Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Dynamic ensemble approach for multi-class classification based on neighborhood rough sets and sequential three-way decisions

Wenyan Xu
Xidian University
Qiang Chen
199422508630163.com

Xidian University
Yangyang Guo
Xidian University

Research Article

Keywords: Neighborhood rough sets, Sequential three-way decisions, One-versus-all decomposition, Multi-class classification, Ensemble learning

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10792506/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Dynamic ensemble approach for multi-class classification based on neighborhood rough sets and sequential three-way decisions

Wenyan Xu[1], Qiang Chen[1*], Yangyang Guo[1]

[1*]School of Mathematics and Statistics, Xidian University, Xi'an, 710071, Shaanxi, China.

*Corresponding author(s). E-mail(s): 19942250863@163.com;

Abstract

Xu et al. proposed a binary classifier that combines sequence three-way decision (S3WD) with classical rough set (CRS), and demonstrated its promising application prospects in binary classification. However, the reference tuples based on CRS become fixed once generated from the training set, which limits the classifier's adaptability to different data distributions; moreover, its binary classification architecture cannot directly handle multi-class classification tasks. To address these issues, this paper proposes EM-S3WD, an integrated multi-class classification framework based on neighborhood rough set (NRS) and S3WD. This framework uses NRS to construct adaptive reference tuples and combines the "one vs all" (OVA) decomposition strategy to construct multiple S3WD classifiers based on NRS. Additionally, a conditional dynamic integration strategy is introduced to achieve the extension of the binary classifier to multi-class scenarios. Finally, experimental results on public benchmark datasets show that the proposed method improves the adaptability of reference tuple construction and achieves competitive classification performance in multi-class classification tasks.

Keywords: Neighborhood rough sets, Sequential three-way decisions, One-versus-all decomposition, Multi-class classification, Ensemble learning

1 Introduction

The three-way decision (3WD) theory proposed by Yao [1, 2] provides an effective decision-making paradigm for handling uncertain and incomplete information. By dividing the decision space into three disjoint regions, namely the acceptance region, the rejection region, and the delayed decision region, 3WD introduces a delayed decision option into traditional two-way decision-making. This mechanism is especially useful when the available information is insufficient to support an immediate acceptance or rejection decision.

To further enhance the flexibility of uncertain decision-making, granular computing [3, 4] has been incorporated into the 3WD framework, leading to the sequential three-way decision (S3WD) model [5]. S3WD organizes decision-making as a sequential process, in which decisions can be progressively made under different information granularities or decision levels. Existing studies on S3WD have mainly developed along two directions. The first direction focuses on theoretical extensions based on decision-theoretic rough sets (DTRS) [2], including granularity selection [6–8] and threshold calculation [9, 10]. The second direction applies S3WD to practical tasks, such as stock prediction [11], medical diagnosis [12], cluster analysis [13], and classification problems [14, 15].

In recent years, S3WD has been further developed for classification and multi-class decision-making. Some studies have investigated sequential multi-class three-way decisions from the perspective of cost-sensitive learning [16], while others have attempted to improve S3WD classifiers through more flexible threshold learning, information-gain-based decision mechanisms, or automatic threshold optimization [17, 18]. These studies show that S3WD remains an active research topic in uncertain classification. However, most of them focus on threshold determination, risk modeling, or general decision processes, while the construction of reference representations in a specific S3WD classifier has received relatively less attention.

Among existing S3WD-based classification studies, the binary classifier proposed by Xu et al. [19] is the direct basis of this paper. This classifier adopts a two-stage structure. In the first stage, classical rough set

theory (CRS) [20] is used to select reference tuples. In the second stage, a sequential three-way decision process with adaptive threshold calculation is employed to classify testing samples. This reference-tuple-based design provides an interpretable binary classification mechanism and has achieved promising performance in binary classification tasks.

However, when Xu et al.'s binary S3WD classifier is considered from the perspective of model construction and multi-class extension, two specific limitations can be observed. First, the reference tuples in this model are generated based on CRS. Since CRS relies on strict equivalence relations, the reference tuples are completely determined once the training set is fixed. Therefore, in the subsequent model selection process, the original model mainly adjusts the similarity threshold σ, while the reference tuple set itself remains unchanged. In other words, the reference representation is static rather than adaptively selected according to data distribution or validation performance. This may restrict the flexibility of the classifier, especially for continuous-valued data where strict equivalence relations are often too rigid to capture local data structures.

Neighborhood rough sets (NRS) provide a natural way to alleviate the rigidity of CRS for continuous-valued data. Different from CRS, NRS introduces neighborhood relations to characterize data granules, making the lower and upper approximations dependent on neighborhood granularity [21]. Recent NRS-based studies have further demonstrated the usefulness of neighborhood granules in feature selection, attribute reduction, uncertainty measurement, and continuous data classification [22–24]. Nevertheless, these studies mainly aim to construct general NRS-based learning models, feature selection methods, or attribute reduction methods. They do not specifically investigate how neighborhood granularity can be used to reconstruct the reference tuples of an existing S3WD classifier. Therefore, in this paper, NRS is introduced not as a general replacement of rough set theory, but as a mechanism to make the reference tuple construction in Xu et al.'s S3WD classifier adaptive.

Second, Xu et al.'s model was originally designed for binary classification [19]. A natural way to extend this binary classifier to multi-class classification is to use the one-versus-all (OVA) decomposition strategy, which has been widely adopted to transform multi-class problems into multiple binary classification problems [25–27]. However, directly applying OVA to the S3WD-based binary classifier may introduce a confidence conflict problem. Specifically, multiple binary S3WD classifiers may assign the same or very close maximum confidence value to a testing sample. In such cases, the final decision becomes ambiguous. Random tie-breaking may reduce decision stability, while prior-based or majority-based strategies may ignore the sample-specific evidence contained in the current confidence outputs.

Ensemble learning has been widely used to improve classification robustness by aggregating multiple models [28–30]. Recent studies on dynamic classifier selection and dynamic ensemble selection show that the competence of classifiers can be evaluated according to each query sample, and that different classifiers or classifier subsets may provide complementary information for difficult samples [31–33]. However, the goal of this paper is not to construct a general ensemble classifier. The main decision mechanism should still be the S3WD classifier extended from Xu et al.'s model. Therefore, auxiliary classifiers are used only when the OVA-based S3WD outputs produce confidence conflicts. When no conflict occurs, the original S3WD decision is preserved. When a conflict occurs, auxiliary classifiers provide additional evidence for the conflicting classes, and Kullback–Leibler divergence is used to measure the discrepancy between the main S3WD outputs and auxiliary outputs.

Based on the above analysis, existing studies provide useful foundations for this work, including S3WD-based classification, NRS-based modeling for continuous data, OVA-based multi-class decomposition, and ensemble learning. However, these components have usually been studied independently. They do not directly solve the two model-specific problems considered in this paper: the static reference tuple construction caused by CRS in Xu et al.'s binary S3WD classifier, and the confidence conflict problem caused by directly extending this classifier to multi-class classification through OVA.

Motivated by these observations, this paper proposes a dynamic ensemble multi-class classification framework based on neighborhood rough sets and sequential three-way decisions, termed EM-S3WD. The proposed method is a targeted model-level extension of Xu et al.'s CRS-based binary S3WD classifier. By introducing NRS, the reference tuples become dependent on neighborhood granularity and can be selected through validation-guided parameter learning. By introducing OVA, the original binary S3WD classifier is extended to multi-class classification. By introducing a conditional dynamic integration strategy, confidence conflicts among OVA-based S3WD classifiers can be resolved without replacing the main S3WD decision mechanism.

The main contributions of this paper are summarized as follows:

1. An NRS-based adaptive reference tuple construction mechanism is developed for the S3WD classifier proposed by Xu et al. [19]. In contrast to the original CRS-based construction, the resulting reference tuples depend on the neighborhood granularity. A validation-guided procedure is further introduced to jointly select the neighborhood radius and the similarity threshold.

2. The formation mechanism of confidence conflicts in the OVA-based multi-class extension is analyzed. The confidence scores generated by the binary S3WD classifiers are shown to belong to a finite set determined by the matched-attribute count and the decision level. This finite-valued property provides a structural explanation for why different class-specific classifiers may produce the same maximal confidence score.
3. A conditional dynamic integration strategy is proposed to resolve confidence conflicts. Auxiliary classifiers are activated only for conflict samples, and their outputs are dynamically weighted according to their discrepancies from the main S3WD outputs and their validation performance. For non-conflict samples, the original OVA-based S3WD decision is directly preserved.
4. The proposed framework is evaluated through theoretical analysis and experiments. The theoretical results establish finite-step termination, decision preservation in non-conflict cases, and conditional prediction complexity. Experiments on nine benchmark datasets examine parameter sensitivity, confidence-conflict resolution, adaptive reference tuple construction, and overall classification performance.

The remainder of this paper is organized as follows. In Section 2, we briefly review the basic concepts used in this paper. In Section 3, we present the proposed EM-S3WD framework. In Section 4, experimental results and analysis are reported. Finally, Section 5 concludes the paper.

2 Preliminaries

2.1 Classical and neighborhood rough sets

We first review the two rough set formalisms used in this paper: Pawlak's classical rough sets (CRS)[20] and neighborhood rough sets (NRS) [21].

Definition 2.1. ([20]) An information table is a tuple

$$I = (U, A, V_A, r),$$

where U is a finite nonempty set of objects, $A = C \cup D$ is a finite nonempty set of attributes, C and D are the condition and decision attribute sets, respectively, $V_A = \{V_a \mid a \in A\}$ contains the attribute domains, and r is an information function. For an attribute subset $Y \subseteq A$, the indiscernibility relation R_Y is defined as

$$xR_Y y \Longleftrightarrow a(x) = a(y), \quad \forall a \in Y. \tag{1}$$

The equivalence class containing x is denoted by

$$[x]_Y = \{y \in U \mid xR_Y y\}.$$

For a concept $X \subseteq U$, its lower and upper approximations with respect to Y are respectively defined as

$$\begin{cases} \underline{\text{apr}}_Y(X) = \{x \in U \mid [x]_Y \subseteq X\}, \\ \overline{\text{apr}}_Y(X) = \{x \in U \mid [x]_Y \cap X \neq \varnothing\}. \end{cases} \tag{2}$$

The boundary region of X is given by

$$\text{BND}_Y(X) = \overline{\text{apr}}_Y(X) - \underline{\text{apr}}_Y(X). \tag{3}$$

Definition 2.2. ([21]) Let $I = (U, A, V_A, r)$ be an information table, $B \subseteq A$, and $\delta \geq 0$. For an object $x_i \in U$, its δ-neighborhood in the attribute subspace B is defined as

$$\delta_B(x_i) = \{x_j \in U \mid \Delta_B(x_i, x_j) \leq \delta\}, \tag{4}$$

where Δ_B is a distance function and δ is the neighborhood radius. For numerical data, the Euclidean distance is defined as

$$\Delta_B(x_i, x_j) = \sqrt{\sum_{a \in B} (a(x_i) - a(x_j))^2}.$$

Definition 2.3. ([21]) Let $X \subseteq U$. The neighborhood lower and upper approximations of X are respectively defined as

$$\begin{cases} \underline{\text{apr}}_{\delta}(X) = \{x \in U \mid \delta_B(x) \subseteq X\}, \\ \overline{\text{apr}}_{\delta}(X) = \{x \in U \mid \delta_B(x) \cap X \neq \varnothing\}. \end{cases} \tag{5}$$

Proposition 1. ([21]) Let δ_1 and δ_2 be two nonnegative neighborhood radii. If $\delta_1 \geq \delta_2$, then

$$\delta_{1,B}(x) \supseteq \delta_{2,B}(x), \quad \forall x \in U,$$

and

$$\underline{\text{apr}}_{\delta_1}(X) \subseteq \underline{\text{apr}}_{\delta_2}(X), \quad \forall X \subseteq U.$$

Thus, the neighborhood becomes coarser and the lower approximation becomes smaller as the neighborhood radius increases.

2.2 Sequential three-way decision proposed in [19]

Thirdly, we review the main concepts of sequential three-way decision (S3WD) proposed by Xu et al. in [19]. One will see that we extend this S3WD model to NRS-version and, combining the "one-versus-all" and conditional dynamic integration strategy, propose a novel ensemble approach for multi-classification tasks in this paper.

