15], nonparallel hyperplane

The associate editor coordinating the review of this manuscript and approving it for publication was Chunbo Xiu.

SVM (NHSVM) [16], maximum margin of twin spheres support vector machine (MMTSSVM) [17], novel twin SVM (NTSVM) [18], wavelet twin SVM (WTWSVM) [19], angle-based twin SVM (ATSVM) [20], sparse pinball TSVM (SPTWSVM) [21], Pin-GTSVM [22] and improved universum TSVM (IUTSVM) [23]. In order to deal with complex XOR problem, inspired by multi-weight vector projection SVM (MVSVM) [24], Chen *et al.* [25] proposed the projection twin SVM (PTSVM), which seeks a projection direction rather than a hyperplane for each class. Followed by, many variants of PTSVM were proposed, such as LSPTSVM [26], [27], LIWLSPTSVM [28], RPTSVM [29], L1-TPSVM [30], PTSVR [31], NPTSVM [32] and L21-EMVSVM [33]. Moreover, some overviews on twin support vector machine can be found in references [34]–[36].

On the one hand, as we knew, TWSVM implements the empirical risk minimization principle and the generalization ability of TWSVM is weak. In order to address this problem, TBSVM implements the structural risk minimization (SRM) principle by introducing a regularization term to the objective function and resulting in better generalization ability. In other words, by introducing the regularization term, TBSVM can not only eliminate the singularity but also have a deep theoretical basis from the statistical learning theory. Thus, regularization technique is widely used in various TWSVM methods such as RLS-TWSVM [37], RELS-TSVM [38], RLTSVR [39], LTPMSVM [40], KNNUPWTSVR [41] and others [42]–[44]. Specifically, we can find that RPTSVM is an improved version of PTSVM with a regularization term and RLS-TWSVM is an improved version of LS-TWSVM with a regularization term. The rest are similar. On the other hand, SVM treats all the training samples with the same importance and ignores differences between positive and negative classes, which results in the learned decision surface biasing toward the majority class. To deal with this problem, Lin and Wang [45] proposed Fuzzy SVM (FSVM) based on fuzzy membership values such that different training samples have different contributions to the final decision surface. However, it assigns the smaller fuzzy memberships to support vectors which might decrease the effects of support vectors on the construction of decision surface. In order to overcome this problem in FSVM, a new fuzzy SVM for imbalance datasets was proposed to reduce the misclassification accuracy of the minority class in FSVM [46]. In addition, based on utilizing fuzzy membership, various Fuzzy SVMs were presented such as NFSVM [47], Bilateral-weighted FSVM (BFSVM) [48], WCS-FSVM [49], FTSVM [50], NFTSVM [51] and FLTPMSVM [52], etc.

Recently, Fan *et al.* [53] proposed an entropy-based fuzzy SVM (EFSVM) for class imbalance problem in which fuzzy membership is computed based on class certainty of samples. Following EFSVM, Gupta *et al.* [54] proposed a fuzzy twin support vector machine based on information entropy which is termed as EFTWSVM-CIL. At the same time, Gupta *et al.* [55] proposed a new fuzzy least squares twin support vector machine (EFLSTWSVM-CIL) for class imbalance learning. Moreover, Richhariya and Tanveer [56] proposed a robust fuzzy least squares twin SVM (RFLSTSVM-CIL) for class imbalance learning. We can find that the choice of fuzzy membership is very important for classification problems. Based on the above analysis and inspired by TBSVM and EFTWSVM-CIL, in this paper, we propose a new entropy-based fuzzy twin bounded support vector machine, namely EFTBSVM. The contributions of our proposed EFTBSVM can be highlighted as follows. Firstly, entropy-based fuzzy membership is given to evaluate the class certainty of all training samples. Similar to EFTWSVM-CIL, our proposed EFTBSVM utilizes entropy to evaluate the class certainty of each sample and then computes the corresponding fuzzy membership based on the class certainty. Thus, it pays more attention to the samples with higher class certainty to result in a more robust decision surface, which can improve the classification accuracy with extended generalization ability. Secondly, different from EFTWSVM-CIL, our proposed EFTBSVM implements the structural risk minimization (SRM) principle instead of the empirical risk minimization (ERM) principle by adding the regularization term into the objective functions. Finally, the experimental results obtained on various synthetic datasets and benchmark datasets can illustrate the effectiveness of the proposed EFTBSVM over other state-of-the-art methods.

The rest of this paper is organized as follows. Section II gives a brief review of TWSVM, TBSVM, FTSVM, and EFTWSVM-CIL. The details of linear EFTBSVM and its nonlinear version are proposed in Section III, respectively. In Section IV, experimental results and further discussions are given to show the effectiveness of the proposed EFTBSVM. Finally, the conclusion is drawn in Section V.

II. RELATED WORKS

In this section, we briefly explain the basics of TWSVM, TBSVM, FTSVM and EFTWSVM-CIL. Let us consider a binary classification problem in the n-dimensional real space R^n and training data is represented by $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$, where $x_i \in R^n$ and $y_i \in \{-1, 1\}, i = 1, 2, \cdots, m$. For simplicity, we organize the m_1 inputs of positive class by matrix $A \in R^{m_1 \times n}$ and the m_2 inputs of negative class by matrix $B \in R^{m_2 \times n}$, respectively.

A. TWSVM

Different from SVM, twin support vector machine (TWSVM) [10] aims to find a pair of nonparallel hyperplanes

$$w_1^T x + b_1 = 0 \quad \text{and } w_2^T x + b_2 = 0 \tag{1}$$

such that each hyperplane is close to the training samples of one class and as far as possible from the samples of the other class. Therefore, the primal optimization problem of linear TWSVM is expressed as

$$\begin{aligned} \min_{w_1, b_1, \xi_2} \quad & \frac{1}{2}\|Aw_1 + e_1 b_1\|_2^2 + c_1 e_2^T \xi_2 \\ s.t. \quad & -(Bw_1 + e_2 b_1) + \xi_2 \geq e_2, \quad \xi_2 \geq 0, \end{aligned} \tag{2}$$

$$\min_{w_2,b_2,\xi_1} \frac{1}{2}\|Bw_2+e_2b_2\|_2^2+c_2e_1^T\xi_1$$
$$s.t.\ (Aw_2+e_1b_2)+\xi_1 \ge e_1,\quad \xi_1 \ge 0, \tag{3}$$

where $c_1 \ge 0$ and $c_2 \ge 0$ are penalty parameters, ξ_1 and ξ_2 are slack variables, e_1 and e_2 are vectors with each element of the value of 1.

By introducing the method of Lagrangian multipliers, the corresponding Wolfe dual of QPPs (2) and (3) are represented as

$$\max_{\alpha}\ e_2^T\alpha - \frac{1}{2}\alpha^T G(H^TH)^{-1}G^T\alpha$$
$$s.t.\ 0 \le \alpha \le c_1e_2, \tag{4}$$
$$\max_{\gamma}\ e_1^T\gamma - \frac{1}{2}\gamma^T H(G^TG)^{-1}H^T\gamma$$
$$s.t.\ 0 \le \gamma \le c_2e_1, \tag{5}$$

where $G=[B\ e_2],\ H=[A\ e_1]$ and $\alpha \in R^{m_2},\ \gamma \in R^{m_1}$ are Lagrangian multipliers.

The nonparallel hyperplanes (1) can be obtained from the solutions α and γ of (4) and (5) by

$$\left[w_1^T b_1\right]^T = -\left(H^TH\right)^{-1} G^T\alpha, \tag{6}$$
$$\left[w_2^T b_2\right]^T = \left(G^TG\right)^{-1} H^T\gamma, \tag{7}$$

Once the solutions of w_1, b_1 and w_2, b_2 are obtained from (6) and (7), the two nonparallel hyperplanes (1) are known. Then, the new data point $x \in R^n$ is assigned to positive class W_1 or negative class W_2 by

$$x \in W_k,\ k = \arg\min_{k=1,2}\{|w_1^Tx+b_1|, |w_2^Tx+b_2|\}, \tag{8}$$

where $|\cdot|$ is the absolute value.

Although $(H^TH)^{-1}$ and $(G^TG)^{-1}$ are always positive semidefinite, they are possible that they may not be well conditioned in some situations. Therefore, (6) and (7) are modified to

$$\left[w_1^T b_1\right]^T = -\left(\varepsilon I+H^TH\right)^{-1} G^T\alpha, \tag{9}$$
$$\left[w_2^T b_2\right]^T = \left(\varepsilon I+G^TG\right)^{-1} H^T\gamma, \tag{10}$$

where ε is a very small number and it is set to be 10^{-5} in TWSVM, I is an identity matrix of appropriate dimensions.

B. TBSVM

To address the above problem in TWSVM, by adopting the regularization term, Shao *et al.* [12] proposed twin bounded support vector machine (TBSVM). It implements the structural risk minimization (SRM) principle and the primal problem is expressed as

$$\min_{w_1,b_1,\xi_2} \frac{1}{2}c_3(\|w_1\|^2+b_1^2)+\frac{1}{2}\|Aw_1+e_1b_1\|_2^2+c_1e_2^T\xi_2$$
$$s.t.\ -(Bw_1+e_2b_1)+\xi_2 \ge e_2,\quad \xi_2 \ge 0, \tag{11}$$
$$\min_{w_2,b_2,\xi_1} \frac{1}{2}c_4(\|w_2\|^2+b_2^2)+\frac{1}{2}\|Bw_2+e_2b_2\|_2^2+c_2e_1^T\xi_1$$
$$s.t.\ (Aw_2+e_1b_2)+\xi_1 \ge e_1,\quad \xi_1 \ge 0, \tag{12}$$

where $c_1 \ge 0$ and $c_2 \ge 0$ are penalty parameters, $c_3 \ge 0$ and $c_4 \ge 0$ are trade-off parameters, ξ_1 and ξ_2 are slack variables, e_1 and e_2 are vectors with each element of the value of 1. Similar to TWSVM, the Wolfe dual of QPPs (11) and (12) can be represented as follows.

$$\max_{\alpha}\ e_2^T\alpha - \frac{1}{2}\alpha^T G(H^TH+c_3I)^{-1}G^T\alpha$$
$$s.t.\ 0 \le \alpha \le c_1e_2, \tag{13}$$
$$\max_{\gamma}\ e_1^T\gamma - \frac{1}{2}\gamma^T H(G^TG+c_4I)^{-1}H^T\gamma$$
$$s.t.\ 0 \le \gamma \le c_2e_1, \tag{14}$$

The nonparallel hyperplanes (1) can be obtained from the solutions α and γ to the optimization problems (13) and (14) by

$$\left[w_1^T b_1\right]^T = -\left(c_3I+H^TH\right)^{-1} G^T\alpha, \tag{15}$$
$$\left[w_2^T b_2\right]^T = \left(c_4I+G^TG\right)^{-1} H^T\gamma, \tag{16}$$

