Pattern Recognition 170 (2026) 111985

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Supervised incremental feature selection using regularization vector for dynamic multi-scale interval valued datasets

Zihan Feng, Xiaoyan Zhang *

College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China

ARTICLE INFO

Keywords:
Affiliation contradictory state
Incremental feature selection
Multi-scale fuzzy decision
Regularization vector
Review rate

ABSTRACT

Feature selection is pivotal for enhancing machine learning and data mining models, where its accuracy directly affects model performance and applicability. Traditional methods often overlook the dynamic nature of data and the multi-scale aspect of high-dimensional datasets, leading to limitations in real-world applications. This paper introduces a novel incremental feature selection method using a regularization vector (RV) tailored for dynamic multi-scale interval valued fuzzy decision systems (D-$MIvFD$). The paper first establishes the concepts of object affiliation relation and class, providing a theoretical basis for integrating replay and regularization. It then introduces the affiliation contradictory state (ACS) and RV, broadening the application of contradictory state (CS) in dynamic settings and enabling efficient feature selection. The integration of regularization and replay strategies is realized through four algorithms designed for different update patterns. Empirical results across various datasets show that the proposed method significantly outperforms multiple conventional techniques, highlighting its practical potential for real-world deployments.

1. Introduction

Incremental feature selection is a continuously evolving method within the fields of dynamic data analysis and machine learning, serving a fundamental role in the dynamic identification and optimization [1]. This iterative process not only boosts data processing efficiency and alleviates overfitting risks, but also ameliorates the model's capacity to generalize, particularly when handling extensive and high-dimensional [2] datasets, such as pictures and videos [3]. By selectively introducing or excluding features over time, incremental feature selection [4] enables models to adjust to evolving data patterns, which is a crucial capability in real-time data analysis and online learning scenarios [5]. With the integration of deep learning [6] technologies, incremental feature selection methods are being further optimized to accommodate more intricate data structures and extract profound feature insights, thus indicating expanded application prospects in areas such as image object detection [7], natural language processing [8], concept-cognitive learning [9], pattern recognition [10].

One notable distinction between incremental feature selection [11] and conventional feature selection methods lies in its ability to effectively and adaptively respond to the evolving peculiarity of real-world information [12]. In various realistic fields such as autonomous driving [13] and remote sensing monitoring [14] research, datasets usually exhibit dynamic characteristics, with features and data points fluctuating over time based on the concrete context, which necessitates a superior standard of adaptability from the model. Incremental feature selection provides the advantage of adjusting and optimizing feature subsets [15] in real-time, enabling it to seamlessly accommodate these dynamic variations. Numerous studies have been conducted on the topic of incremental feature selection. An incremental feature selection approach to multi-dimensional variation based on matrix dominance conditional entropy was expounded by Xu et al. [16]. This paper introduces two algorithms IFS-A and IFS-D that leverage prior reduction results to efficiently decrease the required computation time. An incremental reduction method of imbalanced distributed mixed data based on k-nearest neighbor rough set was formulated by Xu et al. [17]. A novel incremental feature selection approach based on fuzzy rough set theory was delineated by Zhao et al. [4]. A rough set theory-based group incremental feature selection approach was elaborated by Zhao et al. [18]. The triple nested equivalence class rough set theory ($TNECRST$) was introduced in this article as a refinement to the incremental feature selection technique based on rough set theory (RST), addressing the issues of excessive redundancy and limited efficiency associated with the extant approach. A novel incremental feature selection based on sub-tolerance relation for dynamic incomplete data was illuminated by Zhao et al. [19]. The sub-tolerance relation class (STC) was suggested to promote the incremental feature selection algorithm for managing incomplete stream data.

* Corresponding author.
E-mail addresses: fengzihan1218@163.com (Z. Feng), zxy19790915@163.com (X. Zhang).

https://doi.org/10.1016/j.patcog.2025.111985
Received 23 September 2024; Received in revised form 5 June 2025; Accepted 9 June 2025
Available online 23 June 2025

The prevalence of uncertainty and imprecision in information presents challenges in the decision-making process [20], such as object boundary perception [21] and multi-label classification, rendering traditional decision models inadequate. The exploration of $D\text{-}MIvFD$ provides a more adaptable and comprehensive framework for addressing tanglesome decision-making scenarios [22]. This area of study holds theoretical and practical significance in intensifying decision quality and streamlining decision processes [23]. The ongoing researches in this domain [24] are displayed below. An attribute ranking method based on rough sets and interval-valued fuzzy sets was proposed by BK Vo et al. [25]. This study introduces a novel rough-fuzzy hybrid ($RAFAR$) approach that employs an interval-valued fuzzy matrix to articulate the preference relations among attribute pairs. An overlap function-based three-way intelligent decision model under interval-valued fuzzy information systems ($IFISs$) was suggested by Wang et al. [26]. This study conceives a $3WD$ model rooted in prospect theory (PT) within the framework of $IFISs$ to resolve decision-making challenges effectively. In addition, an embedded feature selection model that includes both feature fusion and feature enhancement between views is proposed by Hao et al. [27]. In this paper, the comprehensive information of features is preserved by minimizing the difference between the weights of the two types of features, so as to reduce the impact of noise on multi-view multi-label learning. The literature classification related to this study is summarized in Table 13.

However, the current work does not incorporate a relatively full-fledged and comprehensive incremental feature selection model or algorithm for $D\text{-}MIvFD$ format data. Furthermore, several aspects, including the efficiency, accuracy and adaptability of feature selection, still require further promotion. Finally, the task of integrating the ideas of other research fields into this discipline and thereby facilitating the progress of the theoretical framework remains to be further accomplished. To summarize, in the light of these extant issues, the following contributions are presented in this article.

- First and foremost, the notion of $D\text{-}MIvFD$ is procured, and the conceptions of object affiliation relation and affiliation class are raised, which lays a theoretical foundation for the implementation of the replay tactic.
- Secondly, the concepts of affiliation first contradictory index and RV are elucidated in the context of dynamic data update, which furnishes a possibility for the combination of regularization and replay ideas.
- Finally, in accordance with four dynamic update modes, four high-efficiency incremental feature selection strategies are manifested. Experimental results demonstrate the superior performance of the proposed methods over traditional approaches across multiple datasets.

The rest of the article is organized as follows. Interval valued decision information system ($IVDIS$) and multi-scale decision information system ($MDIS$) are illuminated in Section 2. Furthermore, the concepts of object affiliation relation, affiliation class, ACS and RV are elaborated in Section 3. After that, four incremental feature selection techniques are expounded in Section 4. In Section 5, numerous numerical experiments are conducted to appraise the efficacy of the aforementioned methods in conjunction with comparative experiments. Finally, the results and conclusions of this paper are summarized and further directions for future research are pointed out.

2. Interval valued decision information system and multi-scale decision information system

In this section, the concept of $IVDIS$ is presented initially, followed by a comprehensive interpretation of $MDIS$. These principles serve as the foundation for the investigation conducted in this study.

2.1. Interval valued decision information system

In practical decision-making scenarios, data frequently does not exist in the form of a single value and instead exhibits a degree of uncertainty, which can be represented through interval values. An $IVDIS$ [28] can be picturesquely demonstrated by a tuple $S = (O, C \cup \{d\})$. $O = \{o_1, o_2, \dots, o_n\}$ is a non-empty finite set of objects, which is known as the universe of discourse. $C = \{c_1, c_2, \dots, c_m\}$ is a non-empty restricted set of conditional attributes, where c_j can be referred to as a mapping, $c_j : O \to W(V_j)$, for all $c_j \in C$, i.e. $c_j(o_i) = [a_j^L(o_i), a_j^U(o_i)]$, $o_i \in O$. Where $a_j^L(o_i) \le a_j^U(o_i)$, $a_j^L(o_i) \in V_j$, $a_j^U(o_i) \in V_j$, and V_j is the domain of conditional attribute c_j, $W(V_j)$ is the set of all interval values over V_j. Besides, $d \notin \{c_j | j = 1, 2, \dots, m\}$ the decision such that $d : O \to V_d$, where V_d is the domain of decision attribute d.

2.2. Multi-scale decision information system

$MDIS$ operates on the principle of multi-granularity, enabling the examination and manipulation of data from various perspectives and in-depth levels. An $MDIS$ [29] can be pictorially exhibited by a tuple $S = (O, C \cup \{d\})$. $O = \{o_1, o_2, \dots, o_n\}$ is a non-empty limited set of objects, which is deemed as the universe of discourse. $C = \{c_1, c_2, \dots, c_m\}$ is a non-empty numbered set of conditional attributes, each conditional attribute possesses L scales. Where c_j^l can be considered as a mapping, $c_j^l : O \to V_j^l$, for all $c_j \in C$. And V_j^l is the domain of conditional attribute c_j on the lth scale. Furthermore, $d \notin \{c_j^l | j = 1, 2, \dots, m; l = 1, 2, \dots, L\}$ the decision such that $d : O \to V_d$.

3. Dynamic multi-scale interval valued Fuzzy decision information system

This section provides a meticulous explanation of the multi-scale interval valued fuzzy decision information system ($MIvFD$). Subsequently, in consideration of the characteristics of dynamic issues, four patterns of updating information system are delineated. Drawing from the principles of incremental learning grounded in regularization and replay, the notions of object affiliation relation and affiliation class are introduced. Where regularization and replay methodologies are significant techniques employed in machine learning to enhance the generalization capacity and stability of models. Regularization mitigates the risk of overfitting by incorporating a penalty term into the loss function, whereas the replay strategy refines the learning process of the current model by leveraging previously acquired data or experiences. When these two concepts are applied to the domain of information systems research, certain adaptations are implemented to enhance their relevance and applicability, and the particular modifications are expounded in the following sections. In addition, in accordance with the illustration of CS and object affiliation relation, ACS and affiliation contradictory state sequence ($ACSS$) that are appropriate for addressing dynamic issues are presented. The primary model of this research is expounded in Fig. 1.

