A Novel Cost-Sensitive Three-Way Intuitionistic Fuzzy Large Margin Classifier

Shuangyi Fan[a], Heng Li[a], Cong Guo[b], Dun Liu[c] and Libo Zhang[a,*]

[a]College of Artificial Intelligence, Southwest University, Chongqing 400715, China
[b]College of Management and Engineering, Nanjing University, Nanjing 210093, China
[c]College of Economics and Management, Southwest Jiaotong University, Chengdu 61003, China

ARTICLE INFO

Keywords:
Three-way decision
Large margin classifier
Intuitionistic fuzzy
Cost sensitive

ABSTRACT

Three-way decision (3WD) has been widely applied in diverse fields in tackling uncertainty, particularly in classification domain. As a discriminative learning algorithm, Large Margin Distribution Machine (LDM) aims to maximize the inter-class margin by leveraging the marginal distribution of samples, which captures the underlying structure and achieves superior classification performance. However, existing LDMs still suffer from some inherent flaws, including: i) the neglect of fuzzy decision items, leading to the inadequate handling of uncertainty; ii) the utilization of a cost-insensitive mechanism, resulting in high misclassification costs; and iii) the disregard of sample credibility, contributing to the noise susceptibility. To address these limitations, we introduce the intuitionistic fuzzy (IF) and cost-sensitive 3WD (CS3WD), presenting an innovative model known as the CS3W-IFLMC model. The IF theory is employed to quantify sample confidence levels, augmenting noise suppression in our proposed model. Moreover, the integration of the CS3WD method effectively reduces the overall decision cost and further improves the handling of uncertainty in datasets. Consequently, the CS3W-IFLMC model demonstrates superior noise resilience and generalization performance. Our comparative experiments validate the efficacy of the proposed CS3W-IFLMC model in achieving robustness to noise while upholding a competitive classification performance.

1. Introduction

Three-way decision (3WD), put forward by Yao, effectively handles complexity and uncertainty [1]. It partitions the universe into positive, negative and boundary regions, each linked to a specific action [2]. Unlike traditional approaches, 3WD assigns uncertain objects to the boundary area, such as delay decision [3]. 3WD has gained prominence in classification [4, 5, 6], clustering [7, 8], attribute reduction [9, 10] , and other fields [11, 12]. It reduces misclassifications and enhances deterministic decisions [13]. Extensive research shows the superiority of 3WD models over traditional classifiers in accuracy and reliability [14]. In recent years, new research progress has been made in 3WD fields. Han [15] proposed a 3WD model based on fuzzy decision trees. Subhashini [16] and Savchenko [17] innovatively integrated convolutional neural networks into the model, effectively improving the robustness of the neural network model. In addition, Du [18] integrates grey correlation analysis and TOPSIS grey multi-criteria into the 3WD model, which further extends the totally ordered set evaluation-based three-way decisions. To minimize cost, the cost-sensitive three-way classification (CS3WC) method integrates 3WD with cost-sensitive strategies [19, 20]. CS3WC addresses imbalanced misclassification costs and data imbalance, prominent in classification tasks [21, 22]. It aligns with human decision-making and expands the 3WD framework in classification [21, 22]. Although the related 3WD methods are widely used, they are less integrated with machine learning. Considering that machine learning is an indispensable artificial intelligence model in the era of big data, the integration of the three decision-making models with machine learning is of great research value [23].

The Support Vector Machine (SVM), introduced by Vapnik et al. [24, 25], is widely used in face recognition [26], spam filtering [27], and financial risk assessment [28]. However, empirical evidence suggests that margin distribution is more important than the minimum margin for classifier performance [29]. In 2014, Zhou et al. introduced margin information into SVM and proposed the Large Margin Distribution Machine (LDM) [30]. The LDM maximizes the

*Corresponding author: Libo Zhang
E-mail address: ShuangyiFan111@163.com (S. Fan),
hengliswu@foxmail.com (H. Li), gc577162881@163.com (C. Guo), liudun@swjtu.edu.cn (D. Liu), lbzhang@swu.edu.cn (L. Zhang).
ORCID(s): 0000-0001-5992-0790 (L. Zhang)

average margin while minimizing margin variance, aiming to enhance robustness and accuracy. LDM consistently outperforms SVM in various tasks [31, 32, 33]. Afterwards, researchers have developed variants like Unconstrained Large Margin Distribution Machines (ULDM) [34], Cost-Sensitive Large Margin Distribution Machine (CS-LDM) [35], and Twin Large Margin Distribution Machine (TLDM) [36]. The improvement to the LDM model described above enables it to handle various complex scenarios. Specifically, in CS-LDM, the introduction of a cost-sensitive strategy increases the margin weight for minority class samples, effectively addressing imbalanced data. However, it fails to effectively identify noise and overlooks the issue of misclassification costs. In the research on the noise resistance of LDM, despite progress, quantifying data uncertainty remains challenging, leading to potential errors [37]. In many applications, the cost of indecision is lower than that of erroneous decisions [38]. For medical diagnosis, delayed decisions considering uncertain domains reduce costs and misdiagnosis risks. The existing LDM models struggle to distinguish outliers from normal samples, limiting their resilience to noisy data [39].

The cost-sensitive 3WD (CS3WD) optimizes decision cost. In 2022, Shen et al. improved credit evaluation in unsupervised learning by using 3WD to filter rejected samples and enhance SVM performance [40]. However, they overlooked unthresholded output information and failed to unify mathematical forms of the 3WD and semi-supervised support vector machine (S4VM) model. Additionally, S4VM lacks consideration of marginal distributions, a feature in LDM. Moreover, converting LDM output to probability has yielded commendable results [41] and establishes a foundation for decision-making in LDM classification using CS3WD mechanisms. Integrating the CS3WD into LDM handles complex scenarios and increases classification credibility. To enhance noise resistance, evaluating the credibility and uncertainty of individual samples is crucial [25]. Intuition Fuzzy set (IF) theory is extensively used for managing real-world uncertainties [42]. The IF model considers membership and non-membership degrees of samples, indicating their association with a specific class [43, 44]. Rezvani et al. and Zhang et al. incorporated different IF strategies into SVM, significantly improving noise handling [45, 39]. The IF strategy proposed by Rezvani et al. [45] is a single-center strategy, which results in inaccurate confidence and reduces the classification accuracy of the model. Additionally, its non-membership degree relies on the local neighborhood set of samples, making it time-consuming and difficult to determine a threshold. Building upon this, Zhang et al. [39] proposed a new dual-center IF (DC-IF) function, incorporating marginal distribution into the classifier, effectively addressing the aforementioned issues. However, distinguishing between samples remains a challenge for the existing IF strategy in LDM.

Based on the aforementioned analysis, this study presents a novel CS3W-IFLMC model, which combines the CS3WD method and IF theory to enhance the LDM framework. To enhance the model's robustness, a novel IF method is devised for evaluating the credibility of samples. Specifically, the membership and non-membership degrees are determined by the sample's distance to both positive and negative class centers. Then, a comprehensive sample score is computed as a measurement of the sample's credibility, which is incorporated into the optimization objective function to optimize the classification hyperplane. Subsequently, a logistic regression model based on the comprehensive margin information is optimized using the cross-entropy loss function to obtain the posterior probability of the sample. Finally, the CS3WD framework is introduced to attain optimal decision-making results, which further enhances the capability to deal with uncertainty. Note that, the proposed CS3W-IFLMC is a general model that can be degraded into the CS2W-IFLMC, IFLMC and traditional LDM under corresponding specific conditions.

Our comparative experiments on UCI and Kaggle benchmark datasets, along with datasets containing different levels of noise, validate the efficacy of the proposed model in achieving robustness to noise while upholding a competitive classification performance in cost-sensitive and cost-insensitive classification tasks. The main contributions of this paper are outlined as follows:

- By incorporating the CS3WD and IF theory into LDM, we propose a novel CS3W-IFLMC model that exhibits superior classification performance.
- The combination of 3WD with LDM provides a novel and feasible approach to effectively address uncertainty in classification problems.
- A novel IF strategy is proposed to characterize the credibility of each given sample, thereby fortifying the model's resistance to noise.

This paper is organized as follows: Section 2 provides a comprehensive literature review to establish a robust theoretical foundation. Section 3 introduces the CS3W-IFLMC model, the proposed IF method and the solving process of the model. Section 4 rigorously conducts experiments with UCI and Kaggle datasets to empirically validate our

superior model performance. Section 5 offers a forward-looking summary, highlighting potential implications and valuable insights for future advancements.

2. Preliminaries

In this section, we expound upon the fundamental principles of CS3WD (see Section 2.1), SVM (see Section 2.2), and LDM (see Section 2.3).

2.1. Cost-sensitive Three-way Classification

Fig. 1 illustrates the framework of 3WD. The whole area is classified into three regions, each with a strategy. A general definition of 3WD is presented in Definition 1.

Definition 1. [1] Let U be a set of objects, L be the set of partial order relations, and $\upsilon : U \to L$ be an evaluation function, which value is the decision status value of x. For an object $x \in U$, $\upsilon(x)$ is its evaluation status value (ESV). Let $(L, \succeq)$ indicates a set that is completely ordered, (δ, η) denotes a pair of thresholds which satisfies $\delta \succ \eta$ (i,e., $\delta \succeq \eta \wedge \neg(\eta \succeq \delta)$). Using this evaluation-based model as a foundation, three pairwise disjoint regions are defined as follows:

$$\begin{aligned} &\text{Region I } (\upsilon) = \{x \in U | \upsilon(x) \succeq \delta\}, \\ &\text{Region II } (\upsilon) = \{x \in U | \delta \succ \upsilon(x) \succ \eta\}, \\ &\text{Region III } (\upsilon) = \{x \in U | \eta \succeq \upsilon(x)\}, \end{aligned} \tag{1}$$

where Region I $\cup$ Region II $\cup$ Region III $= U$.