Once the solutions of w_1, b_1 and w_2, b_2 are obtained from (15) and (16), the nonparallel hyperplanes (1) are known. A new data point $x \in R^n$ is then assigned to positive class W_1 or negative class W_2 by

$$x \in W_k,\ k = \arg\min_{k=1,2}\{\frac{|w_1^Tx+b_1|}{\|w_1\|}, \frac{|w_2^Tx+b_2|}{\|w_2\|}\}, \tag{17}$$

C. FTSVM

Different from TWSVM, a weighting parameter is utilized to construct the classifier based on fuzzy membership values in the case of linear FTSVM [50]. The formulation of linear FTSVM can be written as

$$\min_{w_1,b_1,\xi_2} \frac{1}{2}\|Aw_1+e_1b_1\|_2^2+c_1s_2^T\xi_2$$
$$s.t.\ -(Bw_1+e_2b_1)+\xi_2 \ge e_2,\quad \xi_2 \ge 0, \tag{18}$$
$$\min_{w_2,b_2,\xi_1} \frac{1}{2}\|Bw_2+e_2b_2\|_2^2+c_2s_1^T\xi_1$$
$$s.t.\ (Aw_2+e_1b_2)+\xi_1 \ge e_1,\quad \xi_1 \ge 0, \tag{19}$$

where $c_1 \ge 0$ and $c_2 \ge 0$ are penalty parameters, ξ_1 and ξ_2 are slack variables, e_1 and e_2 are vectors with each element of the value of 1, s_1 and s_2 represent fuzzy membership of each type of sample points.

By introducing the method of Lagrangian multipliers, the corresponding Wolfe dual of QPPs (18) and (19) can be represented as

$$\max_{\alpha}\ e_2^T\alpha - \frac{1}{2}\alpha^T G(H^TH)^{-1}G^T\alpha$$
$$s.t.\ 0 \le \alpha \le c_1s_2, \tag{20}$$
$$\max_{\gamma}\ e_1^T\gamma - \frac{1}{2}\gamma^T H(G^TG)^{-1}H^T\gamma$$
$$s.t.\ 0 \le \gamma \le c_2s_1, \tag{21}$$

where $G=[B\ e_2],\ H=[A\ e_1]$ and $\alpha \in R^{m_2},\ \gamma \in R^{m_1}$ are Lagrangian multipliers.

The nonparallel hyperplanes (1) can be obtained from the solutions α and γ of (20) and (21) by

$$\left[w_1^T b_1\right]^T = -\left(H^T H\right)^{-1} G^T \alpha, \tag{22}$$

$$\left[w_2^T b_2\right]^T = \left(G^T G\right)^{-1} H^T \gamma, \tag{23}$$

Once w_1, b_1 and w_2, b_2 are obtained from (22) and (23), the two nonparallel hyperplanes (1) are known. A new data point $x \in R^n$ is then assigned to positive class W_1 or negative class W_2 by

$$x \in W_k,\ k = \arg\min_{k=1,2}\{|w_1^T x + b_1|, |w_2^T x + b_2|\}, \tag{24}$$

where $|\cdot|$ is the absolute value.

D. EFTWSVM-CIL

Motivated by entropy-based fuzzy support vector machine (EFSVM) [53], Gupta *et al.* [54] proposed an entropy-based fuzzy twin support vector machine for imbalanced datasets (EFTWSVM-CIL). In EFTWSVM-CIL, in order to enhance the participation of the minority class in decision classifier, the samples of majority class with lower entropy get larger fuzzy membership values. And the entropy of any sample x_i is calculated as

$$E_i = -p_{pos\_x_i} \cdot \ln(p_{pos\_x_i}) - p_{neg\_x_i} \cdot \ln(p_{neg\_x_i}), \tag{25}$$

where $p_{pos\_x_i}$ and $p_{neg\_x_i}$ are the probability of minority class and majority class of sample x_i.

Further, the data points of the majority class are divided into n subsets based on increasing order of entropy. Then, the fuzzy membership of samples in each subset are calculated as

$$F_q = 1.0 - \beta \cdot (q-1), \quad q = 1, 2, \cdots, n, \tag{26}$$

The fuzzy membership function is written as

$$s_i = \begin{cases} 1 - \beta \cdot (q-1), & \text{if } y_i = -1 \ \&\ x_i \in qth \text{ subset} \\ 1, & \text{if } y_i = 1, \end{cases} \tag{27}$$

Thus, the primal optimal problem of linear EFTWSVM-CIL is expressed as

$$\begin{aligned} &\min_{w_1,b_1,\xi_2} \frac{1}{2}\|Aw_1 + e_1 b_1\|_2^2 + c_1 s_2^T \xi_2 \\ &s.t.\ -(Bw_1 + e_2 b_1) + \xi_2 \geq e_2, \quad \xi_2 \geq 0, \end{aligned} \tag{28}$$

$$\begin{aligned} &\min_{w_2,b_2,\xi_1} \frac{1}{2}\|Bw_2 + e_2 b_2\|_2^2 + c_2 s_1^T \xi_1 \\ &s.t.\ (Aw_2 + e_1 b_2) + \xi_1 \geq e_1, \quad \xi_1 \geq 0, \end{aligned} \tag{29}$$

where c_1, c_2 are penalty parameters, ξ_1, ξ_2 are slack variables, e_1, e_2 are vectors with each element of the value of 1, s_1, s_2 are vectors containing the entropy-based fuzzy membership values of minority as well as majority, respectively.

By introducing the method of Lagrangian multipliers, the corresponding dual of QPPs (28) and (29) are represented as

$$\begin{aligned} &\min_{\alpha_1} \frac{1}{2}\alpha_1^T G(H^T H)^{-1} G^T \alpha_1 - e_2^T \alpha_1 \\ &s.t.\ 0 \leq \alpha_1 \leq c_1 s_2, \end{aligned} \tag{30}$$

$$\begin{aligned} &\min_{\alpha_2} \frac{1}{2}\alpha_2^T H(G^T G)^{-1} H^T \alpha_2 - e_1^T \alpha_2 \\ &s.t.\ 0 \leq \alpha_2 \leq c_2 s_1, \end{aligned} \tag{31}$$

where $G = [B\ e_2]$, $H = [A\ e_1]$ and $\alpha_1 \in R^{m_2}, \alpha_2 \in R^{m_1}$ are Lagrangian multipliers.

The nonparallel hyperplanes (1) can be obtained from the solutions α_1 and α_2 of (30) and (31) by

$$\left[w_1^T b_1\right]^T = -\left(H^T H\right)^{-1} G^T \alpha, \tag{32}$$

$$\left[w_2^T b_2\right]^T = \left(G^T G\right)^{-1} H^T \gamma, \tag{33}$$

Once w_1, b_1 and w_2, b_2 are obtained from (32) and (33), the two nonparallel hyperplanes (1) are known. A new data point $x \in R^n$ is assigned to positive class W_1 or negative class W_2 by

$$x \in W_k,\ k = \arg\min_{k=1,2}\{|w_1^T x + b_1|, |w_2^T x + b_2|\}, \tag{34}$$

where $|\cdot|$ is the absolute value.

III. ENTROPY-BASED FUZZY TWIN BOUNDED SUPPORT VECTOR MACHINE

As the evaluation of fuzzy membership is the key issue of FSVM, in this section, we introduce the entropy-based fuzzy membership at first. After that, by adopting the entropy-based fuzzy membership, entropy-based fuzzy twin bounded support vector machine (EFTBSVM) for binary classification is presented.

A. ENTROPY-BASED FUZZY MEMBERSHIP

In information theory, entropy is the average amount of information contained in each received message [57], which characterizes the certainty about the source of information. As you know, the smaller entropy indicates the information is more certain. Thus, by utilizing entropy, we can evaluate the class certainty of training samples and compute the fuzzy membership of the training samples based on their class certainty. It means that a sample with higher class certainty will be assigned to larger fuzzy membership, and vice versa. Therefore, we suppose the probabilities of training samples x_i belonging to the positive class and negative class are p_{+i} and p_{-i}, respectively. The entropy of x_i is defined as

$$H_i = -p_{+i} \cdot \ln(p_{+i}) - p_{-i} \cdot \ln(p_{-i}), \tag{35}$$

where *ln* represents the natural logarithm operator. The key point of calculating H_i by (35) is to evaluate the probability of each sample belong to the positive class and negative class. Specifically, we calculate the K nearest neighbors of sample x_i and compute the values p_{+i} and p_{-i} by (36) as follows.

$$p_{+i} = \frac{num_{+i}}{k}, p_{-i} = \frac{num_{-i}}{k}, \tag{36}$$

where num_{+i} and num_{-i} represent the number of positive and negative samples in the selected K nearest neighbors, and $num_{+i} + num_{-i} = k$.

According to (35) and (36), we compute the entropy of the positive samples and represent as $H_+ = \{H_{+1}, H_{+2}, \cdots, H_{+m_1}\}$. Then, the data points of positive class are divided into N_+ subsets based on increasing order of entropy and the fuzzy membership of positive samples in each subset are calculated as

$$FM_{+j} = 1.0 - \beta \times (j-1), \quad j = 1, 2, \cdots, N_+, \tag{37}$$

where FM_{+j} is the fuzzy membership for positive samples distributed in jth subset with fuzzy membership parameter $\beta \in (0, \frac{1}{N_+ - 1}]$ which controls the scale of the fuzzy values of positive samples.

Finally, the fuzzy membership of positive samples are defined as

$$s_{+i} = 1 - \beta \times (j-1), \quad if\ x_i \in jth\ subset\ (i = 1, 2, \cdots, m_1) \tag{38}$$

In the same way, the entropy of negative samples are obtained and indicated as $H_- = \{H_{-1}, H_{-2}, \cdots, H_{-m_2}\}$. Data points of negative class are divided into N_- subsets based on increasing order of entropy and the fuzzy membership of negative samples in each subset are calculated as

$$FM_{-j} = 1.0 - \beta \times (j-1), \quad j = 1, 2, \cdots, N_-, \tag{39}$$

where FM_{-j} is the fuzzy membership for negative samples distributed in jth subset with fuzzy membership parameter $\beta \in (0, \frac{1}{N_- - 1}]$.