3.1. Multi-scale interval valued Fuzzy decision information system

An $MIvFD$ [30] can be held up by a tuple $S = (O, C \cup \{d\})$. $O = \{o_1, o_2, \dots, o_n\}$ is a non-empty restricted set of objects, which is referred to as the universe of discourse. $C = \{c_1, c_2, \dots, c_m\}$ is a non-empty finite set of conditional attributes, each conditional attribute possesses L scales. Where c_j^l can be deemed as a mapping, $c_j^l : O \to W(V_j^l)$, i.e. $c_j^l(o_i) = [a_j^{l,L}(o_i), a_j^{l,U}(o_i)]$, $o_i \in O$. And $a_j^{l,L}(o_i) \le a_j^{l,U}(o_i)$, $a_j^{l,L}(o_i) \in V_j^l$, $a_j^{l,U}(o_i) \in V_j^l$, $W(V_j^l)$ is the set of all interval values over V_j^l, for all $c_j \in C$. In addition, $d \notin \{c_j^l | j = 1, 2, \dots, m; l = 1, 2, \dots, L\}$ the decision such that $d : O \to [0, 1]$, where $[0, 1]$ is the domain of decision attribute d with a single scale.

In an $MIvFD$, there are two surjective relation between contiguous scales. One is $\phi_j^{l,l+1} : V_j^l \to V_j^{l+1}$, i.e. $v_j^{l+1}(o_i) = \phi_j^{l,l+1}(v_j^l(o_i))$, where

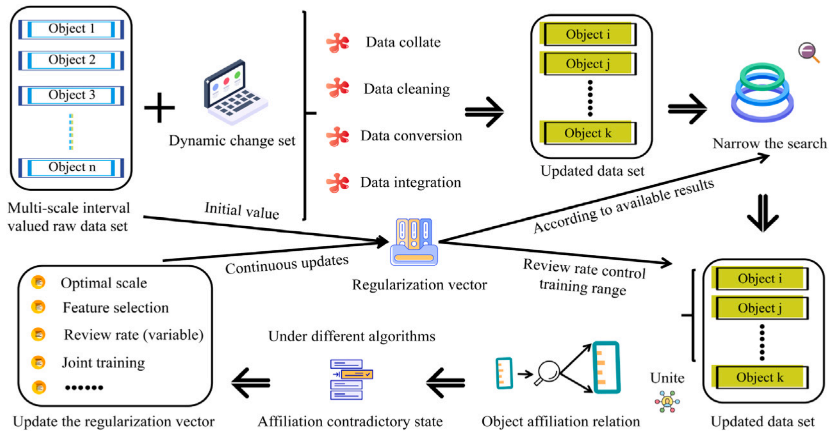

Fig. 1. Model flow chart.

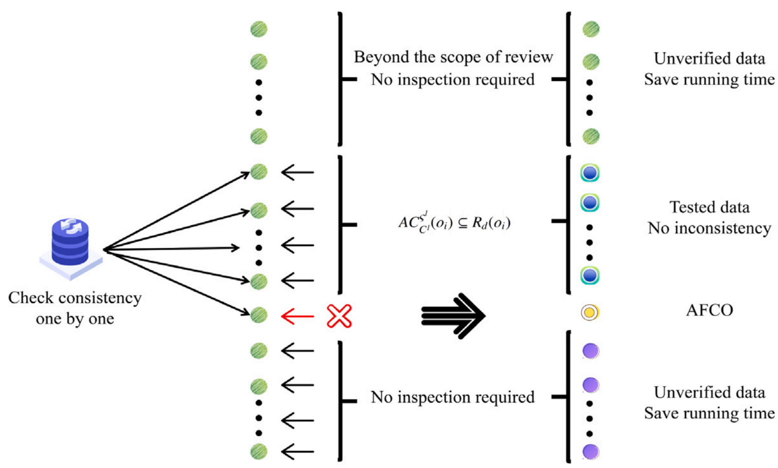

Fig. 2. The calculation process of $AFCO$ and $AFCI$.

$o_i \in O$, $v_j^l \in V_j^l$, $v_j^{l+1} \in V_j^{l+1}$, $l \in \{1,2,\dots,L-1\}$, $j \in \{1,2,\dots,m\}$. The another is $\Phi_j^{l,l+1} : W(V_j^l) \to W(V_j^{l+1})$, i.e. $c_j^{l+1}(o_i) = \Phi_j^{l,l+1}(c_j^l(o_i)) = \Phi_j^{l,l+1}([a_j^{l,L}(o_i), a_j^{l,U}(o_i)]) = [a_j^{l+1,L}(o_i), a_j^{l+1,U}(o_i)]$, where $o_i \in O$, $l \in \{1,2,\dots,L-1\}$. And $\Phi_j^{l,l+1}([a_j^{l,L}(o_i), a_j^{l,U}(o_i)]) = [\phi_j^{l,l+1,L}(a_j^{l,L}(o_i)), \phi_j^{l,l+1,U}(a_j^{l,U}(o_i))]$, where $\phi_j^{l,l+1}$ is considered as scale transformation mapping, and $\Phi_j^{l,l+1}$ is deemed as domain transformation mapping.

A fuzzy equivalence relation R_d on objects can divide object set O into t decision classes based on mapping $d : O \to [0,1]$ and given t, demonstrated as $O/R_d = \{R_d(o)|o \in O\} = \{D_1, D_2, \dots, D_t\}$, where $R_d(o) = \{p|(o,p) \in R_d, p \in O\}$, $D_i = \{o|\frac{i-1}{t} \le d(o) < \frac{i}{t}\}$, $i \in \{1,2,\dots,t\}$, specially $D_t = \{o|\frac{t-1}{t} \le d(o) \le 1\}$.

To expedite the process of feature selection within $MIvFD$ framework, CS is introduced as a metric for evaluating the consistency of information systems. It can be divided into L $sub-IVDISs$ for an $MIvFD$, written as $S^l = (O, C^l \cup \{d\})$, $|O| = n$, $l \in \{1,2,\dots,L\}$. We judge whether $o_i \in O$ satisfy $SC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$ in line with the order of subscripts, where $SC_{C^l}^{\varsigma^l}(o_i)$ is the similarity class of object o_i w.r.t. C^l and ς^l. The first object that does not gratify $SC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$ is called the first contradictory object (FCO) w.r.t. $S^l = (O, C^l \cup \{d\})$, denoted by $FCO(S^l) = o_i$. And i is the first contradictory index (FCI) w.r.t. $S^l = (O, C^l \cup \{d\})$, recorded by $FCI(S^l) = i$. If $\forall o_i \in O$, $SC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$, then $FCI(S^l) = n+1$. At last $CS(S^l)$ [30] is propounded as follows.

$$CS(S^l) = CS(O, C^l \cup \{d\}) = \begin{cases} 0, & FCI(S^l) = n+1 \\ 1, & FCI(S^l) \ne n+1 \end{cases}. \tag{1}$$

3.2. Relevant definitions in dynamic context

Initially, it is substantial to emphasize that the dynamic operations illuminated in this research encompass object increase, object decrease, attribute increase and attribute decrease. The information system capable of performing these four types of dynamic operations based on $MIvFD$ is referred to as D-$MIvFD$.

As previously stated, it is indispensable to establish an indicator that appraises the level of resemblance among the interval values. Supposing two interval values $E = [e^L, e^U]$ and $F = [f^L, f^U] \in W(R)$, and $E \cap F = [g^L, g^U]$, the pivot inclusion degree (PI) of E and F can be depicted as follows [30], where $0 \le \zeta_1, \zeta_2 \le 1$ and $\zeta_1 + \zeta_2 = 1$.

$$PI(E,F) = \zeta_1 \frac{g^U - g^L}{e^U - e^L} + \zeta_2 \frac{g^U - g^L}{f^U - f^L}. \tag{2}$$

Subsequently, in consideration of the current context of dynamic issues and the objective of expeditious feature selection, the object affiliation relation is introduced.

Definition 1. Let $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1,2,\dots,m; l = 1,2,\dots,L\} \cup \{d\})$ be a D-$MIvFD$, $\varsigma^l \in [0,1]$, as well $B \subseteq C$, object affiliation relation w.r.t. ς^l and B^l on $U \times Y$ is delineated as follows, where $o \in U$, $p \in Y$ and $(o,p) \in U \times Y$.

$$AR_{B^l}^{\varsigma^l} = \{(o,p) \in U \times Y | PI(c^l(o), c^l(p)) \ge \varsigma^l, \ \forall c \in B\}, \ U \subseteq Y. \tag{3}$$

The affiliation class w.r.t. B^l and ς^l on $U \times Y$ is proclaimed as follows.

$$AC_{B^l}^{\varsigma^l}(o) = \{p | (o,p) \in AR_{B^l}^{\varsigma^l},\ o \in U,\ p \in Y\}. \tag{4}$$

To guarantee that the coarser the scale, the larger the affiliation class, the threshold ranges [29] for diverse scales are specified as follows, $l \in \{1, 2, \dots, L-1\}$.

$$\varsigma^{l+1} \le \kappa^{l+1} \le \varsigma^l,\ where\ \kappa^{l+1} = \min\left(\min_{c\in C}\left(\min_{(o,p)\in AR_{C^l}^{\varsigma^l}} (PI(c^{l+1}(o), c^{l+1}(p)))\right),\ \varsigma^l\right). \tag{5}$$

In accordance with Eq. (5) declaration, it can be acquired that $AR_{C^l}^{\varsigma^l} \subseteq AR_{C^{l+1}}^{\varsigma^{l+1}}$.

