A Novel Dual-Center Based Intuitionistic Fuzzy Twin Bounded Large Margin Distribution Machines

Libo Zhang, *Member, IEEE*, Qiang Jin, *Student Member, IEEE*, Shuangyi Fan and Dun Liu*, *Member, IEEE*

***Abstract*—Intuitionistic fuzzy (IF) set theory combined with twin support vector machines (TSVM) has shown highly advantageous performance in robust and fast classification. However, the existing methods have some inherent flaws: i) The inaccurate confidence caused by the single-center based IF strategy degrades the classification performance of the model. ii) The non-membership degree is time-consuming and the threshold is difficult to determine, as it depends on sample's local neighborhood set. iii) Recent theoretical studies have demonstrated that optimizing the margin distribution achieve better generalization performance than maximizing the minimum margin in SVM-based models (SVMs). In view of the aforementioned shortcomings, we propose a novel dual-center based IF (DC-IF) function, and introduce the margin distribution into IFTSVM, forming a DC-IF twin bounded large margin distribution machines (DC-IFTBLDM). Initially, the DC-IF strategy references the contribution of one sample to both categories, which is time efficient and more accurately reflects the credibility of the samples. Moreover, the margin mean and variance are used to optimize the margin distribution, and a regularization term is utilized to minimize structural risk. Therefore, DC-IFTBLDM has a good performance in noise immunity and classification capability. Comparative experiments based on synthetic datasets, UCI benchmark datasets, and datasets with different noise ratios verify the effectiveness of the proposed model. The parameter sensitivity analysis further proves the high accuracy and stability of IFTBLDM.**

***Index Terms*—Intuitionistic fuzzy, margin distribution, twin support vector machines, anti-noise performance.**

I. Introduction

SUPPORT vector machine (SVM) is a well-known supervised learning method based on statistical theory [1]–[3]. It constructs two parallel hyperplanes and reduces the generalization error by maximizing the margin between the planes [4]. More specifically, SVM maximizes the minimum distance of the sample to the classification boundary, which has received much attention [5]–[7]. To alleviate the large computational burden of SVM, Mangasarian et al. [8] proposed the generalized eigenvalue proximal support vector machine (GEPSVM), which employs two non-parallel hyperplanes. On the basis of GEPSVM, Jayadeva et al. [9] proposed the twin support vector machines (TSVM). By transforming one large quadratic programming problem (QPP) in SVM into two smaller QPPs, TSVM is almost 4 times faster than SVM [10]. Therefore, various modified TSVM models have been developed based on different considerations, such as the Universum based Lagrangian Twin Bounded Support Vector Machine (ULTBSVM) [11], Regularized Universum twin support vector machine for classification of EEG Signal (RUTWSVM) [12], and facial expression recognition based on iterative optimization and dual support vector machines (IUTWSVM) [13], other latest advanced methods are detailed in [14]–[16]. This paper studies how to enhance the performance of the algorithm without knowing the sample distribution. Among them, the noise sensitivity problem is the most profound one, as noise and outliers are unavoidable in real applications [17]–[21]. Specifically, TSVM treats all training samples equally and obtains poor classification accuracy in the noise datasets.

To mitigate the impact of noise and outliers, Lin et al. [22] proposed fuzzy support vector machine (FSVM), which computes the membership degree of an input sample according to its contribution. Gao et al. [23] combined fuzzy membership and TSVM to develop a coordinate descent fuzzy twin support vector machine with high efficiency and reduced noise pollution. On the basis of the probabilistic neighborhood of each sample, Chen et al. [24] constructed an entropy-based fuzzy TSVM model, which achieves better classification results. As a generalization of fuzzy set, intuitionistic fuzzy (IF) set is a more effective tool to describe the vagueness in a range of domains [25]–[27]. Following these works, Rezvani et al. [28] proposed IF twin support vector machines (IFTSVM), which further reduces the influence of noise by utilizing the non-membership. However, the non-membership of a sample is only determined by its membership degree and local neighborhood set. It is difficult to set an appropriate distance threshold for the neighborhood. If the threshold is too small, the neighborhood set is too small to distinguish between support vectors and noise points. Instead, if the threshold is too large, the computation of the neighborhood set will be significantly large. In addition, due to the single-center based strategy, samples with the same degree of membership may still be assigned the same score value, despite their different distances from the center of the other category.

From the perspective of structural risk, Wang and Gao et al. [29], [30] have verified that the margin distribution is more important than minimum margins in optimizing generalization performance. By characterizing the margin distribution in terms of margin mean and margin variance, Zhou et al. [31] proposed the large margin distribution machine (LDM) on

*Corresponding author.

Libo Zhang, Qiang Jin, Shuangyi Fan are with the School of Aritificial Intelligence, Southwest University, Chongqing, 400715, China; (e-mail: lbzhang@swu.edu.cn; qiangjincs@163.com; ShuangyiFan111@163.com) Dun Liu is with the School of Economics and Management, Southwest Jiaotong University, Chengdu, 610031, China. (e-mail: newton83@163.com)

This work is supported by the National Nature Science Foundation of China (Nos. 62276217, 62106205 and 61876157), Natural Science Foundation of Chongqing (No. cstc2021jcyj-msxmX0824), the Youth Project of Science and Technology Research Program of Chongqing Education Commission of China (No. KJQN202100207), the Science Fund for Distinguished Young Scholars of Sichuan Province (No. 2022JDJQ0034).

the basis of SVM. The theoretical advantages and excellent classification performance of LDM have been demonstrated in [32]–[34]. Subsequently, the margin distribution was used in a variety of excellent SVM algorithms, such as double distribution support vector machine (DDSVM) [35], unconstrained large profit distribution machine (ULDM) [36], and optimal profit distribution machine (ODM) [37]. However, these modifications inevitably have a large computational burden. Combining the time advantage of nonparallel support vectors, various LDM models have been developed, such as Nonparallel support vector machine with large margin distribution (LMD-NPSVM) [32], weighted least squares twin LDM (WLSTLDM) [38], twin support vector machines based on adjustable LDM (ALD-TSVM) [39], and twin bounded large margin distribution machine (TBLDM) [40], etc. However, in the existing LDM methods, samples contribute the same to the hyperplanes, while samples in real datasets have different confidence levels because of the noise and other factors.

Inspired by the aforementioned studies, we propose a novel dual-center based intuitionistic fuzzy twin bounded large margin distribution machine (DC-IFTBLDM), which considers the contribution of sample points to both classes. In the proposed dual-center based IF (DC-IF) strategy, the non-membership degree of a sample is determined by its distance to the centers of both categories, which is significantly more efficient than the local and single-based strategy in the IFTSVM. The score function is a weighted sum of the membership and the non-membership function, indicating the importance level of the sample, which helps to improve the noise immunity. Moreover, by adding a regular term, the structure of DC-IFTBLDM is more rigorous. Specifically, the relevant inverse matrix in the dual problem is a non-singular matrix, which can be derived without any additional assumptions. A series of experiments on 30 UCI datasets from multiple domains demonstrate the superiority of the proposed model. In summary, the main contributions of this work are as follows:

- The proposed DC-IF strategy is not only time efficient, but also effectively mitigates the influence of noise and outliers. Specifically, the time complexity of DC-IF is $O(m)$, while the IF strategy in IFTSVM is $O(m^2)$.
- To the best knowledge of the authors, this is the first work to combine IF theory with LDM-based models (LDMs). DC-IFTBLDM utilizes the proposed DC-IF strategy to reflect the confidence levels of samples, thus substantially improving the noise immunity.
- DC-IFTBLDM improves the structure of IFTSVM by adding regular terms and optimizing the margin distribution, which further reduces the structural risk and improves the classification accuracy.

The arrangement of this paper is as follows. In Section II, we review some related work. In Section III, we introduce our intuitionistic fuzzy strategy, as well as the specific derivation and optimization process of DC-IFTBLDM. In Section IV, we conduct a series of experiments using artificial and UCI datasets to demonstrate the superiority of our model. In Section V, we summarize and look forward to our work.

II. Related Work

In this section, TSVM, the existing IF strategies in SVMs and LDM are briefly introduced.

A. TSVM

Suppose $S=\{(x_i,y_i)\mid i=1,2,\cdots,m\}$ is a set of binary classification training samples, where $x_i\in\mathbb{R}^n$ and the class label $y_i=\{-1,+1\}$. Here, data points x_i with label $y_i=1$ and -1 are represented by matrices $A\in\mathbb{R}^{n\times m_+}$ and $B\in\mathbb{R}^{n\times m_-}$, where m_+ and m_- are the numbers of positive and negative classes, respectively. m is the total number of samples. Let $\phi(x_i)$ as the feature maping of x_i. Denote the whole set of input matrix $X=[x_1,\cdots,x_m]$, the sample label column vector $y=[y_1,\cdots,y_m]^T$, a diagonal matrix $Y\in\mathbb{R}^{m\times m}$ with $y_1,\ldots,y_m$ as the diagonal elements and the high-dimensional samples matrix $\phi(X)=[\phi(x_1),\cdots,\phi(x_m)]$. y_A and y_B are the label vectors corresponding to matrix A and B, respectively. Besides, we set the kernel matrix $K=\phi(X)^T\phi(X)\in\mathbb{R}^{m\times m}$, $K_A=\phi(A)^T\phi(X)\in\mathbb{R}^{m_+\times m}$, $K_B=\phi(B)^T\phi(X)\in\mathbb{R}^{m_-\times m}$.

TSVM generates two non-parallel hyperplanes, which ensures each plane is close to one of the two datasets and far from the other. Two non-parallel hyperplanes are denoted as

$$x^Tw_1+b_1=0 \quad \text{and} \quad x^Tw_2+b_2=0. \tag{1}$$

TSVM transforms a large-scale QPP in SVM into solving the following pair of QPPs [9]:

$$\begin{aligned}\min_{w_1,b_1,q}\ & \frac{1}{2}(Aw_1+e_1b_1)^T(Aw_1+e_1b_1)+c_1e_2^Tq\\ s.t.\ & -(Bw_1+e_2b_1)+q\ge e_2,\quad q\ge 0,\end{aligned} \tag{2}$$

and

$$\begin{aligned}\min_{w_2,b_2,q}\ & \frac{1}{2}(Bw_2+e_2b_2)^T(Bw_2+e_2b_2)+c_2e_1^Tq\\ s.t.\ & (Aw_2+e_1b_2)+q\ge e_1,\quad q\ge 0,\end{aligned} \tag{3}$$

where c_1 and c_2 are the penalty parameters; e_1 and e_2 are all-ones vectors of the appropriate size; q is a slack variable of the appropriate size. And according to the Kuhn-Tucker conditions (KKT), the dual problem of TSVM is

$$\begin{aligned}\min_{\alpha}\ & \frac{1}{2}\alpha^TQ\left(P^TP\right)^{-1}Q^T\alpha-e_2^T\alpha\\ s.t.\ & 0\le\alpha\le c_1,\end{aligned} \tag{4}$$

and

$$\begin{aligned}\min_{\beta}\ & \frac{1}{2}\beta^TP\left(Q^TQ\right)^{-1}P^T\beta-e_2^T\beta\\ s.t.\ & 0\le\beta\le c_2,\end{aligned} \tag{5}$$

where α and β are the Lagrangian parameters; $P=[A\quad e_1]$, $Q=[B\quad e_2]$.

B. The IF Strategies in SVMs

According to the IF theory, there are three parameters in an IF number (IFN): the membership degree, non-membership degree and the underlying hesitation degree, which is denoted as μ $(0\le\mu\le 1)$, ν $(0\le\nu\le 1)$ and $\pi=1-\mu-\nu$, respectively [28]. If the hesitation degree is $\pi = 0$, an IFN will degenerate into an ordinary fuzzy number (FN) [41].