Finally, the fuzzy membership of negative samples are defined as

$$s_{-i} = 1 - \beta \times (j-1), \quad if\ x_i \in jth\ subset\ (i = 1, 2, \cdots, m_2). \tag{40}$$

B. LINEAR EFTBSVM

In this subsection, inspired by TBSVM and EFTWSVM-CIL, we propose the entropy-based fuzzy twin bounded support vector machine (EFTBSVM) for binary classification. The structural risk minimization principle is exploited by adding a regularization term in our proposed EFTBSVM. Thus, the primal optimization problems of EFTBSVM are expressed as follows.

$$\begin{aligned}
\min_{w_1, b_1, \xi_2} \ & \frac{c_2}{2}(\|w_1\|^2 + b_1^2) + \frac{1}{2}(Aw_1 + e_1 b_1)^T (Aw_1 + e_1 b_1) \\
& + c_1 S_-^T \xi_2 \\
s.t. \ & -(Bw_1 + e_2 b_1) + \xi_2 \geq e_2, \ \xi_2 \geq 0, \qquad (41) \\
\min_{w_2, b_2, \xi_1} \ & \frac{c_4}{2}(\|w_2\|^2 + b_2^2) + \frac{1}{2}(Bw_2 + e_2 b_2)^T (Bw_2 + e_2 b_2) \\
& + c_3 S_+^T \xi_1 \\
s.t. \ & Aw_2 + e_1 b_2 + \xi_1 \geq e_1, \quad \xi_1 \geq 0, \qquad (42)
\end{aligned}$$

where $c_i \geq 0$ are penalty parameters, ξ_1 and ξ_2 are slack variables, $S_+ = (s_{+1}, s_{+2}, \cdots, s_{+m_1})^T$ and $S_- = (s_{-1}, s_{-2}, \cdots, s_{-m_2})^T$ are the entropy-based fuzzy membership values of positive class and negative class, e_1 and e_2 are the vectors of 1 with appropriate dimensions.

To obtain the solutions to the problems (41) and (42), we need to derive their dual problems, respectively. So, at first, we consider the primal problem (41), by introducing the Lagrangian multipliers, the Lagrangian function is given by

$$\begin{aligned}
L(w_1, b_1, \xi_2, \alpha, \beta) \\
= \frac{c_2}{2}(\|w_1\|^2 + b_1^2) + \frac{1}{2}(Aw_1 + e_1 b_1)^T (Aw_1 + e_1 b_1) \\
+ c_1 S_-^T \xi_2 - \alpha^T[-(Bw_1 + e_2 b_1) + \xi_2 - e_2] - \beta^T \xi_2,
\end{aligned} \tag{43}$$

where $\alpha \in R^{m_2}$ and $\beta \in R^{m_2}$ are Lagrangian multipliers. By using Karush-Kuhn-Tucker (KKT) conditions, we obtain

$$\nabla_{w_1} = c_2 w_1 + A^T(Aw_1 + e_1 b_1) + B^T \alpha = 0, \tag{44}$$

$$\nabla_{b_1} = c_2 b_1 + e_1^T(Aw_1 + e_1 b_1) + e_2^T \alpha = 0, \tag{45}$$

$$\nabla_{\xi_2} = c_1 S_- - \alpha - \beta = 0, \tag{46}$$

$$-(Bw_1 + e_2 b_1) + \xi_2 \geq e_2, \quad \xi_2 \geq 0, \tag{47}$$

$$\alpha^T(Bw_1 + e_2 b_1 - \xi_2 + e_2) = 0, \quad \beta^T \xi_2 = 0, \tag{48}$$

$$\alpha \geq 0, \quad \beta \geq 0, \tag{49}$$

Since $\beta \geq 0$, from (46) and (49), we can get

$$0 \leq \alpha \leq c_1 S_-, \tag{50}$$

Let $H = [A \ \ e_1] \in R^{m_1 \times (n+1)}, G = [B \ \ e_2] \in R^{m_2 \times (n+1)}$, $u_+ = [w_1^T \ \ b_1]^T \in R^{n+1}$, (44) and (45) imply that

$$(c_2 I_+ + H^T H) u_+ = -G^T \alpha, \tag{51}$$

where I_+ is an identity matrix.

Thus, we can obtain the augmented vector

$$u_+ = -(c_2 I_+ + H^T H)^{-1} G^T \alpha, \tag{52}$$

Then, putting (52) into (43) and using (44)-(46), we can get the Wolfe dual problem of (41) as follows.

$$\begin{aligned}
\max_{\alpha} \ & e_2^T \alpha - \frac{1}{2}\alpha^T G(H^T H + c_2 I_+)^{-1} G^T \alpha \\
s.t. \ & 0 \leq \alpha \leq c_1 S_-,
\end{aligned} \tag{53}$$

Similarly, we obtain the augmented vector $u_- = [w_2^T \ \ b_2]^T \in R^{n+1}$ and the Wolfe dual problem of (42) as follow.

$$u_- = (c_4 I_- + G^T G)^{-1} H^T \gamma, \tag{54}$$

$$\begin{aligned}
\max_{\gamma} \ & e_1^T \gamma - \frac{1}{2}\gamma^T H(G^T G + c_4 I_-)^{-1} H^T \gamma \\
s.t. \ & 0 \leq \gamma \leq c_3 S_+,
\end{aligned} \tag{55}$$

where $\gamma \in R^{m_1}$ is Lagrangian multiplier and I_- is an identity matrix.

Once $u_+ = [w_1^T \ \ b_1]^T \in R^{n+1}$ and $u_- = [w_2^T \ \ b_2]^T \in R^{n+1}$ are obtained from (52) and (54), the nonparallel hyperplanes

(1) are known. Then, a new data point $x \in R^n$ is assigned to positive class W_1 or negative class W_2 by

$$x \in W_k,\ k = \arg\min_{k=1,2}\{\frac{|w_1^T x + b_1|}{\|w_1\|}, \frac{|w_2^T x + b_2|}{\|w_2\|}\}, \quad (56)$$

where $|\cdot|$ is the absolute value.

Finally, we summarize the pipeline of linear EFTBSVM classifier in Algorithm 1.

Algorithm 1 Linear EFTBSVM Classifier

Step 1. Give the positive training samples $A \in R^{m_1 \times n}$ and negative training samples $B^{m_2 \times n}$.

Step 2. Select the appropriate neighborhood size k and then construct the entropy-based fuzzy membership value $S_+ = (s_{+1}, s_{+2}, \cdots, s_{+m_1})^T$ and $S_- = (s_{-1}, s_{-2}, \cdots, s_{-m_2})^T$ by using (38) and (40).

Step 3. Choose the proper penalty parameters c_i $(i = 1, 2, 3, 4)$ and then obtain the solutions $u_+ = [w_1^T\ b_1]^T \in R^{n+1}$ and $u_- = [w_2^T\ b_2]^T \in R^{n+1}$ form (52) and (54), respectively.

Step 4. For a new sample $x \in R^n$, calculate the distances $\frac{|w_1^T x+b_1|}{\|w_1\|}$ and $\frac{|w_2^T x+b_2|}{\|w_2\|}$, and assign it with positive or negative class label using (56).

C. NONLINEAR EFTBSVM

For the nonlinear case, firstly, we define $C = [A; B]$ and choose an appropriate kernel function K. Following the same idea, the linear EFTBSVM classifier can be extended to its nonlinear case by considering the following kernel-generated surfaces.

$$K(x^T, C^T)w_1+b_1 = 0 \quad \text{and } K(x^T, C^T)w_2+b_2 = 0, \quad (57)$$

The primal problems of the nonlinear EFTBSVM can be expressed as follows.

$$\begin{aligned}\min_{w_1,b_1,\eta_2}\ & \frac{1}{2}(K(A, C^T)w_1 + e_1 b_1)^T(K(A, C^T)w_1 + e_1 b_1) \\ & + \frac{c_2}{2}(\|w_1\|^2 + b_1^2) + c_1 S_-^T \eta_2 \\ s.t.\ & -(K(B, C^T)w_1 + e_2 b_1) + \eta_2 \geq e_2, \quad \eta_2 \geq 0, \quad (58)\end{aligned}$$

$$\begin{aligned}\min_{w_2,b_2,\eta_1}\ & \frac{1}{2}(K(B, C^T)w_2 + e_2 b_2)^T(K(B, C^T)w_2 + e_2 b_2) \\ & + \frac{c_4}{2}(\|w_2\|^2 + b_2^2) + c_3 S_+^T \eta_1 \\ s.t.\ & K(A, C^T)w_2 + e_1 b_2 + \eta_1 \geq e_1, \quad \eta_1 \geq 0, \quad (59)\end{aligned}$$

Similar to the linear case, by using the Lagrangian method and KKT conditions, we can obtain Wolfe dual problems as follows.

$$\begin{aligned}\max_{\alpha}\ & e_2^T\alpha - \frac{1}{2}\alpha^T\tilde{G}(\tilde{H}^T\tilde{H} + c_2\tilde{I}_+)^{-1}\tilde{G}^T\alpha \\ s.t.\ & 0 \leq \alpha \leq c_1 S_-, \quad (60)\end{aligned}$$

$$\begin{aligned}\max_{\gamma}\ & e_1^T\gamma - \frac{1}{2}\gamma^T\tilde{H}(\tilde{G}^T\tilde{G} + c_4\tilde{I}_-)^{-1}\tilde{H}^T\gamma \\ s.t.\ & 0 \leq \gamma \leq c_3 S_+, \quad (61)\end{aligned}$$

where $\tilde{H} = [K(A, C^T)\quad e_1] \in R^{m_1\times(m+1)}$, $\tilde{G} = [K(B, C^T)\ \ e_2]^{m_2\times(m+1)}$, $\tilde{I}_+$ and $\tilde{I}_-$ are identity matrices.

According to (58)-(61), the augmented vectors can be obtained by

$$v_+ = -(c_2\tilde{I}_+ + \tilde{H}^T\tilde{H})^{-1}\tilde{G}^T\alpha, \quad (62)$$
$$v_- = (c_4\tilde{I}_- + \tilde{G}^T\tilde{G})^{-1}\tilde{H}^T\gamma, \quad (63)$$

where $v_+ = [w_1^T\ b_1]^T \in R^{m+1}$, $v_- = [w_2^T\ b_2]^T \in R^{m+1}$.

Once $v_+ = [w_1^T\ b_1]^T$ and $v_- = [w_2^T\ b_2]^T$ are obtained from (62) and (63), the two nonparallel hyperplanes (57) are known. Then, a new data point $x \in R^n$ is assigned to positive class W_1 or negative class W_2 by

$$x \in W_k, k = \arg\min_{k=1,2}\{\frac{|K(x, C^T)w_1+b_1|}{\sqrt{w_1^T K(C, C^T)w_1}}, \frac{|K(x, C^T)w_2+b_2|}{\sqrt{w_2^T K(C, C^T)w_2}}\} \quad (64)$$

where $|\cdot|$ is the absolute value.

Similar to linear EFTBSVM, we summarize the nonlinear EFTBSVM classifier in Algorithm 2.