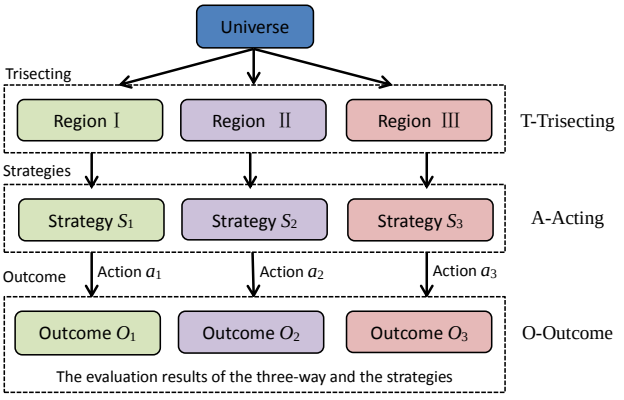


Figure 1: The framework of 3WD model.

The category set can be represented as $\{C_P, C_N\}$. Despite the two deterministic decisions in traditional methods, 3WD offers another option: boundary decision C_B. For an instance x in category C_N, if it is classified into category C_P, the decision cost is denoted as $\tau_{PN} = \tau(\phi(x) = a_P | x \in C_N)$. Similarly, $\tau_{NN}, \tau_{BN}, \tau_{PP}, \tau_{NP}$ and τ_{BP} are defined, which are shown in Table 1.

The conditional probability $Pr(C_i|x)$ is always adopted as the evaluation status function in Definition 1. Given the conditional probability $Pr(C_P|x)$ and $Pr(C_N|x)$, the anticipated losses of three decisions are presented below:

$$R(a_k|x) = \tau_{kN} Pr(C_N|x) + \tau_{kP} Pr(C_P|x), k \in \{P, N, B\}. \tag{2}$$

Based on Bayesian risk rule, the optimal result for x is given by:

$$\phi^*(x) = \underset{k \in \{P,N,B\}}{\arg\min} \; R(a_k|x). \tag{3}$$

The parameters δ and η can be induced based on the above decision rules [11]. Compared with traditional cost-sensitive classifiers, the extra boundary decision further reduces the total losses [46].

Table 1
Cost matrix of 3WD in 2-Class classification problem

Action \ Category	C_P	C_N
a_P	τ_{PP}	τ_{PN}
a_B	τ_{BP}	τ_{BN}
a_N	τ_{NP}	τ_{NN}

2.2. Support Vector Machine

In binary classification problem, for each sample $\boldsymbol{x_i}(i = 1, \ldots, m)$, we can define the label of y_i as follows: $y_i = 1$ if x_i belongs to the positive sample set $\mathcal{I}_1$, or $y_i = -1$ if x_i belongs to the negative sample set $\mathcal{I}_2$. The representation of SVM's hyperplane is as follows:

$$\boldsymbol{w}^T \boldsymbol{x} + b = \pm 1, \tag{4}$$

where $\boldsymbol{w}$ represents the normal vector of the classification hyperplane and b represents the bias term. They are determined by support vectors which are the minimum distance between two sample sets. To obtain the hyperplanes, it needs to solve the following quadratic programming (QP) problem:

$$\begin{aligned} &\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m}\xi_i, \\ &\text{s.t. } y_i\left(\boldsymbol{w}^T\phi(\boldsymbol{x_i}) + b\right) \geq 1 - \xi_i, i = 1, \ldots, m, \\ &\quad \xi_i \geq 0, i = 1, \ldots, m, \end{aligned} \tag{5}$$

where ξ is the slack variable; C is the trade-off parameter between the margin width and misclassification loss; $\phi(\boldsymbol{x})$ is a feature mapping.

2.3. Large Margin Distribution Machine

According to the Structural Risk Minimization (SRM) theory, marginal distribution is more important to model generalization performance than minimum margin [29]. Based on SVM, Zhou et al. proposed LDM which maximizes margin mean and minimizes margin variance simultaneously. The optimization function of LDM is as follows:

$$\begin{aligned} &\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + \lambda_1\hat{\gamma} - \lambda_2\bar{\gamma} + C\sum_{i=1}^{m}\xi_i, \\ &\text{s.t. } y_i\left(\boldsymbol{w}^T\phi\left(\boldsymbol{x_i}\right) + b\right) \geq 1 - \xi_i, i = 1, \ldots, m, \\ &\xi_i \geq 0, i = 1, \ldots, m, \end{aligned} \tag{6}$$

where λ_1 and λ_2 are weight hyperparameters. $\hat{\gamma}$ and $\bar{\gamma}$ are margin variance and margin mean respectively, and the calculation formula is as follows:

$$\begin{aligned} \hat{\gamma} &= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\left(y_i\boldsymbol{w}^T\phi(\boldsymbol{x_i}) - y_j\boldsymbol{w}^T\phi(\boldsymbol{x_j})\right)^2, \\ \bar{\gamma} &= \frac{1}{m}\sum_{i=1}^{m} y_i\boldsymbol{w}^T\phi(\boldsymbol{x_i}). \end{aligned} \tag{7}$$

In summary, SVM effectively achieves optimal classification by minimizing the shortest distance and thereby mitigating the impact of uncertainty on decision results. However, the corresponding hyperplane is susceptible to the influence of noise. On this basis, LDM further reduces the impact of uncertainty by introducing margin mean and margin variance to the objective function. From a classification viewpoint, the 3WD model is designed for binary classification [2], which is the same with SVMs and LDMs. The 3WD model provides a feasible way for SVM and LDM to improve the ability to handle uncertainty, which is explored in this work.

3. CS3W-IFLMC Model

In this section, we present a novel IF method, introduce the CS3WD method, and propose a unified CS3W-IFLMC model with noise immunity and low decision cost in Section 3.1. Subsequently, the proposed IF method is expounded and substantiated in Section 3.2. Additionally, the solution process of the model is elucidated in Section 3.3. Finally, the decision boundary threshold of the model is derived in Section 3.4, and the pseudocode of CS3W-IFLMC is provided to further elucidate the algorithmic logic of CS3W-IFLMC.

3.1. Model Construction

To enhance the noise resistance and decision reliability of LDM, an innovative IF function is constructed and introduced into the objective function of LDM, forming a classification method with the ability to recognize sample confidence. Subsequently, to further reduce decision costs, the Platt scaling method is employed to process the output information from LDM, yielding the probability of the sample belonging to the positive class. Finally, the CS3WD framework is introduced to obtain the optimal decision results.

To enhance the noise robustness of LDM, we introduce the IF method and formulate a novel IF function. The membership and non-membership degrees of a sample are computed based on its distance from the class center. The membership degree of a sample is higher if it is closer to the center of the class, and lower if it is farther away. Taking a positive sample $\boldsymbol{x_i}$ as an example, we calculate the membership degree of the sample using the following formula.

$$\varpi(\boldsymbol{x_i}) = 1 - D_P(\boldsymbol{x_i})/(E_P + \varepsilon), \tag{8}$$

where $D_P(\boldsymbol{x_i})$ represents the distance between the sample $\boldsymbol{x_i}$ and the center of positive class, and the specific calculation formula is shown in Eq. (9), and the detailed proof process can be seen in reference [39]. And E_P denotes the maximum distance between positive samples in the training set and the center of positive class, commonly referred to as the scatter hyper-sphere radius.

$$D_P\left(\boldsymbol{x_i}\right) = k\left(\boldsymbol{x_i}, \boldsymbol{x_i}\right) - \frac{2}{n_+} k\left(\boldsymbol{x_i}, \boldsymbol{x_i}\right) + \frac{1}{n_+^2} \sum_{y_a=1} \sum_{y_b=1} k\left(\boldsymbol{x_a}, \boldsymbol{x_b}\right), \tag{9}$$

where $k(\boldsymbol{x_i}, \boldsymbol{x_i}) = \phi(\boldsymbol{x_i})\phi(\boldsymbol{x_i})$ is the kernel matrix, while n_+ and n_- denote the quantities of positive and negative class samples, respectively. As evident from Eq. (9), we can efficiently determine the distance from the sample to the class center by computing the kernel function matrix among samples.

To compute the non-membership degree of the sample $\boldsymbol{x_i}$, we use the following formula.

$$\varsigma(\boldsymbol{x_i}) = \frac{D_P\left(\boldsymbol{x_i}\right) - D_N\left(\boldsymbol{x_i}\right) - D_g}{D_{PN} - D_g}, \tag{10}$$

where $D_g = \min(D_P(x_i) - D_N(x_i)), i = 1, ..., m$ represents the minimum distance difference and D_{PN} represents the distance between positive and negative classes, which can be obtained by calculating the inter-class distance between samples of the two classes. Then the subsequent comprehensive scoring formula is presented as follows:

$$s_i = \left(w_1 \varpi(\boldsymbol{x_i})^r + \left(1 - w_1\right)\left(1 - \varsigma(\boldsymbol{x_i})^r\right)\right)^{\frac{1}{r}}, \tag{11}$$

where r is a positive parameter that can be adjusted to differentiate between the two intuitive fuzzy values. The parameter w_1 lies within the range of $[0, 1]$ and serves as a weight for balancing the membership and non-membership degrees in accordance with the decision maker's preference.