The fuzzy membership function in FTSVM [23] is related to the distance between the training sample and its own class center, which is described as

$$\mu(x_i) = \begin{cases} 1 - \dfrac{\|\phi(x_i) - C_+\|}{r_+ + \delta}, y_i = +1 \\ 1 - \dfrac{\|\phi(x_i) - C_-\|}{r_- + \delta}, y_i = -1, \end{cases} \tag{6}$$

where $\delta > 0$ is defined as a small constant. $r_+ (r_-)$ and $C_+ (C_-)$ are the radius and class center of the positive (negative) class, which are represented as follows:

$$\begin{cases} C_{\pm} = \dfrac{1}{m_{\pm}} \sum\limits_{y_i = \pm 1} \phi(x_i) \\ r_{\pm} = \max\limits_{y_i = \pm 1} \left\|\phi(x_i) - C^{\pm}\right\|, \end{cases} \tag{7}$$

where $m_+ (m_-)$ is total number of positive (negative) samples.

IFTSVM constructs the IF strategy by utilizing the relationship between all inharmonious points and the total number of training samples in their local neighborhood [28]. The non-membership function $\nu(x_i)$ and the local neighbourhood set $\rho(x_i)$ are described below:

$$\begin{cases} \nu(x_i) = (1 - \mu(x_i))\,\rho(x_i) \\ \rho(x_i) = \dfrac{|\{x_j \mid \|\phi(x_i) - \phi(x_j)\| \le \alpha, y_j \ne y_i\}|}{|\{x_j \mid \|\phi(x_i) - \phi(x_j)\| \le \alpha\}|}. \end{cases} \tag{8}$$

The score function in IFTSVM is calculated as follows:

$$s_i = \begin{cases} \mu_i, & \nu_i = 0 \\ 0, & \mu_i \le \nu_i \\ \dfrac{1 - \nu_i}{2 - \mu_i - \nu_i}, & \text{others.} \end{cases} \tag{9}$$

C. LDM

The margin mean can be expressed as

$$\bar{\gamma} = \frac{1}{m} \sum_{i=1}^{m} \gamma_i = \frac{1}{m} (Xy)^T w. \tag{10}$$

Calculate the margin variance as explained below:

$$\hat{\gamma} = \frac{2}{m} w^T X X^T w - \frac{2}{m^2} w^T X y y^T X^T w. \tag{11}$$

The original LDM with soft margin is as follows:

$$\begin{aligned} \min_{w,\xi}\ & \frac{1}{2}\omega^T\omega + \lambda_1\hat{\gamma} - \lambda_2\bar{\gamma} + C\sum_{i=1}^{m}\xi_i \\ s.t.\ \ & y_i\omega^T\phi(x_i) + \xi_i \ge 1, \xi_i \ge 0, i = 1, \ldots, m, \end{aligned} \tag{12}$$

where C is a parameter that weighs the total error; ξ_i is the slack variable. By substituting Eqs. (10) and (11) into Eq. (12), the original problem of LDM is rewritten as follows:

$$\begin{aligned} \min_{w,\xi}\ & \frac{1}{2}w^T w + \frac{2\lambda_1}{m^2}\left(m w^T X X^T w - w^T X y y^T X^T w\right) \\ & - \lambda_2 \frac{1}{m}(Xy)^T w + C\sum_{i=1}^{m}\xi_i \\ s.t.\ \ & y_i w^T\phi(x_i) + \xi_i \ge 1, \xi_i \ge 0, i = 1, \ldots, m. \end{aligned} \tag{13}$$

LDM achieves better generalization performance by optimizing margin distribution, rather than the minimum margin in SVM. By maximizing the margin mean and margin variance, all samples contribute to the optional hyperplane.

III. DC-IFTBLDM

In this section, a novel DC-IFTBLDM model is proposed. Specifically, an analysis of the existing IF strategies are provided in Section III-A; A novel dual-center IF strategy is introduced in Section III-B; The model construction is presented in detail in Section III-C; The optimization procedure is given in Section III-D; Finally, Section III-E presents the algorithm and the time complexity analysis.

A. The Insufficiency of the Existing IF Strategies

Fig. 1 shows two types of points, where points A, B, and C are the same distance from the center of the positive class center. According to Eq. (6), points A, B, and C have the same membership, which indicates that they should have the same level of credibility. However, points B and C are closer to other categories than point A, and they are more likely to be noise spots. Points D and E also suffer from the same problem.

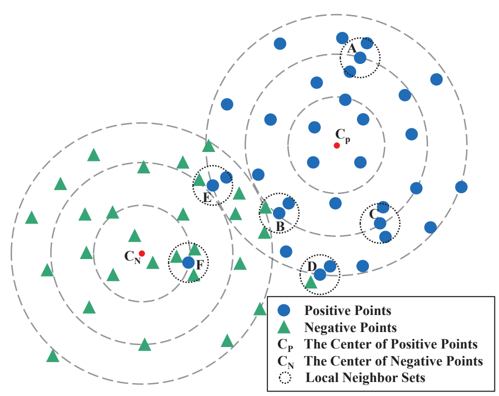


Fig. 1. Analysis of the IF strategies in existing models. In FTSVM, A, B, and C have the same membership degree, and so do D and E. In IFTSVM, the non-membership degrees of A and B are different. However, A and C still have the same non-membership degree, as do D and E.

According to the score function in Eq. (9) and Fig. 1, A and B have different score values, due to the diverse inharmonious proportions in their local neighborhood set. Specifically, F is classified as noise, because the proportion of heterogeneous points around it is high, and B is considered a support vector since there are many homogeneous points around it. However, if the threshold α is comparatively small, A and C will have the same score values, since there are no outliers in the local neighborhood set. Compared to A, C is closer to the negative class. In addition, D and E may be both identified as support vectors, and assigned the same score value, although E is mixed with negative points while D is not.

Remark 1. *Due to the uncertainty and inevitable noise in practical applications, samples always have different confidence levels. IF set theory plays a key role in robust classification learning by assigning different confidence levels* *to samples. The IF strategy in IFTSVM further alleviates the influence of noise by introducing the non-membership degree, which is the state-of-the-art strategy. However, the non-membership degree in existing IF methods adopts time-consuming local strategies, and the threshold is difficult to determine. Furthermore, the single-center strategy also limits the anti-noise performance of IFTSVM.*

B. Dual-Center Based IF Strategy

In this subsection, we propose a novel DC-based IF strategy that considers the contribution to both classes.

1) Membership Function: The membership degree of a sample is computed based on the distance to the center of its own category. The class centers of the positive class and the negative class are denoted as C_P and C_N respectively. The distance of the samples to the positive or negative class center are presented as follows:

$$\begin{cases} P_i(C_P) = \|\phi(x_i) - C_P\|^2, y_i = +1 \\ P_i(C_N) = \|\phi(x_i) - C_N\|^2, y_i = +1 \\ N_i(C_N) = \|\phi(x_i) - C_N\|^2, y_i = -1 \\ N_i(C_P) = \|\phi(x_i) - C_P\|^2, y_i = -1. \end{cases} \tag{14}$$

The defined square of the Euclidean distance has two advantages: enlarging the distance between sample points to obtain a more accurate membership degree and constructing a kernel function in the formula to facilitate calculation.

Then the scattering hypersphere radii can be obtained by

$$\begin{cases} R_P = \max P_i(C_P), y_i = +1 \\ R_N = \max N_i(C_N), y_i = -1. \end{cases} \tag{15}$$

For each sample, the membership degree can be described as

$$\mu(x_i) = \begin{cases} 1 - P_i(C_P)/(R_P + \delta), y_i = +1 \\ 1 - N_i(C_N)/(R_N + \delta), y_i = -1, \end{cases} \tag{16}$$

where $\delta > 0$ is a parameter avoiding the vanishing of $\mu(x_i)$.

2) Non-membership Function: The non-membership degree is computed based on the membership degree and the distance to the center of the other category, rather than the local strategy in the IFTSVM. The formula is as follows:

$$\nu(x_i) = \begin{cases} \dfrac{1/P_i(C_N)}{(1/P_i(C_P) + 1/P_i(C_N))}, y_i = +1 \\ \dfrac{1/N_i(C_P)}{(1/N_i(C_N) + 1/N_i(C_P))}, y_i = -1. \end{cases} \tag{17}$$

The score function is defined as a weighted sum of the membership and the non-membership function, indicating the credibility of the sample. The score function is expressed as

$$s_i = \left(w_1 \mu\left(x_i\right)^p + \left(1 - w_1\right)\left(1 - v\left(x_i\right)^p\right)\right)^{\frac{1}{p}}, \tag{18}$$

where p is a positive parameter, which is adjusted to distinguish two intuitive fuzzy values. $w_1 \in [0, 1]$, which is a parameters that weigh the membership and non-membership degree, depending on the preference of the decision maker.

In this paper, the membership and non-membership functions are established based on the inner product distance in the high-dimensional space mapped by the kernel function k.

Theorem 1. *In Eq. (14), the distance of the sample to the center of the positive or negative class is calculated by the kernel function as follows:*

$$P_i(C_P) = k\left(x_i, x_i\right) - \frac{2}{m_-} k\left(x_i, x_i\right) + \frac{1}{{m_+}^2} \sum_{y_a=1} \sum_{y_b=1} k\left(x_a, x_b\right).$$

Proof.

$$P_i(C_P) = \|\phi(x_i) - C_P\|^2 = \left\|\phi(x_i) - \frac{1}{m_+} \sum_{y_j=1} \phi(x_j)\right\|^2$$
$$= \left(\phi(x_i) - \frac{1}{m_+} \sum_{y_j=1} \phi(x_j)\right) \cdot \left(\phi(x_i) - \frac{1}{m_+} \sum_{y_j=1} \phi(x_j)\right)$$
$$= \phi\left(x_i\right) \cdot \phi\left(x_i\right) - \frac{2}{m_+} \phi\left(x_i\right) \cdot \sum_{y_j=1} \phi\left(x_j\right) +$$
$$\frac{1}{{m_+}^2} \sum_{y_j=1} \phi\left(x_j\right) \cdot \sum_{y_j=1} \phi\left(x_j\right)$$
$$= k\left(x_i, x_i\right) - \frac{2}{m_+} \sum_{y_j=1} k\left(x_i, x_j\right) + \frac{1}{{m_+}^2} \sum_{y_a=1} \sum_{y_b=1} k\left(x_a, x_b\right).$$

□

Similarly, the other distances can also be computed by using Theorem 1. Therefore, k can be utilized to compute IFNs.