Algorithm 2 Nonlinear EFTBSVM Classifier

Step 1. Give the positive training samples $A \in R^{m_1\times n}$ and negative training samples $B^{m_2\times n}$, and choose an appropriate kernel function K based on cross-validation.

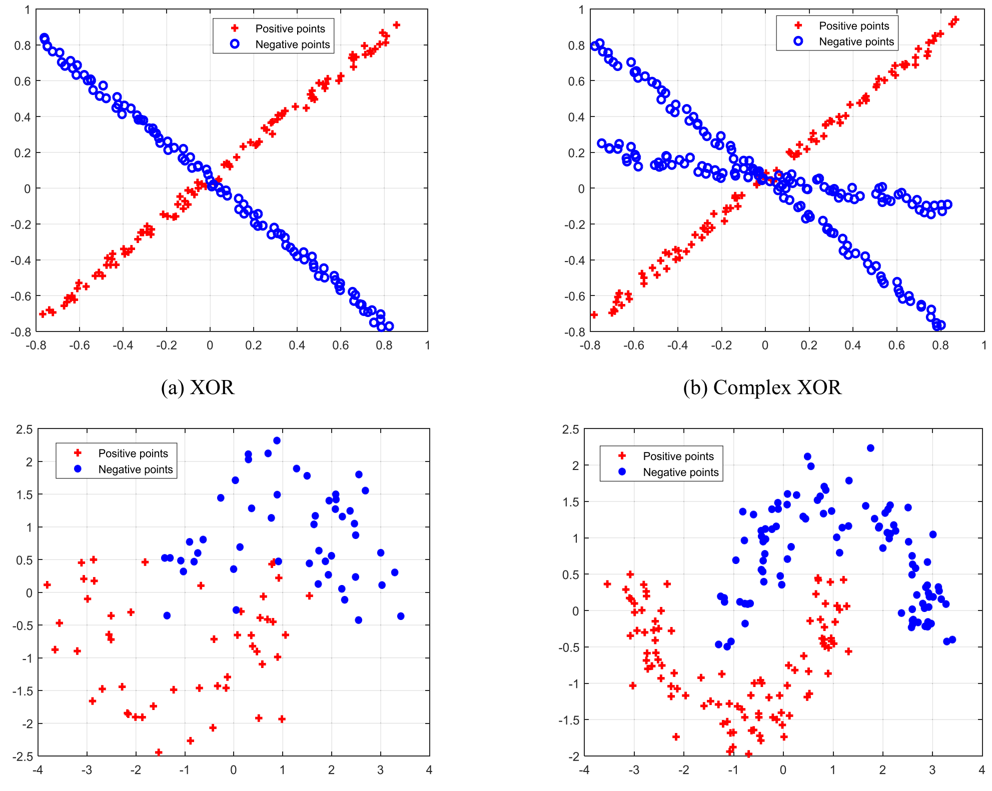
Step 2. Select the appropriate neighborhood size k and then construct the entropy-based fuzzy membership value $S_+ = (s_{+1}, s_{+2}, \cdots, s_{+m_1})^T$ and $S_- = (s_{-1}, s_{-2}, \cdots, s_{-m_2})^T$ by using (38) and (40).

Step 3. Choose the proper penalty parameters c_i $(i = 1, 2, 3, 4)$ and then obtain the solutions $v_+ = [w_1^T\ b_1]^T \in R^{m+1}$ and $v_- = [w_2^T\ b_2]^T \in R^{m+1}$ form (62) and (63), respectively.

Step 4. For a new sample $x \in R^n$, calculate the distances $\frac{|K(x,C^T)w_1+b_1|}{\sqrt{w_1^T K(C,C^T)w_1}}$ and $\frac{|K(x,C^T)w_2+b_2|}{\sqrt{w_2^T K(C,C^T)w_2}}$, and then assign it with positive or negative class label using (64).

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

To validate the classification performance of the proposed EFTBSVM, we investigate its classification accuracies on various synthetic datasets and benchmark datasets. In our experiments, we concentrate on the comparison between the proposed EFTBSVM and several related classifiers, including TWSVM [10], TBSVM [12], FTSVM [50], EFSVM [53], EFTWSVM-CIL [54] and RFLSTSVM-CIL [56]. All methods are implemented in MATLAB R2018a on a personal computer (PC) with an Intel (R) Core (TM) i7-7700CPU (3.60GHz × 8) and 32 GB random-access memory. The "Accuracy", which is used to evaluate the classification performance of all classifiers, defined as Accuracy = (TP+TN)/(TP+FP+TN+FN), where TP, TN, FP, and FN are the number of true positives, true negatives, false positives and false negatives, respectively. The QPP in EFSVM and QPPs in TWSVM, TBSVM, FTSVM, EFTWSVM-CIL, and



(a) XOR (b) Complex XOR (c) Two-moons-1 (d) Two-moons-2

FIGURE 1. Four synthetic datasets. (a) XOR, (b) Complex XOR, (c) Two-moons-1 and (d) Two-monns-2.

our EFTBSVM are solved by the SOR algorithm, which is also used to solve QPPs in reference [18]. And the 5-fold cross-validation approach and grid search method are used to find the optimal parameters such as the penalty parameters c_i, the kernel wide parameter σ and the neighborhood size k. Specifically, the penalty parameters $c_i\,(i = 1, 2, 3, 4)$ and the kernel wide parameter σ of Gaussian kernel function $K(x, y) = e^{-\|x-y\|^2/2\sigma^2}$ in all methods are selected from the set $\{2^i | i = -8, -7, \cdots, 7, 8\}$. The neighborhood size k in EFSVM, EFTWSVM-CIL and our EFTBSVM are chosen from the set $\{1, 3, 5, \cdots, 17, 19\}$. Besides, the parameter C_0 in RFLSTSVM-CIL is chosen from the set $\{0.5, 1, 1.5, 2, 2.5\}$. For simplicity, the number of separated subsets N_+ and N_- are set to 10 and the fuzzy membership parameter β is set to 0.05, respectively.

A. SYNTHETIC DATASETS

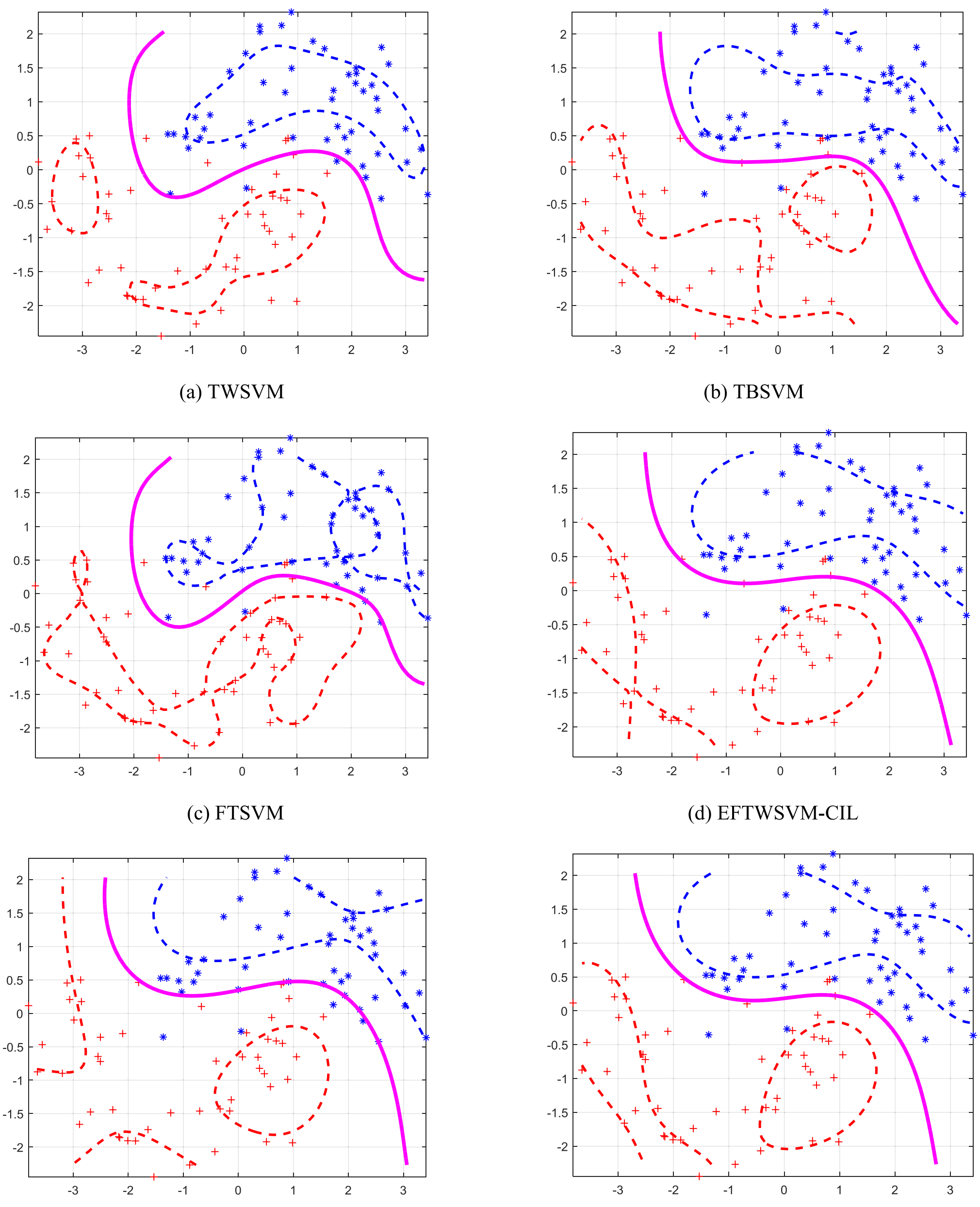
In this subsection, four synthetic datasets, including the XOR dataset, complex XOR dataset [18], Two-moons manifold dataset [58], [59] have been used to demonstrate that the proposed EFTBSVM can well deal with linearly inseparable problems. In our experiments, XOR dataset contains 200 samples (100 positives and 100 negatives), complex XOR dataset contains 260 samples (100 positives and 160 negatives), Two- moons-1 manifold dataset contains 100 samples (50 positives and 50 negatives) and Two-moons-2 manifold dataset contains 200 samples (100 positives and 100 negatives). Figure 1 shows the XOR, complex XOR and two kinds of Two-moons datasets, respectively. Specifically, for XOR and complex XOR datasets, we validate the performance of linear classifiers of TWSVM, TBSVM, EFSVM, FTSVM, EFTWSVM-CIL, RFLSTSVM-CIL and our EFTBSVM. For Two-moons manifold datasets, we validate the performance of all nonlinear methods with Gaussian kernel function. And we randomly select 40% for training sets and 60% for testing sets, each experiment repeat 10 times and the average results are listed in Table 1. From Table 1, we can conclude that our proposed EFTBSVM achieves the best performance on these four synthetic datasets. We take Two-moons-1 dataset for example, the accuracy of nonlinear EFTBSVM is 95.00%, while TWSVM is 93.83%, TBSVM is 94.67%, EFSVM is 93.67%, FTSVM is 92.17%, EFTWSVM-CIL is 93.50%, and RFLSTSVM-CIL is 94.00%, respectively. Also, the decision hyperplanes of the six nonparallel hyperplane classifiers, i.e. TWSVM, TBSVM, FTSVM, EFTWSVM-CIL, RELSTSVM-CIL, and our proposed EFTBSVM on Two-moon-1 dataset are shown in Figure 2, respectively.