In LDM, the slack variable ξ_i in Eq. (6) serves as a measure of misclassification, representing the difference in vertical coordinates before and after translation. A larger ξ_i implies more movement of the decision boundary, indicating greater sensitivity to noise and a higher susceptibility to errors. As s_i signifies the likelihood that sample $\boldsymbol{x_i}$ belongs to a particular class, the product $s_i\xi_i$ emerges as a suitable metric to comprehensively gauge decision errors. We update the optimal hyperplane partition strategy of LDM and name it the IFLMC model. The objective function can be reformulated as follows:

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

To minimize the decision cost of IFLMC, we implement the CS3WD method, allowing the model to postpone decisions or choose not to decide. The output of IFLMC is the distance from a sample point to the hyperplane, referred to as the margin. In binary classification problems, the sign of the margin indicates the class to which the sample belongs, and the magnitude of the margin reflects the proximity of the sample point to the decision boundary. We employ the Platt scaling method to transform the output of IFLMC into a threshold output, aiming to depict the probability of a sample belonging to the positive class.

As for the input sample $\boldsymbol{x_i}$, denote $f(\boldsymbol{x_i})$ as the threshold-free output of IFLMC. Following the formulation proposed by Platt [41], the probability of the sample $\boldsymbol{x}_i$ belonging to the positive class is parameterized using the sigmoid function, given by the following expression.

$$P(y_i = 1 \mid f(\boldsymbol{x_i})) = \frac{1}{1 + \exp(Af(\boldsymbol{x_i}) + B)}, \tag{13}$$

where A and B are the parameters to be optimized through the negative log-likelihood estimation on the training data.

Prior to the training process, we introduce a target probability t_i, defined as follows:

$$t_i = \frac{y_i + 1}{2}, \tag{14}$$

where y_i is the label of the sample and takes the value of $\{-1, 1\}$. Considering the characteristics of the sigmoid function, we use Eq. (14) to constrain t_i to $\{0, 1\}$.

To achieve a perfect fit to the target values, the sigmoid input needs to approach both extremes of the real number line. However, the sigmoid function exhibits insensitivity to changes in values near these extremes, making it challenging to distinguish between them. To address this issue, a smoothing process is applied to t_i, as illustrated below:

$$t_i = \begin{cases} \frac{n_+ + 1}{n_+ + 2}, & \text{if} \quad y_i = 1; \\ \frac{1}{n_+ + 2}, & \text{if} \quad y_i = -1; \end{cases} \quad i = 1, \ldots, m, \tag{15}$$

where n_+ represents the number of samples in the positive class. After obtaining the threshold-free output $f(\boldsymbol{x_i})$ and the target probability t_i for the training samples, we optimize the parameters A and B by minimizing the following objective function.

$$\min - \sum_i t_i \log\left(p_i\right) + \left(1 - t_i\right) \log\left(1 - p_i\right), \tag{16}$$

where $p_i = P(y_i = 1 \mid f(\boldsymbol{x_i}))$.

After the training process, we obtain the optimized parameters $\bar{A}$ and $\bar{B}$. To calculate the probability ${p_i}^z$ of a test sample $\boldsymbol{x_t}$ belonging to the positive class, we utilize the following equation.

$${p_i}^z = \frac{1}{1 + \exp(\bar{A}f(\boldsymbol{x_t}) + \bar{B})}. \tag{17}$$

Likewise, the confidence level $1 - {p_i}^z$ can be calculated to signify the probability of the test sample being part of the negative class. By utilizing the cost-sensitive three-decision rule, we obtain the following Bayesian risk loss, presented as follows:

$$\begin{aligned} R(a_P|x) &= \tau_{PN}(1 - {p_i}^z) + \tau_{PP}{p_i}^z, \\ R(a_B|x) &= \tau_{BN}(1 - {p_i}^z) + \tau_{BP}{p_i}^z, \\ R(a_N|x) &= \tau_{NN}(1 - {p_i}^z) + \tau_{NP}{p_i}^z, \end{aligned} \tag{18}$$

where $\tau_{PP}, \tau_{PN}, \tau_{NP}, \tau_{NN}, \tau_{BP}, \tau_{BN}$ are cost parameters, as detailed in Table 1. Consequently, the final decision function can be expressed as follows:

$$\phi^*(\boldsymbol{x}) = \underset{k\in\{P,N,B\}}{\arg\min}\ R(a_k|\boldsymbol{x}). \tag{19}$$

Remark 1. *CS3W-IFLMC model is a comprehensive model, and we have the following observations.*

- *Without the boundary decision items, the proposed CS3W-IFLMC will degenerate into CS2W-IFLMC, enabling less flexible to handle uncertainty.*
- *If there are no boundary decision terms, $\tau_{PP} = \tau_{NN} = 0$ and $\tau_{PN} = \tau_{NP}$, the proposed CS3W-IFLMC degenerates into CI2W-IFLMC that is actually the IFLMC (Eq. (12)). Furthermore, if $s = e$, the model degenerates into the LDM model.*
- *Through introducing the IF theory, the model attains a competitive level of noise resistance and generalization performance, as substantiated by experimental validation in Section 4.2.*
- *By combining IFLMC with CS3WC, our CS3W-IFLMC model is more adequate to handle uncertainty and cost-sensitive tasks, which is validated by experimental results in Section 4.3.*

3.2. The property of the proposed IF method

To demonstrate the rationality of the non-membership calculation formula, we provide the following three theorems.

Theorem 1. *The non-membership degree attributed to a sample is 0 when it is positioned at the center of the positive class.*

Proof. When the sample is positioned at the center of the positive class, the following equation can be derived.

$$D_P(\boldsymbol{x}_i) - D_N(\boldsymbol{x}_i) = -PN, \tag{20}$$

where PN represents the actual value of the distance between the two class centers, and the same applies below. Based on the trilateral knowledge of the triangle formed by the sample, positive class center, and negative class center, when the sample is situated on the extension line connecting the positive class center to the negative class center, the minimum value of D_g is $-PN$. Substituting this value into Eq. (10), we arrive at the subsequent expression.

$$\varsigma(\boldsymbol{x}_i) = \frac{D_P(\boldsymbol{x}_i) - D_N(\boldsymbol{x}_i) - D_g}{D_{PN} - D_g} = \frac{-PN - (-PN)}{2PN} = 0. \tag{21}$$

□

Theorem 2. *The non-membership degree of a sample is 1 when it is positioned at the center of the negative class.*

Proof. When the sample is positioned at the center of the negative class, the following equation can be derived.

$$D_P(\boldsymbol{x}_i) - D_N(\boldsymbol{x}_i) = PN. \tag{22}$$

Similarly, by considering the minimum value of D_g as $-PN$ and substituting it into Eq. (10), we can derive the following expression.

$$\varsigma(\boldsymbol{x}_i) = \frac{D_P(\boldsymbol{x}_i) - D_N(\boldsymbol{x}_i) - D_g}{D_{PN} - D_g} = \frac{PN - (-PN)}{2PN} = 1. \tag{23}$$

□

Theorem 3. *If a sample and its class center are located on the same side of the perpendicular bisector between the positive and negative class centers, the non-membership degree of the sample is less than* 0.5. *Conversely, the non-membership degree is greater than* 0.5.

Proof. Assuming $D_P(\boldsymbol{x_i}) = kD_N(\boldsymbol{x_i})$, here k represents a constant, we can utilize Eq. (10) to derive the following equation.

$$\varsigma(\boldsymbol{x_i}) = \frac{D_P(\boldsymbol{x_i}) - D_N(\boldsymbol{x_i}) - D_g}{D_{PN} - D_g} = \frac{k-1}{2} \cdot \frac{D_N(\boldsymbol{x_i})}{PN} + \frac{1}{2}. \tag{24}$$

When the Euclidean distance between a positive sample and the positive class center is smaller than the distance to the negative class center, the corresponding coefficient k is less than 1. This results in a non-membership degree $\varsigma(\boldsymbol{x_i})$ that is less than 0.5. Conversely, the corresponding coefficient k is greater than 1. Consequently, the non-membership degree $\varsigma(\boldsymbol{x_i})$ is greater than 0.5. □

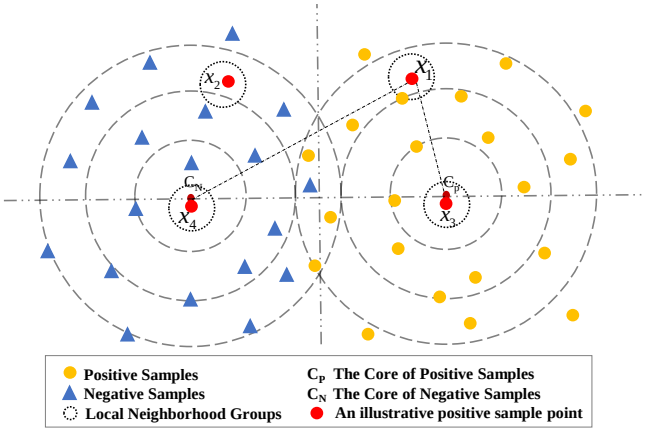


Figure 2: Fuzzy Membership Method Illustration.

From a theoretical perspective, the efficacy of the proposed IF methodology is readily apparent, and in this context, we present numerical illustrations to clarify the Theorems 1 - 3.