C. Model Construction of DC-IFTBLDM

DC-IFTBLDM seeks for a pair of nonparallel hyper planes $f_{(1)}(x) = w_{(1)}^T \phi(x) = 0$ and $f_{(2)}(x) = w_{(2)}^T \phi(x) = 0$ (Inspired by [31], the bias b does not affect the overall derivation process). The margin of a single sample is

$$\begin{cases} \gamma_+^i = y_{(1)}^i f_{(2)}\left(x_{(1)}^i\right) = y_{(1)}^i w_{(2)}^T \phi\left(x_{(1)}^i\right), i = 1, \cdots, m_+ \\ \gamma_-^i = y_{(2)}^i f_{(1)}\left(x_{(2)}^i\right) = y_{(2)}^i w_{(1)}^T \phi\left(x_{(2)}^i\right), i = 1, \cdots, m_-. \end{cases} \tag{19}$$

The margin mean and margin variance are as follows:

$$\begin{cases} \bar{\gamma}_+ = \dfrac{1}{m_+} \displaystyle\sum_{i=1}^{m_+} \gamma_+^i, \hat{\gamma}_+ = \dfrac{1}{m_+} \displaystyle\sum_{i=1}^{m_+} \left(\gamma_+^i - \bar{\gamma}_+\right)^2 \\ \bar{\gamma}_- = \dfrac{1}{m_-} \displaystyle\sum_{i=1}^{m_-} \gamma_-^i, \hat{\gamma}_- = \dfrac{1}{m_-} \displaystyle\sum_{i=1}^{m_-} \left(\gamma_-^i - \bar{\gamma}_-\right)^2. \end{cases} \tag{20}$$

Bring (19) into (20), the matrix form is expressed as

$$\begin{cases} \bar{\gamma}_+ = \dfrac{1}{m_+} y_A^T \phi(A)^T w_{(2)}, \hat{\gamma}_+ = w_{(2)}^T \phi(A) Q_1 \phi(A)^T w_{(2)} \\ \bar{\gamma}_- = \dfrac{1}{m_-} y_B^T \phi(B)^T w_{(1)}, \hat{\gamma}_- = w_{(1)}^T \phi(B) Q_2 \phi(B)^T w_{(1)}, \end{cases} \tag{21}$$

where the symmetric matrices Q_1 and Q_2 are as follows:

$$Q_1 = \frac{m_+ I_{m_+} - y_A y_A^T}{m_+^2}, \quad Q_2 = \frac{m_- I_{m_-} - y_B y_B^T}{m_-^2}, \tag{22}$$

where I_{m_+} and I_{m_-} are the all-ones matrix of the apt size.

To guarantee that the matrices in the DC-IFTBLDM dual problem are non-singular, we add a regularization term to maximize some margin. The formula is described as

$$\frac{c_i}{2}\|w_{(i)}\|^2, i=1,2. \tag{23}$$

By utilizing Eq. (23), the DC-IFTBLDM dual problem can be derived without any additional assumptions and modifications.

Most practical datasets are linearly inseparable in the original space. Using the kernel function to project the samples from the original space to the high-dimensional space, a nonlinear DC-IFTBLDM model is constructed.

$$\begin{aligned} \min_{w_{(1)},\xi_1} \ & \frac{1}{2}\|\phi\left(A\right)^T w_{(1)}\|^2 + \frac{c_1}{2}\|w_{(1)}\|^2 + \lambda_1 \hat{\gamma}_- \\ & - \lambda_3 \bar{\gamma}_- + c_3 s_2^T \xi_2 \\ s.t. \quad & -\phi\left(B\right)^T w_{(1)} + \xi_2 \geq e_2, \xi_2 \geq 0, \end{aligned} \tag{24}$$

and

$$\begin{aligned} \min_{w_{(2)},\xi_2} \ & \frac{1}{2}\|\phi\left(B\right)^T w_{(2)}\|^2 + \frac{c_2}{2}\|w_{(2)}\|^2 + \lambda_2 \hat{\gamma}_+ \\ & - \lambda_4 \bar{\gamma}_+ + c_4 s_1^T \xi_1 \\ s.t. \quad & -\phi\left(A\right)^T w_{(2)} + \xi_1 \geq e_1, \xi_1 \geq 0, \end{aligned} \tag{25}$$

where $\lambda_1, \lambda_2, \lambda_3$, and λ_4 are hyperparameters, which trade-off the complexity of models. The second term of the objective function is the regularization term in Eq. (23), which implements the structural risk minimization principle. c_1 and c_2 are the hyperparameters corresponding to the regularization term. $s_1 \in \mathbb{R}^{m+}$ and $s_2 \in \mathbb{R}^{m-}$ are both the score value vectors. c_3 and c_4 are parameters that weigh the total error, and the variables $\xi_i (i = 1, 2)$ in Eqs. (24) and (25) are constructed for the point x_i that does not satisfy hard-margin partition constraints. Therefore, the term ${s_2}^T \xi_2$ and ${s_1}^T \xi_1$ can be considered as some measure of misclassification.

Remark 2. *Note that the proposed DC-IFTBLDM model is quite general and complete. According to Eqs. (24) and (25), we have the following observations.*

- *If $s_1 = e_1, s_2 = e_2$, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ = 0, the DC-IFTBLDM model degenerates into TBSVM. The proposed model introduces the IF strategy, the margin mean and variance into TBSVM to enhance the classification performance.*
- *If $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ = 0, Eq. (18) is replaced by Eq. (9), the DC-IFTBLDM model degenerates into IFTSVM. The proposed model adopts a global and DC-based IF strategy, which overcomes the inherent shortcomings of the local and single-center based IF strategy in existing IF SVMs.*
- *If $s_1 = e_1, s_2 = e_2$, the DC-IFTBLDM model degenerates into TBLDM. Further, if $c_1, c_2 = 0$, it degenerates into TLDM. In our model, the IF strategy is utilized to capture data uncertainty, and the regularization term reduces the effects of noise and outliers in the dataset, thereby improving generation performance.*

D. The Optimization and Dual Problem

Different from the derivation of traditional TSVMs, we derive the dual formulation of convex models in a more concise and compact way in this subsection.

Since DC-IFTBLDM has perfect symmetry, we only need to derive one set of equations, and the derivation process for the other set is similar. Bring Eq. (21) into Eq. (24), we can get the following:

$$\begin{aligned} \min_{w_{(1)},\xi_2} \ & \frac{1}{2}\|\phi(A)^T w_{(1)}\|^2 + \frac{c_1}{2}\|w_{(1)}\|^2 + c_3 s_2^T \xi_2 \\ & + \lambda_1 w_{(1)}^T \phi(B) Q_2 \phi(B)^T w_{(1)} - \frac{\lambda_2}{m_-} y_B^T \phi(B)^T w_{(1)} \\ s.t. \quad & -\phi(B)^T w_{(1)} + \xi_2 \geq e_2, \xi_2 \geq 0. \end{aligned} \tag{26}$$

Theorem 2. *The optimal solution ${w_{(1)}}^*$ and ${w_{(2)}}^*$ of Eqs. (24) and (25) can be expressed succinctly as $w_{(1)} = \phi(X)\beta_1$, and $\beta_1 \in \mathbb{R}^m$ is a coefficient vector.*

Proof. w^* can be decomposed into a part that lives in the span of $\phi(x_i)$ and an orthogonal part vector, i.e.,

$$w_{(1)} = \sum_{i=1}^{m} \beta_1 \phi\left(x_i\right) + \eta = \phi\left(X\right)\beta_1 + \eta, \tag{27}$$

where η is a vector, which satisfies $\phi\left(X\right)^T \eta = 0$ [31].

We note that the first, fourth and fifth terms in the objective function of Eq. (26) are independent of η, and the third term is independent of η. In addition, the first term in the constraint is also independent of η. Next, the second item in the objective function is concerned.

$$\begin{aligned} \left\|w_{(1)}\right\|^2 &= {w_{(1)}}^T w_{(1)} \\ &= \left(\phi\left(X\right)\beta_1 + \eta\right)^T \left(\phi\left(X\right)\beta_1 + \eta\right) \\ &= {\beta_1}^T \phi\left(X\right)^T \phi\left(X\right)\beta_1 + \eta^T \eta \\ &\geq {\beta_1}^T \phi\left(X\right)^T \phi\left(X\right)\beta_1, \end{aligned} \tag{28}$$

where the equal sign holds if and only if η = 0.

So, setting η to 0 does not affect the other terms, but strictly reduces the second term of the objective function. Therefore, Theorem 2 is accepted. □

Based on Theorem 2, the transformation of the original problem into the Wolfe dual problem can be derived.

Theorem 3. *The original problem Eq. (26) can be transformed into the Wolfe dual problem as follows:*

$$\begin{aligned} \min_{\alpha_1} \ & \frac{1}{2}\alpha_1^T H_1 \alpha_1 - \left(\frac{\lambda_1}{m_-} H_1 y_B + e_2\right)^T \alpha_1 \\ s.t. \ & 0 \leq \alpha_1 \leq c_3 s_2, \end{aligned} \tag{29}$$

where $H_1 = K_B(c_1 K + K_A^T K_A + 2\lambda_2 K_B^T Q_2 K_B)^{-1} K_B^T$; α_1 is a Lagrangian multipliers vector. And β_1 has the following representation:

$$\beta_1 = G_1^{-1}\left(\frac{\lambda_1}{m_-} K_B^T y_B - K_B^T \alpha_1\right). \tag{30}$$

Proof. According to Theorem 2, we can deduce the following formulas:

$$\begin{cases} \left\|w_{(1)}\right\|^2 = \beta_1^T K \beta_1, & \left\|w_{(2)}\right\|^2 = \beta_2^T K \beta_2 \\ \phi(A)^T w_{(1)} = K_A \beta_1, & \phi(B)^T w_{(1)} = K_B \beta_1 \\ \phi(A)^T w_{(2)} = K_A \beta_2, & \phi(B)^T w_{(2)} = K_B \beta_2. \end{cases} \tag{31}$$

Substituting Eq. (31) into Eq. (26), we can obtain the following final matrix form:

$$\begin{aligned}\min_{\beta_1,\xi_2}\ & \frac{1}{2}\beta_1^T K_A^T K_A\beta_1+\frac{c_1}{2}\beta_1^T K\beta_1-\frac{\lambda_2}{m_-}y_B^T K_B\beta_1\\ & +\lambda_1\beta_1^T K_B^T Q_2 K_B\beta_1+c_3 s_2^T\xi_2\\ s.t.\ & -K_B\beta_1+\xi_2\geq e_2,\xi_2\geq 0.\end{aligned}\tag{32}$$

And Eq. (32) can be written in a more concise form as shown below:

$$\begin{aligned}\min_{\beta_1,\xi_2}\ & \frac{1}{2}\beta_1^T G_1\beta_1-\frac{\lambda_1}{m_-}y_B^T K_B\beta_1+c_3 e_2^T\xi_2\\ s.t.\ & -K_B\beta_1+\xi_2\geq s_2,\xi_2\geq 0,\end{aligned}\tag{33}$$

where $G_1=c_1K+K_A^TK_A+2\lambda_2K_B^TQ_2K_B\in\mathbb{R}^{m\times m}$, and is a symmetric non-negative definite matrix.

The Lagrangian function of the optimization problem in Eq. (33) is presented as

$$\begin{aligned}L_1(\beta_1,\xi_2,\alpha_1,\delta_1)=&\frac{1}{2}\beta_1^T G_1\beta_1-\frac{\lambda_1}{m_-}y_B^T K_B\beta_1+c_3 s_2^T\xi_2\\ &-\alpha_1^T(-K_B\beta_1+\xi_2-e_2)-\delta_1^T\xi_2,\end{aligned}\tag{34}$$

where $\alpha_1,\delta_1\in\mathbb{R}^{m_-}$ are Lagrangian multipliers vectors. Let $\partial L_1/\partial\beta_1=\partial L_1/\partial\xi_2=0$, we obtain the following formulas:

$$\begin{aligned}G_1\beta_1&=\frac{\lambda_1}{m_-}K_B^Ty_B-K_B^T\alpha_1,\\ c_3e_2-\alpha_1-\delta_1&=0\Rightarrow 0\leq\alpha_1\leq c_3e_2.\end{aligned}\tag{35}$$

Since G_1 is non-singular, β_1 can be deduced from Eq. (35) as mentioned below:

$$\beta_1=G_1^{-1}\left(\frac{\lambda_1}{m_-}K_B^Ty_B-K_B^T\alpha_1\right).\tag{36}$$

Submitting Eq. (35) and Eq. (30) into the Lagrangian function Eq. (34), the Wolfe dual form of the model Eq. (33) is presented as follows:

$$\begin{aligned}\min_{\alpha_1}\ & \frac{1}{2}\alpha_1^T H_1\alpha_1-\left(\frac{\lambda_1}{m_-}H_1y_B+e_2\right)^T\alpha_1\\ s.t.\ & 0\leq\alpha_1\leq c_3s_2,\end{aligned}\tag{37}$$

where $H_1=K_BG_1^{-1}K_B^T$. □

Similarly, according to the other set of equations Eq. (21) and Eq. (25) of DC-IFTBLDM, we derive the Wolfe dual form of the other model Eq. (25) using Theorem 3 as indicated below:

$$\begin{aligned}\min_{\alpha_2}\ & \frac{1}{2}\alpha_2^T H_2\alpha_2+\left(\frac{\lambda_3}{m_+}H_2y_A-e_1\right)^T\alpha_2\\ s.t.\ & 0\leq\alpha_2\leq c_4e_1,\end{aligned}\tag{38}$$

where $\alpha_2\in\mathbb{R}^{m_+}$ is a non-negative Lagrangian multipliers vector, and $H_2=K_AG_2^{-1}K_A^T$. And the β_2 is expressed as

$$\beta_2=G_2^{-1}\left(\frac{\lambda_2}{m_+}K_A^Ty_A+K_A^T\alpha_2\right),\tag{39}$$

where $G_2=c_2K+K_B^TK_B+2\lambda_4K_A^TQ_1K_A\in\mathbb{R}^{m\times m}$ is a symmetric non-negative definite matrix.