Definition 2.4. ([19]) Let $I = \{U, A, V_A, r\}$ be an information table, $Y = \{a_1, \ldots, a_n\}$ be a subset of A, and $X \subseteq U$ be a concept. A reference tuple $\vec{\tau}$ on Y with respect to X is an n-tuple $\vec{\tau} = (\tau^1, \tau^2, \ldots, \tau^n)$ satisfying that there exists an equivalence class $[x]_Y \subseteq POS_Y(X)$ such that $\tau^i = a_i(x)$ for any $i \in \{1, 2, \ldots, n\}$. A reference tuple $\vec{\tau}$ on Y with respect to $\neg X$ is defined analogously with $[y]_Y \subseteq NEG_Y(X)$. A reference set on Y with respect to X (resp. $\neg X$) is a collection $\Gamma = \{\vec{\tau}_1, \vec{\tau}_2, \ldots, \vec{\tau}_m\}$ of reference tuples on Y with respect to X (resp. $\neg X$).

Definition 2.5. ([19]) Let $I = \{U, A, V_A, r\}$ be an information table, $Y = \{a_1, \ldots, a_n\}$ be a subset of A, $X \subseteq U$ be a concept, $\Gamma = \{\vec{\tau}_1, \ldots, \vec{\tau}_m\}$ be a reference set on Y with respect to X (resp. $\neg X$), and $\sigma \geq 0$. For any object $u \in U$, the correlation degree between u and $\vec{\tau}_j$ on Y is defined as

$$E_Y(u, \vec{\tau}_j) = \frac{\sum_{i=1}^{n} \rho(a_i(u), \tau_j^i)}{|Y|}, \tag{6}$$

where

$$\rho(a_i(u), \tau_j^i) = \begin{cases} 1, & |a_i(u) - \tau_j^i| \leq \sigma \\ 0, & |a_i(u) - \tau_j^i| > \sigma \end{cases} \tag{7}$$

Definition 2.6. ([19]) Let $I = \{U, A, V_A, r\}$ be an information table, $Y = \{a_1, \ldots, a_n\}$ be a subset of A, $X \subseteq U$ be a concept, $\Gamma = \{\vec{\tau}_1, \ldots, \vec{\tau}_m\}$ be a reference set on Y with respect to X (resp. $\neg X$), and k be a positive integer. Let $B_0 = U$, and $\langle P_1, \ldots, P_k \rangle$, $\langle N_1, \ldots, N_k \rangle$, $\langle B_1, \ldots, B_k \rangle$ be three sequences of subsets of U defined by

$$\begin{cases} P_i = \{u \in U_i \mid E_Y(u, \vec{\tau}_j) \geq \alpha_i\}, \\ N_i = \{u \in U_i \mid E_Y(u, \vec{\tau}_j) \leq \beta_i\}, \\ B_i = \{u \in U_i \mid \beta_i < E_Y(u, \vec{\tau}_j) < \alpha_i\}, \end{cases} \tag{8}$$

where, for any $i \in \{1, 2, \ldots, k\}$, $\vec{\tau}_j \in \Gamma$, $U_i = B_{i-1}$, and (α_i, β_i) is a pair of thresholds satisfying $0 \leq \beta_i < \alpha_i \leq 1$ such that $B_1 \supset B_2 \supset \ldots \supset B_k = \emptyset$. Then, for the given Y and Γ, the positive, negative and boundary regions of U on Y with respect to Γ are defined by

$$\begin{cases} POS_Y^{\Gamma}(U) = \bigcup_{i=1}^{k} P_i, \\ NEG_Y^{\Gamma}(U) = \bigcup_{i=0}^{k} N_i, \\ BND_Y^{\Gamma}(U) = \emptyset. \end{cases} \tag{9}$$

The tuple $\Phi_Y^{\Gamma}(I) = (I, Y, \Gamma, \langle P_1, \ldots, P_k \rangle, \langle N_1, \ldots, N_k \rangle, \langle B_1, \ldots, B_k \rangle)$ is called a sequential three-way decision model of I on Y with respect to Γ.

Theorem 2.1. ([19]) Let $\Phi_Y^{\Gamma}(I)$ be a S3WD model with k levels. For any given $i \in \{1, \ldots, k\}$ and reference tuple $\vec{\tau}_j \in \Gamma$, the following three functions

$$\begin{aligned} f_{1i}(u, \vec{\tau}_j) &= \frac{E_Y(u, \vec{\tau}_j) \cdot (nE_Y(u, \vec{\tau}_j) + 1)}{n+1}, \\ f_{2i}(u, \vec{\tau}_j) &= \frac{(1 - E_Y(u, \vec{\tau}_j))[n(1 - E_Y(u, \vec{\tau}_j)) + 1]}{n+1}, \\ f_{3i}(u, \vec{\tau}_j) &= \frac{2nE_Y(u, \vec{\tau}_j) \cdot (1 - E_Y(u, \vec{\tau}_j))}{n+1} \end{aligned} \tag{10}$$

compute the probabilities of u's ($u \in U_i$) being in P_i, N_i and B_i, respectively.

Xu et al. proposed the corresponding algorithm in [19] for binary classification based on the above S3WD theory and demonstrated its effectiveness, which we will not elaborate on further (see [19] for all the details).

2.3 Kullback–Leibler divergence

The Kullback–Leibler (KL) divergence measures the difference between two probability distributions [34]. In this paper, KL divergence is used to measure the discrepancy between the main S3WD output and each auxiliary classifier's output, which determines the auxiliary weights used in conflict resolution.

Definition 2.7. ([34]) Let $P = (p_1, p_2, \ldots, p_m)$ and $Q = (q_1, q_2, \ldots, q_m)$ be two discrete probability distributions defined on the same space. The KL divergence of P from Q is defined as

$$D_{\mathrm{KL}}(P \| Q) = \sum_{i=1}^{m} p_i \log \frac{p_i}{q_i}. \tag{11}$$

The KL divergence is nonnegative and satisfies

$$D_{\mathrm{KL}}(P \| Q) = 0$$

if and only if $P = Q$.

2.4 The "one-versus-all" decomposition strategy

The "one-versus-all" (OVA) decomposition strategy [35–37] is widely used in multiclass classification in machine learning. Its core idea is dividing an m-class problem into m binary classification problems. Each binary classification problem is handled by a binary classifier, which is responsible for distinguishing one class from all the others. So, in the rest of this paper, we agree: when we say "class i", we refer to "classifier i" for an m-class problem. For a testing sample, each binary classifier outputs a confidence score $r_i \in [0, 1]$, forming a score vector

$$\boldsymbol{R} = (r_1, r_2, \ldots, r_m). \tag{12}$$

The predicted class is determined by

$$\text{Class} = \arg \max_{i=1,\ldots,m} r_i.$$

If two or more classes attain the same maximal confidence score, the maximum-confidence rule cannot produce a unique class label. In this paper, this situation is referred to as a confidence conflict.

Remark 2.1. Decision ambiguity is not exclusive to OVA, although its form depends on the multi-class strategy. Under OVA, independently constructed binary classifiers may assign the same maximal confidence score to multiple classes. Under one-versus-one decomposition, ambiguity commonly appears as a tie in the aggregated pairwise votes. Direct multi-class classifiers avoid decomposition-induced voting conflicts, but may still produce identical decision scores for different classes. Therefore, replacing OVA does not necessarily eliminate decision ambiguity; rather, it changes its form and the corresponding resolution mechanism.

OVA is adopted in this study because the underlying S3WD classifier in [19] follows a binary target-versus-complement structure. OVA preserves this structure and establishes an explicit correspondence between each binary S3WD classifier and one target class. Accordingly, this paper focuses on the non-unique maximum arising in the OVA-based S3WD confidence vector.

3 The proposed EM-S3WD framework

3.1 NRS-based adaptive reference tuple construction

In the CRS-based S3WD classifier proposed by Xu et al. [19], reference tuples are generated from CRS-induced lower approximations. Once the training set is fixed, the reference tuple set is also fixed. To make the reference representation more flexible for continuous-valued data, this subsection introduces NRS-based reference tuples. The key idea is that different neighborhood radii may induce different lower approximations, and thus different reference tuple sets.

Definition 3.1. Let $I = \{U, A, V_A, r\}$ be an information table, $Y = \{a_1, \ldots, a_n\}$ be a subset of A, $X \subseteq U$ be a concept, δ be any given neighborhood radius and $\underline{apr}_\delta(X)$ be the δ-neighborhood based lower approximation of X defined in NRS. A **δ-reference tuple on Y with respect to X**, denoted by $\vec{\tau}^\delta$, is an n-tuple $\vec{\tau}^\delta = (\tau^1, \tau^2, \ldots, \tau^n)$ satisfying that there exists an object $x \in \underline{apr}_\delta(X)$ such that $\tau^i = a_i(x)$ for any $i \in \{1, 2, \ldots, n\}$.

Definition 3.2. Let $I = \{U, A, V_A, r\}$ be an information table, Y be a subset of A, $X \subseteq U$ be a concept, and δ be any given neighborhood radius. Suppose that $\vec{\tau}_1^\delta, \vec{\tau}_2^\delta, \ldots, \vec{\tau}_m^\delta$ are all δ-reference tuples on Y with respect to X induced by the objects in $\underline{apr}_\delta(X)$. Then the set

$$\Gamma^\delta = \{\vec{\tau}_1^\delta, \vec{\tau}_2^\delta, \ldots, \vec{\tau}_m^\delta\}$$

is called the **δ-reference set on Y with respect to X**.

Definitions 3.1 and 3.2 establish a direct connection between the neighborhood radius δ and the reference tuple set. Different from the CRS-based construction, where reference tuples are determined only by equivalence classes, the NRS-based construction makes the reference tuple set dependent on the neighborhood granularity. Therefore, changing δ may change the lower approximation and the induced reference tuples.

For a fixed neighborhood radius δ, the NRS-based reference tuple set can be generated by Algorithm 1. It should be noted that Algorithm 1 only describes how to generate reference tuples under a given δ. The final reference tuple set used by the proposed classifier will be determined through validation-guided parameter selection, which will be introduced later in this subsection.

Algorithm 1: NRS-based reference tuple generation under a fixed radius

Input: Training set U_{train}, attribute subset $Y = \{a_1, \ldots, a_n\}$, concept $X \subseteq U_{\text{train}}$, neighborhood radius δ
Output: δ-reference set $\Gamma_Y^\delta(X)$

Initialize $R = \emptyset$ and $\Gamma_Y^\delta(X) = \emptyset$;
foreach $x \in U_{\text{train}}$ **do**
  Compute the neighborhood $\delta_Y(x)$;
  if $\delta_Y(x) \subseteq X$ **then**
    $R = R \cup \{x\}$;
  end
end
foreach $x \in R$ **do**
  Construct $\vec{\tau}^\delta(x) = \big(a_1(x), a_2(x), \ldots, a_n(x)\big)$;
  $\Gamma_Y^\delta(X) = \Gamma_Y^\delta(X) \cup \{\vec{\tau}^\delta(x)\}$;
end
return $\Gamma_Y^\delta(X)$

Remark 3.1. The term "adaptive" in the proposed reference tuple construction has two meanings. First, under a fixed neighborhood radius δ, the reference tuple set is induced by the NRS lower approximation rather than by CRS equivalence classes. Second, the final radius δ is not manually fixed in advance, but selected according to validation performance together with the similarity threshold. Therefore, the reference representation is adaptive to the data distribution through validation-guided parameter selection. This does not mean that reference tuples are updated online during testing. Although Algorithm 1 can generate reference tuples under a fixed δ, a fixed radius still leads to a fixed reference tuple set. The key issue, therefore, is not only how to generate δ-reference tuples, but also which δ should be used for a given dataset.

Moreover, the similarity threshold σ plays an equally important role, as it directly controls the correlation degree between a testing sample and the selected reference tuples.

To address these issues, δ and σ are jointly selected on the validation set. For each candidate pair (δ, σ), the corresponding reference tuple sets are generated from the training set, and the induced classifiers are evaluated on the validation set. The optimal pair is selected by maximizing the validation accuracy:

$$(\delta^*, \sigma^*) = \arg \max_{\delta \in \Delta,\ \sigma \in \Sigma} Acc_{\text{val}}(\delta, \sigma).$$

The final reference tuple sets are those induced by δ^*, and they remain fixed during testing. Algorithm 2 summarizes this procedure. It should be noted that the grid search itself is not a methodological contribution; its role is to operationalize the NRS-induced adaptive reference representation by selecting the neighborhood radius that determines the final reference tuple sets.

After the optimal parameter pair (δ^*, σ^*) is obtained, the reference tuple sets induced by δ^* are fixed and used in the subsequent OVA-based S3WD classifiers. Therefore, the proposed method differs from the CRS-based construction in that the reference tuples are not determined solely by the training set, but are selected through the validation performance of the classifier.