Proof. For any $(o,p) \in AR_{C^l}^{\varsigma^l}, x \in C$, $PI(x^{l+1}(o), x^{l+1}(p)) \ge \kappa^{l+1} \ge \varsigma^{l+1}$. Consequently, $(o,p) \in AR_{C^{l+1}}^{\varsigma^{l+1}}$ and $AR_{C^l}^{\varsigma^l} \subseteq AR_{C^{l+1}}^{\varsigma^{l+1}}$ can be procured.

$$PI(x^{l+1}(o), x^{l+1}(p)) \ge \min_{c\in C}\left(\min_{(o,p)\in AR_{C^l}^{\varsigma^l}} (PI(c^{l+1}(o), c^{l+1}(p)))\right). \tag{6}$$

Definition 2. Let $S = (O, C \cup \{d\})$ be an $IVDIS$, $\varsigma^* \in [0,1]$, moreover $B, H \subseteq C$. If $AC_B^{\varsigma^*}(o_i) \subseteq AC_H^{\varsigma^*}(o_i)$, for all $o_i \in U$, we proclaim that $AC_B^{\varsigma^*}$ is finer than $AC_H^{\varsigma^*}$, and then affirm it as $AC_B^{\varsigma^*} \preceq AC_H^{\varsigma^*}$.

Proposition 1. *Let $S = (O, C \cup \{d\})$ be a D-$MIvFD$, for $l \in \{1, 2, \dots, L-1\}$.*

$$AC_{C^l}^{\varsigma^l}(o_i) \subseteq AC_{C^{l+1}}^{\varsigma^{l+1}}(o_i),\ \forall o_i \in U,\ AC_{C^l}^{\varsigma^l} \preceq AC_{C^{l+1}}^{\varsigma^{l+1}}. \tag{7}$$

The above constitutes the monotonicity of affiliation class in a D-$MIvFD$, which implies that, as the scale increases, the affiliation class becomes larger.

Definition 3. Let $S = (O, C \cup \{d\})$ be a D-$MIvFD$, furthermore $\varsigma^l \in [0,1]$, for $l \in \{1, 2, \dots, L\}$. S is said to be consistent if $S^1 = (O, C^1 \cup \{d\})$ is consistent, i.e. $AC_{C^1}^{\varsigma^1}(o_i) \subseteq R_d(o_i)$, for all $o_i \in U$. On the contrary, S is inconsistent. Similarly, $S^l = (O, C^l \cup \{d\})$ is consistent, if and only if $AC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$, for all $o_i \in U$.

Evidently, as the scale increases, the consistency of S^l will progressively diminish. In order to refrain from the demand for continual assessment of the consistency of information subtable during feature selection, it is suggested to ensure that U undergoes a consistency processing prior to the execution of any algorithms, i.e. delete o_i, if $AC_{C^1}^{\varsigma^1}(o_i) \nsubseteq R_d(o_i)$, for all $o_i \in U$.

Proposition 2. *Let $S = (O, C \cup \{d\})$ be an $IVDIS$, $\varsigma^* \in [0,1]$, in addition $B, H \subseteq C$. If $H \subseteq B$, it follows that $AC_B^{\varsigma^*} \preceq AC_H^{\varsigma^*}$.*

When $S^l = (O, C^l \cup \{d\})$, $l \in \{1, 2, \dots, L\}$ in a D-$MIvFD$, check whether $AC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$ is satisfied from the smallest to largest subscripts for all $o_i \in U = \{o_1, o_2, \dots, o_x\}$, $|U| = x$. The first object o_i that does not gratify $AC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$ is called affiliation first contradictory object ($AFCO$) w.r.t. S^l and $U \times Y$, denoted by $AFCO(S^l, U \times Y) = o_i$. And then i is affiliation first contradictory index ($AFCI$) w.r.t. S^l and $U \times Y$, recorded by $AFCI(S^l, U \times Y) = i$. If S^l is consistent, then $AFCI(S^l, U \times Y) = x + 1$.

The procedure of computation $AFCO$ and $AFCI$ within an information system is illustrated in Fig. 2.

Definition 4. In a D-$MIvFD$, $S = (O, C \cup \{d\})$, $S^l = (O, C^l \cup \{d\})$, $l \in \{1, 2, \dots, L\}$. The format of ACS w.r.t. S^l and $U \times Y$ is signified as follows, where $|U| = x$ and $U = \{o_1, o_2, \dots, o_x\}$.

$$ACS(S^l, U \times Y) = \begin{cases} 0, & AFCI(S^l, U \times Y) = x + 1 \\ 1, & AFCI(S^l, U \times Y) \neq x + 1 \end{cases}. \tag{8}$$

There is no doubt that $ACS(S^l, U \times Y)$ is a Boolean value. When S^l is consistent, i.e. $AC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$, for all $o_i \in U$, $ACS(S^l, U \times Y) = 0$, otherwise $ACS(S^l, U \times Y) = 1$. $ACS(S^l, U \times Y)$ can be gained by inspecting whether $AC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$ in order, and the scrutiny can be terminated when $AFCO(S^l, U \times Y)$ is found. When multiple $ACSs$ are lined up together, a series of Boolean values could embody a transformation in the quantitative characteristics of a specific procedure. A monotonically undecreasing ordered sequence consisting of denumerable restricted $ACSs$ is stipulated as $ACSS$ to fulfill prompt and efficient feature selection in D-$MIvFD$. Several especial sequences will be revealed and implemented in the process of feature and optimal scale selection, such as optimal scale selection affiliation contradictory state sequence ($OSACS$), feature selection affiliation contradictory state sequence ($FSACS$).

Then introducing an indicator to document crucial intermediate calculation outcomes during the D-$MIvFD$ update process, and employing it to direct subsequent experiments could dramatically intensify speed and reduce computational resource consumption.

Definition 5. In the dynamic update process of D-$MIvFD$ $S = (O, C \cup \{d\})$, the format of RV is proposed as follows.

$$RV(S) = (COS, CFS, CTNO, CTNA, LOU, LAU, Y, \Lambda, OI). \tag{9}$$

Table 1
Symbol interpretation in RV.

Symbol	Explanation	Value range
COS	Current optimal scale	$COS \in \{1, 2, \dots, L\}$
CFS	Current feature selection	–
$CTNO$	Current total number of objects	–
$CTNA$	Current total number of attributes	–
LOU	Last object update	–
LAU	Last attribute update	–
Y	Review rate	$Y \in [0, 1]$
Λ	Joint training indicator	$\Lambda \in \{0, 1\}$
OI	Other information	–

The notations in RV are formulated in Table 1 accordingly. LOU and LAU monitor the latest updates of objects and attributes, respectively. It is imperative to highlight that Y can vary during the dynamic update process, and serious-minded consideration is required to determine the appropriate setting of Y for excellent outcomes. Furthermore, Y serves to restrict the percentage of data that will be incorporated into the new training phase, i.e. the percentage of replay.

In the domain of information system research, the omission of historical data in dynamic feature selection algorithms can result in catastrophic forgetting, which may lead to a range of significant repercussions. These include a diminished comprehension of historical patterns, the neglect of time series data, an increased risk of overfitting to current datasets, the oversight of long-term trends and seasonal variations, and a decrease in model robustness. Consequently, historical data is essential in the incremental feature selection process, as it enables the model to effectively capture the long-term dynamics and changes within the system, thereby ensuring stable predictive performance and enhanced generalization capabilities. This necessity underscores the importance of integrating the concept of replay into the research framework.

To streamline algorithm design, this study opts to involve the latter portion of dataset in the new training, denoted as $O \otimes Y$. Then $\Lambda = 1$ implies that the complete data table is trained to obtain new feature selection. When $Y = 1$ (i.e. $\Lambda = 1$), it requires the longest execution time but offers the most comprehensive training range.

Table 2
A dynamic updated multi-scale interval valued fuzzy decision information system.