Theorem 4. *By utilizing Theorem 3, we obtain the Wolfe dual problems and the parameter vectors β_1 and β_2 related to the optimal hyperplanes. Then, the expression for the decision function of DC-IFTBLDM can be constructed as follows:*

$$f(x)=\arg\min_{i=1,2}\frac{|K(x,X)\beta_i|}{\sqrt{\beta_i^TK\beta_i}},\tag{40}$$

where the kernel matrix $K(x,X)=[k(x,x_1),\cdots,k(x,x_m)]\in\mathbb{R}^{1\times m}$.

Proof. Obviously, the decision function is defined as the class of the hyperplane that is closer to the input point, and the concrete proof is similar to the traditional TSVMs [9]. □

Finally, the label value of the test sample point $\boldsymbol{x}\in\mathbb{R}^n$ can be predicted using the decision function in the Theorem 4.

Remark 3. *After deriving the Wolfe dual problems, we have the following analysis:*

- *The DC-IF strategy more accurately reflects the confidence of the samples and reduces the effect of noise. To the best knowledge of authors, this is the first time that the large margin distribution is embedded in IFTSVM, which further improves the generalization performance and classification accuracy. In Section IV-C and IV-D, DC-IFTBLDM achieves the best classification accuracy, which demonstrates its excellent anti-noise performance.*
- *According to the Eqs. (24) and (25), the parameters c_3 and c_4 are multiplied by ${s_2}^T\xi_2$ and ${s_1}^T\xi_1$, weighing the overall loss of all samples. The proposed DC-IF strategy is based on a more accurate score value of each sample, which is less influenced by the parameters c_3 and c_4. In Section IV-F, the DC-IFTBLDM model achieves better stability and higher classification accuracy than other algorithms regardless of parameter tuning of c_3 and c_4.*

E. Complexity Analysis of the DC-IFTBLDM

The detailed steps of DC-IFTBLDM are shown in Algorithm 1. Suppose the number of samples in each class $m_+=m_-=m/2$. We analyze the complexity of DC-IFTBLDM by utilizing the big-O notation [42].

TABLE I
THE COMPUTATIONAL COMPLEXITY OF DIFFERENT IF STRATEGIES

time complexity	IF [28]		DC-IF	
	single point	all points	single point	all points
membership	$O(m)$	$O(m)$	$O(m)$	$O(m)$
non-membership	$O(m)$	$O(m^2)$	$\boldsymbol{O(1)}$	$\boldsymbol{O(m)}$
score	$O(1)$	$O(m)$	$O(1)$	$O(m)$
total	$O(m)$	$O(m^2)$	$\boldsymbol{O(m)}$	$\boldsymbol{O(m)}$

The best result is shown in bold (similarly hereinafter).

The computational complexity of IF strategy in IFTSVM [28] and our method are shown in Table I. We can learn that the time complexity of the membership degree and the score function in both methods is the same, while the complexity of the non-membership degree in IFTSVM is much higher than

Algorithm 1 DC-IFTBLDM

Input: samples matrix A and B, label matrix y_A and y_B, kernel matrix K, K_A and K_B.
Output: predict label of data x.

Compute R_P and R_N by Eq. (15).
$\mu(x_i) \leftarrow$ Eqs. (16), $\nu(x_i) \leftarrow$ Eqs. (17), $\forall i = 1, 2, ..., m$.
Obtain the score value s_i by utilizing Eq. (18).
while Parameters are not fully traversed **do**
  Update $c_i (i = 1, 2, 3, 4)$ and $\lambda_i (i = 1, 2, 3, 4)$.
  Calculate Q_1 and Q_2 by using Eq. (22).
  $H_1 \leftarrow K_B(c_1 K + K_A^T K_A + 2\lambda_2 K_B^T Q_2 K_B)^{-1} K_B^T$.
  $H_2 \leftarrow K_A(c_2 K + K_B^T K_B + 2\lambda_4 K_A^T Q_1 K_A)^{-1} K_A^T$.
  repeat
    Update variables α_1 in Eq. (29) and α_2 in Eq. (37) using the QPP toolkit.
  until convergenc
  $\beta_1 \leftarrow \alpha_1$ by Eq. (30), $\beta_2 \leftarrow \alpha_2$ by Eq. (39).
  Calculate the label value by Eq. (40) and accuracy.
end while
Acquire the best parameter combination and receive the optimal model.
Calculate the label value of the test sample x by Eq. (40).

that in our DC-IF method. Therefore, the constructed DC-IF strategy significantly more time-efficient than that in IFTSVM.

In traditional SVM, the computational complexity of converting the original optimization problem into a dual form is $O(m^2)$, while that of solving the dual problem is $O(m^3)$ [43]. The complexity of FSVM is slightly larger, as it needs extra time to compute the fuzzy membership [22]. By converting the whole optimization problem into two small-scale problems, the computational complexity of TSVM is reduced to $O(1/4 \times m^3)$ [9]. For LDM, it takes extra time to compute the first and second-order statistics [31]. Specifically, it takes $O(m^3)$ to turn the original problem into a dual form, while the complexity of solving the dual problem is the same with SVM. The complexity of FLDM is slightly larger than LDM, as it needs extra time to compute the fuzzy membership [44]. By utilizing a twin strategy, the time complexity of our method is roughly 1/4 of the above two LDM algorithms.

IV. Experimental Results and Analysis

In this section, we conduct a series of experiments on artificial datasets and UCI benchmark datasets from several domains. The experimental settings are presented in Section IV-A. In Section IV-B, the effect of the proposed IF strategy in recognizing noise is visually demonstrated on artificial datasets. Experimental results on 30 UCI datasets are analyzed in Section IV-C. In order to further verify the anti-noise performance of the model, we conduct comparative experiments on UCI datasets polluted by different scales of white Gaussian noise, shown in Section IV-D. In Section IV-E, we explore the improvement effect of different IF strategies for TBLDM. Finally, the sensitivity of the model to parameters is analyzed in Section IV-F. The code of this paper can be obtained from "https://github.com/Keys015/DC-IFTBLDM".

A. Experimental Settings

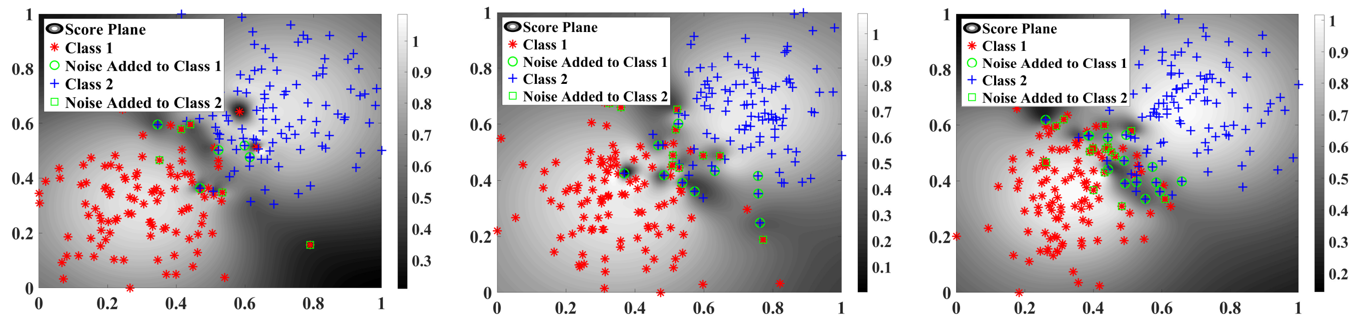
To explore the performance of the proposed DC-IFTBLDM, we conducted experiments on 30 UCI datasets from different domains, which is shown in Table II. For the regular datasets, when randomly divided, 70% is used as the training set, and the remaining 30% is the test set. For 6 larger-scale datasets, inspired by [45], we select 10% of the datasets to choose the best parameters. Five-fold cross-validation and grid optimization were used to select optimal parameters. For experiments on UCI datasets, Gaussian kernel is adopted. All algorithms are implemented on a PC with 8 × 4.00 GHz CPU and 32GB memory using MATLAB R2018a.

TABLE II
Details of the UCI Datasets.

Category	Dataset	Samples×Attributes	Abbr.
Algorithms	Monk-1	556 ×7	A_1
	Monk-2	601 × 7	A_2
	Monk-3	554 ×7	A_3
	Votes	435 ×16	A_4
Biological	Promoters	106×57	B_1
	Iris	150×4	B_2
	Wine	178×13	B_3
	Haberman	306×3	B_4
Disease	Breast Cancer	277 ×9	D_1
	WDBC	569×30	D_2
	WPBC	198×33	D_3
	ecoli-data	327×7	D_4
Medical	New-thyroid	215× 5	M_1
	Echocardiogram	131 ×9	M_2
	Statlog(Heart)	270 ×13	M_3
	Hepatitis	155 ×19	M_4
Physics	Ionosphere	351× 34	P_1
	Glass	214 ×9	P_2
	Sonar	208 ×60	P_3
	Spect	267 ×22	P_4
Others	Australian	690 ×14	O_1
	Credit-a	653 ×15	O_2
	Plrx	182 ×12	O_3
	Clean1	476 ×166	O_4
Large	waveform(0-1)	3345 ×40	L_1
	waveform(0-2)	3347 ×40	L_2
	waveform(1-2)	3308 ×40	L_3
	credit-g	1000 ×20	L_4
	kr-vs-kp	3196 ×36	L_5
	CMC	1473 ×9	L_6

Abbr. stands for abbreviation.

DC-IFTBLDM is compared with 9 other algorithms, including benchmark and state-of-the-art algorithms: IFTSVM [28], CDFTSVM [23], TBSVM [16], FLDM [44] and etc. The penalty parameter c (c_1, c_2, c_3, c_4) in the soft margin model has the greatest impact on the SVMs and LDMs. Specifically, there is only one parameter c in SVM and LDM, while there are a pair of soft parameters in twin variant models. Without loss of generality, we assume that $c_1 = c_2, c_3 = c_4, \lambda_1 = \lambda_3, \lambda_2 = \lambda_4$ in DC-IFTBLDM. Similarly, in IFTSVM, CDFTSVM and TBSVM, we set $c_1 = c_2, c_3 = c_4$. In TSVM and FTSVM, $c_1 = c_2$ is set. The model parameters



(a) Data with 10 noise points (b) Data with 20 noise points (c) Data with 30 noise points

Fig. 2. Score plane diagram for different noise quantities (Confidence plane plots for different amounts of noise. The darker the color of the area around the point in the figure indicates that it has a lower score value, which indicates its identity as a noisy point. As the amount of noise added increases, the more black areas in the graph).