B. UCI DATASETS

To further compare our EFTBSVM with TWSVM, TBSVM, EFSVM, FTSVM, EFTWSVM-CIL, and RFLSTSVM-CIL, we choose 11 datasets from UCI machine learning

TABLE 1. Classification accuracy on synthetic datasets.

Dataset	TWSVM	TBSVM	EFSVM	FTSVM	EFTWSVM-CIL	RFLSTSVM-CIL	EFTBSVM
	Acc±std	Acc±Std	Acc±Std	Acc±Std	Acc±Std	Acc±Std	Acc±Std
XOR	98.76±1.07	98.92±0.88	73.68±1.12	98.87±0.97	99.00±0.95	98.92±0.48	**99.58±0.44**
Complex XOR	90.32±2.28	91.67±2.36	71.25±2.32	90.83±2.16	94.10±1.70	93.28±1.87	**94.74±2.30**
Two-moons-1	93.83±2.94	94.67±2.33	93.67±1.53	92.17±2.16	93.50±2.54	94.00±3.22	**95.00±1.57**
Two-moons-2	99.42±0.40	99.92±0.26	99.75±0.40	99.68±0.32	99.83±0.35	99.92±0.26	**100.0±0.00**



(a) TWSVM (b) TBSVM

(c) FTSVM (d) EFTWSVM-CIL

(e) RFLSTSVM CIL (f) EFTBSVM

FIGURE 2. Classification hyperplanes of nonlinear methods on Two-moons-1 datasets. (a) TWSVM, (b) TBSVM, (c) FTSVM, (d) EFTWSVM-CIL (e) RFLSTSVM-CIL and (f) EFTBSVM.

repository [60]. They are Australian, Bupa-Liver, House-Votes, Heart-c, Heart-Statlog, Ionosphere, Musk, PimaIndian, Sonar, Spect, and Wpbc, respectively. Table 2 shows the characteristics of above-selected datasets.

Note that, we use the standard 5-fold cross-validation method to evaluate the performance of seven algorithms. That means the dataset is divided randomly into five subsets, one of those sets is reserved as a test set, and the others

TABLE 2. The characteristics of benchmark datasets.

Datasets	#Samples	#Features	Datasets	#Samples	#Features
Australian	690	14	Musk	476	166
Bupa-Liver	345	6	PimaIndian	768	8
House-Votes	435	16	Sonar	208	60
Heart-c	303	13	Spect	267	44
Heart-Statlog	270	13	Wpbc	198	34
Ionosphere	351	34			

are regarded as a training set. This process is repeated five times, and then the average of five testing results is used as the performance measure. Specifically, the experimental results of their linear versions are listed in Table 3 and the best accuracy is shown in boldface for each dataset. From Table 3, we could find that the accuracies of the proposed EFTBSVM are better than that of TWSVM, TBSVM, EFSVM, FTSVM, EFTWSVM-CIL, and RFLSTSVM-CIL in most of the datasets. We take Spect dataset for example, the accuracy of linear EFTBSVM is 83.32%, while TWSVM is 81.24%, TBSVM is 81.64%, EFSVM is 83.15%, FTSVM is 81.67%, EFTWSVM-CIL is 81.68% and RFLSTSVM-CIL is 82.66%, respectively. In addition, the experimental results of their nonlinear versions are listed in Table 4 and the best accuracy is shown in boldface for each dataset. From Table 4, we also take Spect dataset for example, the nonlinear EFTBSVM obtains the recognition rate 84.30%, which is 1.88% higher than TWSVM, 0.36% higher than TBSVM, 1.52% higher than EFSVM, 1.15% higher than FTSVM, 0.78% higher than EFTWSVM-CIL, and 0.76% higher than RFLSTSVM-CIL, respectively. Thus, the results in Table 4 are similar to those in Table 3, which confirms the effectiveness of our EFTBSVM.

Moreover, in order to make a statistic comparison on the effectiveness of the compared algorithms, the Friedman test [61] is carried out. For this test, the average ranks of the compared algorithms on the selected datasets are given in the last row of Table 3 and Table 4. Here, we consider $L(=7)$ number of compared algorithms and $n\ (=11)$ number of datasets. Let r_i^j be the rank of the j-th algorithm on the i-th datasets. Thus, we assume all the methods are equivalent under null hypothesis and the average rank of the j-th algorithm is calculated as $R_j = \frac{1}{n}\sum_{i=1}^{n} r_i^j$. And the Friedman statistic is defined as

$$\chi_F^2 = \frac{12n}{L(L+1)}\left[\sum_j R_j^2 - \frac{L(L+1)^2}{4}\right], \tag{65}$$

In fact, the Friedman statistic is distributed according to χ_F^2 with $(L-1)$ degrees of freedom, when n and L are reasonable large. In addition, Iman and Davenport [62] has showed that Friedman's χ_F^2 presents a pessimistic behavior, and they have derived a better statistic as follows.

$$F_F = \frac{(n-1)\chi_F^2}{n(L-1)-\chi_F^2}, \tag{66}$$

which is distributed according to the F-distribution with $(L-1)$ and $(L-1)(n-1)$ degrees of freedom.

For the linear case, in Table 3, it is noticed that the proposed EFTBSVM ranks the first with an average score of 1.8636. To demonstrate that the measured average ranks are significantly different from the mean rank by the null hypothesis, according to (65) and (66), we can get

$$\chi_F^2 = \frac{12\times 11}{7\times 8}[(5.5909^2+4.1364^2+5.3636^2+4.5^2 +2.6818^2+3.8636^2+1.8636^2)-\frac{7\times 8^2}{4}] = 25.8782$$

$$F_F = \frac{10\times 25.8782}{11\times 6 - 25.8782} = 6.4499$$

In addition, for 7 algorithms and 11 datasets, F_F is distributed according to the F-distribution with $(7-1)=6$ and $(7-1)\times(11-1) = 60$ degrees of freedom. Thus, we find that the critical value of $F(6, 60)$ is 2.254 for the level of significance $\alpha = 0.05$ and it is less than the value of $F_F = 6.4499$, which indicates the null hypothesis is rejected. It means that the compared algorithms are significantly different on selected datasets.

In Table 4, it is noticed that the nonlinear EFTBSVM ranks the first with an average score of 1.7727. According to (65)

FIGURE 3. Illustration of 10 subjects in the USPS database.

FIGURE 4. Illustration of 20 subjects in the COIL-20 database.

TABLE 3. Test results of linear TWSVM, TBSVM, EFSVM, FTSVM, EFTWSVM-CIL, RFLSTSVM-CIL and EFTBSVM.

Datasets	TWSVM	TBSVM	EFSVM	FTSVM	EFTWSVM-CIL	RFLSTSVM-CIL	EFTBSVM
	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)
Australian	87.54±2.87	87.39±2.74	86.52±3.64	87.83±2.53	87.68±1.98	87.82±2.01	**87.97±0.40**
Bupa-Liver	71.01±6.56	71.59±4.43	69.28±4.49	71.29±5.08	**73.33±5.48**	71.86±4.74	72.17±3.01
House-Votes	96.09±1.31	96.55±1.15	95.40±2.70	96.32±1.49	**96.78±0.51**	96.55±2.21	96.32±1.26
Heart-c	85.81±2.48	85.81±2.30	84.18±2.33	82.89±3.25	85.80±3.83	85.49±4.51	**86.15±4.70**
Heart-Statlog	85.56±3.04	85.93±3.12	85.56±4.01	85.60±2.75	85.93±2.81	85.56±3.04	**86.67±3.46**
Ionosphere	93.74±5.09	93.44±3.73	89.18±2.16	94.02±3.69	**94.87±2.97**	93.73±1.65	94.01±3.11
Musk	84.66±4.76	86.55±4.10	86.98±4.49	84.26±4.16	85.08±1.76	85.57±3.81	**87.81±2.75**
PimaIndian	77.86±2.92	77.99±2.50	77.21±3.13	78.13±3.19	**78.52±1.54**	77.87±3.03	78.26±3.75
Sonar	78.85±1.93	80.26±3.75	**81.74±4.45**	80.75±4.85	81.27±4.19	81.36±3.31	81.28±3.81
Spect	81.24±4.73	81.64±2.50	83.15±4.35	81.67±4.15	81.68±4.86	82.66±2.83	**83.32±4.21**
Wpbc	83.36±4.48	84.13±4.46	82.37±4.12	83.83±3.16	84.20±4.61	83.86±2.28	**84.38±3.12**
Average	84.16±3.65	84.66±3.16	83.60±3.62	84.24±3.48	85.01±3.14	84.76±3.04	**85.30±3.05**
Average rank	5.5909	4.1364	5.3636	4.5000	2.6818	3.8636	**1.8636**

TABLE 4. Test results of nonlinear TWSVM, TBSVM, EFSVM, FTSVM, EFTWSVM-CIL, RFLSTSVM-CIL and EFTBSVM.