O	c_5		c_6		c_1		c_2		c_3		c_4		d
	c_5^1	c_5^2	c_6^1	c_6^2	c_1^1	c_1^2	c_2^1	c_2^2	c_3^1	c_3^2	c_4^1	c_4^2	
o_1	[96, 96]	[9, 10]	[45, 83]	[4, 9]	[73, 82]	[7, 9]	[35, 57]	[3, 6]	[8, 10]	[0, 1]	[59, 76]	[5, 8]	0.3
o_2	[30, 36]	[3, 4]	[1, 19]	[0, 2]	[62, 86]	[6, 9]	[44, 87]	[4, 9]	[32, 74]	[3, 8]	[36, 67]	[3, 7]	0.1
o_3	[20, 66]	[2, 7]	[32, 55]	[3, 6]	[32, 54]	[3, 6]	[22, 46]	[2, 5]	[40, 82]	[4, 9]	[12, 24]	[1, 3]	0.6
o_4	[18, 52]	[1, 6]	[55, 89]	[5, 9]	[12, 46]	[1, 5]	[27, 53]	[2, 6]	[34, 86]	[3, 9]	[3, 20]	[0, 2]	0.9
o_5	[24, 35]	[2, 4]	[5, 36]	[0, 4]	[79, 80]	[7, 8]	[77, 88]	[7, 9]	[76, 96]	[7, 10]	[64, 83]	[6, 9]	0.2
o_6	[15, 59]	[1, 6]	[1, 72]	[0, 8]	[26, 48]	[2, 5]	[41, 73]	[4, 8]	[77, 98]	[7, 10]	[48, 94]	[4, 10]	0.1
o_7	[47, 50]	[4, 5]	[22, 77]	[2, 8]	[17, 37]	[1, 4]	[18, 37]	[1, 4]	[5, 38]	[0, 4]	[51, 84]	[5, 9]	0.6
o_8	[41, 98]	[4, 10]	[44, 85]	[4, 9]	[68, 71]	[6, 8]	[2, 27]	[0, 3]	[29, 33]	[2, 4]	[31, 40]	[3, 4]	0.7
o_9	[31, 98]	[3, 10]	[36, 78]	[3, 8]	[3, 24]	[0, 3]	[3, 73]	[0, 8]	[60, 91]	[6, 10]	[68, 86]	[6, 9]	0.4
o_{10}	[14, 44]	[1, 5]	[20, 53]	[2, 6]	[9, 24]	[0, 3]	[4, 39]	[0, 4]	[34, 55]	[3, 6]	[47, 86]	[4, 9]	0.3
o_{11}	[20, 66]	[2, 7]	[32, 55]	[3, 6]	[96, 96]	[9, 10]	[45, 83]	[4, 9]	[85, 96]	[8, 10]	[36, 86]	[3, 9]	0.8
o_{12}	[40, 46]	[4, 5]	[69, 82]	[6, 9]	[34, 89]	[3, 9]	[17, 58]	[1, 6]	[32, 99]	[3, 10]	[11, 52]	[1, 6]	0.6
o_{13}	[54, 87]	[5, 9]	[26, 89]	[2, 9]	[23, 45]	[2, 5]	[12, 50]	[1, 5]	[37, 98]	[3, 10]	[27, 100]	[2, 10]	0.2
o_{14}	[42, 59]	[4, 6]	[18, 28]	[1, 3]	[7, 66]	[0, 7]	[69, 86]	[6, 9]	[31, 95]	[3, 10]	[26, 29]	[2, 3]	0.4
o_{15}	[47, 70]	[4, 7]	[94, 96]	[9, 10]	[76, 79]	[7, 8]	[64, 94]	[6, 10]	[40, 61]	[4, 7]	[50, 71]	[5, 8]	0.7

This chapter lays the theoretical foundation for four feature selection algorithms argued in the following section. Subsequently, a micromesh portrait of feature selection will be provided for the four update patterns of $D\text{-}MI\upsilon FD$.

4. Incremental feature selection for dynamic multi-scale interval valued Fuzzy decision information system

Dynamic updating is a continual procedure that necessitates prior knowledge pertaining to the initial dataset $S=(O,C\cup\{d\})=(O,\{c_j^l|j=1,2,\dots,m;l=1,2,\dots,L\}\cup\{d\})$, $O=\{o_1,o_2,\dots,o_n\}$. The approach detailed in [30] is employed to determine the optimal scale and feature selection for the original dataset, and $RV(S)=(OS,FS,n,m,0,0,Y,0,null)$. It is requisite to emphasize that prior to feature selection, the optimal scale must be notarized in $D\text{-}MI\upsilon FD$. The optimal scale refers to the coarsest scale that guarantees the consistency of information subtable.

4.1. Incremental feature selection for dynamic object increase

Appending h_1 objects to the original dataset, the h_1 objects constitute O_1, denoted as $O_1=\{o_{n+1},o_{n+2},\dots,o_{n+h_1}\}$. O_1 and its corresponding multi-scale interval value and fuzzy decision value are added to S, and the organization and arrangement of the data are demonstrated in Fig. 3. When $U=O$, $Y=O$ and $U=O\cup O_1$, $Y=O\cup O_1$ respectively, corresponding to $AC_{C^l}^{\varsigma^l}(o)$, $\hat{AC}_{C^l}^{\varsigma^l}(p)$, $\forall o\in O, p\in O\cup O_1$, then $AC_{C^l}^{\varsigma^l}(o)\subseteq \hat{AC}_{C^l}^{\varsigma^l}(o)$. Further, it can be deduced that, when $S^l=(O,C^l\cup\{d\})$ is inconsistent, then $\hat{S}^l=(O\cup O_1,C^l\cup\{d\})$ is also inconsistent. Evidently, the new optimal scale $\hat{OS}$ only needs to be searched within $\{1,2,\dots,OS\}$. Meanwhile, in order to economize on training time, $\hat{OS}$ will be sought within the range of information table regulated by Y, i.e. $U=(O\otimes Y)\cup O_1$ and $Y=O\cup O_1$, $\hat{AC}_{C^l}^{\varsigma^l}(o)$, $\forall o\in U$. Additionally, as the scale becomes coarser, the inconsistency of information table gradually amplifies, corresponding to the monotonically increasing $ACS(\hat{S}^l,U\times Y)$, where $\hat{S}^l=(O\cup O_1,C^l\cup\{d\})$, $l\in\{1,2,\dots,OS\}$, $U=(O\otimes Y)\cup O_1$ and $Y=O\cup O_1$, which constitutes $OSACS=[ACS(\hat{S}^1,U\times Y),ACS(\hat{S}^2,U\times Y),\dots,ACS(\hat{S}^{OS},U\times Y)]$. Subsequently, a half-search is executed on $OSACS$ to chase down the sequence of final $ACS(\hat{S}^l,U\times Y)=0$, thereby identifying $\hat{OS}$. The next step involves confirming new feature selection set $\hat{FS}$ through a bond of $\hat{OS}$ and FS. This article employs a top-down approach to feature selection, i.e. delete the first attribute of C successively to form a conditional attribute subset sequence. The last j attributes in C are used to constitute the conditional attribute subset, $j\in\{m,m-1,m-2,\dots,1,0\}$.

$$\{\{c_1,c_2,\dots,c_m\},\{c_2,\dots,c_m\},\{c_3,\dots,c_m\},\dots,\{c_m\},\varnothing\}=\{C_m,C_{m-1},C_{m-2},\dots,C_1,C_0\}. \tag{10}$$

The smallest subset C_j that guarantees the consistency of information subtable at $\hat{OS}$ is $\hat{FS}$. When $\hat{OS}=OS$, it is imperative to seek $\hat{FS}$ within $\{C_m,C_{m-1},\dots,C_{FS}\}$; otherwise the scope of inquiry extends to $\{C_m,C_{m-1},C_{m-2},\dots,C_1,C_0\}$. As the feature subset dwindles, $ACS(\hat{S}_j^{\hat{OS}},U\times Y)$ emerges gradual aggrandization, where $\hat{S}_j^{\hat{OS}}=(O\cup O_1,C_j^{\hat{OS}}\cup\{d\})$, $U=(O\otimes Y)\cup O_1$ and $Y=O\cup O_1$. Then, only one half-search in either $FSACS=[ACS(\hat{S}_m^{\hat{OS}},U\times Y),ACS(\hat{S}_{m-1}^{\hat{OS}},U\times Y),\dots,ACS(\hat{S}_{FS}^{\hat{OS}},U\times Y)]$ or $FSACS=[ACS(\hat{S}_m^{\hat{OS}},U\times Y),ACS(\hat{S}_{m-1}^{\hat{OS}},U\times Y),\dots,ACS(\hat{S}_0^{\hat{OS}},U\times Y)]$ is indispensable to obtain the order in which the last $ACS(\hat{S}_j^{\hat{OS}},U\times Y)=0$, $\hat{FS}$ can then be inferred from this order. Following the completion of feature selection in this phase, $RV(S)$ is subsequently updated to $(\hat{OS},\hat{FS},n+h_1,m,+h_1,0,Y,0,null)$.

Example 1. $S=(O,C\cup\{d\})=(O,\{c_j^l|j=1,2,3,4;l=1,2\}\cup\{d\})$ is the information system corresponding to part of Table 2, obviously S is consistent. When $\varsigma^1=0.8$ and $\varsigma^2=0.5$, $RV(S)=(2,\{c_2,c_3,c_4\},10,4,0,0,0.7,0,null)$. Next, add five extra objects to S and position them subsequent to the current items, the relevant data of these objects are presented in Table 2. Suppose $t=2$, then $D_1=\{o_1,o_2,o_5,o_6,o_9,o_{10},o_{13},o_{14}\}$ and $D_2=\{o_3,o_4,o_7,o_8,o_{11},o_{12},o_{15}\}$. Following the accomplishment of consistency processing, $\hat{OS}$ is determined through a half-search. Since $ACS(\hat{S}^1,U\times Y)=0$ and $ACS(\hat{S}^2,U\times Y)=1$, then $\hat{OS}=1$. $\hat{FS}$ is subsequently calculated through a half-search operation and based on $ACS(\hat{S}_0^1,U\times Y)=1$, $ACS(\hat{S}_1^1,U\times Y)=1$, $ACS(\hat{S}_2^1,U\times Y)=0$, $ACS(\hat{S}_3^1,U\times Y)=0$ and $ACS(\hat{S}_4^1,U\times Y)=0$. Ultimately, $\hat{FS}=\{c_3,c_4\}$ and $RV(S)=(1,\{c_3,c_4\},15,4,+5,0,0.7,0,null)$.