$c_i(i = 1, 2, 3, 4)$ are chosen from from $\{2^i | i = -8, ..., 4\}$, and the kernel parameter σ is from the $\{2^i | i = -5, ..., 2\}$. For uniformity, the QP technique is chosen to solve the optimization problem of all algorithms.

Statistical analysis forms an important tool to compare performances of multiple algorithms over multiple datasets and check if the performance differences are random or statistically significant [46]. The use of performance values in the statistical analysis may create limitations and doubts regarding the commensurability of the performance measure [47]. It is a more feasible approach to rank the classifiers based on the performance of each dataset and then use these rankings for statistical analysis. In this paper, to further analyze the performance of DC-IFTBLDM, we performed two types of statistical analysis, Friedman test [47] and Wilcoxon signed-rank test (WST) [48] with post-hoc analysis.

TABLE III
PARAMETER SETTINGS FOR SYNTHETIC DATASETS WITH GAUSSIAN NOISE.

Type	Parameter	Setting
Normal	Mean of Class 1	[0.4, 0.4]
	Mean of Class 2	[1.7, 1.7]
	Covariances of Class 1 or 2	[0.2, 0; 0, 0.3]
	Numbers of Class 1 or 2	100
Noise	Mean of Class 1 or 2	[1, 1]
	Covariances of Class 1 or 2	[0.15, -0.1; -0.1, 0.15]
	Numbers of Class 1 or 2	5; 10; 15

B. Experiments on Artificial Datasets

The Gaussian distribution is utilized to generate a two-dimensional artificial datasets : x_i $(i \in \mathbf{P}) \sim \mathcal{N}(\mu_1, \Sigma_1)$, x_i $(i \in \mathbf{N}) \sim \mathcal{N}(\mu_2, \Sigma_2)$. To exhibit the effectiveness of the DC-IF strategy, we add noise points to the artificial dataset. The specific parameter settings are tabulated in Table III. All sample points and the score planes are shown in Fig. 2.

From Fig. 2, the colour around some sample points is darker, which indicates that it has been assigned lower score value and has higher probability of being noise. We note that the proposed IF strategy can identify the noise points in the original artificial datasets and the artificially added noise points. Intuitively, Fig. 2 demonstrates the anti-noise ability of the proposed IF strategy.

C. Experiments on UCI Datasets

In this subsection, to demonstrate the classification performance of the proposed DC-IFTBLDM, we conduct experiments on 30 UCI datasets and perform statistical analysis using two different methods. The comparative experimental results of DC-IFTBLDM and 9 different algorithms on the Gaussian kernel are shown in Table IV.

From Table IV, we can learn that our proposed model achieves the best classification accuracy on the 27 datasets. Note that DC-IFTBLDM obtains better results than the IFTSVM algorithm on 29 datasets, showing the same accuracy on the remaining 1 dataset. To further analyze the performance of the algorithms, we analysis the classification accuracy rankings of the classifiers on all datasets. The average classification ranking result is tabulated in Table IV (See the *Supporting Document* for more details). From Table IV, the average accuracy ranking of the proposed DC-IFTBLDM algorithm is 1.43, which is the top ranking. In addition, the second-tier algorithms such as IFTSVM and LDM are lower than DC-IFTBLDM by more than 3.6 ranking metrics. LDM ranks third after DC-IFTBDLM and IFTSVM, which indicates the superiority of the LDM variant structure. To sum up, compared with the 9 algorithms including IFTSVM, the DC-IFTBLDM algorithm has better classification performance and generalization performance. To further analyze the performance of DC-IFTBLDM, we performed two types of statistical analysis on the experimental result.

1) Friedman test with post-hoc analysis: From the classification accuracy ranking results of each algorithm on all data sets, we obtain $n = 30$ and $k = 10$. The results of Friedman's test and post-hoc analysis on the UCI dataset and noisy UCI dataset are shown in Table V.

It can be learned from Table V that $\chi_F^2 = 71.7545$ and $F_F = 10.4965$, where F_F is distributed according to F-distribution with $(9, 261)$ degrees of freedom. The critical value of $F(9, 261)$ is 1.916 when the level of significance $\alpha = 0.05$. Note that $F_F = 10.4965$, which is much larger than 1.916. It can be concluded that there are significant differences among the 10 algorithms.

TABLE IV
EXPERIMENTAL RESULT OF DC-IFTBLDM AND 9 OTHER ALGORITHMS.

Dataset	SVM	LDM	FSVM	FLDM	TSVM	FTSVM	TBSVM	CDFTSVM	IFTSVM	DC-IFTBLDM
	Acc/c Time(s)	Acc/c Time(s)	Acc/c Time(s)	Acc/c Time(s)	Acc/c Time(s)	Acc/c Time(s)	Acc/c Time(s)	Acc/c Time(s)	Acc/c Time(s)	Acc/c Time(s)
A_1	98.19/16 0.03108	**100/0.25** 0.04925	95.18/16 0.02940	75.3/0.016 0.01721	99.4/0.004 1.28751	**100/0.004** 0.03178	98.19/0.063 0.04302	96.99/0.004 0.04437	96.39/0.004 0.04174	**100/16** 0.04920
A_2	89.5/16 0.03609	95.03/0.5 0.05585	85.64/16 0.02892	65.75/0.004 0.02978	**95.58/0.016** 0.32880	**95.58/0.008** 0.04442	91.71/16 0.05223	90.06/16 0.05174	90.06/16 0.05344	**95.58/0.5** 0.15516
A_3	95.18/2 0.03009	98.19/0.25 0.04637	97.59/16 0.03178	81.33/0.125 0.02065	95.78/0.5 0.13742	96.39/2 0.03217	95.78/0.125 0.04358	95.78/4 0.04134	96.39/4 0.03799	**99.4/16** 0.20669
A_4	93.85/16 0.01795	93.08/0.063 0.01914	92.31/16 0.01776	95.38/0.5 0.01358	92.31/0.004 0.04245	92.31/0.004 0.01310	94.62/0.031 0.02110	93.08/8 0.02014	93.08/0.125 0.02296	**96.15/16** 0.05904
B_1	87.5/1 0.00202	90.63/0.125 0.00245	87.5/8 0.00254	87.5/0.063 0.00224	90.63/1 0.02025	90.63/1 0.00420	87.5/0.125 0.00741	87.5/0.125 0.00863	81.25/0.125 0.00740	**93.75/16** 0.01471
B_2	93.33/4 0.00389	93.33/1 0.01116	93.33/16 0.00263	**95.56/16** 0.00228	**95.56/0.031** 0.02996	**95.56/16** 0.00560	**95.56/0.063** 0.00936	91.11/0.063 0.00902	93.33/4 0.00892	**95.56/16** 0.04119
B_3	94.34/1 0.00436	96.23/0.25 0.00524	96.23/4 0.00393	96.23/1 0.00326	92.45/0.004 0.01918	94.34/0.016 0.00407	90.57/4 0.01108	92.45/0.004 0.00983	96.23/4 0.01077	**98.11/16** 0.03289
B_4	70.65/2 0.01152	73.91/0.125 0.01179	73.91/8 0.00927	75/16 0.00792	70.65/0.5 0.03382	73.91/0.25 0.00781	72.83/0.008 0.01329	75/0.063 0.01575	**76.09/0.031** 0.01417	**76.09/16** 0.07609
D_1	73.49/2 0.00813	77.11/0.5 0.00829	74.7/16 0.00776	75.9/0.125 0.00489	72.29/16 0.07130	72.29/0.008 0.00872	73.49/4 0.01327	75.9/0.25 0.01308	77.11/2 0.01445	**78.31/16** 0.07449
D_2	96.49/16 0.02935	97.08/1 0.04786	97.08/8 0.03316	96.49/8 0.02959	94.15/0.031 0.06552	94.74/0.5 0.02149	96.49/2 0.03860	95.32/1 0.04112	97.08/8 0.04716	**99.42/16** 0.09993
D_3	76.27/16 0.00520	76.27/0.031 0.00514	77.97/8 0.00401	77.97/2 0.00404	77.97/8 0.01980	79.66/1 0.00650	79.66/0.5 0.01229	83.05/4 0.01149	81.36/4 0.01133	**84.75/16** 0.05562
D_4	95.92/0.125 0.01061	95.92/4 0.01055	96.94/16 0.00864	93.88/4 0.00646	93.88/0.125 0.03533	96.94/0.25 0.00786	94.9/2 0.01468	96.94/0.5 0.01538	94.9/2 0.01428	**97.96/0.125** 0.17123
M_1	95.38/16 0.00514	96.92/4 0.00658	**98.46/16** 0.00375	87.69/16 0.00394	92.31/2 0.02698	96.92/2 0.00559	95.38/0.063 0.01047	93.85/8 0.01142	96.92/0.063 0.01036	**98.46/16** 0.06450
M_2	82.05/4 0.00312	**87.18/0.063** 0.00291	76.92/8 0.00268	**87.18/8** 0.00174	82.05/0.125 0.02392	79.49/0.25 0.00357	79.49/0.5 0.00688	84.62/8 0.00666	82.05/0.5 0.00741	**87.18/0.5** 0.05414
M_3	76.54/8 0.00693	80.25/0.031 0.00658	77.78/8 0.00867	**82.72/0.5** 0.00366	79.01/0.031 0.03802	80.25/0.063 0.00961	80.25/0.5 0.01270	76.54/16 0.01109	81.48/8 0.01209	**82.72/16** 0.04244
M_4	65.96/1 0.00318	70.21/0.5 0.00504	68.09/4 0.00298	70.21/16 0.00322	61.7/0.125 0.03033	68.09/2 0.00414	65.96/0.008 0.00829	68.09/8 0.00685	63.83/4 0.00830	**72.34/16** 0.03483
P_1	90.57/8 0.01135	89.62/0.5 0.01351	90.57/16 0.01012	94.34/16 0.00867	91.51/0.125 0.03010	92.45/0.125 0.00922	93.4/0.016 0.02017	90.57/0.25 0.01764	93.4/0.5 0.01647	**95.28/16** 0.02824
P_2	90.63/8 0.00536	87.5/2 0.00563	90.63/16 0.00485	92.19/4 0.00382	93.75/0.063 0.04342	90.63/0.016 0.00598	92.19/0.008 0.01167	90.63/0.125 0.01100	93.75/0.063 0.01003	**95.31/16** 0.15636
P_3	83.87/8 0.00491	87.1/1 0.00552	88.71/16 0.00465	75.81/8 0.00481	82.26/0.004 0.02087	85.48/0.5 0.00489	80.65/0.25 0.00910	88.71/4 0.00927	90.32/8 0.00932	**91.94/4** 0.03935
P_4	77.78/16 0.00643	80.25/0.25 0.00682	81.48/1 0.00934	81.48/0.5 0.00484	76.54/0.5 0.02844	79.01/16 0.00662	77.78/0.25 0.01220	76.54/0.004 0.01434	80.25/0.25 0.01193	**82.72/16** 0.03777
O_1	84.54/4 0.04958	87.92/0.031 0.06343	86.47/16 0.04221	87.44/0.008 0.04378	85.99/0.125 0.16620	85.99/0.5 0.02795	87.92/0.016 0.04501	87.44/0.031 0.03969	86.96/0.004 0.04673	**88.89/16** 0.16998
O_2	87.24/0.031 0.04145	85.2/2 0.06324	86.22/8 0.03871	87.76/0.008 0.03016	85.71/0.125 0.14907	88.27/0.125 0.02451	86.22/0.004 0.04254	86.73/0.031 0.04654	87.24/0.063 0.04137	**89.29/0.004** 0.10155
O_3	70.91/0.125 0.00338	70.91/0.031 0.05939	70.91/0.125 0.00332	70.91/0.004 0.00391	60/8 0.06278	70.91/4 0.00475	**72.73/0.004** 0.00961	70.91/0.063 0.00937	69.09/0.016 0.01018	70.91/16 0.05790
O_4	86.01/8 0.01570	89.09/0.25 0.02917	92.31/16 0.02045	84.62/16 0.01790	**93.01/0.008** 0.06693	90.21/0.008 0.01407	90.21/8 0.02463	91.61/16 0.02685	89.51/16 0.02849	92.31/16 0.05646
L_1	89.51/1 0.01003	90.97/0.063 0.01558	90.47/1 0.00909	90.73/1 0.01056	90/0.25 0.03050	90/0.25 0.01066	91.8/0.031 0.01292	91.5/0.008 0.01188	92.06/0.031 0.01362	**92.39/16** 0.07276
L_2	90.54/0.5 0.02148	90.24/0.031 0.02163	90.71/4 0.01915	89.81/2 0.01316	90.21/4 0.03685	90.34/4 0.02059	90.84/0.008 0.01927	91.14/0.031 0.01922	91.37/0.125 0.02707	**91.64/16** 0.03938
L_3	94.22/0.125 0.08783	93.22/2 0.03441	93.92/0.5 0.08508	92.78/0.016 0.02573	93.22/0.008 0.03487	93.22/0.008 0.04544	93.35/0.125 0.03971	93.69/0.016 0.03882	92.98/0.008 0.02833	**94.66/16** 0.06667
L_4	72/4 0.00369	72.11/0.125 0.00348	70.56/8 0.00355	71.22/16 0.00265	69.56/0.016 0.02772	70.89/16 0.00473	73.33/1 0.00852	72.33/0.125 0.00825	72.67/0.063 0.00859	**73.56/16** 0.0212
L_5	95.69/8 0.01804	96.31/0.031 0.02193	95.86/16 0.01869	95.27/16 0.01842	95.34/8 0.07785	95.34/8 0.0148	96.31/8 0.02082	95.79/8 0.02047	95.76/16 0.02101	**96.94/4** 0.0678
L_6	62.22/16 0.00509	63.050.25 0.00527	61.31/16 0.00356	**63.42/16** 0.00492	61.39/0.5 0.03672	60.11/1 0.00516	62.67/1 0.00722	61.76/4 0.00739	62.07/2 0.00773	63.12/16 0.01974
Average Acc Win/Tie/Loss	85.35 29/1/0	86.83 27/3/0	85.99 27/3/0	84.06 25/4/1	85.14 27/2/1	86.33 26/4/0	86.06 28/1/1	86.19 29/1/0	86.50 29/1/0	**89.16** -
Average rank	6.98	5.12	5.93	5.88	7.32	5.97	5.43	5.77	5.07	**1.43**