Algorithm 2: Validation-guided parameter selection and reference tuple determination

Input: Training set U_{train}, validation set U_{val}, concept $X \subseteq U_{\text{train}}$, attribute subset Y, candidate radius set Δ, candidate similarity threshold set Σ

Output: Optimal parameter pair (δ^*, σ^*) and optimal reference tuple set Γ^{δ^*}

```
Initialize BestAcc = −∞;
foreach δ ∈ Δ do
    Generate the δ-reference tuple set Γ^δ with respect to X by Algorithm 1;
    foreach σ ∈ Σ do
        Construct the NRS-based S3WD classifier Φ_Y^{Γ^δ}(I) using Γ^δ and σ;
        Classify the samples in U_val using Φ_Y^{Γ^δ}(I);
        Compute the validation accuracy Acc_val(δ, σ);
        if Acc_val(δ, σ) > BestAcc then
            BestAcc = Acc_val(δ, σ);
            (δ*, σ*) = (δ, σ);
            Γ^{δ*} = Γ^δ;
        end
    end
end
return (δ*, σ*) and Γ^{δ*}
```

3.2 OVA-based multi-class extension

After the validation-guided selection in Section 3.1, the optimal parameter pair (δ^*, σ^*) and the corresponding reference tuple sets are obtained. Based on them, the original binary S3WD classifier is extended to multi-class classification through the OVA strategy. For an m-class problem, m binary NRS-based S3WD classifiers are constructed, where the p-th classifier distinguishes class p from all the remaining classes.

Let $X_p \subseteq U_{\text{train}}$ denote the set of training samples belonging to class p, and let $X_p^c = U_{\text{train}} \setminus X_p$ denote the set of samples belonging to all other classes. In the p-th OVA classifier, X_p is treated as the positive concept. According to Section 3.1, the corresponding NRS-induced reference tuple set is determined by the selected neighborhood radius δ^*.

In Section 2.2, the binary S3WD classifier proposed by Xu et al. [19] is denoted by $\Phi_Y^{\Gamma}(I)$, where the reference set Γ is generated based on CRS. Since the CRS-induced reference set is independent of the neighborhood radius, the reference representation is fixed once the training set is given. In this paper, by replacing Γ with the NRS-induced reference set Γ^{δ}, the CRS-based classifier $\Phi_Y^{\Gamma}(I)$ is extended to an NRS-based binary S3WD classifier, denoted by $\Phi_Y^{\Gamma^{\delta}}(I)$. For $\Phi_Y^{\Gamma^{\delta}}(I)$, the correlation degree between an object and a δ-reference tuple is redefined as follows.

Definition 3.3. Let $I = \{U, A, V_A, r\}$ be an information table, $Y = \{a_1, \ldots, a_n\}$ be a subset of A, $X \subseteq U$ be a concept, and $\vec{\tau}^{\delta} = (\tau^1, \tau^2, \ldots, \tau^n)$ be a δ-reference tuple on Y with respect to X. For any object $u \in U$,

the **correlation degree between** u **and** $\vec{\tau}^{\delta}$ **on** Y, denoted by $E_Y(u, \vec{\tau}^{\delta})$, is defined as follows:

$$E_Y(u, \vec{\tau}^{\delta}) = \frac{\sum_{i=1}^{n} \rho(a_i(u), \tau^i)}{|Y|}, \tag{13}$$

where

$$\rho(a_i(u), \tau^i) = \begin{cases} 1, & \sqrt{|a_i(u) - \tau^i|} \leq \sigma, \\ 0, & \sqrt{|a_i(u) - \tau^i|} > \sigma. \end{cases} \tag{14}$$

For multi-class classification, the m NRS-based binary S3WD classifiers independently generate confidence scores for each testing sample. The confidence score is designed to reflect both the decision region and the decision level of the S3WD process. A sample assigned to the positive region at an earlier level should obtain a higher confidence score, while a sample assigned to the negative region should obtain a lower confidence score. The formal definition is given as follows.

Definition 3.4. Let $I = \{U, A, V_A, r\}$ be an information table with $U = \{u_1, \ldots, u_n\}$, $Y \subseteq A$ be an attribute subset, and $\Phi_1, \ldots, \Phi_m$ be m NRS-based binary S3WD classifiers constructed under the OVA strategy. Let P_p and N_p denote the positive region and the negative region of Φ_p, respectively, where $p \in \{1, \ldots, m\}$. For $u_j \in U$, the **confidence level of** u_j **in the output of** Φ_p, denoted by r_p^j, is defined by

$$r_p^j = \begin{cases} \frac{k-(i-1)}{k} f_{1i}(u_j, \vec{\tau}_q^{\delta}), & u_j \in P_p, \\ \frac{i}{k} f_{1i}(u_j, \vec{\tau}_q^{\delta}), & u_j \in N_p, \end{cases} \tag{15}$$

where $\vec{\tau}_q^{\delta} \in \Gamma^{\delta}$, i is the output level of u_j in Φ_p, and k is the total number of decision levels.

Proposition 2. Let $d = |Y|$, and suppose that an NRS-based binary S3WD classifier contains k decision levels. For any testing sample u_j, the confidence level r_p^j defined in Eq. (15) belongs to a finite set containing at most $kd + 1$ elements.

Proof According to Eqs. (17) and (18), the attribute-wise matching function satisfies

$$\rho(a_t(u_j), \tau_t) \in \{0, 1\}.$$

Let

$$h = \sum_{t=1}^{d} \rho(a_t(u_j), \tau_t),$$

where $h \in \{0, 1, \ldots, d\}$. It follows that

$$E_Y(u_j, \vec{\tau}_q^{\delta}) = \frac{h}{d}.$$

Substituting this result into the positive probability function gives

$$f_{1i}(u_j, \vec{\tau}_q^{\delta}) = \frac{h(h+1)}{d(d+1)}.$$

According to Definition 3.4, the confidence level can be expressed as

$$r_p^j = \frac{s}{k} \frac{h(h+1)}{d(d+1)},$$

where

$$s = \begin{cases} i, & u_j \in N_p, \\ k - i + 1, & u_j \in P_p. \end{cases}$$

Since

$$s \in \{1, \ldots, k\}, \qquad h \in \{0, \ldots, d\},$$

the confidence level belongs to a finite set. When $h = 0$, the confidence level is zero for every s. For each $h \in \{1, \ldots, d\}$, there are at most k possible values. Therefore, the number of distinct confidence values is at most

$$kd + 1.$$

□

This finite-valued property increases the possibility that different OVA-based S3WD classifiers produce identical confidence scores. A confidence conflict occurs when such an identical score is also the maximal component of the confidence vector.

Definition 3.5. Let $\Phi_1, \ldots, \Phi_m$ be m NRS-based binary S3WD classifiers, and let $r_1^j, \ldots, r_m^j$ be the confidence levels of $u_j \in U$ generated by $\Phi_1, \ldots, \Phi_m$, respectively. Then the **confidence vector** of u_j is defined as

$$R_j = (r_1^j, r_2^j, \ldots, r_m^j). \tag{16}$$

Suppose that $|U| = n$. The **confidence matrix** composed of n samples is defined by

$$M_c = (R_1, R_2, \ldots, R_n)^T. \tag{17}$$

If the maximum component of R_j is unique, the preliminary prediction label of u_j is determined by the maximum confidence rule:

$$Label_{u_j} = \arg \max_{i=1,\ldots,m} r_i^j. \tag{18}$$

Remark 3.2. Proposition 2 provides a structural explanation for confidence conflicts in the OVA-based S3WD framework. Each confidence score is determined by the discrete matched-attribute count h and the decision-level coefficient s/k, and therefore belongs to a finite set. Consequently, different binary classifiers may produce identical confidence scores.

If such identical scores are also the maximal components of R_j, the maximum-confidence rule cannot determine a unique class label. In particular, if $h = 0$ for every classifier, all confidence scores are zero. Thus, confidence conflicts can arise naturally from the finite-valued confidence definition combined with the OVA decomposition.

3.3 Conditional dynamic integration for confidence conflicts

As discussed in Remark 3.2, the OVA-based S3WD classifiers may produce confidence conflicts when two or more classes obtain the same maximal confidence score for a testing sample. In such cases, the maximum confidence rule cannot determine a unique class label. To address this problem, this subsection introduces a conditional dynamic integration strategy.

It should be emphasized that the proposed strategy is not a full-time ensemble classifier. The NRS-based S3WD classifiers constructed under the OVA strategy are regarded as the main models, while the auxiliary classifiers are used only when confidence conflicts occur. If the confidence vector generated by the main models has a unique maximum component, the original OVA-S3WD decision is directly preserved. If a confidence conflict occurs, auxiliary classifiers are invoked to provide additional evidence for the conflicting classes.

In this paper, SVM [38] and Naive Bayes (NB) [38] are used as auxiliary classifiers. Let S_{main} be the score vector generated by the main OVA-based S3WD classifiers for a testing sample u_j, and let S_k be the score vector generated by the k-th auxiliary classifier, where $k = 1, \ldots, K$. The normalized score vectors of S_{main} and S_k are denoted by P_{main} and P_k, respectively.

For any m-dimensional score vector R, the conflict set is defined as

$$\text{Conflict}(R) = \left\{ c_i \mid R(c_i) = \max_{1 \le l \le m} R(c_l) \right\}.$$

If $|\text{Conflict}(R)| \ge 2$, then the corresponding sample is called a conflict sample, and each class $c \in \text{Conflict}(R)$ is called a conflicting class. For any class c_j, $R(c_j)$ denotes the confidence score of class c_j.

Definition 3.6. Given the main score vector S_{main} for u_j, if $|\text{Conflict}(S_{\text{main}})| \ge 2$, the discrepancy between the main models and the k-th auxiliary model on the conflicting classes is defined by

$$D_{\text{KL}}(P_{\text{main}} \| P_k) = \sum_{c \in \text{Conflict}(S_{\text{main}})} P_{\text{main}}(c) \log \left(\frac{P_{\text{main}}(c)}{P_k(c)} \right). \tag{19}$$

In Eq. (19), the KL divergence is calculated only on the conflicting classes, because the proposed integration strategy aims to resolve the ambiguity among the classes sharing the maximal confidence score. In implementation, a small positive constant can be added to each normalized score to avoid zero probabilities in the logarithm.

Next, two types of weights are introduced. The first one controls the contribution of the main S3WD models, and the second one controls the relative contribution of each auxiliary classifier.

Definition 3.7. Let $\beta > 0$ be a moderating factor and K be the number of auxiliary classifiers. The **weight coefficient of the main models** is defined by

$$\alpha = \exp\left(-\beta \cdot \min_{1\le k\le K} D_{\mathrm{KL}}(P_{\mathrm{main}} \| P_k)\right). \tag{20}$$

According to Eq. (20), a smaller discrepancy between the main models and at least one auxiliary classifier leads to a larger α, which means that more information from the main S3WD models is preserved. Conversely, when the auxiliary outputs differ more from the main outputs on the conflicting classes, the auxiliary classifiers are allowed to play a stronger role in breaking the conflict.

Definition 3.8. Suppose that K is the number of auxiliary classifiers, $\mathrm{Accuracy}_k$ is the validation accuracy of the k-th auxiliary classifier, and $\beta > 0$ is the moderating factor. The **weight coefficient of the k-th auxiliary classifier** is defined by

$$w_k = \frac{\exp\left(-\beta \cdot D_{\mathrm{KL}}(P_{\mathrm{main}} \| P_k)\right) \cdot \mathrm{Accuracy}_k}{\sum_{j=1}^{K} \exp\left(-\beta \cdot D_{\mathrm{KL}}(P_{\mathrm{main}} \| P_j)\right) \cdot \mathrm{Accuracy}_j}. \tag{21}$$

Thus, an auxiliary classifier receives a larger weight when it has a higher validation accuracy and its confidence distribution is more consistent with the main S3WD outputs on the conflicting classes.

Definition 3.9. For any class c_j, let $S_{\mathrm{main}}(c_j)$ be the original confidence score of c_j generated by the main models, and let $S_k(c_j)$ be the confidence score of c_j generated by the k-th auxiliary classifier. If $c_j \in \mathrm{Conflict}(S_{\mathrm{main}})$, the **temporary confidence score** of c_j is defined by

$$r^{\mathrm{temp}}(c_j) = \alpha \cdot S_{\mathrm{main}}(c_j) + (1-\alpha) \sum_{k=1}^{K} w_k S_k(c_j). \tag{22}$$

For $c_j \notin \mathrm{Conflict}(S_{\mathrm{main}})$, its score remains unchanged, i.e.,

$$r^{\mathrm{temp}}(c_j) = S_{\mathrm{main}}(c_j).$$

Eq. (22) updates only the scores of the conflicting classes. However, when the conflict set is a proper subset of all classes, it is possible that the maximal temporary score of the conflicting classes becomes no larger than the score of a non-conflicting class. In that case, the adjustment would change the decision scope from resolving the conflict among the originally maximal classes to selecting a class that was not involved in the original conflict. To avoid this situation, when $\mathrm{Conflict}(S_{\mathrm{main}}) \neq \{c_1, \ldots, c_m\}$, the following preservation constraint is imposed if necessary:

$$\text{if } \max_{c_j \in \mathrm{Conflict}(S_{\mathrm{main}})} r^{\mathrm{temp}}(c_j) \le \max_{c_i \notin \mathrm{Conflict}(S_{\mathrm{main}})} S_{\mathrm{main}}(c_i),$$

then for each $c_j \in \mathrm{Conflict}(S_{\mathrm{main}})$, the final confidence score is adjusted as

$$r^{\mathrm{final}}(c_j) = r^{\mathrm{temp}}(c_j) \cdot \frac{S_{\mathrm{main}}(c_j)}{\max_{c_l \in \mathrm{Conflict}(S_{\mathrm{main}})} r^{\mathrm{temp}}(c_l)}. \tag{23}$$

Otherwise, we set

$$r^{\mathrm{final}}(c_j) = r^{\mathrm{temp}}(c_j), \quad c_j \in \mathrm{Conflict}(S_{\mathrm{main}}).$$

When all classes belong to $\mathrm{Conflict}(S_{\mathrm{main}})$, there is no non-conflicting class to preserve. In this case, all classes are updated by Eq. (22), and we directly set

$$R^{\mathrm{final}} = R^{\mathrm{temp}}.$$

The final label is then determined by the maximum value of R^{final}.