4.2. Incremental feature selection for dynamic object decrease

The initial dataset is subjected to object deletion operation, the h_2 objects that require to be removed form O_2, $O_2\subseteq O$, as well as the organization and arrangement of the data are illustrated in Fig. 4. When $U=O$, $Y=O$ and $U=O-O_2$, $Y=O-O_2$ severally, corresponding to $AC_{C^l}^{\varsigma^l}(o)$, $\hat{AC}_{C^l}^{\varsigma^l}(p)$, $\forall o\in O,p\in O-O_2$, then $\hat{AC}_{C^l}^{\varsigma^l}(p)\subseteq AC_{C^l}^{\varsigma^l}(p)$. Further, when $S^l=(O,C^l\cup\{d\})$ is consistent, then $\hat{S}^l=(O-O_2,C^l\cup\{d\})$ is also consistent. Clearly, it is essential to search for $\hat{OS}$ within $\{OS,OS+1,\dots,L\}$. To save time, $\hat{OS}$ will be sought within the range of information table directed by Y, i.e. $U=(O-O_2)\otimes Y$ and $Y=O-O_2$, $\hat{AC}_{C^l}^{\varsigma^l}(o)$, $\forall o\in U$. Besides, as the scale becomes coarser, the inconsistency of information table gradually reinforces,

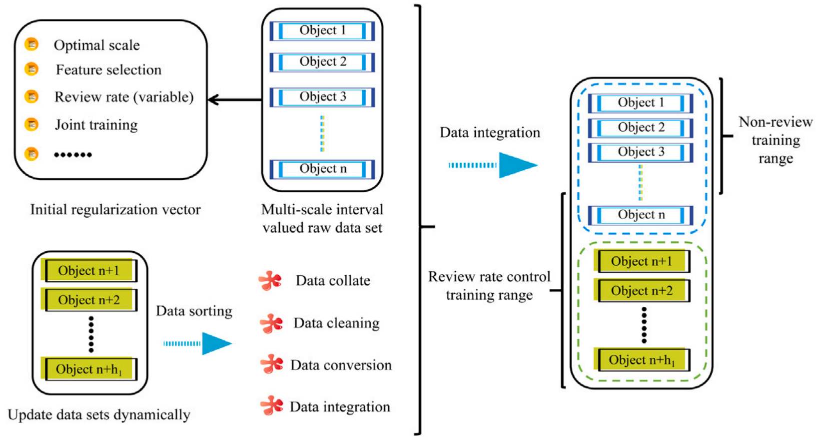

Fig. 3. Data organization and arrangement with dynamic object increase.

Algorithm 1: Feature Selection for Dynamic Object Increase ($ObDI-FS$).

Input: A $D\text{-}MIvFD$ $S=(O,C\cup\{d\})$, O, $RV(S)$, O_1.
Output: New optimal scale $\hat{OS}$, new feature selection set $\hat{FS}$, new $RV(S)$.
1: Calculating $\{D_1,D_2,...,D_t\}$ about the updated dataset
2: $left=1, right=OS$ /*Finding new optimal scale $\hat{OS}$.*/
3: **while** $left\le right$ **do**
4: $mid=(left+right)//2$
5: Calculating $AFCI(\hat{S}^{mid},U\times Y)$ and $ACS(\hat{S}^{mid},U\times Y)$
6: **if** $ACS(\hat{S}^{mid},U\times Y)=1$ **then**
7: $right=mid-1$
8: **else**
9: $left=mid+1$
10: **end if**
11: **end while**
12: $\hat{OS}=left-1$
13: **if** $\hat{OS}=OS$ **then**
14: $left=FS, right=m$
15: **else**
16: $left=0, right=m$
17: **end if**
18: **while** $left\le right$ **do**
19: $mid=(left+right)//2$
20: Calculating $AFCI(\hat{S}_{mid}^{\hat{OS}},U\times Y)$ and $ACS(\hat{S}_{mid}^{\hat{OS}},U\times Y)$
21: **if** $ACS(\hat{S}_{mid}^{\hat{OS}},U\times Y)=0$ **then**
22: $right=mid-1$
23: **else**
24: $left=mid+1$
25: **end if**
26: **end while**
27: $\hat{FS}=C_{left}$
28: **return** $\hat{OS}$, $\hat{FS}$, $RV(S)=(\hat{OS},\hat{FS},n+h_1,m,+h_1,0,Y,0,null)$

---

corresponding to the monotonically increasing $ACS(\hat{S}^l,U\times Y)$, where $\hat{S}^l=(O-O_2,C^l\cup\{d\})$, $l\in\{OS,OS+1,\dots,L\}$, $U=(O-O_2)\otimes Y$ and $Y=O-O_2$, which constitutes $OSACS=[ACS(\hat{S}^{OS},U\times Y),ACS(\hat{S}^{OS+1},U\times Y),\dots,ACS(\hat{S}^L,U\times Y)]$. Then a half-search is proceeded on $OSACS$ to chase down the sequence of final $ACS(\hat{S}^l,U\times Y)=0$, thereby ascertaining $\hat{OS}$.

When $\hat{OS}=OS$, search for $\hat{FS}$ within $\{C_{FS},C_{FS-1},\dots,C_0\}$; otherwise the search area will be expanded to $\{C_m,C_{m-1},\dots,C_0\}$. As the feature subset shrinks, $ACS(\hat{S}_j^{\hat{OS}},U\times Y)$ exhibits gradual expansion, where $\hat{S}_j^{\hat{OS}}=(O-O_2,C_j^{\hat{OS}}\cup\{d\})$. Then, a half-search in either $FSACS=[ACS(\hat{S}_{FS}^{\hat{OS}},U\times Y),ACS(\hat{S}_{FS-1}^{\hat{OS}},U\times Y),\dots,ACS(\hat{S}_0^{\hat{OS}},U\times Y)]$ or $FSACS=[ACS(\hat{S}_m^{\hat{OS}},U\times Y),ACS(\hat{S}_{m-1}^{\hat{OS}},U\times Y),\dots,ACS(\hat{S}_0^{\hat{OS}},U\times Y)]$ is requisite to acquire the order in which the last $ACS(\hat{S}_j^{\hat{OS}},U\times Y)=0$, $\hat{FS}$ can then be inferred from this order. $RV(S)$ is subsequently updated upon the completion of this procedure, and $RV(S)=(\hat{OS},\hat{FS},n-h_2,m,-h_2,0,Y,0,null)$.

Algorithm 2: Feature Selection for Dynamic Object Decrease ($ObDD-FS$).

Input: A $D\text{-}MIvFD$ $S=(O,C\cup\{d\})$, O, $RV(S)$, O_2, $|O_2|=h_2$, $O_2\subseteq O$.
Output: New optimal scale $\hat{OS}$, new feature selection set $\hat{FS}$, new $RV(S)$.
1: Calculating $\{D_1,D_2,...,D_t\}$ about the updated dataset
2: $left=OS, right=L$ /*Finding new optimal scale $\hat{OS}$.*/
3: Steps 3-11 of Algorithm 1
4: $\hat{OS}=left-1$
5: **if** $\hat{OS}=OS$ **then**
6: $left=0, right=FS$
7: **else**
8: $left=0, right=m$
9: **end if**
10: Steps 18-26 of Algorithm 1
11: $\hat{FS}=C_{left}$
12: **return** $\hat{OS}$, $\hat{FS}$, $RV(S)=(\hat{OS},\hat{FS},n-h_2,m,-h_2,0,Y,0,null)$

4.3. Incremental feature selection for dynamic attribute increase

Incorporating h_3 attributes into initial dataset, the h_3 attributes make up A_1, recorded as $A_1=\{c_{m+1},c_{m+2},\dots,c_{m+h_3}\}$, and the organization and arrangement of the data are manifested in Fig. 5. $\forall o\in O$, $AC_{C^l}^{\varsigma^l}(o)$, $\hat{AC}_{(C\cup A_1)^l}^{\varsigma^l}(o)$, then $\hat{AC}_{(C\cup A_1)^l}^{\varsigma^l}(o)\subseteq AC_{C^l}^{\varsigma^l}(o)$. According to the statement of consistency in information table, when $S^l=(O,C^l\cup\{d\})$ is consistent, then $\hat{S}^l=(O,(C\cup A_1)^l\cup\{d\})$ is also consistent, thus simply search for $\hat{OS}$ within $\{OS,OS+1,\dots,L\}$. To save training time, $\hat{OS}$ will be sought within the range of information table controlled by Y, i.e. $U=O\otimes Y$ and $Y=O$, $\hat{AC}_{(C\cup A_1)^l}^{\varsigma^l}(o)$, $\forall o\in U$. In addition, as the scale becomes coarser, the inconsistency of information table gradually aggrandizes, corresponding to the monotonically increasing $ACS(\hat{S}^l,U\times Y)$, where $\hat{S}^l=(O,(C\cup A_1)^l\cup\{d\})$, $l\in\{OS,OS+1,\dots,L\}$, $U=O\otimes Y$ and $Y=O$, which constitutes $OSACS=[ACS(\hat{S}^{OS},U\times Y),ACS(\hat{S}^{OS+1},U\times Y),\dots,ACS(\hat{S}^L,U\times Y)]$. Subsequently, a half-search is performed on $OSACS$ to chase down the sequence of final $ACS(\hat{S}^l,U\times Y)=0$, therefore recognizing $\hat{OS}$. Whereafter $\hat{FS}$ can be notarized within $\{C_{m+h_3},C_{m+h_3-1},\dots,C_0\}$. As the feature subset lessens, $ACS(\hat{S}_j^{\hat{OS}},U\times Y)$ appears gradual amplification,

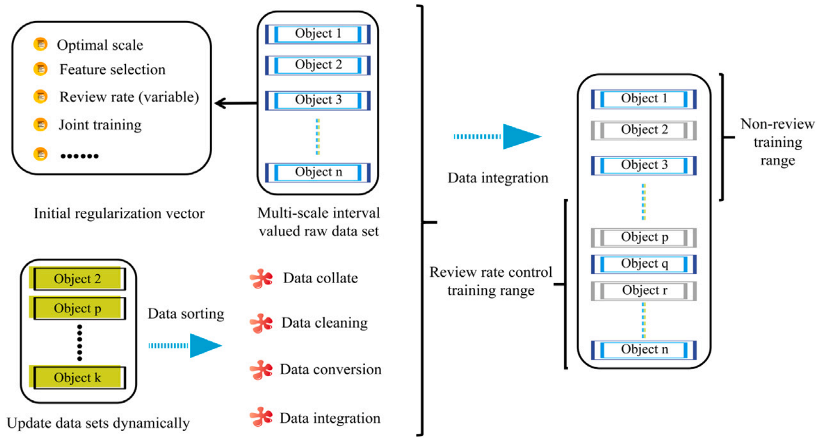

Fig. 4. Data organization and arrangement with dynamic object decrease.