Acc denotes classification accuracy (similarly hereinafter). The parameter c denotes c_3 in DC-IFTBLDM, IFTSVM, CDFTSVM, and TBSVM; corresponds to c_1 in TSVM and FTSVM (similarly hereinafter). Win/Tie/Loss denotes the number of datasets in which the DC-IFTBLDM algorithm is superior/same/inferior to the corresponding algorithm (similarly hereinafter).

TABLE VI
WST PAIRWISE COMPARISON BETWEEN DC-IFTBLDM AND THE OTHER CLASSIFIERS.

		# vs SVM	# vs LDM	# vs FSVM	# vs FLDM	# vs TSVM	# vs FTSVM	# vs TBSVM	# vs CDFTSVM	# vs IFTSVM
N	Positive Rank(PR)	29^W	27^W	27^W	25^W	27^W	26^W	28^W	29^W	29^W
	Negative Rank(NR)	0^L	0^L	0^L	1^L	1^L	0^L	1^L	0^L	0^L
	Tie(T)	1^T	3^T	3^T	4^T	2^T	4^T	1^T	1^T	1^T
	Total	30	30	30	30	30	30	30	30	30
	Sum of PRs	464	459	459	450	458	455	453	464	464
	Sum of NRs	0	0	0	5	3	0	11	0	0
	z-values	-4.7030	-4.5411	-4.5407	-4.4319	-4.5771	-4.4573	-4.4869	-4.7032	-4.7030
	p-values	0.00000256***	0.00000560***	0.00000561***	0.00000934***	0.00000472***	0.00000830***	0.00000723***	0.00000256***	0.00000256***

N denotes the total count. # denotes the proposed algorithm DC-IFTBLDM. W, L, and T represent winning, losing, and tying, respectively. p denotes the level of significance, which is usually set as 0.01 ($***$).

TABLE V
RESULTS OF FRIEDMAN TEST WITH POST-HOC ANALYSIS.

Dataset	χ^2_F	F_F	CD
UCI (original)	71.7545	10.4965	2.4734
UCI (r = 0.1)	41.4382	7.6798	4.2841
UCI (r = 0.5)	52.2919	12.4808	4.2841

r is the signal-to-noise ratio parameter in white Gaussian noise. (similarly hereinafter)

In the Nemenyi test, the CD value corresponding to the average rank is 2.4734 by taking α = 0.05. The comparison result is shown in Fig. 3. If the horizontal line segments of the two algorithms overlap, it means that there is no significant difference between the two algorithms, otherwise there is a significant difference. It can be seen that the red line where DC-IFTBLDM is located does not overlap with all other lines, which indicates that the average ranking gap between DC-IFTBLDM and other algorithms is greater than the CD threshold. Therefore, it can be concluded that DC-IFTBLDM is significantly different from the other 9 algorithms and that DC-IFTBLDM is statistically better than the others.

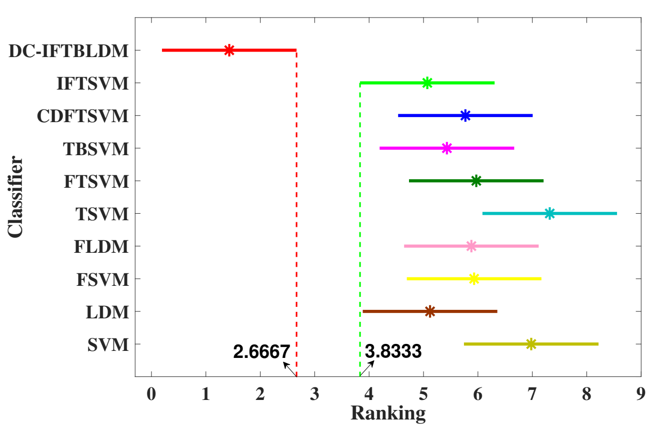


Fig. 3. Result of the Nemenyi test (where $*$ denotes the average rank and line segment centered on the $*$ represents the critical range CD when $\alpha = 0.05$).

2) Wilcoxon signed-rank test: In order to further seek the significant difference between DC-IFTBLDM and the other algorithms, the non-parametric Wilcoxon signed-rank test (WST) and related post-hoc tests are adopted. The test results are shown in Table VI.

Taking the first column of Table VI as an example, the classification accuracy of DC-IFTBLDM is better than SVM in all datasets. Because the calculated probability value p = 0.00000256 is less than the significance level 0.01, there is a significant difference between DC-IFTBLDM and SVM. Similarly, the same conclusions can be drawn from the comparison of DC-IFTBLDM with other algorithms in the remaining columns in Table VI. Therefore, it can be concluded that the accuracy distribution of our DC-IFTBLDM is significantly different from that of other algorithms.

D. Experiments on Datasets with added Gaussian Noise

To further validate the effectiveness and anti-noise performance of DC-IFTBLDM, we process the 10 UCI benchmark datasets in Section IV-B with Gaussian white noise and conduct comparative experiments. Without loss of generality, different scales of white Gaussian noise are considered, which can provide the datasets with different levels of pollution. Specifically, the mean value of white Gaussian noise is 0, and the signal-to-noise ratio r is set to 0.1 and 0.5, respectively. As with the experiments in Section IV-B, DC-IFTBLDM is compared with 9 algorithms. The classification accuracy and ranking result are shown in Table VII.

From Table VII, when r = 0.1, the classification accuracy of DC-IFTBLDM is better than all the other algorithms on 9 datasets. Compared with IFTSVM, DC-IFTBLDM has better accuracy on 8 datasets. Furthermore, the average ranking of DC-IFTBLDM is 1.35, which is the highest-ranking accuracy, and 2.25 rankings ahead of the second. Similarly, when r = 0.5, DC-IFTBLDM also has excellent performance, it achieves the best classification accuracy on 9 datasets, with an average ranking of 1.51 as the highest ranking, ahead of the second place by 2.04 rankings. Specifically, compared with IFTSVM, DC-IFTBLDM has 9 datasets with better accuracy. In conclusion, DC-IFTBLDM shows better anti-noise performance compared to other algorithms.

To further check the performance of the model on the noisy UCI dataset, the Friedman test was used, as shown in Table V. The critical value of $F(9, 81)$ is 2.0 when the level of significance α = 0.05. Whether r = 0.1 or 0.5, the value of F_F is significantly greater than 2.0. Comparative experiments on datasets processed with different proportions of noise also proves the excellent anti-noise performance of the proposed DC-IFTBLDM model in statistics.

TABLE VII
THE EXPERIMENTAL RESULTS OF DC-IFTBLDM AND OTHER 9 ALGORITHMS ON UCI DATASET WITH GAUSSIAN WHITE NOISE.