Example 1. *In Fig. 2 we observe that x_3 is positioned at the centroid of the positive-class samples. The discrepancy between its distance from the positive-class centroid and the negative-class centroid can be calculated as $-PN$. Moreover, D_{PN} and $D_g$are set to PN and $-PN$, respectively. Substituting these values into the provided non-membership formula yields a non-membership degree of* 0 *for x_3.*

In the case of sample x_4 in Fig. 2, it is located at the center of the negative-class samples. In this scenario, the difference between its distance from the positive-class center and the negative-class center can be quantified as PN. Keeping all other values the same as those of x_3, applying these values in the formula results in a non-membership degree of 1 *for x_4.*

The data points x_1 and x_2 represent more general instances. x_1 is closer to the positive-class center than the negative-class center, leading to a k-value smaller than 1. Consequently, the first term in Eq. (24) is negative, assigning a non-membership degree for x_1 less than 0.5. *On the other hand, x_2 is farther away from the positive-class center compared to the negative-class center, resulting in a greater than 1. As a result, the first term in Eq. (24) becomes positive, yielding a non-membership degree for x_2 greater than* 0.5.

3.3. Solver for the model

Model optimization comprises three key steps. Initially, the optimal classification hyperplane is derived by optimizing the dual problem of IFLMC. Subsequently, the cross-entropy loss function is employed to train and optimize the logistic regression model for estimating the posterior probability of the IFLMC classification result. Finally, the obtained probability is input into the CS3WD decision-making framework to attain the optimal decision-making result.

By incorporating IF logic into LDM, both linear and nonlinear problems can be effectively addressed. In real-world scenarios, linear inseparability is a common occurrence, and thus our main emphasis here lies in deriving IFLMC formulas for linearly inseparable data. The score s_i signifies the probability of a sample belonging to a particular class,

rendering $s_i\xi_i$ a suitable metric for the comprehensive measurement of decision errors. Consequently, the optimization objective function of IFLMC can be expressed as the following equation:

$$\begin{aligned} &\min_{\boldsymbol{w},\xi} \frac{1}{2}\|\boldsymbol{w}\|^2 + \lambda_1\hat{\gamma} - \lambda_2\bar{\gamma} + C\sum_{i=1}^{m} s_i\xi_i, \\ &\text{s.t. } y_i\boldsymbol{w}^T\phi\left(\boldsymbol{x_i}\right) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \ldots, m. \end{aligned} \tag{25}$$

Denote $\boldsymbol{X}$ as the matrix whose $i-$th column is $\phi(\boldsymbol{x_i})$, i.e., $\boldsymbol{X} = [\phi(\boldsymbol{x_1})...\phi(\boldsymbol{x_m})]$ and $\boldsymbol{y} = [y_1, ..., y_m]^T$ is a column vector. Substituting Eq. (7) into Eq. (25) yields the following equation:

$$\begin{aligned} &\min_{\boldsymbol{w},\xi} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{2\lambda_1}{m^2}\left(m\boldsymbol{w}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{w} - \boldsymbol{w}^T\boldsymbol{X}\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{X}^T\boldsymbol{w}\right) \\ &- \lambda_2\frac{1}{m}(\boldsymbol{X}\boldsymbol{y})^T\boldsymbol{w} + C\sum_{i=1}^{m} s_i\xi_i, \\ &\text{s.t. } y_i\boldsymbol{w}^T\varphi\left(\boldsymbol{x_i}\right) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \ldots, m. \end{aligned} \tag{26}$$

Inspired by Zhou et al. [30], the optimal solution for this problem can be expressed by $\varphi(\boldsymbol{x_i})$. Let $\boldsymbol{R} = \boldsymbol{X}\boldsymbol{X}^T$. According to this theory, the following expression is derived:

$$\begin{aligned} &\boldsymbol{w} = \boldsymbol{X}\boldsymbol{\alpha}, \\ &\boldsymbol{X}^T\boldsymbol{w} = \boldsymbol{R}\boldsymbol{\alpha}, \\ &\boldsymbol{w}^T\boldsymbol{w} = \boldsymbol{\alpha}^T\boldsymbol{R}\boldsymbol{\alpha}, \end{aligned} \tag{27}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_m]^T$ are the coefficients. Use $\boldsymbol{R_{:i}}$ denote the i-th column of $\boldsymbol{R}$, then the matrix representation of Eq. (26) is obtained as:

$$\begin{aligned} &\min_{\boldsymbol{\alpha},\xi} \quad \tfrac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{Q}\boldsymbol{\alpha} + \boldsymbol{p}^T\boldsymbol{\alpha} + C\sum_{i=1}^{m} s_i\xi_i, \\ &s.t. \quad y_i\boldsymbol{\alpha}^T\boldsymbol{R_{:i}} \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \ldots, m, \end{aligned} \tag{28}$$

where

$$\begin{aligned} &\boldsymbol{Q} = 4\lambda_1(m\boldsymbol{R}^T\boldsymbol{R} - (\boldsymbol{R}\boldsymbol{y})(\boldsymbol{R}\boldsymbol{y})^T)/m^2 + \boldsymbol{R}, \\ &\boldsymbol{p} = -\lambda_2\boldsymbol{R}\boldsymbol{y}/m. \end{aligned} \tag{29}$$

Theorem 4. *The dual problem of Eq. (28) is*

$$\begin{aligned} &\min_{\beta} \quad \tfrac{1}{2}\boldsymbol{\beta}^T\boldsymbol{Z}\boldsymbol{\beta} + (\tfrac{\lambda_2}{m}\boldsymbol{Z}\boldsymbol{e} - \boldsymbol{e})^T\boldsymbol{\beta}, \\ &s.t. \quad 0 \leq \beta_i \leq Cs_i, i = 1, \ldots, m, \end{aligned} \tag{30}$$

where $\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}\boldsymbol{Y}$, $\boldsymbol{Y}$ is a $m \times m$ diagonal matrix with $y_1, ..., y_m$ as the diagonal elements, $\boldsymbol{Q}^{-1}$refers to the inverse matrix of $\boldsymbol{Q}$ and $\boldsymbol{e} = [1, 1, ..., 1]^T$ stands for the all-one vector.

Proof. By introducing the Lagrange method, we can obtain the following equation:

$$\begin{aligned} &L(\boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \tfrac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{Q}\boldsymbol{\alpha} + \boldsymbol{p}^T\boldsymbol{\alpha} + C\sum_{i=1}^{m} s_i\xi_i \\ &- \sum_{i=1}^{m}\beta_i(y_i\boldsymbol{\alpha}^T\boldsymbol{R_{:i}} + s_i\xi_i - 1) - \sum_{i=1}^{m}\theta_i s_i\xi_i, \end{aligned} \tag{31}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are the constraint values of two marginal distribution parameters.

By setting the partial derivations of $\{\boldsymbol{\alpha}, \boldsymbol{\theta}\}$to zero, we have

$$\begin{aligned} &\frac{\partial L}{\partial \boldsymbol{\alpha}} = \boldsymbol{Q}\boldsymbol{\alpha} + \boldsymbol{p} - \sum_{i=1}^{m}\beta_i y_i\boldsymbol{R_{:i}} = 0, \\ &\frac{\partial L}{\partial \xi_i} = s_i C - \beta_i - \theta_i = 0, i = 1, \ldots, m. \end{aligned} \tag{32}$$

Substituting Eq. (31) and Eq. (32) into Eq. (28) yields the Eq. (30). □

For prediction, according to Eq. (32), one can obtain the coefficients $\boldsymbol{\alpha}$ from the optimal $\boldsymbol{\beta}^*$ as

$$\boldsymbol{\alpha} = \boldsymbol{Q}^{-1}\boldsymbol{R}\boldsymbol{Y}\left(\frac{\lambda_2}{m}\boldsymbol{e} + \boldsymbol{\beta}^*\right). \tag{33}$$

Therefore, for a testing instance $\boldsymbol{z}$, its corresponding label can be derived by

$$\operatorname{sgn}\left(\boldsymbol{w}^{\top}\phi(\boldsymbol{z})\right) = \operatorname{sgn}\left(\sum_{i=1}^{m}\alpha_i k\left(\boldsymbol{x}_i, \boldsymbol{z}\right)\right). \tag{34}$$

The above describes the optimization process of IFLMC, but its decision-making approach is immediate. We transform IFLMC into threshold output using Eq. (13) and (15), obtaining the probability of a sample belonging to the positive class. For Eq. (16), we optimize it using the gradient descent method to obtain $\bar{A}$ and $\bar{B}$. Finally, the output probability and the decision result with the minimum cost for test samples can be obtained through Eq. (17) and (19).

3.4. Derivation of the Decision Threshold

To streamline the decision-making process, we can utilize the decision rule provided in Theorem 5 to directly classify the sample into a specific domain based on its confidence level.