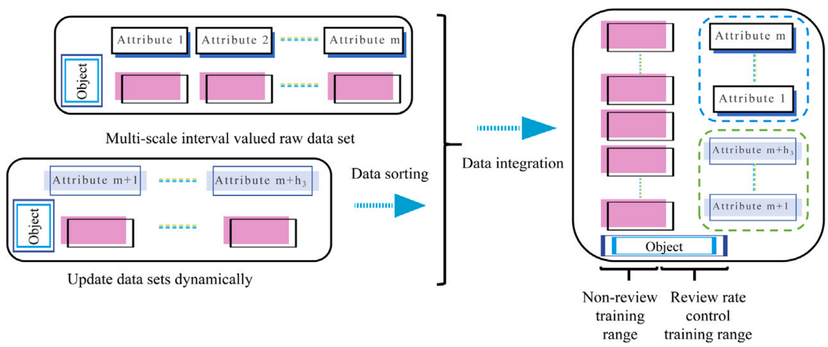

Fig. 5. Data organization and arrangement with dynamic attribute increase.

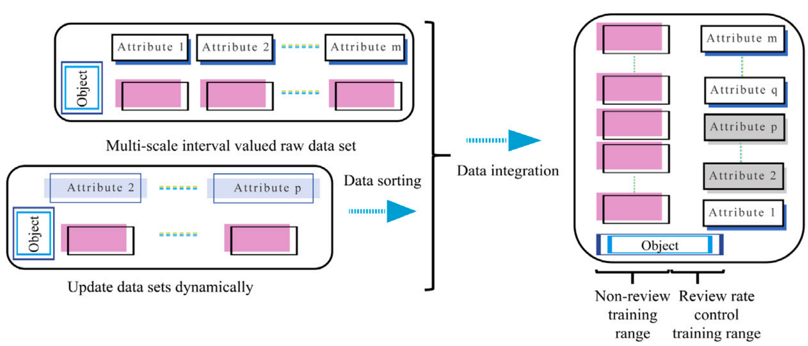

Fig. 6. Data organization and arrangement with dynamic attribute decrease.

where $\hat{S}_j^{\hat{OS}} = (O, (C \cup A_1)_j^{\hat{OS}} \cup \{d\})$, $U = O \otimes Y$ and $Y = O$. Then, only one half-search in $FSACS = [ACS(\hat{S}_{m+h_3}^{\hat{OS}}, U \times Y), ACS(\hat{S}_{m+h_3-1}^{\hat{OS}}, U \times Y), \dots, ACS(\hat{S}_0^{\hat{OS}}, U \times Y)]$ is necessary to obtain the order in which the last $ACS(\hat{S}_j^{\hat{OS}}, U \times Y) = 0$, and this order can furnish implications for $\hat{FS}$.

Example 2. $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1,2,3,4; l = 1,2\} \cup \{d\})$ is the information system corresponding to part of Table 2, S is consistent apparently. When $\varsigma^1 = 0.8$ and $\varsigma^2 = 0.5$, $RV(S) = (2, \{c_2, c_3, c_4\}, 10, 4, 0, 0, 0.7, 0, null)$. Subsequently two additional attributes $\{c_5, c_6\}$ are appended to S and positioned before the original attributes. The corresponding values for these attributes can be found in Table 2. Assuming $t = 2$, then $D_1 = \{o_1, o_2, o_5, o_6, o_9, o_{10}\}$ and $D_2 = \{o_3, o_4, o_7, o_8\}$. In the light of half-search algorithm and $ACS(\hat{S}^1, U \times Y) = 0$ and $ACS(\hat{S}^2, U \times Y) = 0$, it is toilless to verify $\hat{OS} = 2$. In

Algorithm 3: Feature Selection for Dynamic Attribute Increase ($AtDI - FS$).

Input: A $D\text{-}MIvFD$ $S = (O, C \cup \{d\})$, $RV(S)$, A_1.
Output: New optimal scale $\hat{OS}$, new feature selection set $\hat{FS}$, new $RV(S)$.
Calculating $\{D_1, D_2, ..., D_t\}$ about the updated dataset
$left = OS, right = L$ /*Finding new optimal scale $\hat{OS}$.*/
Steps 3-11 of Algorithm 1
$\hat{OS} = left - 1$
$left = 0, right = m + h_3$ /*Finding new feature selection set $\hat{FS}$.*/
Steps 18-26 of Algorithm 1
$\hat{FS} = C_{left}$
return $\hat{OS}$, $\hat{FS}$, $RV(S) = (\hat{OS}, \hat{FS}, n, m + h_3, 0, +h_3, Y, 0, null)$

Table 3
Time complexity of the proposed algorithms.

Step	Function	Best	Worst	Average
1	Calculating decision class	$O(y)$	$O(yt)$	$O(yt)$
2–11	Finding new optimal scale $\hat{OS}$	$O(y)$	$O(uyalog_2(OS))$	$O(uyalog_2(OS))$
5	Calculating $AFCI(\hat{S}^l, U \times Y)$	$O(y)$	$O(uya)$	$O(uya)$
5	Calculating $ACS(\hat{S}^l, U \times Y)$	$O(1)$	$O(1)$	$O(1)$
13–26	Finding new feature selection set $\hat{FS}$	$O(y)$	$O(uyalog_2(a+1))$	$O(uyalog_2(a+1))$
20	Calculating $AFCI(\hat{S}_j^{\hat{OS}}, U \times Y)$	$O(y)$	$O(uya)$	$O(uya)$
20	Calculating $ACS(\hat{S}_j^{\hat{OS}}, U \times Y)$	$O(1)$	$O(1)$	$O(1)$

addition, due to $ACS(\hat{S}_0^2, U\times Y) = 1$, $ACS(\hat{S}_1^2, U\times Y) = 1$, $ACS(\hat{S}_2^2, U\times Y) = 1$, $ACS(\hat{S}_3^2, U\times Y) = 0$, $ACS(\hat{S}_4^2, U\times Y) = 0$, $ACS(\hat{S}_5^2, U\times Y) = 0$ and $ACS(\hat{S}_6^2, U\times Y) = 0$, it can be deduced that $\hat{FS} = \{c_2, c_3, c_4\}$. To sum up, it can be known that $RV(S) = (2, \{c_2, c_3, c_4\}, 10, 6, 0, +2, 0.7, 0, null)$.

4.4. Incremental feature selection for dynamic attribute decrease

Remove h_4 attributes from the initial dataset, with these h_4 attributes collectively constituting A_2, $A_2 \subseteq C$, as well as the organization and arrangement of the data are revealed in Fig. 6. $\forall o \in O$, $AC_{C^l}^{\varsigma^l}(o)$, $\hat{AC}_{(C-A_2)^l}^{\varsigma^l}(o)$, then $AC_{C^l}^{\varsigma^l}(o) \subseteq \hat{AC}_{(C-A_2)^l}^{\varsigma^l}(o)$. Ulteriorly when $S^l = (O, C^l \cup \{d\})$ is inconsistent, then $\hat{S}^l = (O, (C-A_2)^l \cup \{d\})$ is also inconsistent. It is apparent that $\hat{OS}$ can be found within $\{1, 2, \ldots, OS\}$. To expedite the training process, $\hat{OS}$ will be sought within the range of information table managed by Y, i.e. $U = O \otimes Y$ and $Y = O$, $\hat{AC}_{(C-A_2)^l}^{\varsigma^l}(o)$, $\forall o \in U$. Additionally, as the scale becomes coarser, the inconsistency of information table gradually amplifies, corresponding to the monotonically increasing $ACS(\hat{S}^l, U \times Y)$, where $\hat{S}^l = (O, (C-A_2)^l \cup \{d\})$, $l \in \{1, 2, \ldots, OS\}$, $U = O \otimes Y$ and $Y = O$, which constitutes $OSACS = [ACS(\hat{S}^1, U\times Y), ACS(\hat{S}^2, U\times Y), \ldots, ACS(\hat{S}^{OS}, U\times Y)]$. Subsequently, a half-search is conducted on $OSACS$ to chase down the sequence of final $ACS(\hat{S}^l, U \times Y) = 0$, hence notarizing $\hat{OS}$. Subsequently, it is imperative to search for $\hat{FS}$ within $\{C_{m-h_4}, C_{m-h_4-1}, C_{m-h_4-2}, \ldots, C_1, C_0\}$. As the feature subset narrows, $ACS(\hat{S}_j^{\hat{OS}}, U \times Y)$ displays gradual magnification, where $\hat{S}_j^{\hat{OS}} = (O, (C-A_2)_j^{\hat{OS}} \cup \{d\})$, $U = O \otimes Y$ and $Y = O$. Then, only one half-search in $FSACS = [ACS(\hat{S}_{m-h_4}^{\hat{OS}}, U \times Y), ACS(\hat{S}_{m-h_4-1}^{\hat{OS}}, U \times Y), \ldots, ACS(\hat{S}_0^{\hat{OS}}, U \times Y)]$ is needful to gain the order in which the last $ACS(\hat{S}_j^{\hat{OS}}, U \times Y) = 0$, $\hat{FS}$ can be deduced from the order spontaneously. Assuming number of attributes $= a$, $|U| = u$ and $|Y| = y$, the time complexities of these four algorithms are presented in Table 3.

Algorithm 4: Feature Selection for Attribute Decrease ($AtDD-FS$).