atasets	TWSVM	TBSVM	EFSVM	FTSVM	EFTWSVM-CIL	RFLSTSVM-CIL	EFTBSVM
	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)	Acc+Std(%)
Australian	87.83±1.57	87.68±3.07	87.10±1.80	87.39±3.46	87.68±1.98	87.97±2.88	**88.12±3.22**
Bupa-Liver	75.07±4.39	75.65±2.59	73.91±4.39	74.78±3.92	74.49±4.18	75.36±4.23	**76.23±4.65**
House-Votes	96.32±2.21	96.09±1.54	95.86±3.11	96.32±2.74	96.09±1.92	96.32±2.74	**96.55±1.26**
Heart-c	85.14±3.52	85.48±3.62	84.84±4.64	83.83±2.17	**86.13±1.57**	85.48±2.02	85.82±3.39
Heart-Statlog	85.56±3.46	85.19±3.86	85.19±4.54	85.86±2.81	**86.30±2.81**	85.93±2.12	85.80±3.23
Ionosphere	94.32±3.32	**96.58±1.28**	95.73±2.26	94.60±3.67	94.60±3.51	95.31±2.22	**96.58±2.97**
Musk	95.79±2.11	95.38±2.03	94.54±0.88	**96.43±1.89**	95.58±2.41	94.75±1.82	95.59±2.27
PimaIndian	78.52±4.27	78.64±2.79	77.74±4.42	78.13±0.54	78.64±2.64	78.67±2.37	**79.03±2.90**
Sonar	88.95±2.68	90.81±3.86	90.88±2.57	89.94±4.52	91.86±3.58	**92.29±3.18**	91.39±3.69
Spect	82.42±4.15	83.94±4.16	82.78±5.42	83.15±3.49	83.52±4.22	83.54±4.36	**84.30±2.95**
Wpbc	83.86±6.18	82.36±3.82	82.83±3.23	83.33±5.18	84.36±3.74	83.83±1.44	**84.86±2.43**
Average	86.71±3.44	87.07±2.97	86.49±3.39	86.71±3.13	87.20±2.96	87.22±2.67	**87.66±2.30**
Average rank	4.6364	4.1818	6.0455	4.7727	3.5455	3.0455	**1.7727**

and (66), we can get

$$\chi_F^2 = \frac{12 \times 11}{7 \times 8}[(4.6364^2 + 4.1818^2 + 6.0455^2 + 4.7727^2 + 3.5455^2 + 3.0455^2 + 1.7727^2) - \frac{7 \times 8^2}{4}] = 26.6322$$

$$F_F = \frac{10 \times 26.6322}{11 \times 6 - 26.6322} = 6.7650$$

Similarly, for 7 nonlinear algorithms and 11 selected datasets, the critical value of $F(6, 60)$ is equal to 2.254 for the level of significance $\alpha = 0.05$ and it is also less than the value of $F_F = 6.7650$. Thus, the null hypothesis is rejected and then the compared nonlinear algorithms are significantly different.

C. IMAGE RECOGNITION

In this subsection, we apply our proposed EFTBSVM to image recognition problems. Three well-known and publicly available databases corresponding to image classification, *i.e.*, handwritten digit dataset (USPS), object dataset (COIL-20) and recognition of face dataset (AR) are adopted to validate our EFTBSVM with TWSVM, TBSVM,

TABLE 5. The classification performance comparison on the USPS, COIL-20 and AR datasets.

Datasets	TWSVM	TBSVM	EFSVM	FTSVM	EFTWSVM-CIL	RFLSTSVM-CIL	EFTBSVM
	Acc+Std	Acc+Std	Acc+Std	Acc+Std	Acc+Std	Acc+Std	Acc+Std
USPS 1 vs. 7	99.79±0.12	99.85±0.11	99.85±0.09	99.87±0.08	99.82±0.13	99.85±0.11	**99.90±0.07**
USPS 2 vs. 3	98.43±0.19	98.39±0.25	99.18±0.23	**99.19±0.21**	99.15±0.19	99.17±0.18	**99.19±0.21**
USPS 2 vs. 7	99.68±0.12	99.61±0.13	99.62±0.17	99.63±0.14	99.59±0.18	**99.73±0.18**	99.69±0.11
USPS 3 vs. 8	98.49±0.40	98.52±0.37	98.90±0.32	**99.01±0.30**	98.99±0.34	99.00±0.32	99.00±0.32
USPS 4 vs. 7	99.79±0.14	99.81±0.10	99.81±0.12	99.82±0.11	99.81±0.12	99.83±0.10	**99.85±0.07**
COIL-20	99.35±0.35	99.65±0.46	99.58±0.44	99.59±0.28	**99.71±0.28**	99.61±0.27	99.65±0.46
AR	96.84±0.93	98.37±0.51	98.29±0.53	98.33±0.37	**98.93±0.49**	98.35±0.24	98.39±0.35

FIGURE 5. Illustration of 14 images of one person from the AR database.

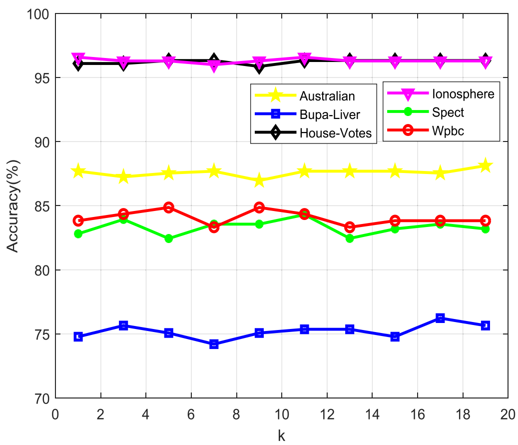


FIGURE 6. Accuracy of EFTBSVM with respect to *k* on the selected datasets.

EFSVM, FTSVM, EFTWSVM-CIL and RFLSTSVM-CIL. The USPS database [63] consists of gray-scale handwritten digit images from 0 to 9, where each digit contains 1100 images and the size of each image is 16 × 16 pixels with 256 gray levels. An illustration of 10 subjects in the USPS database is shown in Figure 3. COIL-20 [64] is a database of grayscale images of 20 objects. The objects are

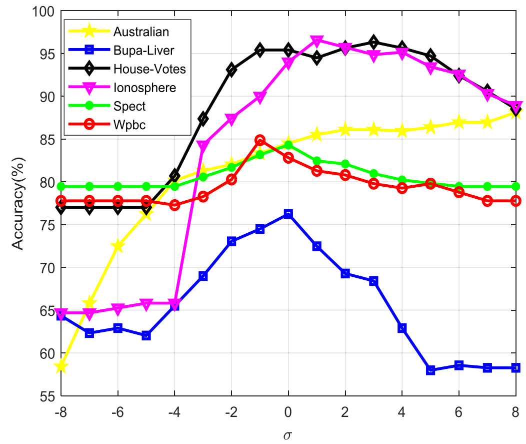


FIGURE 7. Accuracy of EFTBSVM with respect to σ on the selected datasets.

placed on a motorized turntable against a black background. Figure 4 shows an illustration of 20 subjects in the COIL-20 dataset. And the images of the objects are taken at pose intervals 5°, which corresponds to 72 images per object. Specifically, in our experiments, we have resized each of the original 1440 images into 32 × 32 pixels. AR database contains 100 subjects and each subject has 26 face images taken in two sessions [65]. For simplicity, the 1400 images are all cropped into the same size of 40 × 30 pixels in our experiment. Figure 5 shows 14 images of one person from the AR database.

For these datasets, we randomly partition the images of each project into two parts with the same size such that one part is selected for training and the remaining part is utilized for testing. We consider the Gaussian kernel for these nonlinear methods and the process is repeated 10 times. Table 5 lists the experimental results of these nonlinear methods in USPS, COIL-20 and AR datasets. We can conclude that, from USPS to AR, the proposed EFTBSVM achieves the best classification performance than the other six methods in most of the cases. Although our EFTBSVM achieves lower classification

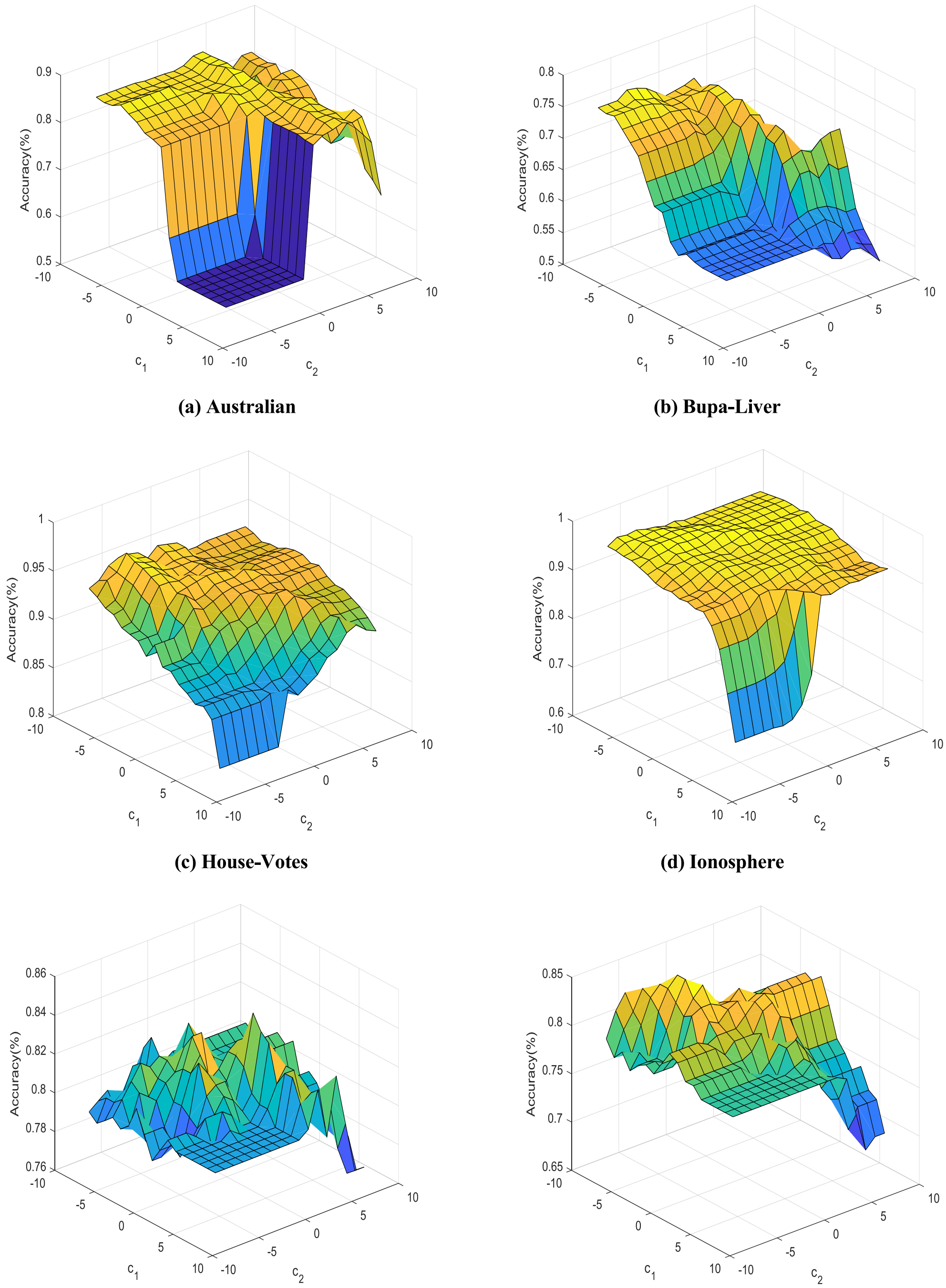


FIGURE 8. Accuracy of EFTBSVM with respect to $c_1 = c_3, c_2 = c_4$ on the selected datasets. (a) Australian, (b) Bupa-Liver, (c) House-Votes, (d) Ionosphere, (e) Spect and (f) Wpbc.

accuracy than EFTWSVM-CIL on COIL-20 and AR datasets, it also obtains higher classification accuracy than the other five methods.