Theorem 5. *In practice, suppose $\tau_{PP} = \tau_{NN} = 0$, $\tau_{PP} \leq \tau_{BP} \leq \tau_{NP}$, $\tau_{NN} \leq \tau_{BN} \leq \tau_{PN}$, the decision rules are presented as follows:*

$$x \in \begin{cases} C_P, & if \quad {p_i}^z \geq \delta, \\ C_B, & if \quad \eta < {p_i}^z < \delta, \\ C_N, & if \quad {p_i}^z \leq \eta, \end{cases} \tag{35}$$

where the parameters δ and η can be computed as follows:

$$\delta = \frac{\tau_{BN}}{\tau_{BN} + \tau_{NP} - \tau_{BP}}, \tag{36}$$

$$\eta = \frac{\tau_{PN} - \tau_{BN}}{\tau_{PN} - \tau_{BN} + \tau_{BP}}. \tag{37}$$

Proof. When a decision is made in the boundary domain, the cost associated with the boundary domain is lower than that of both the positive and negative domains, satisfying the following equation:

$$\begin{cases} R\left(a_B \mid x\right) \leq R\left(a_P \mid x\right), \\ R\left(a_B \mid x\right) \leq R\left(a_N \mid x\right). \end{cases} \tag{38}$$

Expanding the two equations separately, we obtain the following expressions:

$$\begin{aligned} & R\left(a_B \mid x\right) \leq R\left(a_P \mid x\right) \\ \Rightarrow & \tau_{BN}\left(1 - p_i^z\right) + \tau_{BP}p_i^z \leq \tau_{PN}\left(1 - p_i^z\right) + \tau_{PP}p_i^z \\ \Rightarrow & \left(\tau_{PN} - \tau_{BN} + \tau_{BP} - \tau_{PP}\right)p_i^z \leq \tau_{PN} - \tau_{BN} \\ \because \quad & \tau_{PN} - \tau_{BN} + \tau_{BP} - \tau_{PP} \geq 0 \\ \therefore \quad & p_i^z \leq \frac{\tau_{PN} - \tau_{BN}}{\tau_{PN} - \tau_{BN} + \tau_{BP} - \tau_{PP}} = \eta, \end{aligned} \tag{39}$$

and

$$\begin{aligned} & R\left(a_B \mid x\right) \leq R\left(a_N \mid x\right) \\ \Rightarrow & \tau_{BN}\left(1 - p_i^z\right) + \tau_{BP}p_i^z \leq \tau_{NN}\left(1 - p_i^z\right) + \tau_{NP}p_i^z \\ \Rightarrow & \left(\tau_{NN} - \tau_{BN} + \tau_{BP} - \tau_{NP}\right)p_i^z \leq \tau_{NN} - \tau_{BN} \\ \because \quad & \tau_{NN} - \tau_{BN} + \tau_{BP} - \tau_{NP} \leq 0 \\ \therefore \quad & p_i^z \geq \frac{\tau_{NN} - \tau_{BN}}{\tau_{NN} - \tau_{BN} + \tau_{BP} - \tau_{NP}} = \delta. \end{aligned} \tag{40}$$

By using the same method, the decision rule for Eq. (35) can be obtained. □

Algorithm 1 CS3W-IFLMC

Input: Training data set $\mathbf{X_i}$, test data $\mathbf{X_z}$, parameters λ_1, λ_2, C, kernel function K and the Cost-sensitive Matrix.
Output: Decision options for test data $\mathbf{X_z}$
for each combination of parameters **do**
    Perform 5-fold cross-validation:
    for each fold **do**
        Divide the training set into new training and validation sets.
        Calculate distances from the sample to positive and negative class centers.
        Compute sample's membership and non-membership using Eq. (8) and (10).
        Calculate sample scores using Eq. (11).
        Optimize using Eq. (30) to obtain β^*.
        Calculate α using Eq. (33).
        Obtain validation set labels and calculate accuracy.
    end for
    Calculate the average accuracy over the 5 folds.
end for
Choose parameters with the highest average accuracy.
Train the optimal model with the selected parameters on the training set $\mathbf{X_i}$.
Predict using the optimal model on test set $\mathbf{X_z}$ to obtain raw outputs.
Optimize using Eq. (16) to obtain parameters $\bar{A}$ and $\bar{B}$.
Calculate the probability outputs for test set $\mathbf{X_z}$ using Eq. (17).
Determine the final decision boundaries using Eq. (35).

Table 2
Dataset details

Size	Dataset	Samples×Attributes
Small	German	1000×25
	Iris	150×5
	Wine	178×14
	heart_statlog	270×14
	Glass	214×10
	Sonar	208×61
	clean1	476×167
	cmc_0_2	1140×10
	transfusion_new	748×5
Large	creditcard_new	1872×30
	waveform-5000_0_1	3345×41
	kr-vs-kp	3196×37

The "transfusion_new" and "creditcard_new" datasets are official Kaggle repositories, while the remaining datasets are UCI benchmark datasets.

To provide a clearer explanation of the algorithm logic behind CS3W-IFLMC, we present the pseudocode, presented as Algorithm 1.

4. Experiment

In this section, we present a thorough experimental evaluation of the CS3W-IFLMC model on benchmark datasets from diverse fields of UCI and official Kaggle repositories. We provide detailed explanations of our experimental settings in Section 4.1. In Section 4.2, we validate the effectiveness of the proposed IF method, across 12 benchmark

Table 3
Parameter value range

Parameter	Value set
λ_1	$\{2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}\}$
λ_2	$\{1\}$
C	$\{1, 10, 50, 100, 500\}$
γ	$\{2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^{0}, 2^{1}, 2^{2}, 2^{3}, 2^{4}\}$

Table 4
Comparison of accuracy between original data

Dataset	SVM	FSVM	LDM	IFLMC
	acc	acc	acc	acc
German	74.10	75.20	**75.70**	75.27
Iris	96.00	**97.33**	96.00	96.89
Wine	**98.30**	97.73	97.16	97.73
heart_statlog	82.22	82.96	82.22	**83.95**
Glass	**91.04**	91.04	91.04	89.94
Sonar	79.61	79.61	77.18	**81.23**
clean1	88.40	**89.45**	86.71	87.20
cmc-0-2	96.92	96.63	95.89	**96.87**
kr-vs-kp	94.19	94.61	94.47	**95.07**
creditcard_new	95.76	95.76	95.73	**95.77**
waveform-5000_0_1	91.00	90.42	91.13	**91.98**
transfusion_new	74.59	72.36	74.22	**77.37**
Ave.	88.51	88.59	88.12	**89.11**

In the presented table, the best accuracy on each dataset is bolded. The "acc" denotes classification accuracy while "Ave." represents average classification accuracy (similarly hereinafter).

and noise-added datasets. Moreover, in Section 4.3.3, we conduct a comprehensive cost analysis of our CS3W-IFLMC model compared to CS2W-IFLMC and traditional IFLMC models using the same datasets. Additionally, in Section 4.3.1 and Section 4.3.2, we investigate the effect of boundary costs on the total cost and number of boundary samples. Finally, in Section 4.4, we summarize the experimental results.

4.1. Experimental Settings

In order to conduct a thorough assessment of the proposed 3W-IFLMC model, we performed experiments on 12 distinct datasets obtained from UCI and Kaggle. These datasets were carefully selected from various fields, as detailed in Table 2. For standard datasets, we divided them into 70% for training and 30% for testing using a random approach. For the larger-scale datasets, following the approach in [47], we selected 10% of the datasets to perform parameter selection using five-fold cross-validation and grid optimization. During model construction, we assessed the impact of varying λ_2 on the output. Analysis revealed that λ_2 has minimal influence on the results, indicating low sensitivity. To simplify the model and reduce complexity, we fixed λ_2 at 1. Other parameters, with significant impact on performance, were set with a wide range to fully explore optimal configurations during optimization. This strategy enhances model adaptability, accuracy, and generality. The parameters and their traversal ranges used in the experiment are shown in Table 3. To evaluate the model's performance, we employed a Gaussian kernel for experiments on UCI and Kaggle datasets. All experiments were conducted on a PC equipped with an 8 × 4.00 GHz CPU and 32GB memory using MATLAB R2020a. These experimental settings were carefully controlled to ensure reliable and reproducible results.

4.2. Experimental Analysis on the Proposed IF Strategy

From the above analysis, it can be concluded that if there are no boundary decision terms, $\tau_{PP} = \tau_{NN} = 0$ and $\tau_{PN} = \tau_{NP}$, the proposed CS3W-IFLMC degenerates into IFLMC. In order to demonstrate the classification performance of the IFLMC model, we conducted a comparison with three benchmark algorithms: SVM, LDM, and

Table 5
Comparison of accuracy between data with noise

Dataset	Noise ratio	SVM	FSVM	LDM	IFLMC
		acc	acc	acc	acc
German	r=0.05	68.11	64.11	69.33	**70.52**
	r=0.1	**69.60**	68.30	69.50	69.00
Iris	r=0.05	80.00	68.89	78.52	**93.33**
	r=0.1	87.16	87.84	85.14	**88.74**
Wine	r=0.05	93.01	92.74	92.74	**93.28**
	r=0.1	91.57	91.57	90.45	**92.51**
heart_statlog	r=0.05	75.49	**76.90**	75.31	75.66
	r=0.1	70.90	70.90	70.90	**71.89**
Glass	r=0.05	87.11	87.11	**87.56**	86.89
	r=0.1	83.18	82.24	83.64	**84.42**
Sonar	r=0.05	77.93	77.93	**78.85**	78.16
	r=0.1	72.33	72.33	**74.27**	73.46
clean1	r=0.05	81.28	81.28	72.27	**81.68**
	r=0.1	**79.41**	77.94	77.10	79.27
cmc-0-2	r=0.05	88.70	86.62	89.92	**89.90**
	r=0.1	87.10	**87.54**	87.10	87.39
kr-vs-kp	r=0.05	88.70	86.62	**89.92**	89.90
	r=0.1	84.22	84.22	84.29	**85.97**
creditcard_new	r=0.05	90.68	90.54	90.70	**90.92**
	r=0.1	85.34	85.16	85.64	**86.23**
waveform-5000_0_1	r=0.05	87.65	87.63	**87.74**	87.39
	r=0.1	81.35	81.23	81.31	**83.01**
transfusion_new	r=0.05	72.81	72.26	72.61	**74.00**
	r=0.1	73.11	72.96	73.18	**73.68**
Ave.	r=0.05	82.62	81.05	82.12	**84.30**
	r=0.1	80.44	80.19	80.21	**81.30**