Input: A $D\text{-}MI\upsilon FD$ $S = (O, C \cup \{d\})$, $RV(S)$, A_2, $|A_2| = h_4$, $A_2 \subseteq C$.
Output: New optimal scale $\hat{OS}$, new feature selection set $\hat{FS}$, new $RV(S)$.

1: Calculating $\{D_1, D_2, ..., D_t\}$ about the updated dataset
2: $left = 1, right = OS$ /*Finding new optimal scale $\hat{OS}$.*/
3: Steps 3-11 of Algorithm 1
4: $\hat{OS} = left - 1$
5: $left = 0, right = m - h_4$ /*Finding new feature selection set $\hat{FS}$.*/
6: Steps 18-26 of Algorithm 1
7: $\hat{FS} = C_{left}$
8: **return** $\hat{OS}$, $\hat{FS}$, $RV(S) = (\hat{OS}, \hat{FS}, n, m - h_4, 0, -h_4, Y, 0, null)$

5. Experimental analysis

In this section, the practicality and efficacy of the four algorithms are validated through experimental procedures. All experimental hardware setups are configured as Windows 11, Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz and 16.0 GB memory. The software environment for executing the algorithms is Python 3.7. Twelve datasets from the University of California, Irvine (https://archive.ics.uci.edu/) are selected to verify the performance of the proposed algorithms, and the details of twelve datasets are displayed in Table 4. The technique expounded in [30] is implemented to convert the standard dataset into the $MI\upsilon FD$, where each feature possesses five scales.

Then a series of constraints will be imposed on these static tables to enable them to execute various update patterns. In this study, the dataset's object set and attribute set are segmented into four equivalent portions, each representing 25% of the whole. These segments are denoted as Ω_1, Ω_2, Ω_3 and Ω_4 for the object set, and Ψ_1, Ψ_2, Ψ_3 and Ψ_4 for the attribute set. Initially, the object set comprises $\Omega_1 \cup \Omega_2$, while the attribute set consists of $\Psi_3 \cup \Psi_4$. The object set becomes $\Omega_1 \cup \Omega_2 \cup \Omega_3$ when the objects are aggrandized, and Ω_1 when the objects are decreased. The attribute set is transformed to $\Psi_2 \cup \Psi_3 \cup \Psi_4$ when the attribute increases, and to Ψ_4 when the attribute decreases (see Table 7).

5.1. Experimental design

Based on the findings [30] of prior experiments, the parameters illuminated in this study are established as follows, $\zeta_1 = \frac{1}{2}, \zeta_2 = \frac{1}{2}, \varsigma^1 = 0.9, \varsigma^2 = 0.7, \varsigma^3 = 0.5, \varsigma^4 = 0.3, \varsigma^5 = 0.1$. In [30], extensive experimental evidence has demonstrated that optimal accuracy and efficiency, in a general context, can be attained when the parameters are configured as above specified. Where ζ_1 and ζ_2 represent the two parameters of PI, while ς^1, ς^2, ς^3, ς^4 and ς^5 denote the associated thresholds for the five scales. Initially, the outcomes of the initial dataset are derived utilizing the technique elucidated in [30], as manifested in Table 5. Then experiments are conducted on twelve datasets, and the results are documented in Table 6 ∼ Table 8. (Table Notation Description:PT-Consistency processing time, TSS-Training set size, T_1-Finding optimal scale time, T_2-Finding feature selection time)

Based on the exploration of algorithm performance and review rate, it is distinct that higher review rates generally correspond to broader learning scopes and ameliorated algorithm accuracy. Nevertheless, there exists a specific scenario where an increase in learning scope may introduce more abnormal data, leading to a decrease in accuracy. Furthermore, alterations in algorithm accuracy are observed with the addition or removal of attributes. It is logical to deduce that as attributes are augmented, the algorithm's perception of the object becomes more comprehensive, resulting in higher classification accuracy, or else the accuracy levels will be dramatically impacted.

5.2. Contrast experiment

In this subsection, seven relevant algorithms in this field, individually namely $AE\text{-}RAR$ [31], $HKCMI$ [32], $IGUFS(\alpha = 0.5, \beta = 0.1)$ [33], $INF\text{-}UFS$ [34], UM [35], $GDGRN(\alpha = 0.7, \beta = 0.001)$ [24] and $UHDFK$ [2] are compared with the algorithms of this study. These comparison algorithms are detailed and classified in Table 13, and all other parameters are initialized with default values. It is important to note that the incremental feature selection model presented in this study, along with several other comparative algorithms, can fundamentally be classified as feature selection systems. These systems are capable of executing feature selection processes and producing

Table 4
Details of the datasets.

Order	Datasets	Abbreviations	Objects	Features
1	Toxicity	Toxicity	171	**1203**
2	Period Changer	PC	90	1177
3	DARWIN	DARWIN	174	451
4	Tunadromd	Tunadromd	4465	242
5	NATICUSdroid Dataset	NATICUSdroid	29 333	86
6	Predict Students Dropout and Academic Success	PSDAS	4424	36
7	Tarvel Review Ratings	TRR	5456	24
8	Glioma Grading Clinical and Mutation Features	GGCMF	839	23
9	Diferentiated Thyroid Cancer Recurrence	DTCR	383	16
10	Shill Bidding Dataset	SBD	6321	13
11	Accelerometer Gyro Mobile Phone Dataset	AGMPD	31 991	8
12	Sepsis Survival Minimal Clinical Records	SSMCR	**110 341**	3

Table 5
Experimental results on the original datasets.

Datasets	PT(s)	TSS	T_1(s)	OS	T_2(s)	FS
DTCR	0.88	190	**0.02**	1	**0.70**	C_7
Toxicity	0.11	85	**0.41**	5	2.67	C_{360}
PC	0.03	45	**0.24**	5	1.01	C_{252}
DARWIN	0.13	87	1.09	5	5.58	C_{201}
GGCMF	7.34	68	**0.01**	1	**0.27**	C_9
PSDAS	55.59	1824	**0.38**	1	32.13	C_{21}
Tunadromd	286.32	2231	1436.98	5	1.00	C_1
TRR	53.24	1797	**0.11**	1	**0.48**	C_{12}
SBD	52.52	1807	**0.06**	1	**0.02**	C_5
NATICUSdroid	8533.04	11 230	4.57	1	11 782.64	C_{28}
AGMPD	1810.22	15 995	2106.50	5	48.75	C_1
SSMCR	13 894.54	0	0.00	5	0.00	C_1

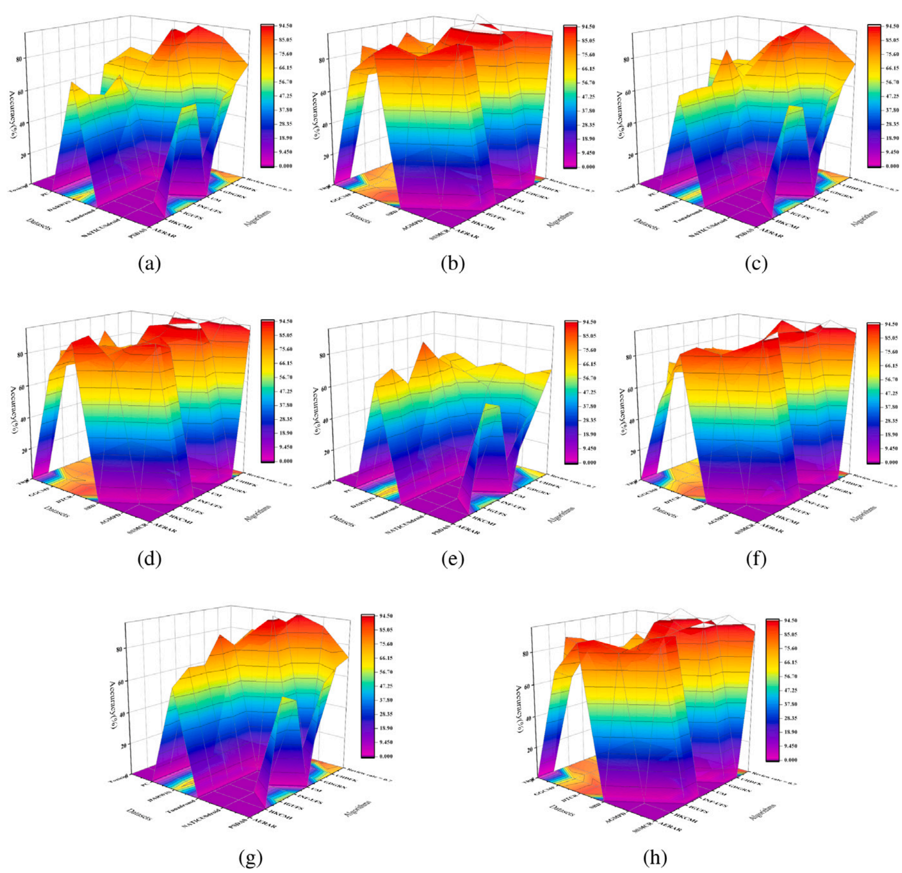

Fig. 7. Comparison of accuracy about different algorithms under four classifiers. Separately (a) and (b) are the accuracy of the KNN classifier, (c) and (d) are the accuracy of the SVM classifier, (e) and (f) are the accuracy of the Bayes classifier, (g) and (h) are the accuracy of the Random Forest classifier.

relevant outcomes following four distinct updates to the dataset: Object Increase, Object Decrease, Attribute Increase and Attribute Decrease.

Following a data update, each system is assigned an accuracy value and a processing time for the feature selection outcomes. To conduct

Table 6
Experimental results of different algorithms on twelve datasets when review rate=0.3.