D. FURTHER DISCUSSIONS

In the proposed EFTBSVM, there are so many parameters, such as the number of nearest neighbor k, the penalty parameters c_1, c_2, c_3, c_4 and the kernel wide parameter σ for nonlinear cases. In fact, these parameters significantly impact the classification performance of the proposed EFTBSVM. In order to demonstrate the influence of these parameters on EFTBSVM, we select 6 datasets from Table 4 and discuss their effects on the classification performance of EFTBSVM. For simplicity, we assume the penalty parameters $c_1 = c_3$ and $c_2 = c_4$. Specifically, when we discuss the influence of parameter k, other parameters are set to the best parameters

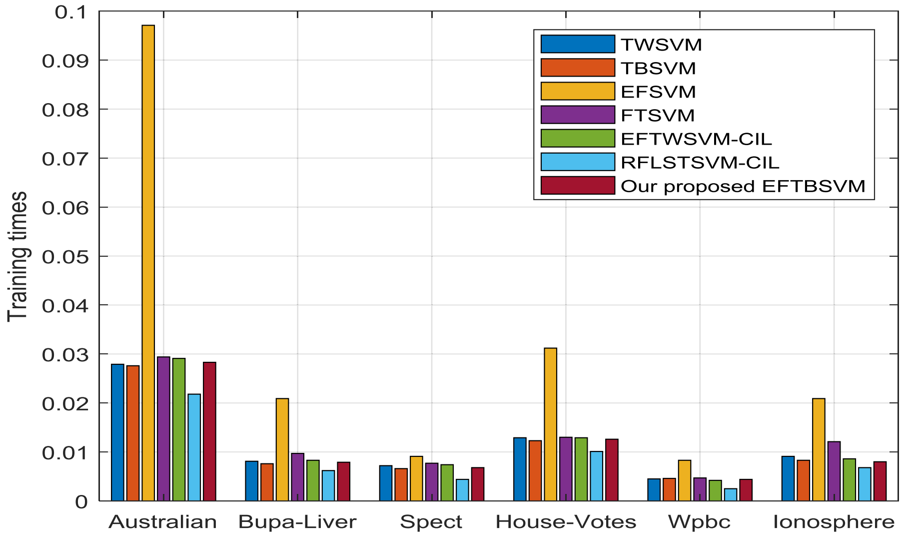


FIGURE 9. Training times of nonlinear methods on six selected UCI datasets.

which are selected by 5-fold cross-validation. The influences of other parameters are also discussed in the similar way. Take Wpbc dataset for example, the best parameters of Wpbc are $k = 5, c_1 = c_3 = 2^{-3}, c_2 = c_4 = 2^{-5}$ and $\sigma = 2^{-1}$. When discussing on k, k is selected from set $\{1, 3, \cdots, 17, 19\}$ and other parameters are set to $c_1 = c_3 = 2^{-3}, c_2 = c_4 = 2^{-5}, \sigma = 2^{-1}$, respectively. When discussing on σ, σ is selected from set $\{2^{-8}, 2^{-7}, \cdots, 2^7, 2^8\}$ and other parameters are set to $c_1 = c_3 = 2^{-3}, c_2 = c_4 = 2^{-5}, k = 5$. In addition, when discussing on c_i, the parameter k is set to 5, σ is set to 2^{-1} and $c_1 = c_3$, $c_2 = c_4$ are selected from set $\{2^{-8}, 2^{-7}, \cdots, 2^7, 2^8\}$. Therefore, the classification accuracy of our EFTBSVM with respect to k, σ and c_i on the selected datasets is shown in Figure 6, Figure 7 and Figure 8, respectively.

Furthermore, from Table 1, Table 3, Table 4 and Table 5 we can see, the training times of all classifiers are not given. However, it does not mean that the training time is not important. In fact, for the 7 classifiers, RFLSTSVM-CIL has the least training time, EFSVM has the most and the other five classifiers have almost the same training times. The main reason may be that RFLSTSVM-CIL solves two systems of linear equations, EFSVM solves a large size quadratic programming problem (QPP) and the other five methods only solve two small sizes QPPs. Thus, to further illustrate this problem, we also select the above 6 datasets from Table 4 and discuss the training times of all classifiers. The training times of all nonlinear classifiers on the selected datasets are shown in Figure 9. In specific, take Australian dataset for example, the training times of nonlinear TWSVM, TBSVM, EFSVM, FTSVM, EFTWSVM-CIL, RFLSTSVM-CIL, and our EFTBSVM are 0.0279 second, 0.0276 second, 0.0971 second, 0.0294 second, 0.0291 second, 0.0218 second, and 0.0283 second, respectively. From Figure 9, we can conclude that our EFTBSVM has no obvious advantage in training time. Thus, how to construct a fast and effective algorithm may be our future work.

V. CONCLUSION

In this paper, a novel entropy-based fuzzy twin bounded support vector machine (EFTBSVM) is proposed by utilizing the entropy to evaluate the class certainty of each sample and assigning the corresponding fuzzy membership value based on the class certainty. Our EFTBSVM pays more attention to the samples with higher class certainty to result in a more robust decision surface, which improves the classification accuracy and generalization ability. The experimental results obtained on synthetic and benchmark datasets demonstrate the effectiveness of EFTBSVM. However, we need to point out that there are so many parameters in EFTBSVM. Thus, parameter selection is a practical problem and should be investigated in the future. Moreover, the extension of the proposed EFTBSVM to multi-class [66]–[68], multi-label [69], [70] and multi-view [71], [72] classification problems are also interesting.

REFERENCES

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[2] V.N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.

[3] M. Schmidt and H. Gish, "Speaker identification via support vector classifiers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Conf. Proc.*, May 1996, pp. 105–108.

[4] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, p. 130.

[5] W. S. Noble, "Kernel methods in computational biology," in *Support Vector Machine Applications in Computational Biology*, MIT, Cambridge, U.K.:, 2004.

[6] D. Isa, L. H. Lee, V. P. Kallimani, and R. RajKumar, "Text document preprocessing with the Bayes formula for classification using the support vector machine," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 9, pp. 1264–1272, Sep. 2008.