3.4 Overall algorithm of EM-S3WD

Based on the above definitions, the proposed EM-S3WD framework integrates NRS-based adaptive reference tuple construction, OVA-based multi-class extension, and conditional dynamic integration. The validation-guided selection of the optimal parameter pair (δ^*, σ^*) and the corresponding reference tuple sets has been

Algorithm 3: Multi-class prediction of EM-S3WD

Input: Test set $U_{\text{test}} = \{u_1, u_2, \ldots, u_n\}$, number of classes m, optimal parameters (δ^*, σ^*), optimal reference sets $\{\Gamma_1^{\delta^*}, \ldots, \Gamma_m^{\delta^*}\}$, auxiliary classifiers $\{Aux_1, \ldots, Aux_K\}$

Output: Predicted labels $\{\hat{y}_1, \ldots, \hat{y}_n\}$

for $i = 1$ **to** m **do**
  Construct S3WD$_i$ = S3WD-Algorithm$(\Gamma_i^{\delta^*}, \sigma^*)$;
  Compute $R_i = \text{S3WD}_i(U_{\text{test}})$;
end
Construct $M_c = [R_1, R_2, \ldots, R_m]$;
for $p = 1$ **to** n **do**
  Extract $S_{\text{main}} = M_c(p, :)$;
  if $|\text{Conflict}(S_{\text{main}})| \geq 2$ **then**
    for $k = 1$ **to** K **do**
      Compute $S_k = \text{predict}(Aux_k, u_p)$;
    end
    Compute $D_{\text{KL}}(P_{\text{main}} \| P_k)$, α, and w_k by Eqs. (19)–(21);
    foreach $c_j \in \text{Conflict}(S_{\text{main}})$ **do**
      Compute $r^{\text{temp}}(c_j)$ by Eq. (22);
      if $\max\limits_{c_j \in \text{Conflict}(S_{\text{main}})} r^{\text{temp}}(c_j) \leq \max\limits_{c_i \notin \text{Conflict}(S_{\text{main}})} S_{\text{main}}(c_i)$ **then**
        Compute $r^{\text{final}}(c_j)$ by Eq. (23);
      end
      else
        $r^{\text{final}}(c_j) = r^{\text{temp}}(c_j)$;
      end
    end
    Update the conflicting entries of S_{main} by r^{final};
  end
  Update $M_c(p, :) = S_{\text{main}}$;
end
for $p = 1$ **to** n **do**
  $\hat{y}_p = \arg\max_{1 \leq i \leq m} M_c(p, i)$;
end
return $\{\hat{y}_1, \ldots, \hat{y}_n\}$

described in Algorithm 2. Under the optimal parameter configuration, the multi-class prediction procedure is summarized in Algorithm 3.

For a testing sample, the m NRS-based S3WD classifiers first generate a confidence vector under the OVA strategy. If the confidence vector has a unique maximum component, the corresponding class is directly selected according to the maximum confidence rule. Otherwise, the conditional dynamic integration strategy is activated to adjust the confidence scores of the conflicting classes by incorporating auxiliary classifiers. As shown in Fig. 1, EM-S3WD consists of three main stages: NRS-based adaptive reference tuple selection, OVA-based multi-class S3WD classification, and conditional dynamic integration for confidence conflicts.

3.5 Theoretical properties and complexity analysis

This subsection analyzes the finite-step termination, decision-preservation property, and prediction complexity of EM-S3WD. These properties show that the proposed dynamic integration is a conditional conflict-resolution procedure rather than an iterative optimization process.

Proposition 3 (Finite-step termination)**.** For any testing sample, the prediction procedure of EM-S3WD terminates after a finite number of operations.

Proof For a testing sample, the m OVA-based S3WD classifiers are first evaluated once to generate the main confidence vector. If its maximum component is unique, the final label is obtained directly. Otherwise, the K auxiliary classifiers are invoked once, and the scores of at most m conflicting classes are updated once according to Eqs. (19)–(23). Since the updated scores are not repeatedly fed back into the integration procedure, no iterative or recursive operation is involved. Therefore, EM-S3WD terminates in finite steps. □

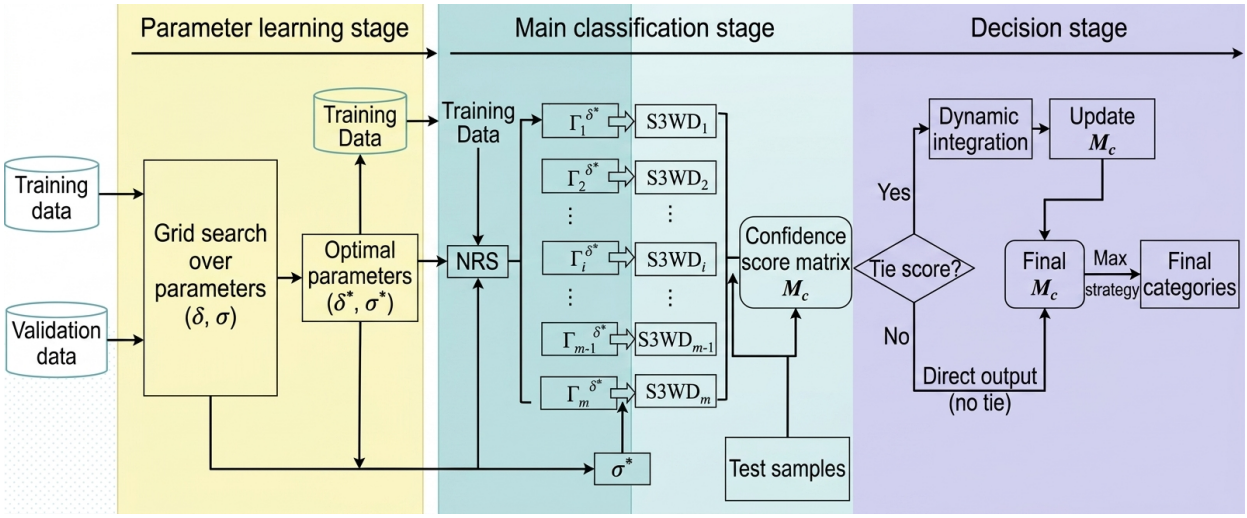


Fig. 1: The EM-S3WD framework

Proposition 4 (Decision preservation in non-conflict cases)**.** Let S_{main} be the confidence vector generated by the OVA-based S3WD classifiers for a testing sample u_j. If S_{main} has a unique maximum component, then the final decision of EM-S3WD is identical to that of the original OVA-based S3WD classifiers.

Proof Suppose that c_q is the unique maximizer of S_{main}. Then

$$\text{Conflict}(S_{\text{main}}) = \{c_q\}.$$

According to Algorithm 3, the auxiliary classifiers are activated only when

$$|\text{Conflict}(S_{\text{main}})| \geq 2.$$

Therefore, no confidence score is modified in this case, and the final label remains

$$\hat{y}_j = \arg \max_{1 \leq i \leq m} S_{\text{main}}(c_i) = c_q.$$

Hence, EM-S3WD preserves the original OVA-based S3WD decision for every non-conflict sample. □

Proposition 5 (Conditional prediction complexity)**.** Let n_{te} be the number of testing samples, m the number of classes, K the number of auxiliary classifiers, and $n_c \leq n_{\text{te}}$ the number of conflict samples. Let C_{S3WD} denote the cost of predicting one sample using one binary S3WD classifier, and let C_{aux} denote the cost of obtaining the score vector of one auxiliary classifier for one sample. The prediction complexity of EM-S3WD is

$$O\left(n_{\text{te}} m C_{\text{S3WD}} + n_c K (C_{\text{aux}} + m)\right).$$

Proof For each testing sample, the m binary S3WD classifiers are evaluated to construct the main confidence vector. The total cost of this stage is

$$O(n_{\text{te}} m C_{\text{S3WD}}).$$

The auxiliary classifiers are invoked only for the n_c conflict samples. For each conflict sample, evaluating the K auxiliary classifiers requires $O(KC_{\text{aux}})$, while calculating the KL-based weights and updating the confidence scores requires at most $O(Km)$. Therefore, the additional cost of conditional integration is

$$O\left(n_c K (C_{\text{aux}} + m)\right).$$

Combining the two stages gives the stated result. □

Remark 3.3. The auxiliary-classifier cost depends on n_c, rather than on the complete test-set size n_{te}. Consequently, when $n_c < n_{\text{te}}$, EM-S3WD avoids

$$(n_{\text{te}} - n_c)K$$

unnecessary auxiliary predictions compared with a full-time integration strategy. This result characterizes the efficiency of the conditional integration mechanism and does not imply that EM-S3WD is universally faster than classifiers with different model structures.

3.6 Illustrative example

Example 3.1. Suppose that we are given the seed data in Table 1, where $U = \{u_1, u_2, ..., u_{14}, u_{15}\}$ is the set of seeds and $A = \{a_1, a_2, ..., a_8\}$ is the set of attributes with a_1 to a_8 representing "Area", "girth", "compactness", "grain length", "grain width", "asymmetry coefficient", "grain groove length", and "category", respectively. Note that the data in Table 1 originates from the seed dataset in the UCI Machine Learning Repository [39]. We select 15 cases from it and replace parts of the values of a_8 with N/A. This is a three-class classification problem, and therefore three binary classifiers are constructed. We take $W = \{u_1, u_2, u_4, u_6, u_7, u_9\}$ as the training set, $V = \{u_3, u_5, u_8\}$ as the validation set for parameter selection, and $\{u_{10}, ..., u_{15}\}$ as the testing set.

Table 1: Information of the seeds

$U\backslash A$	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
u_1	13.37	13.78	0.88	5.32	3.13	4.67	5.09	3
u_2	12.76	13.38	0.90	5.07	3.16	2.83	4.83	3
u_3	12.38	13.44	0.86	5.22	2.99	5.47	5.05	3
u_4	18.96	16.20	0.91	6.05	3.90	4.33	5.75	2
u_5	19.15	16.45	0.89	6.25	3.82	3.08	6.19	2
u_6	18.89	16.23	0.90	6.23	3.77	3.64	5.97	2
u_7	14.43	14.40	0.88	5.59	3.27	3.98	5.14	1
u_8	14.49	14.61	0.85	5.72	3.11	4.12	5.40	1
u_9	14.52	14.60	0.86	5.74	3.11	1.48	5.49	1
u_{10}	12.79	13.53	0.88	5.22	3.05	5.48	4.96	N/A
u_{11}	12.62	13.67	0.85	5.41	2.91	3.31	5.23	N/A
u_{12}	19.06	16.45	0.89	6.42	3.72	2.25	6.16	N/A
u_{13}	20.03	16.90	0.88	6.49	3.86	3.06	6.32	N/A
u_{14}	15.78	14.91	0.89	5.67	3.43	5.59	5.14	N/A
u_{15}	14.33	14.28	0.88	5.50	3.20	3.33	5.22	N/A

(1) **Step 1: Validation-guided parameter selection and reference tuple generation.**

Let $Y = \{a_1, a_2, \ldots, a_7\}$ be the subset of condition attributes, and let $X_1 = \{u_7, u_9\}$ be the concept of seeds for category 1 in the training set $W = \{u_1, u_2, u_4, u_6, u_7, u_9\}$. By NRS, we compute the Euclidean distances between objects, and the distance matrix is as follows:

$$\Delta_B = \begin{bmatrix} 0.00 & 0.72 & 1.85 & 1.90 & 0.50 & 1.30 \\ 0.72 & 0.00 & 2.13 & 2.16 & 0.89 & 1.31 \\ 1.85 & 2.13 & 0.00 & 0.39 & 1.56 & 1.94 \\ 1.90 & 2.16 & 0.39 & 0.00 & 1.56 & 1.76 \\ 0.50 & 0.89 & 1.56 & 1.56 & 0.00 & 0.95 \\ 1.30 & 1.31 & 1.94 & 1.76 & 0.95 & 0.00 \end{bmatrix}.$$

A grid search strategy is adopted on the validation set to jointly determine the optimal parameters (δ^*, σ^*). For each candidate pair (δ, σ), the reference tuples are constructed from the training set using NRS, and the corresponding classifiers are evaluated on the validation set. The parameter pair that achieves the highest validation accuracy is selected as the optimal configuration.