Dataset	SVM	LDM	FSVM	FLDM	TSVM	FTSVM	TBSVM	CDFTSVM	IFTSVM	DC-IFTBLDM
	Acc(Rank)	Acc(Rank)	Acc(Rank)	Acc(Rank)	Acc(Rank)	Acc(Rank)	Acc(Rank)	Acc(Rank)	Acc(Rank)	Acc(Rank)
A_3(r=0.1)	66.87(10)	71.69(5.5)	69.88(9)	70.48(7.5)	70.48(7.5)	72.29(3)	72.29(3)	71.69(5.5)	72.29(3)	**73.49(1)**
A_3(r=0.5)	67.47(10)	71.69(4)	68.07(8.5)	71.08((5.5)	68.67(8.5)	**72.89(1.5)**	72.29(3)	68.67(8.5)	71.08(5.5)	**72.89(1.5)**
B_3(r=0.1)	90.57(10)	92.45(8.5)	94.34(6)	96.23(3)	94.34(6)	96.23(3)	92.45(8.5)	94.34(6)	96.23(3)	**98.11(1)**
B_3(r=0.5)	86.79(10)	**96.23(1.5)**	90.57(7.5)	94.34(4)	90.57(7.5)	92.45(6)	88.68(9)	94.34(4)	94.34(4)	**96.23(1.5)**
D_2(r=0.1)	90.64(9)	92.4(6.5)	91.81(8)	94.15(2.5)	89.47(10)	92.4(6.5)	92.98(5)	93.57(4)	94.15(2.5)	**94.74(1)**
D_2(r=0.5)	91.23(8)	93.57(3)	91.23(8)	**94.74(1.5)**	88.89(10)	91.23(8)	91.81(6)	92.4(4.5)	92.4(4.5)	**94.74(1.5)**
M_3(r=0.1)	80.25(3)	75.31(9)	79.01(4.5)	76.54(7.5)	74.07(10)	76.54(7.5)	77.78(6)	79.01(4.5)	**81.48(1.5)**	**81.48(1.5)**
M_3(r=0.5)	74.07(8.5)	76.54(4)	74.07(8.5)	**79.01(1.5)**	75.31(5.5)	75.31(5.5)	74.07(8.5)	74.07(8.5)	77.78(3)	**79.01(1.5)**
P_3(r=0.1)	55.08(8)	59.89(3)	53.48(10)	60.43(2)	54.55(9)	56.68(6)	58.82(5)	59.36(4)	56.15(7)	**61.29(1)**
P_3(r=0.5)	49.73(9)	58.29(2)	48.66(10)	56.15(5)	53.48(7.5)	54.55(6)	53.48(7.5)	56.68(4)	57.22(3)	**59.68(1)**
O_2(r=0.1)	73.47(7)	77.55(3)	74.49(6)	78.06(2)	70.92(10)	71.43(9)	72.96(8)	75(5)	76.53(4)	**79.59(1)**
O_2(r=0.5)	72.84(10)	75.31(7)	79.01(2)	75.31(7)	76.54(5)	77.78(3.5)	74.07(9)	75.31(7)	77.78(3.5)	**79.59(1)**
L_1(r=0.1)	59.77(7)	59.6(8)	60.03(6)	60.24(4)	59.42(10)	59.57(9)	60.75(2)	60.21(5)	60.27(3)	**60.85(1)**
L_1(r=0.5)	59.87(8)	60.81(7)	61.22(4)	62.63(2)	58.33(10)	59.47(9)	60.91(6)	61.11(5)	61.92(3)	**63(1)**
L_2(r=0.1)	55.98(7)	56.68(6)	54.36(10)	56.72(5)	55.35(9)	55.63(8)	57.32(3)	57.42(2)	**57.92(1)**	57.09(4)
L_2(r=0.5)	54.4(9)	56.62(3)	56.36(5)	57.68(2)	53.34(10)	55.49(7)	55.89(6)	55.09(8)	56.46(4)	**57.92(1)**
L_3(r=0.1)	56.62(4)	54.06(10)	56.25(5)	54.83(9)	54.94(8)	55(7)	56.88(2)	56.02(6)	56.76(3)	**57.32(1)**
L_3(r=0.5)	53.9(9)	56.29(2)	54.73(8)	55.86(4)	53.11(10)	55.13(7)	55.56(6)	55.6(5)	56.06(3)	**56.96(1)**
L_5(r=0.1)	67.4(3)	63.78(10)	67.56(2)	66.58(4)	66.09(7)	66.53(5)	66.18(6)	63.91(9)	64.64(8)	**68.78(1)**
L_5(r=0.5)	65.56(6)	65.78(5)	**67.67(1)**	65.22(7)	64.11(9)	66.11(10)	63.11(4)	64.78(8)	67.56(2)	67.33(3)
r=0.1(AVA/AVR)	69.67/6.8	70.34/6.95	70.12/6.65	71.43/4.65	68.96/8.65	70.23/6.4	70.84/4.85	71.05/5.1	71.64/3.6	**73.27/1.35**
r=0.5(AVA/AVR)	67.59/8.68	71.11/3.90	69.16/5.91	71.20/4.49	68.24/8.27	70.04/6.36	68.99/6.51	69.81/6.22	71.26/3.55	**72.74/1.51**
Wie/Tie/Loss(0.1)	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	8/1/1	–
Wie/Tie/Loss(0.5)	10/0/0	9/1/0	9/0/1	8/2/0	10/0/0	9/1/0	10/0/0	10/0/0	10/0/0	–

AVA denotes the average accuracy and AVR represents average ranking. (See the *Supporting Document* for experimental results on more datasets)

TABLE VIII
ACCURACIES OF DIFFERENT COMBINATION ALGORITHMS ON STATLOG (HEART) DATASET.

	Fuzzy [23]	IF [28]	DC-IF
TBLDM (original)	77.16	79.01	**82.72**
TBLDM (r = 0.1)	76.54	78.40	**81.48**
TBLDM (r = 0.5)	71.60	76.54	**79.01**

Note that Fuzzy strategy is a special case of IF strategy.

E. *Experiments on the DC-IF strategy*

In order to independently verify the superior of the proposed DC-IF strategy, we conducted comparative experiments to explore the improvement effect of different IF strategies for TBLDM. Specifically, we combine the TBLDM baseline with the existing two effective IF strategies in CDFTSVM [23] and IF strategy in IFTSVM [28]. It is worth noting that this paper is the first time to introduce IF strategies into LDMs, i.e. there are no existing IF LDM models prior to our model.

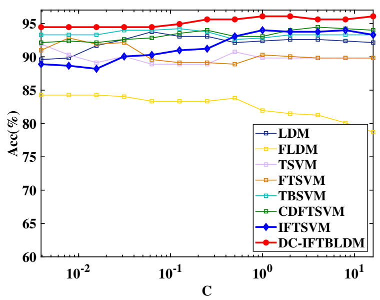
Table VIII shows the classification result on the Statlog (Heart) dataset with and without added Gaussian white noise. We can find out that the DC-IF strategy achieves the best accuracy on all conditions. Specifically, the IF strategy achieves better accuracy than the fuzzy strategy [23] on three datasets. Moreover, DC-IF strategy is superior the IF [28] on both the datasets with different degrees of noise. To further investigate the stability of DC-IFTBLDM, parameter sensitivity tests were performed, and the result is in the following subsection.

F. *Parameter Sensitivity Analysis*

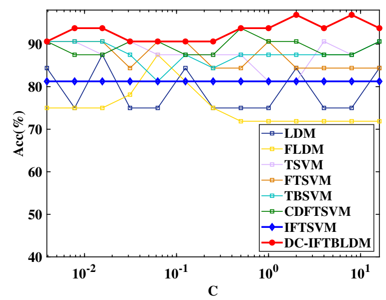
The penalty parameter c in the soft margin model has the greatest impact on the DC-IFTBLDM algorithm because it trades off the importance of points. And this parameter is common to all algorithms. Therefore, the performance of different algorithms is compared by adjusting the penalty parameter c. As the parameter c changes, the variations of the classification accuracy of different algorithms on 12 datasets are shown in Fig. 4. Here, we perform the simple logarithmic processing on the c value for better display. Note that the SVM and FSVM algorithms have the worst effect and are particularly sensitive to the parameter c. In order to compare other algorithms more intuitively, the results of SVM and FSVM are not shown.

From Fig. 4, we can draw the following conclusions. Within the range of parameter variation, DC-IFTBLDM achieved the highest accuracy in almost all datasets. DC-IFTBLDM exhibited excellent stability in almost all datasets with varying c values. For example, in the Waveform-5000 (1-2) dataset, as the value of c changes, the accuracy of DC-IFTBLDM is always the highest and stable at around 94.5%, while the other algorithms have not achieved such high accuracy rate and fluctuate greatly. To conclude, dring the change of c value, DC-IFTBLDM achieved the best performance with smooth changes in most cases.

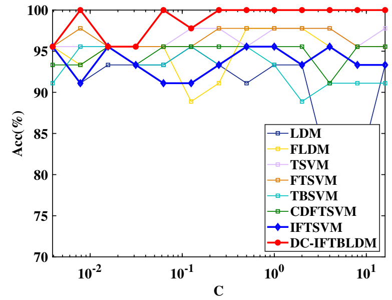
Analyzing all the experimental results, we can draw the following conclusions.



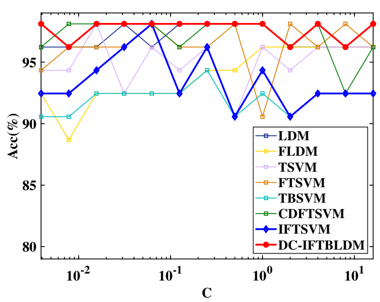
(a) Monks3



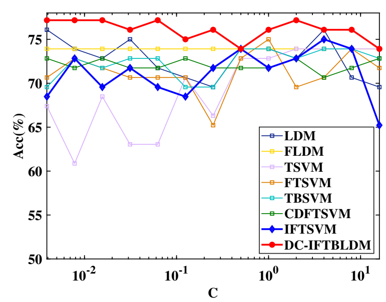
(b) Promoters



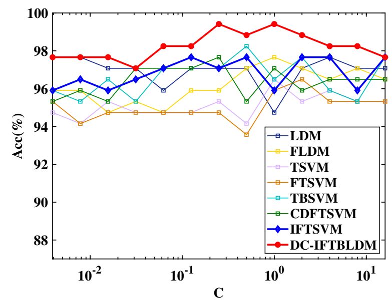
(c) Iris



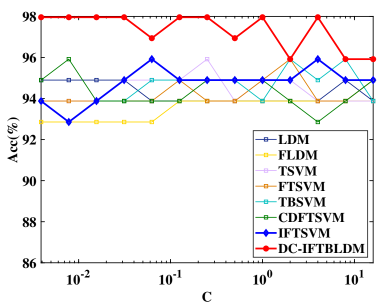
(d) Wine



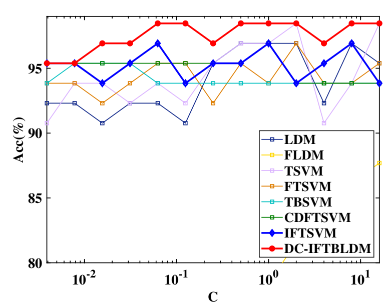
(e) Haberman



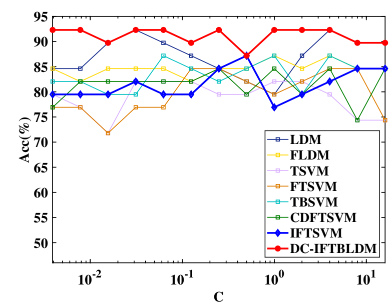
(f) Wdbc



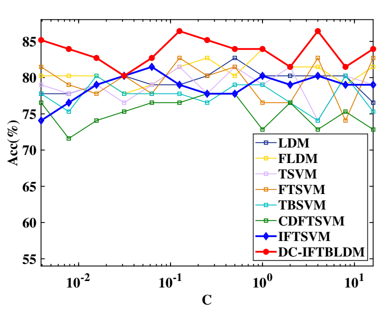
(g) Ecoli-data



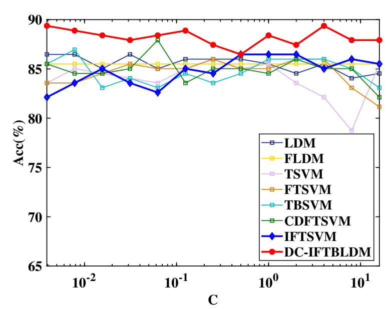
(h) New-thyroid



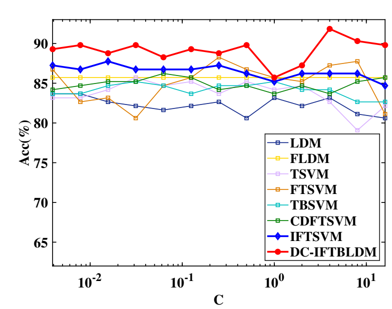
(i) Echocardiogram



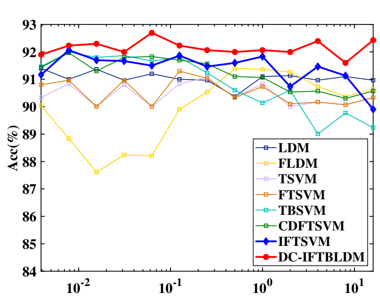
(j) Statlog(Heart)



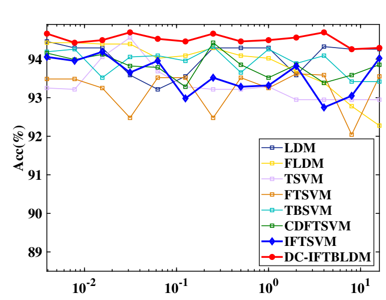
(k) Australian



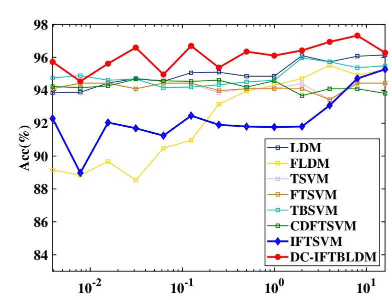
(l) Credit-a



(m) Waveform-5000(0-1)



(n) Waveform-5000(1-2)



(o) Kr-vs-kp

Fig. 4. The influence of parameter c on the classification accuracy of different algorithms.