[7] L. Khan, M. Awad, and B. Thuraisingham, "A new intrusion detection system using support vector machines and hierarchical clustering," *VLDB J.*, vol. 16, no. 4, pp. 507–521, 2007.
[8] O. L. Mangasarian and E. W. Wild, "Proximal support vector machine classifiers," in *Proc. KDD-Knowl. Discovery Data Mining*, May 2001, pp. 77–86.
[9] O. Mangasarian and E. Wild, "Multisurface proximal support vector machine classification via generalized eigenvalues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 69–74, Jan. 2006.
[10] J. R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 905–910, May 2007.
[11] M. A. Kumar and M. Gopal, "Least squares twin support vector machines for pattern classification," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7535–7543, 2009.
[12] Y. H. Shao, C. H. Zhang, X. B. Wang, and N. Y. Deng, "Improvements on twin support vector machines," *IEEE Trans. Neural Netw.*, vol. 22, no. 6, pp. 962–968, May 2011.
[13] X. Peng, "TPMSVM: A novel twin parametric-margin support vector machine for pattern recognition," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2678–2692, 2011.
[14] Y.-H. Shao and N.-Y. Deng, "A coordinate descent margin based-twin support vector machine for classification," *Neural Netw.*, vol. 25, pp. 114–121, Jan. 2012.
[15] Z. Qi, Y. Tian, and Y. Shi, "Robust twin support vector machine for pattern classification," *Pattern Recognit.*, vol. 46, no. 1, pp. 305–316, Jan. 2013.
[16] Y.-H. Shao, W.-J. Chen, and N.-Y. Deng, "Nonparallel hyperplane support vector machine for binary classification problems," *Inf. Sci.*, vol. 263, pp. 22–35, Apr. 2014.
[17] Y. Xu, "Maximum margin of twin spheres support vector machine for imbalanced data classification," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1540–1550, Jun. 2017.
[18] S. Chen, X. Wu, and R. Zhang, "A novel twin support vector machine for binary classification problems," *Neural Process. Lett.*, vol. 44, no. 3, pp. 795–811, Dec. 2016.
[19] S. Ding, Y. An, X. Zhang, F. Wu, and Y. Xue, "Wavelet twin support vector machines based on glowworm swarm optimization," *Neurocomputing*, vol. 225, pp. 157–163, Feb. 2017.
[20] R. Rastogi, P. Saigal, and S. Chandra, "Angle-based twin parametric-margin support vector machine for pattern classification," *Knowl.-Based Syst.*, vol. 139, pp. 64–77, Jan. 2018.
[21] M. Tanveer, A. Tiwari, R. Choudhary, and S. Jalan, "Sparse pinball twin support vector machines," *Appl. Soft Comput.*, vol. 78, pp. 164–175, May 2019.
[22] M. Tanveer, A. Sharma, and P. N. Suganthan, "General twin support vector machine with pinball loss function," *Inf. Sci.*, vol. 494, pp. 311–327, Aug. 2019.
[23] B. Richhariya, A. Sharma, and M. Tanveer, "Improved Universum twin support vector machine," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 2045–2052.
[24] Q. Ye, C. Zhao, N. Ye, and Y. Chen, "Multi-weight vector projection support vector machines," *Pattern Recognit. Lett.*, vol. 31, no. 13, pp. 2006–2011, Oct. 2010.
[25] X. Chen, J. Yang, Q. Ye, and J. Liang, "Recursive projection twin support vector machine via within-class variance minimization," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2643–2655, 2011.
[26] Y. Shao, N. Deng, and Z. Yang, "Least squares recursive projection twin support vector machine for classification," *Pattern Recognit.*, vol. 45, no. 6, pp. 2299–2307, 2012.
[27] S. Ding and X. Hua, "Recursive least squares projection twin support vector machines for nonlinear classification," *Neurocomputing*, vol. 130, pp. 3–9, Apr. 2014.
[28] X. Hua and S. Ding, "Weighted least squares projection twin support vector machines with local information," *Neurocomputing*, vol. 160, pp. 228–237, Jul. 2015.
[29] Y.-H. Shao, Z. Wang, W.-J. Chen, and N.-Y. Deng, "A regularization for the projection twin support vector machine," *Knowl.-Based Syst.*, vol. 37, pp. 203–210, Jan. 2013.
[30] Z. Gu, Z. Zhang, J. Sun, and B. Li, "Robust image recognition by L1-norm twin-projection support vector machine," *Neurocomputing*, vol. 223, pp. 1–11, Feb. 2017.
[31] X. J. Peng and D. Chen, "PTSVRs: Regression models via projection twin support vector machine," *Inf. Sci.*, vol. 435, pp. 1–14, Apr. 2018.
[32] S. Chen, X. Wu, and H. Yin, "A novel projection twin support vector machine for binary classification," *Soft Comput.*, vol. 23, no. 2, pp. 655–668, Jan. 2019.
[33] H. Zhao, Q. Ye, M. A. Naiem, and L. Fu, "Robust L2,1-norm distance enhanced multi-weight vector projection support vector machine," *IEEE Access*, vol. 7, pp. 3275–3286, 2019.
[34] S. Ding, X. Hua, and J. Yu, "An overview on nonparallel hyperplane support vector machine algorithms," *Neural Comput. Appl.*, vol. 25, no. 5, pp. 975–982, 2014.
[35] S. Ding, N. Zhang, X. Zhang, and F. Wu, "Twin support vector machine: Theory, algorithm and applications," *Neural Comput. Appl.*, vol. 28, no. 11, pp. 3119–3130, Nov. 2017.
[36] H. Huajuan, W. Xiuxi, and Z. Yongquan, "Twin support vector machines: A survey," *Neurocomputing*, vol. 300, pp. 34–43, Jul. 2018.
[37] R. Khemchandani and S. Sharma, "Robust least squares twin support vector machine for human activity recognition," *Appl. Soft Comput.*, vol. 47, pp. 33–46, Oct. 2016.
[38] M. Tanveer, M. A. Khan, and S.-S. Ho, "Robust energy-based least squares twin support vector machines," *Appl. Intell.*, vol. 45, no. 1, pp. 174–186, Jul. 2016.
[39] M. Tanveer and K. Shubham, "A regularization on Lagrangian twin support vector regression," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 3, pp. 807–821, Jun. 2017.
[40] P. Borah and D. Gupta, "On Lagrangian twin parametric-margin support vector machine," in *Proc. Int. Conf. Next Gener. Comput. Technol.*, Oct. 2017, pp. 474–487.
[41] D. Gupta, "Training primal K-nearest neighbor based weighted twin support vector regression via unconstrained convex minimization," *Appl. Intell.*, vol. 47, no. 3, pp. 962–991, Oct. 2017.
[42] M. Tanveer, "Robust and sparse linear programming twin support vector machines," *Cogn. Comput.*, vol. 7, no. 1, pp. 137–149, Feb. 2015.
[43] M. Tanveer, "Application of smoothing techniques for linear programming twin support vector machines," *Knowl. Inf. Syst.*, vol. 45, no. 1, pp. 191–214, Oct. 2015.
[44] S. Balasundaram and M. Tanveer, "Smooth Newton method for implicit Lagrangian twin support vector regression," *Int. J. Knowl.-Based Intell. Eng. Syst.*, vol. 17, no. 4, pp. 267–278, Jan. 2013.
[45] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 464–471, Mar. 2002.
[46] R. Batuwita and V. Palade, "FSVM-CIL: Fuzzy support vector machines for class imbalance learning," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 3, pp. 558–571, Jun. 2010.
[47] Q. Tao and J. Wang, "A new fuzzy support vector machine based on the weighted margin," *Neural Process. Lett.*, vol. 20, no. 3, pp. 139–150, Nov. 2004.
[48] Y. Wang, S. Wang, and K. K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 6, pp. 820–831, Dec. 2005.
[49] W. An and M. Liang, "Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises," *Neurocomputing*, vol. 110, no. 7, pp. 101–110, Jun. 2013.
[50] K. Li and H. Ma, "A fuzzy twin support vector machine algorithm," *Int. J. Appl. Innov. Eng. Manage.*, vol. 2, no. 3, pp. 459–465, Mar. 2013.
[51] S.-G. Chen and X.-J. Wu, "A new fuzzy twin support vector machine for pattern classification," *Int. J. Mach. Learn.*, vol. 9, no. 9, pp. 1553–1564, 2018.
[52] D. Gupta, P. Borah, and M. Prasad, "A fuzzy based Lagrangian twin parametric-margin support vector machine (FLTPMSVM)," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–7.
[53] Q. Fan, Z. Wang, D. Li, D. Gao, and H. Zha, "Entropy-based fuzzy support vector machine for imbalanced datasets," *Knowl.-Based Syst.*, vol. 115, pp. 87–99, Jan. 2017.
[54] D. Gupta, B. Richhariya, and P. Borah, "A fuzzy twin support vector machine based on information entropy for class imbalance learning," *Neural Comput. Appl.*, vol. 24, pp. 1–12, May 2018. doi: 10.1007/s00521-018-3551-9.
[55] D. Gupta and B. Richhariya, "Entropy based fuzzy least squares twin support vector machine for class imbalance learning," *Appl. Intell.*, vol. 48, no. 11, pp. 4212–4231, Nov. 2018.
[56] B. Richhariya and M. Tanveer, "A robust fuzzy least squares twin support vector machine for class imbalance learning," *Appl. Soft Comput.*, vol. 71, pp. 418–432, Oct. 2018.
[57] C. E. Shannon, "A mathematical theory of communication," *ACM Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[58] M. Belkin, P. Niyogi, and V. Sindhwani, ''Manifold regularization: A geometric framework for learning from labeled and unlabeled examples,'' *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.

[59] S. Chen, X. Wu, and J. Xu, ''Locality preserving projection twin support vector machine and its application in classification,'' *J. Algorithms Comput. Technol.*, vol. 10, no. 2, pp. 65–72, Jun. 2016.

[60] C. L. Blake and C. J. Merz. (1998). *UCI Repository of Machine Learning Databases, Department of Information and Computer Science*. [Online]. Available: http://www.ics.uci. edu/ mlearn/MLRepository.html

[61] J. Demšar, ''Statistical comparisons of classifiers over multiple data sets,'' *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.

[62] R. L. Iman and J. M. Davenport, ''Approximations of the critical region of the fbietkan statistic,'' *Commun. Statist.-Theory Methods*, vol. 9, no. 6, pp. 571–595, 1980.

[63] *The USPS Database*. Accessed: Nov. 15, 2018. [Online]. Available: http://www.cs.nyu.edu/~roweis/data.html

[64] S. A. Nene, S. K. Nayar, and H. Murase, ''Columbia object image library (COIL-20),'' Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-006-96, Feb. 1996.

[65] A. M. Martinez and R. Benavente, ''The AR face database,'' *CVC Tech. Rep.*, vol. 24, pp. 1–10, Jun. 1998.

[66] Y. Xu, ''K-nearest neighbor-based weighted multi-class twin support vector machine,'' *Neurocomputing*, vol. 205, pp. 430–438, Sep. 2016.

[67] S.-G. Chen and X.-J. Wu, ''Multiple birth least squares support vector machine for multi-class classification,'' *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 6, pp. 1731–1742, Dec. 2017.

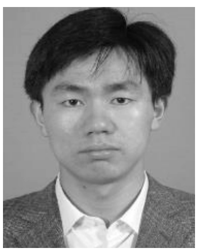
[68] S. Ding, X. Zhang, Y. An, and X. Yu, ''Weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification,'' *Pattern Recognit.*, vol. 67, pp. 32–46, Jul. 2017.

[69] Z. Hanifelou, P. Adibi, S. A. Monadjemi, and H. Karshenas, ''KNN-based multi-label twin support vector machine with priority of labels,'' *Neurocomputing*, vol. 322, pp. 177–186, Dec. 2018.

[70] M. Azad-Manjiri, A. Amiri, and A. S. Sedghpour, ''ML-SLSTSVM: A new structural least square twin support vector machine for multi-label learning,'' *Pattern Anal. Appl.*, vol. 12, pp. 1–14, Feb. 2019. doi: 10.1007/s10044-019-00779-2.

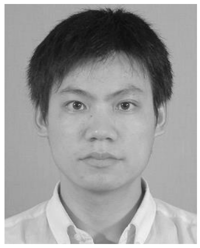
[71] X. Xie, ''Regularized multi-view least squares twin support vector machines,'' *Appl. Intell.*, vol. 48, no. 9, pp. 3108–3115, Sep. 2018.

[72] J. J. Tang, D. W. Li, Y. J. Tian, and D. Liu, ''Multi-view learning based on nonparallel support vector machine,'' *Knowl.-Based Syst.*, vol. 158, pp. 94–108, Oct. 2018.

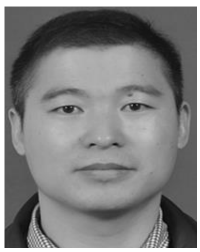


SUGEN CHEN received the B.S. degree in mathematics and applied mathematics from Anqing Normal University, Anqing, Anhui, in 2004, the M.S. degree in computational mathematics from the Hefei University of Technology, Hefei, Anhui, in 2009, and the Ph.D. degree in control science and engineering from Jiangnan University, Wuxi, Jiangsu, in 2016. Since 2015, he has been an Associate Professor with the School of Mathematics and Computational Science, Anqing Normal University. His research interests include pattern recognition and intelligent systems, and machine learning.

JUNFENG CAO received the B.S. degree in mathematical education from Jiangsu Normal University, Xuzhou, China, in 2002, the M.S. degree in fundamental mathematics from Nanjing Normal University, Nanjing, China, in 2008, and the Ph.D. degree in control science and engineering from Jiangnan University, Wuxi, China, in 2017. His research interests include image segmentation and edge detection.



ZHONG HUANG received the B.S. degree in computer science and technology from Anqing Normal University, Anqing, Anhui, China, in 2005, and the M.S. degree in computer software and theory and the Ph.D. degree in computer application technology from the Hefei University of Technology, Hefei, Anhui, China, in 2008 and 2016, respectively. Since 2018, he has been an Associate Professor with the School of Physics and Electrical Engineering, Anqing Normal University. His research interests include humanoid robots, affective computing, and computer vision.



CHUANSHENG SHEN received the Ph.D. degree from the University of Science and Technology of China, in 2012. He was a Visiting Scholar with the Humboldt University of Berlin and also with the Potsdam Institute for Climate Impact Research, sponsored by the China Scholarship Council, from 2016 to 2017. He is currently a Professor of physics and applied mathematics and the Dean of the School of Mathematics and Computational Science, Anqing Normal University, Anhui, China. He has authored more than 30 scientific papers. His research interests include the mesoscopic methods based on statistical mechanics in networked systems, and structure, dynamics, and function of complex networks. Since 2017, he has been a Member Fellow with the China Society for Industrial and Applied Mathematics.