Assume that the optimal neighborhood radius is obtained as $\delta^* = 0.333$. Then, the δ^*-neighborhood-based lower approximation of X_1 induced by Y is

$$\underline{apr}_{\delta^*}(X_1) = \{u_7, u_9\}.$$

Thus, we obtain the reference set $\Gamma_1^{\delta^*} = \{\vec{\tau}_1^{\delta^*}, \vec{\tau}_2^{\delta^*}\}$, where

$$\vec{\tau}_1^{\delta^*} = (14.43, 14.40, 0.88, 5.59, 3.27, 3.98, 5.14),$$
$$\vec{\tau}_2^{\delta^*} = (14.52, 14.60, 0.86, 5.74, 3.11, 1.48, 5.49).$$

Similarly, $\Gamma_2^{\delta^*}$ and $\Gamma_3^{\delta^*}$ can be obtained for X_2 and X_3, respectively.

(2) **Step 2: Construction of binary classifiers and confidence matrix.**

Given the optimal similarity threshold $\sigma^* = 0.085$, we construct the NRS-based binary classifier S3WD_1. Similarly, S3WD_2 and S3WD_3 are constructed. According to Definition 3.4, the confidence score vectors are computed, and the initial confidence matrix M_c is obtained, as shown in Table 2.

(3) **Step 3: Conflict-aware integration and final decision.**

From Table 2, we observe that the fourth row corresponds to a confidence conflict:

$$S_{\text{main}}(u_{13}) = (0.018, 0.018, 0.018).$$

In this case, the decision is refined by incorporating auxiliary classifiers. For example, we obtain

$$S_1(u_{13}) = (0.251, 0.509, 0.239) \quad \text{(SVM)},$$

$$S_2(u_{13}) = (0, 1, 0) \quad \text{(NB)}.$$

Given that $\text{Accuracy}_1 = 1$ and $\text{Accuracy}_2 = 0.667$, the adjusted confidence vector is computed as

$$\begin{aligned} S_{\text{final}}(C) &= 0.9389 \times S_{\text{main}}(C) + 0.0611 \times (1 \times S_1(C) + 0 \times S_2(C)) \\ &= [0.0321, 0.0479, 0.0314]. \end{aligned}$$

Then, the confidence score matrix M_c is updated, as shown in Table 3. Finally, according to the maximum confidence strategy, the resulting labels are obtained as shown in Table 4.

Table 2: The initial confidence matrix

u_i	S3WD_1	S3WD_2	S3WD_3
u_{10}	0.018	0.000	0.536
u_{11}	0.018	0.000	0.054
u_{12}	0.018	0.357	0.018
u_{13}	**0.018**	**0.018**	**0.018**
u_{14}	0.107	0.018	0.036
u_{15}	0.750	0.000	0.054

Table 3: The final confidence matrix

u_i	S3WD_1	S3WD_2	S3WD_3
u_{10}	0.018	0.000	0.536
u_{11}	0.018	0.000	0.054
u_{12}	0.018	0.357	0.018
u_{13}	0.0321	**0.0479**	0.0314
u_{14}	0.107	0.018	0.036
u_{15}	0.750	0.000	0.054

Table 4: The resulting labels obtained by EM-S3WD

$U\backslash A$	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
u_{10}	12.79	13.53	0.88	5.22	3.05	5.48	4.96	3
u_{11}	12.62	13.67	0.85	5.41	2.91	3.31	5.23	3
u_{12}	19.06	16.45	0.89	6.42	3.72	2.25	6.16	2
u_{13}	20.03	16.90	0.88	6.49	3.86	3.06	6.32	2
u_{14}	15.78	14.91	0.89	5.67	3.43	5.59	5.14	1
u_{15}	14.33	14.28	0.88	5.50	3.20	3.33	5.22	1

To further illustrate the effectiveness of the proposed conflict-resolution mechanism, consider the conflict sample u_{13}, whose true label is class 2. As shown in Table 2, the original OVA-based S3WD classifiers generate

$$S_{\text{main}}(u_{13}) = (0.018, 0.018, 0.018),$$

and therefore cannot determine a unique class label. Random selection may assign u_{13} to any of the three classes and consequently produces a non-deterministic decision. The fixed-weight integration strategy predicts u_{13} as class 1, which is inconsistent with its true label. By contrast, EM-S3WD updates the confidence vector to

$$S_{\text{final}}(u_{13}) = (0.0321, 0.0479, 0.0314)$$

and correctly assigns u_{13} to class 2. This result illustrates that the proposed strategy can use sample-specific auxiliary evidence to resolve a confidence conflict that is not correctly handled by fixed integration. Moreover, for non-conflict samples such as u_{15}, EM-S3WD retains the original OVA-based S3WD decision without invoking the auxiliary classifiers. This example illustrates the role of conditional dynamic integration but does not, by itself, imply universal superiority over the comparison strategies.

4 Experimental results and analysis

This section evaluates the proposed EM-S3WD framework from four aspects. First, the sensitivity of the two key parameters, namely the neighborhood radius δ and the similarity threshold σ, is analyzed to show how neighborhood granularity and similarity evaluation affect classification performance. Second, the effectiveness of the conditional dynamic integration strategy is examined by comparing it with different conflict-resolution strategies. Third, the contribution of NRS-based adaptive reference tuple construction is analyzed by comparing it with the CRS-based reference tuple construction. Finally, the overall classification performance of EM-S3WD is evaluated on multiple benchmark datasets. The experiment was conducted using MATLAB (R2024a) on a computer with Windows 10 installed and an Intel(R) Core(TM) i5-10200H CPU @ 2.40GHz. The nine datasets were obtained from the UCI Machine Learning Repository [39], as summarized in Table 5.

Table 5: Information of the data sets

Data title	Attributes	Classes	Objects
iris	4	0,1,2	150
seeds	7	1,2,3	210
Zoo	16	1,2,3,4,5,6,7	101
balance-scale	4	1,2,3	625
Dermatology	34	1,2,3,4,5,6	366
Ecoli	7	1,2,5,6,7,8	332
abalone	8	1,2,3	4177
DryBean	16	1,2,3,4,5,6,7	13611
PageBlocks	10	1,2,3,4,5	5473

4.1 Evaluation metric used in the experiments

Four widely recognized machine learning metrics, namely Accuracy, Recall, Precision, F1 score ($F1$) and $Kappa$ [40], are used to evaluate the effectiveness of our approach (see the formulas (24)-(28)). These metrics provide a comprehensive assessment of classification performance, with higher values indicating superior outcomes. Specifically, Accuracy reflects the degree of correctness of a model's prediction ability as a whole; Recall measures a model's ability to capture positive examples; $F1$ serves as the harmonic mean between Precision and Recall, providing a balanced evaluation of both; $Kappa$ coefficient κ is used to measure the degree of consistency between the predicted results and the actual results of the model. Here, TP represents the total number of true positives, TN represents the total number of true negatives, FP represents the total number of false positives, and FN represents the total number of false negatives.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{24}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{25}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{36}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{27}$$

$$\kappa = \frac{p_o - p_e}{1 - p_e} \text{ with } p_o = \text{Accuracy} \tag{28}$$

$$p_e = \frac{(TP + FP)(TP + FN) + (FN + TN)(FP + TN)}{(TP + TN + FP + FN)^2} \tag{29}$$

4.2 Sensitivity analysis of the proposed EM-S3WD

As described in Sections 3.1 and 3.2, δ and σ are two key hyperparameters of EM-S3WD. The neighborhood radius δ determines the NRS-induced lower approximation and thus affects the generated reference tuple set. The similarity threshold σ controls the correlation degree between a testing sample and a δ-reference tuple in the NRS-based S3WD classifier. Therefore, both parameters directly influence the classification performance.

To investigate their influence, a grid search is conducted on the nine datasets listed in Table 5. For δ, the search range is set to $[0, 3]$ with a step of 0.3 for Zoo and $[0, 2.5]$ with a step of 0.25 for Dermatology.

For the other seven datasets, δ is searched within $[0, 1]$ with a step of 0.1. For σ, the search range is adjusted according to the characteristics of each dataset: $[0.1, 1.0]$ for Dermatology, $[0.01, 0.4]$ for Abalone, and $[0.01, 0.3]$ for the remaining datasets.

Each parameter combination is independently evaluated 100 times, and the average result is used for visualization. To present the joint influence of δ and σ, the following composite index is adopted:

$$\gamma = 0.2 \times (\text{Accuracy} + \text{Precision} + \text{Recall} + \text{F1} + \text{Kappa})$$

The index γ is used only for parameter sensitivity visualization, while the subsequent comparative experiments are evaluated using the individual standard metrics. In Fig. 2, the X-axis, Y-axis, and Z-axis represent δ, σ, and γ, respectively.

The following observations can be obtained from Fig. 2.

1. Firstly, consider the parameter δ. For most datasets, When δ is greater than 1, the index γ becomes 0. This is so because a large number of objects are in the boundary region when $\delta > 1$, hence the lower approximation set is empty, i.e., there are no reference tuples to choose. Moreover, Fig. 2 shows that different datasets have different ranges of δ, and the optimal δ varies among different datasets. Therefore, δ is a highly sensitive parameter and optimization of δ can significantly improve generalization of the model.
2. Secondly, consider the parameter σ. By Fig. 2, its effective range also varies depending on the dataset. In some datasets, σ only has discriminative ability when it is less than a certain threshold, while in others σ needs to be between two specific thresholds. We further analyze this phenomenon and conclude as follows:

   This can be explained by Definition 3.3, where σ governs the correlation degree $E_Y(u, \vec{\tau}^{\delta})$. An overly small σ makes the correlation criterion extremely stringent—only exact matches in all conditional attributes yield a non-zero value—thereby pushing most samples into the negative region and hurting generalization. An overly large σ, on the other hand, makes $E_Y(u, \vec{\tau}^{\delta})$ approach 1, so even dissimilar samples may be accepted into the positive region, weakening discriminative ability and causing overgeneralization. Hence, σ should be selected to strike a balance between discrimination and robustness, with dataset-specific tuning guided by experimental analysis.

4.3 Effectiveness of the conditional dynamic integration strategy

This subsection evaluates the effectiveness of the proposed conditional dynamic integration strategy in resolving confidence conflicts. To this end, EM-S3WD is compared with two conflict-resolution baselines: SM-S3WD, which randomly selects a label from the conflicting classes, and MM-S3WD, which resolves conflicts using fixed weights. Unlike these two methods, EM-S3WD invokes auxiliary classifiers only when a confidence conflict occurs and dynamically revises the confidence scores of the conflicting classes according to the current testing sample.

Since δ and σ affect the outputs of the OVA-based S3WD classifiers and consequently the occurrence of confidence conflicts, a grid search is also adopted in this experiment. For all datasets, δ is searched within $[0, 0.3]$ with a step of 0.1. For σ, a dataset-specific range $[a, b]$ is used with a step of $(b - a)/10$. For each dataset, 30 parameter combinations are evaluated. In Fig. 3, the X-axis represents $\delta + \sigma$ and is used only to distinguish the tested parameter combinations, while the Y-axis represents classification accuracy. According to Fig. 3, the following observations can be made.