Table 6
Comparison of IF strategies

	Fuzzy	IF_a	IF_b	IF
Iris(r=0.05)	92.79	92.34	92.34	**93.33**
Iris(r=0.1)	86.49	87.39	88.29	**88.74**
Glass(r=0.05)	86.79	86.16	86.48	**86.89**
Glass(r=0.1)	81.00	81.00	**84.42**	**84.42**

FSVM. Firstly, we conduct classification experiments on an aggregate of 12 regular datasets to evaluate the predictive accuracy of each algorithm. In addition, to assess the algorithm's capability in handling label noise, we conducted experiments on a dataset where label noise was introduced. Specifically, we extracted a subset of the original data, comprising a proportion of r times the total instances, and modified their corresponding labels to a different class. Here, we set $r = 0.05$ and $r = 0.1$. The experimental results of different algorithms on the original dataset are presented in Table 4. The data set with added noise is shown in Table 5.

According to the results, the IFLMC algorithm has demonstrated superior classification performance compared to other algorithms on both the original dataset and the dataset with label noise. Among the 12 original datasets, IFLMC showed outstanding classification performance, with the highest average accuracy reaching 89.1%. Furthermore, on large-scale datasets, IFLMC outperformed other models, highlighting its superiority in handling large-scale datasets. On the 12 datasets with added noise, regardless of the noise ratio being 0.05 or 0.1, IFLMC achieved higher average accuracies than other models, reaching 84.30% and 81.30% respectively, which was significantly better than the second-ranked model. Specifically, at the noise ratio of 0.05, IFLMC had the highest accuracy on 7 datasets, while at the noise ratio of 0.1, IFLMC had the highest accuracy on 8 datasets. This result indicates that under higher noise ratio conditions,

the IFLMC model exhibits stronger classification performance and stability. These results confirm the effectiveness of the proposed IF strategy and can significantly improve the noise recovery ability of LDM.

To fully demonstrate the advantages of the IF strategy proposed in this paper, we conducted a comparative analysis with the traditional Fuzzy strategy, the IF strategy proposed by Zhang, and the IF strategy proposed by Rezvani on two datasets. The experimental results are summarized in Table 6, where Fuzzy represents the traditional fuzzy strategy, IF_a is the strategy proposed by Rezvani [45], and IF_b is the strategy proposed by Zhang [39]. After carefully examining the experimental results, we found that the fuzzy strategy proposed in this paper significantly outperforms the other comparative strategies in terms of classification accuracy, demonstrating a clear advantage.

4.3. Comprehensive Experiments on CS3W-IFLMC Model

The aforementioned results explicitly demonstrate that the IF logic strategy proposed in this study effectively boosts the classification performance of LDM. However, the conducted experiments in Section 4.2 did not consider the cost of misclassification. To validate the effectiveness of the cost-sensitive 3WD in reducing the decision costs of IFLMC, a comprehensive analysis of the decision costs for CS3W-IFLMC is presented in this section.

4.3.1. Impact of Cost Matrix Variation on Decision Cost

In this subsection, we examine the relationship between the decision cost of the CS3W-IFLMC model and the boundary cost parameters τ_{BP} and τ_{BN}. Without loss of generality, we conducted experiments on 12 original datasets, with the cost parameters set as $\tau_{PP} = 0, \tau_{PN} = 21, \tau_{NP} = 41, \tau_{NN} = 0$. Specifically, the cost parameters for the boundary domains were determined as follows.

$$\begin{cases} \tau_{BP} = \theta_1 \tau_{NP}, \\ \tau_{BN} = \theta_2 \tau_{PN}, \end{cases}$$

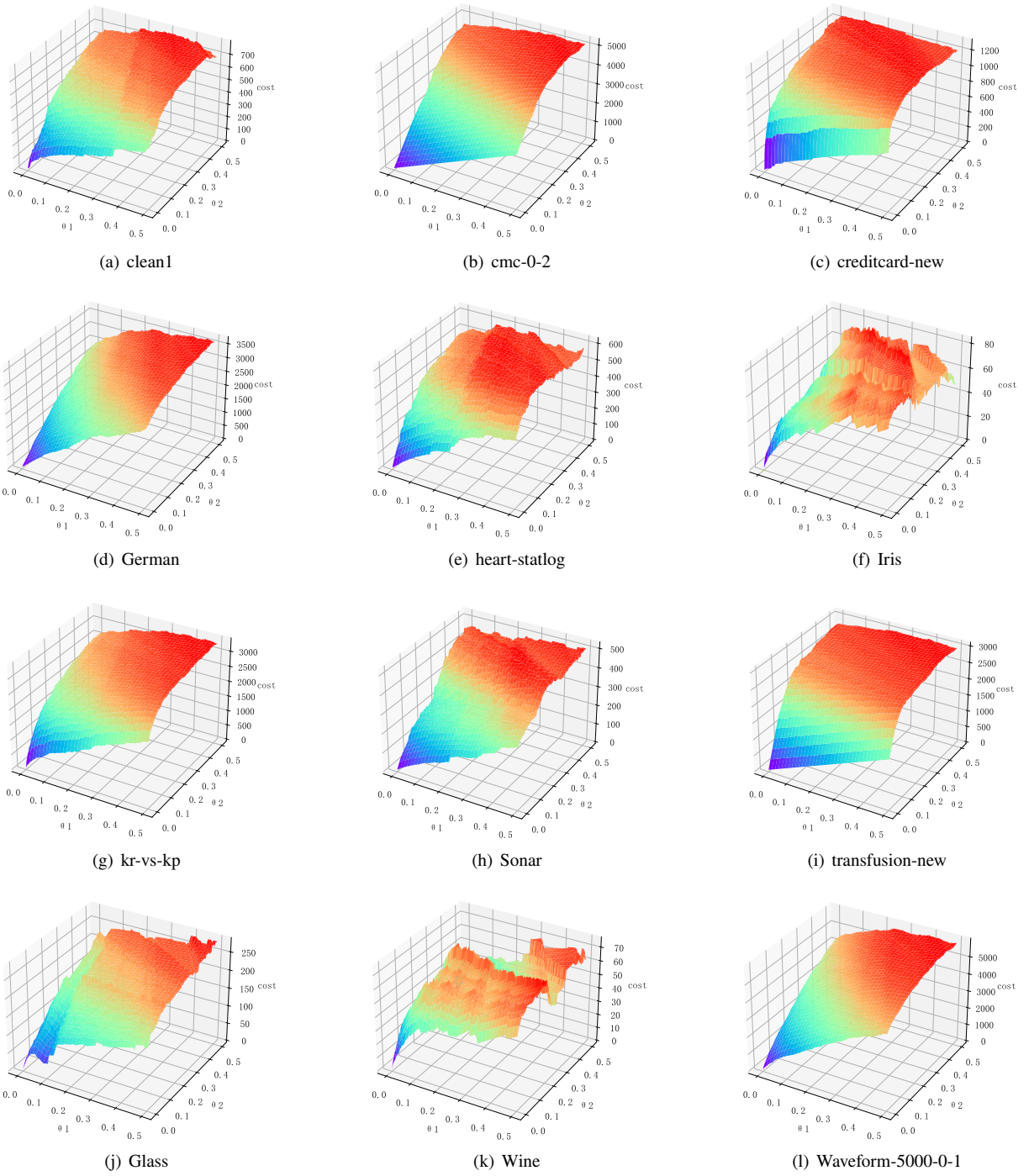
where θ_1 and θ_2 are predefined coefficients with a range between 0 and 1. The results of these experiments are depicted in Fig. 3.

The results of the experiments in this section are presented through three-dimensional surface plots, using a rainbow color scheme. Warmer colors, closer to red, indicate higher costs, while cooler colors, closer to blue, indicate lower costs. From the plots, it is evident that the cost trends across the 12 datasets are generally similar. When θ_1 and θ_2 have smaller values, the costs are lower. Conversely, larger values of θ_1 and θ_2 lead to higher costs. Furthermore, once both θ_1 and θ_2 exceed 0.35, the total cost reaches a stable state, with little to no further increase. This indicates that when the boundary costs in the cost matrix are smaller, the overall cost tends to lower. On the flip side, raising the boundary costs leads to higher total costs. However, once the values of θ_1 and θ_2 surpass a certain threshold (0.35), the total cost stabilizes. These findings provide valuable guidance for selecting appropriate values of θ_1 and θ_2, indicating that values greater than 0.35 are not necessary.

4.3.2. Sample Distribution Analysis with Varying Cost Matrix

In this section, we conducted experiments to explore the relationship between the number of samples classified as boundary domains and the boundary cost coefficients (θ_1, θ_2). We performed experiments on 12 original datasets, where we fixed one parameter and varied the other. The parameter values ranged from 0 to 1 with a step size of 0.1. The results of these experiments are depicted in Fig. 4.