- The proposed DC-IFTBLDM model achieves significantly better performance than other algorithms on the UCI datasets with and without added Gaussian white noise, which reflects its superiority (See Section IV-C and IV-D for details).
- DC-IF effectively reduces the influence of noise and outliers, so its improvement on TBLDM is greater than other IF strategies (See Section IV-E for details).
- DC-IFTBLDM achieves better stability and higher classification accuracy than other algorithms regardless of parameter tuning of c (See Section IV-F for details).

V. Conclusion

This paper proposed a novel dual-center based intuitionistic fuzzy twin bounded large margin distribution machines (DC-IFTBLDM). To the best knowledge of the authors, this is the first time that the IF strategy is introduced into LDM-based models, and it is also the first work that the large margin distribution is embedded in IFTSVM. The novel DC-IF strategy more accurately reflects the credibility of the samples and is time efficient than the IF strategies in the existing SVMs. Specifically, the membership degree of a sample is set to be related to the homogeneous class center, and non-membership is related to the heterogeneous class center. The scoring function is a weighted group sum of membership degree and non-membership degree, which allows for a more accurate assessment of the sample and reduces the effects of noise and outliers. Additionally, we enhance the generalization performance by optimizing margin distribution and a regularization term is used to minimize structural risk. Finally, a series of experiments on the artificial dataset, UCI benchmark datasets and the noisy UCI datasets demonstrate the effectiveness of DC-IFTBLDM. The parameter sensitivity analysis shows that our model has high accuracy and stability. However, the proposed DC-IFTBLDM has some limitations, including the need to optimize multiple hyperparameters, so we are committed to exploring more efficient parameter optimization techniques in the future. Besides, it would also be an interesting direction to introduce Universum points in the DC-IFTBLDM algorithm.

References

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

[2] H. L. Yu, C. Y. Sun, X. B. Yang, S. Zheng, and H. T. Zou, "Fuzzy support vector machine with relative density information for classifying imbalanced data," *IEEE transactions on fuzzy systems*, vol. 27, no. 12, pp. 2353–2367, 2019.

[3] R. Saltos, R. Weber, and S. Maldonado, "Dynamic rough-fuzzy support vector clustering," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1508–1521, 2017.

[4] C. Ding, T. Y. Bao, and H. L. Huang, "Quantum-inspired support vector machine," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[5] Z. Wang, Y. H. Shao, and T. R. Wu, "Proximal parametric-margin support vector classifier and its applications," *Neural Computing and Applications*, vol. 24, no. 3, pp. 755–764, 2014.

[6] Y. H. Shao and N. Y. Deng, "A coordinate descent margin based-twin support vector machine for classification," *Neural networks*, vol. 25, pp. 114–121, 2012.

[7] H. Do and A. Kalousis, "Convex formulations of radius-margin based support vector machines," in *International conference on machine learning*. PMLR, 2013, pp. 169–177.

[8] O. L. Mangasarian and E. W. Wild, "Multisurface proximal support vector machine classification via generalized eigenvalues," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 1, pp. 69–74, 2005.

[9] R. Khemchandani, S. Chandra *et al.*, "Twin support vector machines for pattern classification," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, no. 5, pp. 905–910, 2007.

[10] H. Yan, Q. L. Ye, T. A. Zhang, D. J. Yu, X. Yuan, Y. Q. Xu, and L. Y. Fu, "Least squares twin bounded support vector machines based on l1-norm distance metric for classification," *Pattern recognition*, vol. 74, pp. 434–447, 2018.

[11] B. Kumar and D. Gupta, "Universum based lagrangian twin bounded support vector machine to classify eeg signals," *Computer Methods and Programs in Biomedicine*, vol. 208, p. 106244, 2021.

[12] D. Gupta, H. J. Sarma, K. Mishra, and M. Prasad, "Regularized universum twin support vector machine for classification of eeg signal," in *2019 IEEE international conference on systems, man and cybernetics (SMC)*. IEEE, 2019, pp. 2298–2304.

[13] B. Richhariya and D. Gupta, "Facial expression recognition using iterative universum twin support vector machine," *Applied Soft Computing*, vol. 76, pp. 53–67, 2019.

[14] U. Gupta and D. Gupta, "Kernel-target alignment based fuzzy lagrangian twin bounded support vector machine," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, no. 05, pp. 677–707, 2021.

[15] B. B. Hazarika, D. Gupta, and P. Borah, "An intuitionistic fuzzy kernel ridge regression classifier for binary classification," *Applied Soft Computing*, vol. 112, p. 107816, 2021.

[16] Y. H. Shao, C. H. Zhang, X. B. Wang, and N. Y. Deng, "Improvements on twin support vector machines," *IEEE transactions on neural networks*, vol. 22, no. 6, pp. 962–968, 2011.

[17] C. Zhang, H. X. Li, Y. Gao, and C. L. Chen, "Weakly-supervised enhanced semantic-aware hashing for cross-modal retrieval," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[18] U. Gupta and D. Gupta, "Bipolar fuzzy based least squares twin bounded support vector machine," *Fuzzy Sets and Systems*, vol. 449, pp. 120–161, 2022.

[19] X. Q. Zhu and X. D. Wu, "Class noise vs. attribute noise: A quantitative study," *Artificial intelligence review*, vol. 22, no. 3, pp. 177–210, 2004.

[20] M. Ganaie, M. Tanveer, and C. T. Lin, "Large scale fuzzy least squares twin svms for class imbalance learning," *IEEE Transactions on Fuzzy Systems*, 2022.

[21] X. Q. Ye and D. Liu, "A cost-sensitive temporal-spatial three-way recommendation with multi-granularity decision," *Information Sciences*, vol. 589, pp. 670–689, 2022.

[22] C. F. Lin and S. D. Wang, "Fuzzy support vector machines," *IEEE transactions on neural networks*, vol. 13, no. 2, pp. 464–471, 2002.

[23] B. B. Gao, J. J. Wang, Y. Wang, and C. Y. Yang, "Coordinate descent fuzzy twin support vector machine for classification," in *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*. IEEE, 2015, pp. 7–12.

[24] S. G. Chen, J. F. Cao, Z. Huang, and C. S. Shen, "Entropy-based fuzzy twin bounded support vector machine for binary classification," *IEEE Access*, vol. 7, pp. 86 555–86 569, 2019.

[25] K. Zhang, J. M. Zhan, and W. Z. Wu, "On multicriteria decision-making method based on a fuzzy rough set model with fuzzy $alpha$-neighborhoods," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 9, pp. 2491–2505, 2021.

[26] A. H. Tan, S. W. Shi, W. Z. Wu, J. J. Li, and W. Pedrycz, "Granularity and entropy of intuitionistic fuzzy information and their applications," *IEEE Transactions on Cybernetics*, vol. 52, no. 1, pp. 192–204, 2022.

[27] J. H. Dai, X. T. Zou, and W. Z. Wu, "Novel fuzzy β-covering rough set models and their applications," *Information Sciences*, vol. 608, pp. 286–312, 2022.

[28] S. Rezvani, X. Z. Wang, and F. Pourpanah, "Intuitionistic fuzzy twin support vector machines," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 11, pp. 2140–2151, 2019.

[29] L. W. Wang, M. Sugiyama, Z. X. Jing, C. Yang, Z. H. Zhou, and J. F. Feng, "A refined margin analysis for boosting algorithms via equilibrium margin," *The Journal of Machine Learning Research*, vol. 12, pp. 1835–1863, 2011.

[30] W. Gao and Z. H. Zhou, "On the doubt about margin explanation of boosting," *Artificial Intelligence*, vol. 203, pp. 1–18, 2013.

[31] T. Zhang and Z. H. Zhou, "Large margin distribution machine," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 313–322.

[32] L. M. Liu, M. X. Chu, R. F. Gong, and Y. C. Peng, “Nonparallel support vector machine with large margin distribution for pattern classification,” *Pattern Recognition*, vol. 106, p. 107374, 2020.
[33] Y. C. Li, Y. L. Wang, C. Bi, and X. H. Jiang, “Revisiting transductive support vector machines with margin distribution embedding,” *Knowledge-Based Systems*, vol. 152, pp. 200–214, 2018.
[34] X. G. Zhang, D. X. Wang, Y. C. Zhou, H. Chen, F. Y. Cheng, and M. Liu, “Kernel modified optimal margin distribution machine for imbalanced data classification,” *Pattern Recognition Letters*, vol. 125, pp. 325–332, 2019.
[35] F. Y. Cheng, J. Zhang, Z. Y. Li, and M. Z. Tang, “Double distribution support vector machine,” *Pattern Recognition Letters*, vol. 88, pp. 20–25, 2017.
[36] S. Abe, “Unconstrained large margin distribution machines,” *Pattern Recognition Letters*, vol. 98, pp. 96–102, 2017.
[37] T. Zhang and Z. H. Zhou, “Optimal margin distribution machine,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1143–1156, 2019.
[38] Q. Wu, S. W. Qi, and K. Y. Sun, “Weighted least squares twin large margin distribution machine,” in *Proceedings of the 2017 VI International Conference on Network, Communication and Computing*, 2017, pp. 253–257.
[39] L. M. Liu, M. X. Chu, Y. H. Yang, and R. F. Gong, “Twin support vector machine based on adjustable large margin distribution for pattern classification,” *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 10, pp. 2371–2389, 2020.
[40] H. T. Xu, B. McCane, and L. Szymanski, “Twin bounded large margin distribution machine,” in *Australasian joint conference on artificial intelligence*. Springer, 2018, pp. 718–729.
[41] T. X. Wang, L. B. Zhang, B. Huang, and X. Z. Zhou, “Three-way conflict analysis based on interval-valued pythagorean fuzzy sets and prospect theory,” *Artificial intelligence review*, 2022.
[42] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2022.
[43] D. Gupta, P. Borah, and M. Prasad, “A fuzzy based lagrangian twin parametric-margin support vector machine (fltpmsvm),” in *2017 IEEE symposium series on computational intelligence (SSCI)*. IEEE, 2017, pp. 1–7.
[44] D. H. Dong, M. Y. Feng, J. Kurths, and L. B. Zhang, “Fuzzy large margin distribution machine based on cluster center for classification,” *International Journal of Machine Learning and Cybernetics*, vol. under review, 2022.
[45] Z. Z. Liang and L. Zhang, “Intuitionistic fuzzy twin support vector machines with the insensitive pinball loss,” *Applied Soft Computing*, vol. 115, p. 108231, 2022.
[46] R. K. Sevakula and N. K. Verma, “Compounding general purpose membership functions for fuzzy support vector machine under noisy environment,” *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1446–1459, 2017.
[47] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
[48] F. Wilcoxon, “Individual comparisons by ranking methods,” in *Breakthroughs in statistics*. Springer, 1992, pp. 196–202.