As shown in Fig. 3, EM-S3WD achieves higher overall accuracy for most parameter configurations and exhibits comparatively stable performance on several datasets. Its advantage is particularly evident on Iris, Dermatology, DryBean, and PageBlocks, where it maintains relatively high accuracy over most parameter combinations. Clear improvements are also observed on Seeds, Balance-scale, and Ecoli, although the performance of all three methods fluctuates under some parameter configurations. On Abalone, EM-S3WD still performs better for most combinations, but its advantage is relatively moderate. The results also show that the effectiveness of conflict resolution is dataset and parameter dependent. MM-S3WD and SM-S3WD occasionally approach EM-S3WD under particular parameter combinations, especially on Seeds, Ecoli, Abalone, and PageBlocks. Nevertheless, their performance is generally less stable, and substantial degradation occurs for several configurations. These observations indicate that fixed weighting and random selection cannot consistently resolve the confidence conflicts, whereas the proposed dynamic strategy provides more reliable decisions across different parameter configurations. The overall accuracy in Fig. 3 is calculated over all testing samples. Since the three methods differ only in their treatment of confidence conflicts, their differences may be obscured when conflict samples account for only a small proportion of the test set. Therefore,

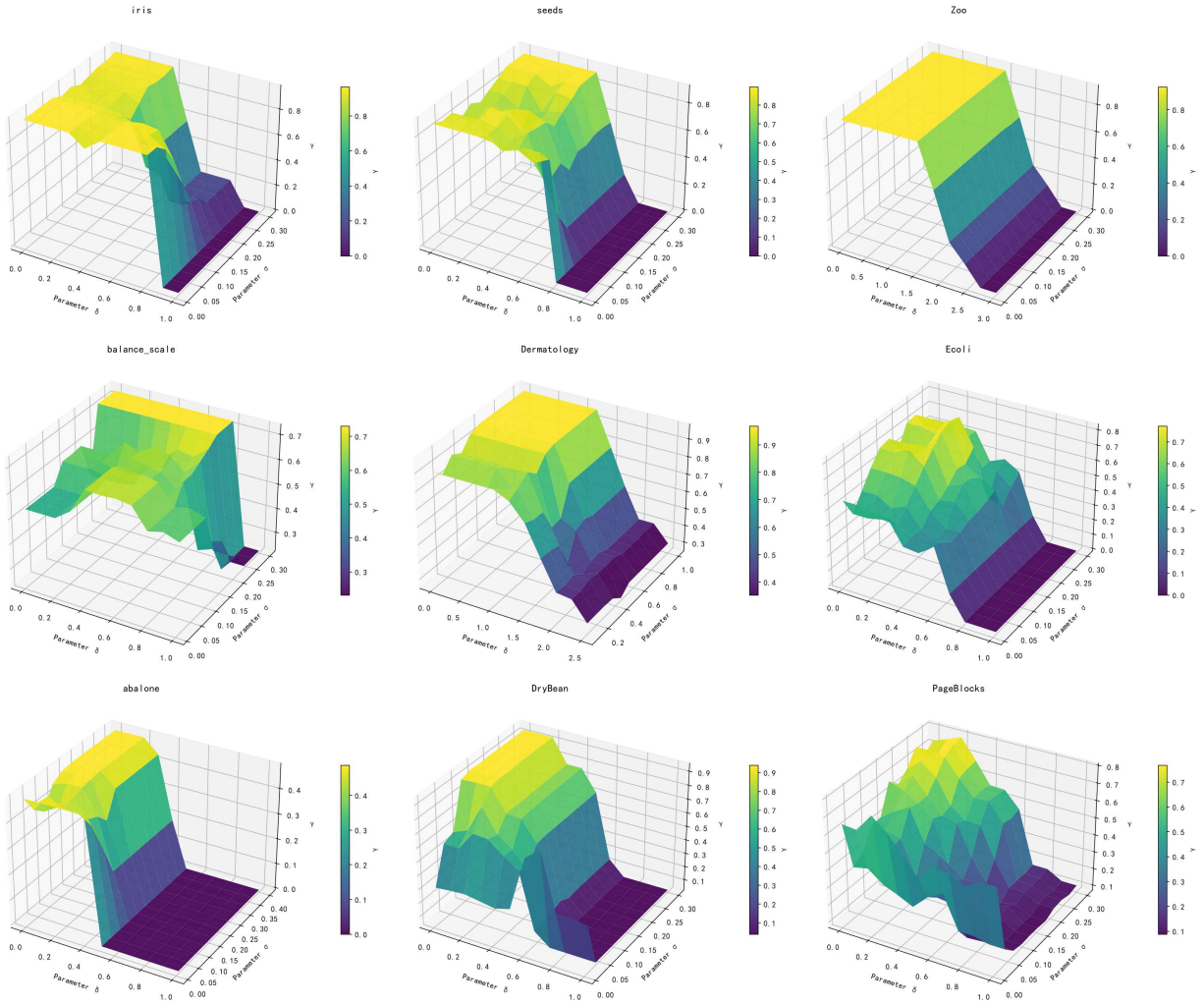


Fig. 2: Experiments on sensitivity to the parameter δ and σ

we further evaluate the three strategies exclusively on the samples for which the main OVA-based S3WD classifiers produce a non-unique maximum confidence score.

Table 6: Conflict rates and classification accuracy on conflict samples across 30 parameter combinations

Dataset	Conflict rate (%)	SM-S3WD (%)	MM-S3WD (%)	EM-S3WD (%)
Iris	43.33 ± 31.62	41.56 ± 19.14	34.00 ± 11.43	$\mathbf{84.91 \pm 21.21}$
Seeds	39.56 ± 31.89	38.41 ± 22.22	43.81 ± 23.09	$\mathbf{83.65 \pm 12.40}$
Zoo	4.88 ± 0.00	50.00 ± 0.00	50.00 ± 0.00	50.00 ± 0.00
Balance-scale	43.81 ± 12.03	33.06 ± 5.75	47.78 ± 18.13	$\mathbf{85.16 \pm 3.54}$
Dermatology	12.45 ± 2.21	44.47 ± 8.89	56.18 ± 14.17	$\mathbf{88.00 \pm 4.29}$
Ecoli	42.43 ± 23.10	31.35 ± 12.80	47.22 ± 14.00	$\mathbf{68.16 \pm 14.66}$
Abalone	63.38 ± 37.70	32.98 ± 6.09	35.19 ± 9.55	$\mathbf{50.38 \pm 9.75}$
DryBean	71.02 ± 38.77	22.11 ± 8.56	28.55 ± 13.14	$\mathbf{81.70 \pm 11.10}$
PageBlocks	47.81 ± 39.13	27.29 ± 9.88	66.70 ± 32.08	$\mathbf{79.55 \pm 23.90}$

Note: Values are reported as mean ± standard deviation. Conflict rate and conflict accuracy are averaged over 30 parameter combinations. The three methods are evaluated on the same conflict samples generated by the main OVA-based S3WD classifiers. SM-S3WD, MM-S3WD, and EM-S3WD denote random tie-breaking, static weighting, and the proposed conditional dynamic integration, respectively.

Table 6 further shows that the mean conflict rate varies substantially across the datasets, from 4.88% on Zoo to 71.02% on DryBean, and is sensitive to the parameter configuration. EM-S3WD obtains the highest mean conflict accuracy on eight of the nine datasets, with particularly clear improvements on Iris, Seeds, Balance-scale, Dermatology, and DryBean. Its advantage is smaller on Abalone and PageBlocks, indicating that the benefit of dynamic integration depends on the discriminative evidence supplied by the auxiliary classifiers. Overall, the conflict-specific results confirm that sample-dependent integration is generally more

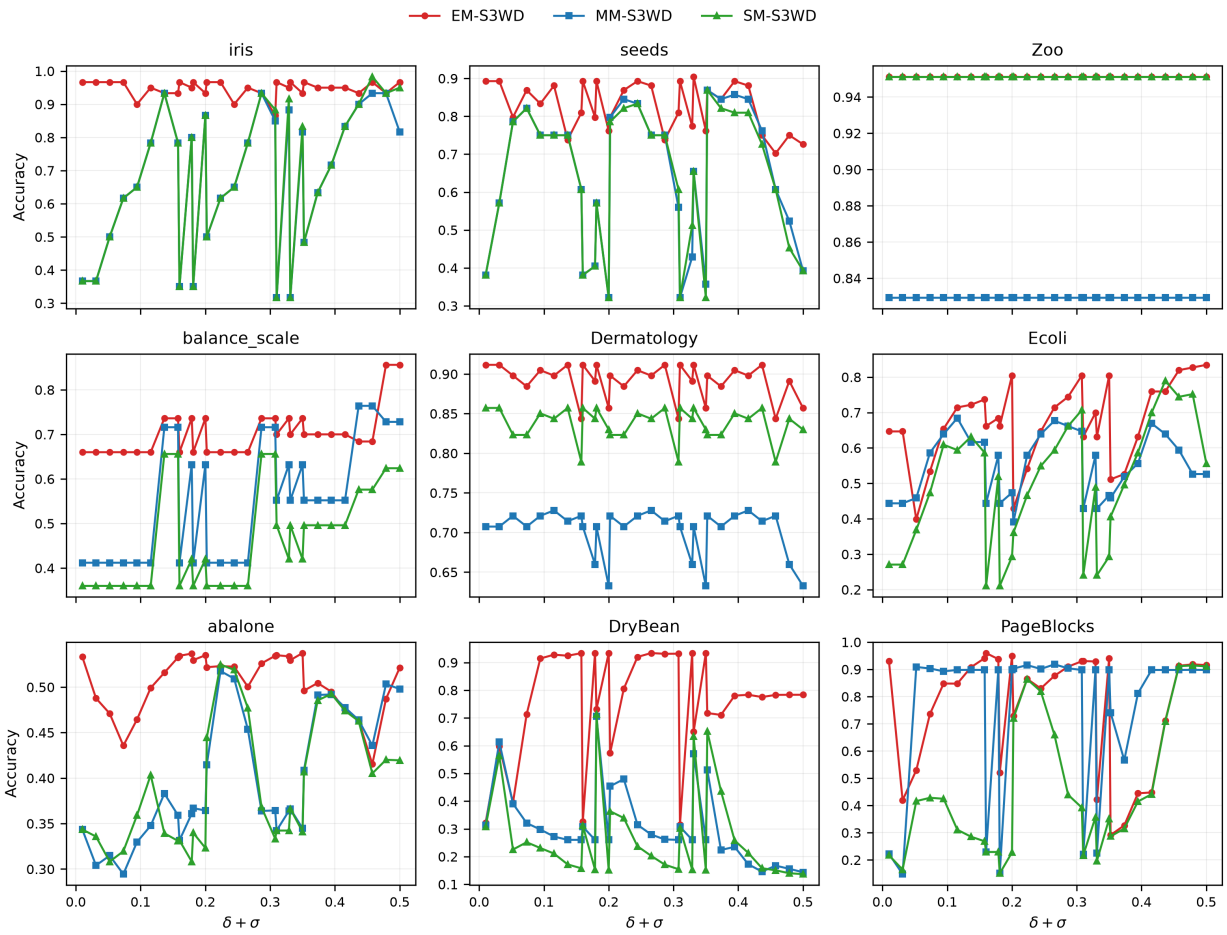


Fig. 3: Comparison of different conflict-resolution strategies

effective than random selection and static weighting, although the magnitude of the improvement remains dataset dependent.

4.4 Effectiveness of NRS-based adaptive reference tuple selection

To evaluate the proposed NRS-based adaptive reference tuple selection mechanism, two model variants are compared: a fixed reference tuple method and a validation-guided adaptive reference tuple method. In the fixed method, the reference tuples remain unchanged after being constructed from the training data. In the adaptive method, the neighborhood radius δ is selected using the validation set, and the corresponding NRS-induced reference tuples are used for classification.

Gaussian noise is added to the input data to simulate different levels of uncertainty. The noise intensity ϵ ranges from 0 to 1.0 with an increment of 0.1. For each noise level, experiments are repeated using different random seeds, and the mean and standard deviation of classification accuracy are recorded. In Fig. 4, the horizontal and vertical axes represent the noise intensity and classification accuracy, respectively. According to Fig. 4, the following observations can be made:

Overall, classification accuracy generally decreases with increasing noise intensity, though the degree of degradation varies across datasets. The adaptive reference tuple method shows a clear advantage on Iris, Seeds, Dermatology, and Abalone, particularly under medium-to-high noise levels; the improvement is most pronounced on Abalone, where the adaptive method consistently outperforms the fixed method across almost all nonzero noise levels. On Balance-scale and PageBlocks, the two methods yield nearly identical performance, indicating that fixed reference tuples are already stable on these datasets. On Zoo, Ecoli, and DryBean, however, no consistent advantage is observed—the two curves alternate or remain close, with the fixed method even slightly better in some cases. These results confirm that the adaptive reference tuple selection mechanism improves noise robustness on several datasets without causing an obvious performance loss on the others. Nevertheless, its effectiveness depends on the dataset, indicating that adaptive neighborhood selection is particularly useful when the fixed reference representation is sensitive to noise.