Fig. 4 presents the distribution of the number of boundary domain samples for four datasets (cmc-0-2, German, Sonar, Glass) under different θ_1 and θ_2 settings. The first column of Fig. 4 represents the scenario where θ_2 is fixed and θ_1 is varied, while the second column corresponds to the scenario where θ_1 is fixed and θ_2 is varied. Upon examining the plots, it becomes evident that the distribution of the number of boundary domain samples exhibits a similar trend when either θ_1 or θ_2 is changed. Furthermore, as θ_1 and θ_2 increase, the number of boundary domain samples also increases. It is worth noting that different datasets demonstrate varying rates of increase in the number of boundary domain samples. These findings suggest that the changes in θ_1 and θ_2 have a comparable effect on the distribution of the number of samples in the boundary domain. Additionally, elevating the boundary cost in the cost matrix leads to a greater number of boundary domain samples. It is noteworthy that the sensitivity of parameters θ_1 and θ_2 varies across different datasets. Specifically, the cmc-0-2 dataset exhibits a significant sensitivity towards these parameters, whereas the Glass dataset demonstrates negligible sensitivity.



(a) clean1 (b) cmc-0-2 (c) creditcard-new

(d) German (e) heart-statlog (f) Iris

(g) kr-vs-kp (h) Sonar (i) transfusion-new

(j) Glass (k) Wine (l) Waveform-5000-0-1

Figure 3: The decision costs of CS3W-IFLMC under different values of θ_1 and θ_2.

4.3.3. Comparison of Decision Costs

In this subsection, we performed experiments on 12 original datasets to compare the final decision costs of three models: IFLMC, CS2W-IFLMC, and CS3W-IFLMC. The CS2W-IFLMC model makes decisions by converting the threshold-free output of IFLMC into probabilities and minimizing the expected cost, considering only positive and negative domains. Without loss of generality, the cost parameters were set as follows: $\tau_{PP} = 0, \tau_{PN} = 21, \tau_{NP} =$

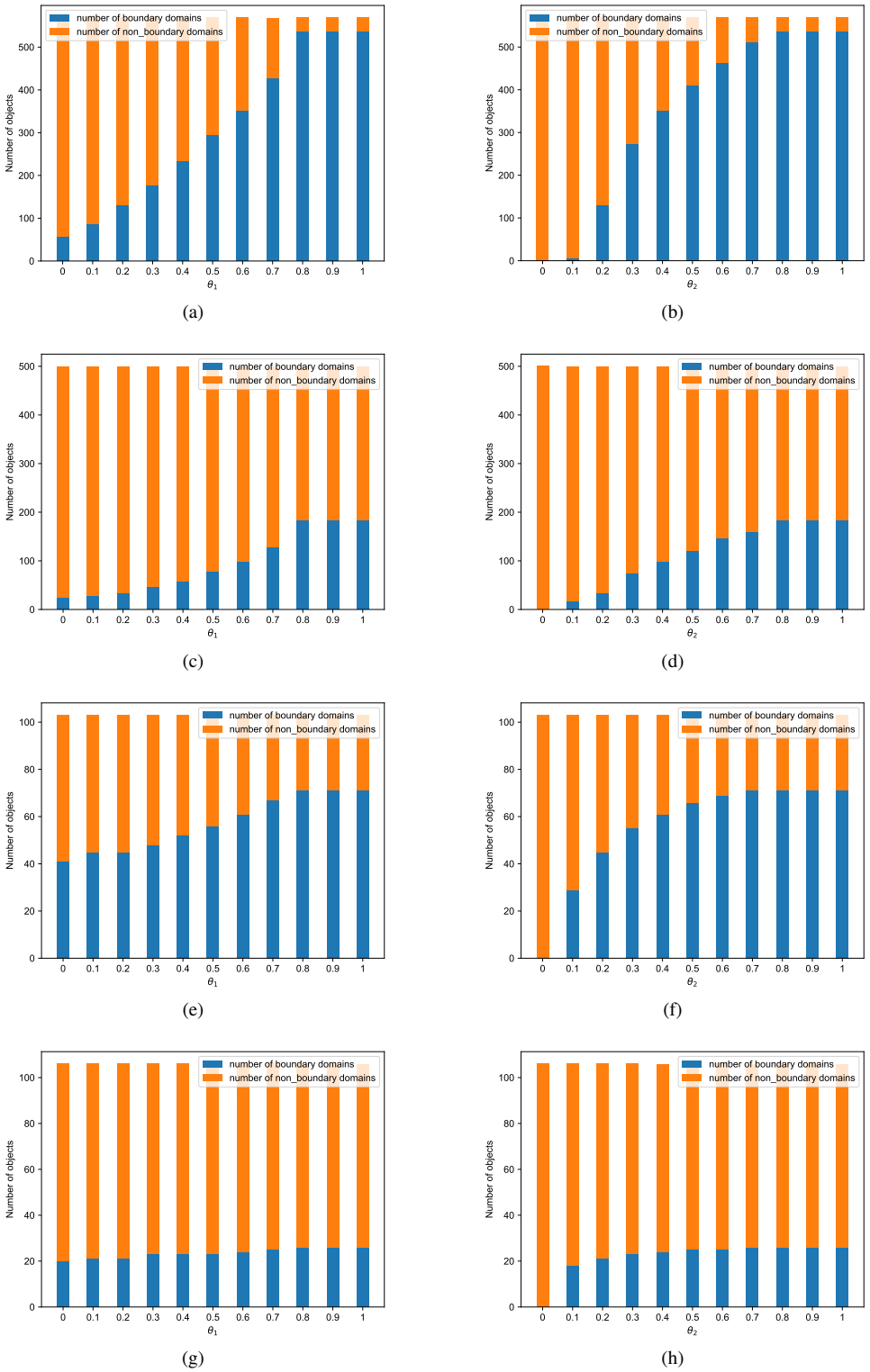


Figure 4: The number of samples in boundary and non-boundary domains under different boundary cost settings: We evaluate our proposed method using multiple datasets. Specifically, datasets (a) and (b) are cmc-0-2 datasets, (c) and (d) are German datasets, (e) and (f) are Sonar datasets, and (g) and (h) are Glass datasets. To analyze the impact of different parameters, we conduct experiments on datasets (a), (c), and (e) by varying the value of θ_1 while keeping θ_2 fixed. Similarly, datasets (b), (d), and (f) are used to examine the effect of changing θ_2 while keeping θ_1 constant.

Table 7
Comparison of the decision costs among different algorithms

Dataset	IFLMC	CS2W-IFLMC	CS3W-IFLMC
German	4703.67	3620.67	**1193.67**
Iris	49.00	**47.00**	48.00
Wine	82.00	62.67	**51.33**
heart_statlog	601.67	583.33	**314.00**
Glass	284.00	283.00	**74.67**
Sonar	492.67	507.67	**181.67**
clean1	843.67	695.00	**321.67**
cmc-0-2	6617.67	5121.00	**1452.00**
kr-vs-kp	3359.33	3300.67	**1786.67**
creditcard_new	1383.33	1214.67	**896.00**
waveform-5000_0_1	7681.33	5873.67	**2201.33**
transfusion_new	3228.67	2978.67	**1585.33**
Ave.	2443.92	2024.00	**842.19**

41, $\tau_{NN} = 0, \tau_{BP} = 2, \tau_{BN} = 3$. The decision cost results of different models on the original datasets are presented in Table 7.

Based on the results presented in Table 7, it is evident that CS3W-IFLMC outperforms other models in terms of decision cost on 11 out of 12 datasets, achieving the lowest average decision cost of 842.19 across all datasets. Comparing CS2W-IFLMC and IFLMC emphasizes the effectiveness of the cost-sensitive approach in reducing the model's decision cost. Furthermore, comparing CS3W-IFLMC with CS2W-IFLMC reveals that incorporating boundary terms and introducing a delayed decision-making process significantly contributes to reducing the decision cost. Lastly, when comparing CS3W-IFLMC with IFLMC, it is evident that CS3W-IFLMC exhibits an average decision cost that is approximately one-third of that of IFLMC. This emphasizes the improved reliability of the model's decision-making process achieved through the conversion of IFLMC's threshold-free output into probabilities and the adoption of a cost-sensitive ternary decision approach.

4.4. A Summary of Experimental Analysis

Upon analyzing all the experimental outcomes, we can derive the following conclusions.

- By incorporating the IF approach, IFLMC demonstrates superior classification performance compared to LDM on both the initial datasets and the datasets with label noise (see Section 4.2).
- The cost parameters τ_{BP} and τ_{BN} for the boundary domains have a consistent effect on the decision cost of the CS3W-IFLMC model. As the values of τ_{BP} and τ_{BN} increase, the overall decision cost of the model also increases, accompanied by a higher number of samples classified as boundary domains. (see Section 4.3.1 and 4.3.2).
- The proposed CS3W-IFLMC model exhibits lower decision costs compared to other algorithms and provides higher reliability in classification results. (see Section 4.3.3).