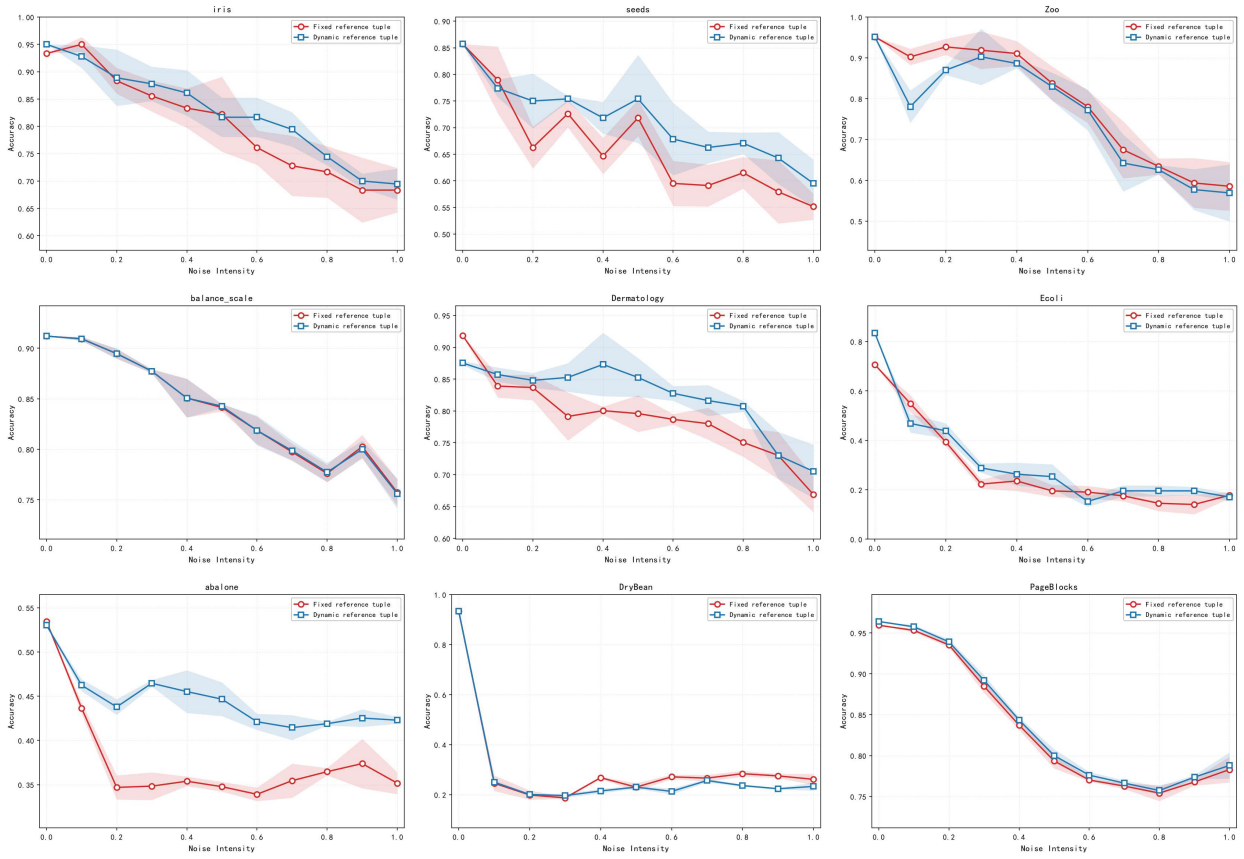


Fig. 4: Noise Robustness Comparison

4.5 Comparison with representative classifiers

This subsection evaluates EM-S3WD against six representative classifiers: 3WD-NRS, CART, KNN, Random Forest (RF), AdaBoost, and Bagging. These methods span neighborhood-rough-set-based three-way classification (3WD-NRS), conventional machine learning (CART and KNN), and ensemble learning (RF, AdaBoost, and Bagging). Parameter settings are listed in Table 7, and the classification results are reported in Table 8.

Table 8 shows that EM-S3WD leads on several datasets. On DryBean it outperforms all six competitors on all five metrics. On Balance-scale it achieves the highest Accuracy and Kappa, and on Abalone it achieves the highest Accuracy, Recall, and Kappa. On Dermatology it is second only to RF on all five metrics, ahead of the remaining baselines.

Table 7: Compared classifiers and related parameters

Classifiers	Parameters
KNN	NumNeighbors = 5
CART	Default parameter
Random Forest (RF)	Decision trees = 50
AdaBoost	n_estimators = 100; algorithm = SAMME
Bagging	n_estimators = 50; estimator = Decision Tree
3WD-NRS	$\omega = 0.1$
SVM	BoxConstraint = 10
Naive Bayes	DistributionNames = kernel

On the remaining datasets, however, the results are more varied. On Iris and Zoo, EM-S3WD stays within a modest margin of the leading classifiers on several metrics. On Seeds and Ecoli, the other classifiers generally outperform EM-S3WD, and on PageBlocks its high Accuracy comes with lower Recall and F1, indicating weaker recognition of minority-class samples. The statistical tests provide a rigorous basis for these conclusions. The Friedman test reveals significant overall differences among the seven classifiers for

Table 8: Performance comparison with representative classifiers (mean±std) over 10 seeds.

Dataset	Metric	EM-S3WD	3WD-NRS	CART	KNN	RF	AdaBoost	Bag
	ACC	0.9417±0.0362	**0.9583±0.0226**	0.9350±0.0441	0.9483±0.0346	0.9517±0.0266	0.9417±0.0432	**0.9583±0.0239**
	REC	0.9417±0.0362	**0.9583±0.0226**	0.9350±0.0441	0.9483±0.0346	0.9517±0.0266	0.9417±0.0432	**0.9583±0.0239**
iris	F1	0.9412±0.0369	0.9581±0.0229	0.9341±0.0459	0.9477±0.0361	0.9515±0.0267	0.9408±0.0452	**0.9582±0.0241**
	KAP	0.9125±0.0543	**0.9375±0.0339**	0.9025±0.0661	0.9225±0.0520	0.9275±0.0399	0.9125±0.0648	**0.9375±0.0358**
	PRE	0.9466±0.0316	**0.9625±0.0191**	0.9411±0.0363	0.9550±0.0254	0.9544±0.0253	0.9482±0.0347	0.9609±0.0217
	ACC	0.8512±0.0436	0.8845±0.0238	0.8750±0.0457	0.8798±0.0228	0.9060±0.0361	0.8952±0.0350	**0.9155±0.0283**
	REC	0.8512±0.0436	0.8845±0.0238	0.8750±0.0457	0.8798±0.0228	0.9060±0.0361	0.8952±0.0350	**0.9155±0.0283**
seeds	F1	0.8517±0.0432	0.8838±0.0248	0.8745±0.0464	0.8775±0.0244	0.9053±0.0366	0.8934±0.0360	**0.9150±0.0290**
	KAP	0.7768±0.0653	0.8268±0.0358	0.8125±0.0685	0.8196±0.0341	0.8589±0.0542	0.8429±0.0524	**0.8732±0.0425**
	PRE	0.8655±0.0420	0.8903±0.0249	0.8805±0.0441	0.8861±0.0218	0.9122±0.0330	0.9037±0.0356	**0.9206±0.0257**
	ACC	0.9244±0.0422	**0.9561±0.0277**	0.9317±0.0321	0.8293±0.0345	0.9366±0.0262	0.8780±0.0000	0.9415±0.0329
	REC	0.8608±0.0737	**0.8821±0.0674**	0.8429±0.0656	0.6639±0.0645	0.8571±0.0476	0.7143±0.0000	0.8714±0.0656
Zoo	F1	0.8450±0.0889	**0.8823±0.0713**	0.8328±0.0712	0.6390±0.0741	0.8574±0.0528	0.6277±0.0000	0.8574±0.0753
	KAP	0.9004±0.0550	**0.9416±0.0367**	0.9096±0.0425	0.7719±0.0462	0.9159±0.0347	0.8382±0.0000	0.9226±0.0435
	PRE	0.8748±0.0903	**0.9074±0.0752**	0.8627±0.0825	0.6866±0.1095	0.9046±0.0534	0.5816±0.0000	0.8881±0.0800
	ACC	**0.9072±0.0065**	0.8260±0.0193	0.7768±0.0153	0.8480±0.0115	0.8320±0.0141	0.8868±0.0209	0.8108±0.0137
	REC	0.6588±0.0083	0.5986±0.0140	0.5849±0.0215	0.6228±0.0087	0.6057±0.0128	**0.7156±0.0349**	0.5958±0.0144
balance-scale	F1	0.6332±0.0130	0.5897±0.0111	0.5902±0.0193	0.6103±0.0119	0.5978±0.0126	**0.7177±0.0391**	0.5959±0.0149
	KAP	**0.8282±0.0120**	0.6895±0.0321	0.6174±0.0259	0.7246±0.0185	0.7004±0.0237	0.8009±0.0351	0.6690±0.0220
	PRE	0.6387±0.1083	0.5845±0.0085	0.5984±0.0197	0.6062±0.0227	0.5910±0.0141	**0.7235±0.0450**	0.5969±0.0166
	ACC	0.9714±0.0160	0.8041±0.0181	0.9306±0.0233	0.8435±0.0242	**0.9769±0.0092**	0.5170±0.0921	0.9544±0.0176
	REC	0.9667±0.0195	0.7065±0.0288	0.9213±0.0321	0.7634±0.0303	**0.9728±0.0109**	0.3626±0.1048	0.9516±0.0198
Dermatology	F1	0.9684±0.0178	0.7244±0.0379	0.9186±0.0289	0.7827±0.0305	**0.9741±0.0099**	0.2884±0.1147	0.9494±0.0188
	KAP	0.9642±0.0200	0.7524±0.0233	0.9133±0.0292	0.8030±0.0306	**0.9710±0.0115**	0.3772±0.1212	0.9430±0.0220
	PRE	0.9721±0.0154	0.8133±0.0595	0.9250±0.0260	0.8633±0.0186	**0.9765±0.0091**	0.3294±0.1313	0.9518±0.0176
	ACC	0.7699±0.0317	0.8496±0.0224	0.7812±0.0295	**0.8744±0.0188**	0.8474±0.0184	0.8045±0.0256	0.8256±0.0253
	REC	0.7478±0.0341	0.8219±0.0316	0.6865±0.0439	**0.8533±0.0209**	0.7693±0.0622	0.7337±0.0456	0.7444±0.0622
Ecoli	F1	0.6827±0.0528	0.8151±0.0405	0.6931±0.0514	**0.8488±0.0203**	0.7957±0.0446	0.7205±0.0556	0.7594±0.0565
	KAP	0.6809±0.0416	0.7900±0.0318	0.6977±0.0406	**0.8253±0.0266**	0.7860±0.0267	0.7279±0.0365	0.7569±0.0348
	PRE	0.6785±0.0786	0.8253±0.0444	0.7270±0.0769	**0.8583±0.0250**	0.8504±0.0274	0.7627±0.0718	0.8003±0.0549
	ACC	**0.5428±0.0115**	0.5151±0.0102	0.4907±0.0085	0.5199±0.0099	0.5417±0.0112	0.5223±0.0136	0.5374±0.0118
	REC	**0.5490±0.0115**	0.5078±0.0102	0.4929±0.0087	0.5250±0.0102	0.5443±0.0112	0.5256±0.0125	0.5400±0.0121
abalone	F1	0.5354±0.0096	0.4900±0.0126	0.4926±0.0083	0.5199±0.0106	**0.5413±0.0116**	0.5102±0.0099	0.5388±0.0119
	KAP	**0.3152±0.0172**	0.2619±0.0156	0.2343±0.0128	0.2810±0.0151	0.3110±0.0169	0.2825±0.0195	0.3046±0.0180
	PRE	0.5332±0.0106	0.5133±0.0102	0.4927±0.0084	0.5183±0.0115	**0.5391±0.0121**	0.5100±0.0137	0.5380±0.0119
	ACC	**0.9180±0.0094**	0.6200±0.0078	0.8884±0.0095	0.6620±0.0066	0.9158±0.0039	0.8180±0.0414	0.9165±0.0048
	REC	**0.9284±0.0098**	0.5936±0.0072	0.9021±0.0094	0.6639±0.0059	0.9260±0.0048	0.7828±0.0817	0.9269±0.0053
DryBean	F1	**0.9306±0.0089**	0.5339±0.0113	0.9027±0.0091	0.6659±0.0058	0.9275±0.0043	0.7755±0.0931	0.9283±0.0047
	KAP	**0.9008±0.0115**	0.5293±0.0101	0.8651±0.0114	0.5888±0.0081	0.8982±0.0047	0.7801±0.0496	0.8990±0.0058
	PRE	**0.9337±0.0078**	0.5317±0.0289	0.9039±0.0090	0.6737±0.0062	0.9293±0.0047	0.8228±0.0774	0.9301±0.0045
	ACC	0.9589±0.0044	0.9006±0.0010	0.9623±0.0047	0.9547±0.0026	**0.9738±0.0018**	0.9525±0.0092	0.9714±0.0026
	REC	0.6261±0.0351	0.2770±0.0261	0.8355±0.0186	0.6740±0.0359	0.8524±0.0265	0.6623±0.0528	**0.8618±0.0219**
PageBlocks	F1	0.6691±0.0438	0.3031±0.0289	0.8208±0.0293	0.7162±0.0292	0.8580±0.0116	0.6962±0.0552	**0.8589±0.0225**
	KAP	0.7619±0.0251	0.0624±0.0151	0.8024±0.0231	0.7356±0.0158	**0.8580±0.0106**	0.7263±0.0723	0.8465±0.0123
	PRE	0.7745±0.0687	0.4941±0.0702	0.8138±0.0378	0.8013±0.0253	**0.8782±0.0260**	0.7725±0.0857	0.8687±0.0374

Table 9: Statistical comparison of the classifiers across nine datasets

Metric	χ_F^2	p_F	EM-S3WD rank	Min. p_{Holm}	Significant
Accuracy	13.6733	0.0335	3.722	1.0000	No
Recall	13.4582	0.0363	3.500	1.0000	No
F1	15.7336	0.0153	4.111	0.4453	No
Kappa	13.6733	0.0335	3.722	1.0000	No
Precision	13.0476	0.0423	4.111	0.6328	No

Note: χ_F^2 and p_F denote the Friedman statistic and its corresponding p-value. Smaller mean ranks indicate better overall performance. Min. p_{Holm} is the minimum Holm-adjusted p-value obtained from the six pairwise Wilcoxon signed-rank tests between EM-S3WD and the competing classifiers for each metric.

all five metrics (all $p < 0.05$). Pairwise Wilcoxon signed-rank tests with Holm correction show that no individual competitor significantly outperforms EM-S3WD, so EM-S3WD is statistically on par with even the strongest baselines.

Overall, EM-S3WD performs competitively with strong ensemble and neighborhood-rough-set classifiers and leads on several datasets and metrics. Its contribution lies in an interpretable multi-class S3WD framework with adaptive reference tuple construction and explicit confidence-conflict resolution, which distinguishes it from the compared classifiers.

5 Conclusions

This paper develops EM-S3WD as a targeted multi-class extension of the binary S3WD classifier proposed by Xu et al. The proposed framework addresses two model-specific issues. First, NRS is used to make the reference tuple construction dependent on neighborhood granularity, and a validation-guided procedure jointly selects the neighborhood radius and similarity threshold. Second, the formation mechanism of confidence conflicts in the OVA-based extension is analyzed through the finite-valued property of the S3WD confidence scores. A conditional dynamic integration strategy is then introduced to resolve these conflicts while preserving the original OVA-based S3WD decisions for non-conflict samples.

The experimental results provide evidence for the two main components of EM-S3WD. The adaptive reference tuple construction improves noise robustness on several datasets, although its effectiveness remains dataset dependent. The conflict-specific experiment shows that the mean conflict rate varies substantially across datasets and parameter configurations. EM-S3WD obtains the highest mean conflict accuracy on eight of the nine datasets, indicating that sample-dependent integration is generally more effective than random selection and static weighting for resolving the identified conflicts. In the comparison with representative classifiers, EM-S3WD achieves competitive overall performance. However, the Holm-adjusted pairwise tests do not establish statistically significant superiority over any individual competitor across all datasets. Therefore, the contribution of EM-S3WD lies mainly in its adaptive reference representation and explicit conflict-aware decision mechanism, rather than in universal performance superiority.

The proposed framework still has several limitations. Its computational cost may increase with the number of samples and classes because of neighborhood computation, parameter search, and the construction of multiple OVA-based classifiers. Moreover, the benefit of conditional integration depends on the occurrence of confidence conflicts and the reliability of the auxiliary confidence outputs. Future work will investigate efficient neighborhood search and parameter selection, alternative auxiliary-model selection mechanisms, and conflict resolution under other multi-class strategies, such as one-versus-one decomposition and direct multi-class S3WD models.

References

[1] Yao, Y.: Three-way decisions with probabilistic rough sets. Information Sciences **180**(3), 341–353 (2010) https://doi.org/10.1016/j.ins.2009.09.021

[2] Yao, Y.: The superiority of three-way decisions in probabilistic rough set models. Information Sciences **181**(6), 1080–1096 (2011) https://doi.org/10.1016/j.ins.2010.11.019

[3] Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems **90**(2), 111–127 (1997) https://doi.org/10.1016/S0165-0114(97)00077-8

[4] Pedrycz, W., Homenda, W.: Building the fundamentals of granular computing: A principle of justifiable granularity. Applied Soft Computing **13**(10), 4209–4218 (2013) https://doi.org/10.1016/j.asoc.2013.06.017

[5] Yao, Y.: Granular computing and sequential three-way decisions. In: Lingras, P., Wolski, M., Cornelis, C., Mitra, S., Wasilewski, P. (eds.) Rough Sets and Knowledge Technology, pp. 16–27. Springer, ??? (2013). https://doi.org/10.1007/978-3-642-41299-8_3

[6] Yang, J., Wang, G., Zhang, Q., Chen, Y., Xu, T.: Optimal granularity selection based on cost-sensitive sequential three-way decisions with rough fuzzy sets. Knowledge-Based Systems **163**, 131–144 (2019) https://doi.org/10.1016/j.knosys.2018.08.019

[7] Qian, J., Han, X., Yu, Y., Liu, C., Yu, J.: Research on multi-granularity sequential three-way decisions based on the fuzzy t-equivalence relation. Applied Soft Computing **149**, 110980 (2023) https://doi.org/10.1016/j.asoc.2023.110980

[8] Zhang, Q., Huang, Z., Wang, G.: A novel sequential three-way decision model with autonomous error correction. Knowledge-Based Systems **212**, 106526 (2021) https://doi.org/10.1016/j.knosys.2020.106526

[9] Qian, J., Liu, C., Yue, X.: Multigranulation sequential three-way decisions based on multiple thresholds. International Journal of Approximate Reasoning **105**, 396–416 (2019) https://doi.org/10.1016/j.ijar.2018.12.007

[10] Liang, P., Cao, W., Hu, J.: A sequential three-way classification model based on risk preference and decision correction. Applied Soft Computing **149**, 110978 (2023) https://doi.org/10.1016/j.asoc.2023.110978

[11] Yang, X., Loua, M.A., Wu, M., Huang, L., Gao, Q.: Multi-granularity stock prediction with sequential three-way decisions. Information Sciences **621**, 524–544 (2023) https://doi.org/10.1016/j.ins.2022.11.077

[12] Hu, J., Cao, W., Liang, P.: A novel sequential three-way decision model for medical diagnosis. Symmetry **14**(5) (2022) https://doi.org/10.3390/sym14051004

[13] Tu, J., Su, S., Xu, J.: A novel grey relational clustering model under sequential three-way decision framework. Information Sciences **663**, 120248 (2024) https://doi.org/10.1016/j.ins.2024.120248

[14] Ju, H., Pedrycz, W., Li, H., Ding, W., Yang, X., Zhou, X.: Sequential three-way classifier with justifiable granularity. Knowledge-Based Systems **163**, 103–119 (2019) https://doi.org/10.1016/j.knosys.2018.08.022

[15] Yang, J., Liu, Z., Xia, S., Wang, G., Zhang, Q., Li, S., Xu, T.: 3wc-gbnrs++: A novel three-way classifier with granular-ball neighborhood rough sets based on uncertainty. IEEE Transactions on Fuzzy Systems **32**(8), 4376–4387 (2024) https://doi.org/10.1109/TFUZZ.2024.3397697

[16] Chen, W., Zhang, Q., Dai, Y.: Sequential multi-class three-way decisions based on cost-sensitive learning. International Journal of Approximate Reasoning **146**, 47–61 (2022) https://doi.org/10.1016/j.ijar.2022.03.009

[17] Liang, P., Lei, D., Gao, X., Hu, J., Chin, K.: A sequential three-way decision model for classification with multilevel information gain and regret value optimization. Information Sciences **658**, 120041 (2024) https://doi.org/10.1016/j.ins.2023.120041

[18] Li, Y., Gao, F., Sha, M., Shao, X.: Sequential three-way decision with automatic threshold learning for credit risk prediction. Applied Soft Computing **165**, 112127 (2024) https://doi.org/10.1016/j.asoc.2024.112127

[19] Xu, W., Ai, S., Yang, H.: Sequential three-way decision with adaptive thresholds and its applications in two binary classifiers. Expert Systems with Applications **295**, 128194 (2026) https://doi.org/10.1016/j.eswa.2025.128194

[20] Pawlak, Z.: Rough sets and intelligent data analysis. Information Sciences **147**(1), 1–12 (2002) https://doi.org/10.1016/S0020-0255(02)00197-4

[21] Hu, Q., Yu, D., Liu, J., Wu, C.: Neighborhood rough set based heterogeneous feature subset selection. Information Sciences **178**(18), 3577–3594 (2008) https://doi.org/10.1016/j.ins.2008.05.024

[22] Sewwandi, M.A.N.D., Li, Y., Zhang, J.: A class-specific feature selection and classification approach using neighborhood rough set and k-nearest neighbor theories. Applied Soft Computing **143**, 110366 (2023) https://doi.org/10.1016/j.asoc.2023.110366

[23] Sewwandi, M.A.N.D., Li, Y., Zhang, J.: Granule-specific feature selection for continuous data classification using neighborhood rough sets. Expert Systems with Applications **238**, 121765 (2024) https://doi.org/10.1016/j.eswa.2023.121765

[24] Zhang, D., Zhu, P.: Variable radius neighborhood rough sets and attribute reduction. International Journal of Approximate Reasoning **150**, 98–121 (2022) https://doi.org/10.1016/j.ijar.2022.08.005

[25] Fürnkranz, J.: Round robin classification. J. Mach. Learn. Res. **2**, 721–747 (2002) https://doi.org/10.1162/153244302320884605

[26] Rifkin, R.M., Klautau, A.: In defense of one-vs-all classification. J. Mach. Learn. Res. **5**, 101–141 (2004)

[27] Hsu, C.-W., Lin, C.-J.: A comparison of methods for multiclass support vector machines. Trans. Neur. Netw. **13**(2), 415–425 (2002) https://doi.org/10.1109/72.991427

[28] Mienye, I.D., Sun, Y.: A survey of ensemble learning: Concepts, algorithms, applications, and prospects. IEEE Access **10**, 99129–99149 (2022) https://doi.org/10.1109/ACCESS.2022.3207287

[29] Mishra, S., Shaw, K., Mishra, D.H., Patil, S., Kotecha, K., Kumar, S., Bajaj, S.: Improving the accuracy of ensemble machine learning classification models using a novel bit-fusion algorithm for healthcare ai systems. Frontiers in Public Health **10** (2022) https://doi.org/10.3389/fpubh.2022.858282

[30] Sun, Y., Li, Z., Li, X., Zhang, J.: Classifier selection and ensemble model for multi-class imbalance learning in education grants prediction. Applied Artificial Intelligence **35**(4), 290–303 (2021) https://doi.org/10.1080/08839514.2021.1877481

[31] Britto, A.S., Sabourin, R., Oliveira, L.E.S.: Dynamic selection of classifiers—a comprehensive review. Pattern Recognition **47**(11), 3665–3680 (2014) https://doi.org/10.1016/j.patcog.2014.05.003

[32] Cruz, R.M.O., Sabourin, R., Cavalcanti, G.D.C.: Dynamic classifier selection: Recent advances and perspectives. Information Fusion **41**, 195–216 (2018) https://doi.org/10.1016/j.inffus.2017.09.010

[33] Ko, A.H.R., Sabourin, R., Britto, Jr., A.S.: From dynamic classifier selection to dynamic ensemble selection. Pattern Recognition **41**(5), 1718–1731 (2008) https://doi.org/10.1016/j.patcog.2007.10.015

[34] Kullback, S., Leibler, R.A.: On information and sufficiency. Annals of Mathematical Statistics **22**, 79–86 (1951)

[35] Xu, X., Li, Y., Yuan, C.: Conditional image generation with one-vs-all classifier. Neurocomputing **434**, 261–267 (2021) https://doi.org/10.1016/j.neucom.2020.12.091

[36] Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. Pattern Recognition **44**(8), 1761–1776 (2011) https://doi.org/10.1016/j.patcog.2011.01.017

[37] Hüllermeier, E., Vanderlooy, S.: Combining predictions in pairwise classification: An optimal adaptive voting strategy and its relation to weighted voting. Pattern Recognition **43**(1), 128–142 (2010) https://doi.org/10.1016/j.patcog.2009.06.013

[38] Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D.: Top 10 algorithms in data mining. Knowledge and Information Systems **14**(1), 1–37 (2008) https://doi.org/10.1007/s10115-007-0114-2

[39] Dua, D., Graff, C.: UCI Machine Learning Repository. http://archive.ics.uci.edu/ml, Irvine, CA (2019)

[40] Ben-David, A.: A lot of randomness is hiding in accuracy. Eng. Appl. Artif. Intell. **20**(7), 875–885 (2007) https://doi.org/10.1016/j.engappai.2007.01.001