  CS3W-IFLMC model code link: https://github.com/HengCode12/CS3W_IFLMC.git

5. Conclusion

By introducing a novel IF strategy and the CS3WD method to the LDM model, we propose a novel CS3W-IFLMC method, which significantly enhances the reliability and noise resilience. This study represents the pioneering effort to integrate 3WD with LDM, offering a viable approach for adeptly managing uncertainty. We discuss the IF strategy and the threshold output of the model from a theoretical perspective. Additionally, we authenticate the effectiveness of the proposed CS3W-IFLMC method through experimental analysis. Comparisons are conducted with SVM, FSVM, and LDM algorithms on 12 benchmark datasets, where IFLMC demonstrates outstanding performance. Furthermore, in the analysis of decision costs, CS3W-IFLMC exhibits the most prominent results. The experimental results indicate that our model has broad application scenarios in fields such as healthcare, financial risk assessment, and intelligent driving. For instance, in tumor recognition, our model can effectively distinguish noise and make delayed decisions

for uncertain samples. By integrating expert knowledge, it produces more reliable decisions, thereby reducing risks for patients. However, as designed for binary classification datasets, CS3W-IFLMC does not apply to multi-class problems, which is one of our future research directions. Moreover, its performance in unbalanced classification tasks needs to be further explored in the future.

CRediT authorship contribution statement

Shuangyi Fan: Writing - original draft, Methodology, Software, Writing - review & editing. **Heng Li:** Writing - review & editing. **Cong Guo:** Writing - review & editing. **Dun Liu:** Conceptualization, Validation, Writing - review & editing. **Libo Zhang:** Conceptualization, Validation, Writing - review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Nos. 62276217, 62106205 and 61876157), the Science Fund for Distinguished Young Scholars of Sichuan Province (No. 2022JDJQ0034).

References

[1] Y. Y. Yao, "Three-way decisions with probabilistic rough sets," *Information Sciences*, vol. 180, no. 3, pp. 341–353, 2010.
[2] W. H. Xu, D. D. Guo, J. S. Mi, Y. H. Qian, K. Y. Zheng, and W. P. Ding, "Two-way concept-cognitive learning via concept movement viewpoint," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2023.
[3] X. Yang, Y. Chen, H. Fujita, D. Liu, and T. R. Li, "Mixed data-driven sequential three-way decision via subjective–objective dynamic fusion," *Knowledge-Based Systems*, vol. 237, p. 107728, 2022.
[4] Y. F. Li, L. B. Zhang, Y. Xu, Y. Yao, R. Y. K. Lau, and Y. T. Wu, "Enhancing binary classification by modeling uncertain boundary in three-way decisions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 7, pp. 1438–1451, 2017.
[5] Y. Zhang and J. T. Yao, "Gini objective functions for three-way classifications," *International Journal of Approximate Reasoning*, vol. 81, pp. 103–114, 2017.
[6] D. Liu, "The effectiveness of three-way classification with interpretable perspective," *Information Sciences*, vol. 567, pp. 237–255, 2021.
[7] H. Yu, Y. Chen, P. Lingras, and G. Y. Wang, "A three-way cluster ensemble approach for large-scale data," *International Journal of Approximate Reasoning*, vol. 115, pp. 32–49, 2019.
[8] M. J. Du, J. Q. Zhao, J. R. Sun, and Y. Q. Dong, "M3w: Multistep three-way clustering," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2022.
[9] Z. Wang, C. J. Shi, L. Wei, and Y. Y. Yao, "Tri-granularity attribute reduction of three-way concept lattices," *Knowledge-Based Systems*, vol. 276, p. 110762, 2023.
[10] X. Y. Zhang and Y. Y. Yao, "Tri-level attribute reduction in rough set theory," *Expert Systems with Applications*, vol. 190, p. 116187, 2022.
[11] G. M. Lang, D. Q. Miao, and H. Fujita, "Three-way group conflict analysis based on pythagorean fuzzy set theory," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 3, pp. 447–461, 2020.
[12] T. X. Wang, H. X. Li, Y. H. Qian, B. Huang, and X. Z. Zhou, "A regret-based three-way decision model under interval type-2 fuzzy environment," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 1, pp. 175–189, 2022.
[13] M. W. Wang, D. C. Liang, and D. F. Li, "A two-stage method for improving the decision quality of consensus-driven three-way group decision-making," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 5, pp. 2770–2780, 2023.
[14] X. D. Yue, Y. F. Chen, D. Q. Miao, and H. Fujita, "Fuzzy neighborhood covering for three-way classification," *Information Sciences*, vol. 507, pp. 795–808, 2020.
[15] X. Han, X. Zhu, W. Pedrycz, and Z. Li, "A three-way classification with fuzzy decision trees," *Applied Soft Computing*, vol. 132, p. 109788, 2023.
[16] L. Subhashini, Y. Li, J. Zhang, and A. S. Atukorale, "Integration of fuzzy logic and a convolutional neural network in three-way decision-making," *Expert Systems with Applications*, vol. 202, p. 117103, 2022.
[17] A. V. Savchenko, "Fast inference in convolutional neural networks based on sequential three-way decisions," *Information Sciences*, vol. 560, pp. 370–385, 2021.
[18] J. Du, S. Liu, and Y. Liu, "A novel grey multi-criteria three-way decisions model and its application," *Computers & Industrial Engineering*, vol. 158, p. 107405, 2021.
[19] H. R. Ju, X. B. Yang, H. L. Yu, T. J. Li, D. J. Yu, and J. Y. Yang, "Cost-sensitive rough set approach," *Information Sciences*, vol. 355-356, pp. 282–298, 2016.
[20] H. Li, L. Zhang, X. Zhou, and B. Huang, "Cost-sensitive sequential three-way decision modeling using a deep neural network," *International Journal of Approximate Reasoning*, vol. 85, pp. 68–78, 2017.

[21] X. Y. Jia, W. W. Li, and L. Shang, "A multiphase cost-sensitive learning method based on the multiclass three-way decision-theoretic rough set model," *Information Sciences*, vol. 485, pp. 248–262, 2019.
[22] Y. B. Zhang, D. Q. Miao, J. Q. Wang, and Z. F. Zhang, "A cost-sensitive three-way combination technique for ensemble learning in sentiment classification," *International Journal of Approximate Reasoning*, vol. 105, pp. 85–97, 2019.
[23] W. Cai, M. Cai, Q. Li, and Q. Liu, "Three-way imbalanced learning based on fuzzy twin svm," *Applied Soft Computing*, vol. 150, p. 111066, 2024.
[24] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[25] C. F. Lin and S. D. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002.
[26] O. Okwuashi and C. E. Ndehedehe, "Deep support vector machine for hyperspectral image classification," *Pattern Recognition*, vol. 103, p. 107298, 2020.
[27] E. G. Dada, J. S. Bassi, H. Chiroma, A. O. Adetunmbi, O. E. Ajibuwa, *et al.*, "Machine learning for email spam filtering: review, approaches and open research problems," *Heliyon*, vol. 5, no. 6, p. e01802, 2019.
[28] H. Zhang, Y. Shi, X. Yang, and R. Zhou, "A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance," *Research in International Business and Finance*, vol. 58, p. 101482, 2021.
[29] W. Gao and Z. H. Zhou, "On the doubt about margin explanation of boosting," *Artificial Intelligence*, vol. 203, pp. 1–18, 2013.
[30] 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*, pp. 313–322, 2014.
[31] D. Gupta, N. Natarajan, and M. Berlin, "Short-term wind speed prediction using hybrid machine learning techniques," *Environmental Science and Pollution Research*, pp. 1–19, 2021.
[32] C. Cui, X. Zhang, and W. Cai, "An energy-saving oriented air balancing method for demand controlled ventilation systems with branch and black-box model," *Applied Energy*, vol. 264, p. 114734, 2020.
[33] F. Cheng, J. Zhang, C. Wen, Z. Liu, and Z. Li, "Large cost-sensitive margin distribution machine for imbalanced data classification," *Neurocomputing*, vol. 224, pp. 45–57, 2017.
[34] S. Abe, "Unconstrained large margin distribution machines," *Pattern Recognition Letters*, vol. 98, pp. 96–102, 2017.
[35] F. Cheng, J. Zhang, and C. H. Wen, "Cost-sensitive large margin distribution machine for classification of imbalanced data," *Pattern Recognition Letters*, vol. 80, pp. 107–112, 2016.
[36] 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.
[37] M. J. Kochenderfer, *Decision making under uncertainty: theory and application*. MIT press, 2015.
[38] H. X. Li, L. B. Zhang, B. Huang, and X. Z. Zhou, "Sequential three-way decision and granulation for cost-sensitive face recognition," *Knowledge-Based Systems*, vol. 91, pp. 241–251, 2016.
[39] L. Zhang, Q. Jin, S. Fan, and D. Liu, "A novel dual-center based intuitionistic fuzzy twin bounded large margin distribution machines," *IEEE Transactions on Fuzzy Systems*, 2023.
[40] F. Shen, Z. Yang, X. Zhao, and D. Lan, "Reject inference in credit scoring using a three-way decision and safe semi-supervised support vector machine," *Information Sciences*, vol. 606, pp. 614–627, 2022.
[41] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
[42] K. Atanassov, "Intuitionistic fuzzy sets," *International journal bioautomation*, vol. 20, p. 1, 2016.
[43] 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.
[44] Q. H. Zhang, C. C. Yang, and G. Y. Wang, "A sequential three-way decision model with intuitionistic fuzzy numbers," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 5, pp. 2640–2652, 2021.
[45] S. Rezvani, X. Wang, and F. Pourpanah, "Intuitionistic fuzzy twin support vector machines," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 11, pp. 2140–2151, 2019.
[46] Y. X. Wu, S. H. Cheng, Y. Li, R. J. Lv, and F. Min, "Stwd-sfnn: Sequential three-way decisions with a single hidden layer feedforward neural network," *Information Sciences*, vol. 632, pp. 299–323, 2023.
[47] 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.