Algorithms	Datasets	PT(s)	TSS	T_1(s)	$\hat{OS}$	T_2(s)	$\hat{FS}$
ObDI-FS	DTCR	0.47	280	**0.00**	1	**0.48**	C_7
	Toxicity	0.14	128	**0.48**	5	3.17	C_{360}
	PC	0.02	67	**0.18**	5	0.93	C_{261}
	DARWIN	0.14	130	1.16	5	2.04	C_{219}
	GGCMF	8.61	343	**0.00**	1	0.35	C_{11}
	PSDAS	37.90	2933	**0.00**	1	**0.00**	C_{21}
	Tunadromd	263.86	3348	2.92	1	928.68	C_{55}
	TRR	89.57	3203	**0.00**	1	**0.00**	C_{12}
	SBD	64.11	3745	**0.00**	1	**0.00**	C_5
	NATICUSdroid	9762.27	13 532	**0.00**	1	1149.52	C_{42}
	AGMPD	2238.09	11 297	**0.28**	1	4.67	C_3
	SSMCR	17 360.70	38 572	0.00	5	0.00	C_1
ObDD-FS	DTCR	0.03	95	**0.02**	1	**0.09**	C_7
	Toxicity	0.02	43	**0.03**	5	**0.17**	C_{16}
	PC	0.00	22	**0.01**	5	**0.04**	C_{146}
	DARWIN	0.01	43	**0.08**	5	**0.56**	C_{129}
	GGCMF	0.33	156	**0.02**	1	**0.02**	C_{10}
	PSDAS	4.31	1069	**0.11**	1	3.94	C_{19}
	Tunadromd	21.93	1115	110.62	5	**0.08**	C_1
	TRR	4.94	1225	**0.06**	1	**0.09**	C_{12}
	SBD	4.19	1385	**0.06**	1	**0.16**	C_5
	NATICUSdroid	650.11	7332	1866.20	5	3.19	C_1
	AGMPD	131.93	7997	160.18	5	4.24	C_1
	SSMCR	1141.73	19 285	0.00	5	0.00	C_1
AtDI-FS	DTCR	0.19	190	**0.23**	2	**0.34**	C_5
	Toxicity	0.03	85	**0.14**	5	2.02	C_{221}
	PC	0.02	45	**0.06**	5	**0.48**	C_{186}
	DARWIN	0.05	87	**0.35**	5	2.27	C_{201}
	GGCMF	2.81	334	**0.08**	1	**0.11**	C_{17}
	PSDAS	9.87	2205	**0.58**	1	10.45	C_{30}
	Tunadromd	104.44	2231	639.18	5	**0.30**	C_1
	TRR	17.39	2688	**0.67**	1	8.45	C_{18}
	SBD	17.71	3146	**0.16**	1	**0.14**	C_7
	NATICUSdroid	3143.88	14 412	7.89	1	14 290.70	C_{65}
	AGMPD	658.79	15 995	838.18	5	14.81	C_1
	SSMCR	5064.72	38 602	**0.66**	4	**0.36**	C_3
AtDD-FS	DTCR	0.11	184	**0.00**	1	**0.08**	C_1
	Toxicity	0.03	85	**0.14**	5	**0.93**	C_{221}
	PC	0.00	45	**0.06**	5	**0.44**	C_{186}
	DARWIN	0.03	87	**0.23**	4	**0.37**	C_{80}
	GGCMF	1.57	301	**0.00**	1	**0.08**	C_4
	PSDAS	9.30	1755	**0.00**	1	**0.03**	C_{11}
	Tunadromd	61.91	2231	214.45	5	**0.31**	C_1
	TRR	16.81	2070	**0.00**	1	**0.17**	C_6
	SBD	14.64	2551	**0.00**	1	**0.02**	C_2
	NATICUSdroid	2274.67	13 210	**0.00**	1	7.45	C_6
	AGMPD	428.50	15 995	516.88	5	17.39	C_1
	SSMCR	3020.00	38 571	0.00	5	0.00	C_1

a thorough assessment of the system's performance, the overall accuracy is calculated as the mean of the accuracy values obtained from four updates. Likewise, the processing times associated with the four updates are averaged to represent the system's time consumption. The comparative analysis of running times is presented in Table 11 and illustrated in Fig. 8.

Furthermore, this study employs K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes ($Bayes$), and Random Forest (RF) algorithms to assess the accuracy of each system, as illustrated in Table 9, Table 10 and Fig. 7. The models are evaluated under varying review rates, specifically Review Rate=0.3 ($\Upsilon = 0.3$), Review Rate=0.5 ($\Upsilon = 0.5$), and Review Rate=0.7 ($\Upsilon = 0.7$).

Through the experimental results, it is not difficult to find that the proposed model has a significant improvement in efficiency, and can deal with the feature selection task of large-scale data sets under the condition of not running out of memory. Subsequently, these algorithms will undergo thorough statistical analyses to yield substantiating estimation findings.

5.3. Statistical analysis

In this subsection, the Friedman test, a nonparametric statistical approach, will be employed to ascertain whether there exist statistically significant disparities in the average performance of the aforementioned algorithms across twelve distinct datasets. The Friedman test is a methodology utilized to compare ranking variances among multiple related sample groups. Prior to the Friedman test, we hypothesize that there is no substantial difference in accuracy performance among all the algorithms. The relevant formulas of Friedman statistics are presented as follows.

$$\chi_F^2 = \frac{12N}{k(k+1)}\left(\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right), \quad F_F = \frac{(N-1)\,\chi_F^2}{N(k-1)-\chi_F^2}. \tag{11}$$

Then χ^2 and $P < 0.05$ are derived through computation, it is evident that the calculation results indicate that the original hypothesis is invalid. The Friedman test can merely be employed to ascertain whether there exists a significant disparity among the experimental

Table 7
Experimental results of different algorithms on twelve datasets when review rate=0.5.

Algorithms	Datasets	PT(s)	TSS	T_1(s)	$\hat{OS}$	T_2(s)	$\hat{FS}$
ObDI-FS	DTCR	0.63	277	**0.00**	1	**0.53**	C_7
	Toxicity	0.18	128	**0.59**	5	4.15	C_{360}
	PC	0.04	67	**0.46**	5	2.70	C_{261}
	DARWIN	0.18	130	1.51	5	3.72	C_{219}
	GGCMF	10.72	274	**0.00**	1	**0.42**	C_{11}
	PSDAS	47.28	2848	**0.00**	1	**0.00**	C_{21}
	Tunadromd	335.39	3348	2.80	1	949.74	C_{55}
	TRR	90.97	3061	**0.00**	1	**0.00**	C_{12}
	SBD	79.94	3464	**0.00**	1	**0.00**	C_5
	NATICUSdroid	12 279.52	11 616	**0.00**	1	1366.76	C_{42}
	AGMPD	2734.31	8159	6.08	1	5.36	C_3
	SSMCR	21 279.81	27 551	0.02	5	0.00	C_1
ObDD-FS	DTCR	0.06	95	**0.02**	1	**0.08**	C_7
	Toxicity	0.01	43	**0.06**	5	**0.25**	C_{16}
	PC	0.01	22	**0.02**	5	**0.05**	C_{252}
	DARWIN	0.01	43	**0.07**	5	**0.20**	C_{147}
	GGCMF	0.95	122	**0.02**	1	**0.07**	C_{10}
	PSDAS	4.31	1041	**0.09**	1	2.31	C_{21}
	Tunadromd	36.83	1115	180.89	5	**0.12**	C_1
	TRR	7.81	1196	**0.09**	1	**0.47**	C_{12}
	SBD	7.46	1265	**0.05**	1	**0.05**	C_5
	NATICUSdroid	1032.14	7332	3108.99	5	5.36	C_1
	AGMPD	223.65	7997	266.37	5	6.69	C_1
	SSMCR	1923.13	13 775	0.00	5	0.00	C_1
AtDI-FS	DTCR	0.30	190	**0.42**	2	**0.51**	C_5
	Toxicity	0.06	85	**0.21**	5	1.47	C_{360}
	PC	0.02	45	**0.13**	5	**0.91**	C_{186}
	DARWIN	0.11	87	**0.60**	5	2.96	C_{201}
	GGCMF	4.73	280	**0.06**	1	**0.27**	C_{17}
	PSDAS	29.14	2203	**0.64**	1	34.55	C_{30}
	Tunadromd	171.83	2231	1043.37	5	**0.61**	C_1
	TRR	28.22	2679	**0.50**	1	1.40	C_{18}
	SBD	27.85	3137	**0.19**	1	**0.08**	C_7
	NATICUSdroid	4649.75	14 256	7.70	1	23 742.60	C_{65}
	AGMPD	1089.73	15 995	1439.31	5	27.23	C_1
	SSMCR	8472.05	27 603	**0.48**	4	**0.27**	C_3
AtDD-FS	DTCR	0.17	181	**0.00**	1	**0.16**	C_1
	Toxicity	0.05	85	**0.25**	4	**0.69**	C_{53}
	PC	0.01	45	**0.11**	5	**0.74**	C_{186}
	DARWIN	0.06	87	**0.32**	4	**0.43**	C_{80}
	GGCMF	2.58	223	**0.00**	1	**0.10**	C_4
	PSDAS	15.97	1447	**0.00**	1	**0.09**	C_{11}
	Tunadromd	102.86	2231	340.32	5	**0.56**	C_1
	TRR	27.22	1679	**0.00**	1	**0.16**	C_6
	SBD	24.04	2143	**0.00**	1	**0.00**	C_2
	NATICUSdroid	3879.48	12 276	**0.00**	1	12.16	C_6
	AGMPD	756.76	15 995	851.25	5	26.86	C_1
	SSMCR	4739.54	27 551	0.00	5	0.00	C_1

outcomes of multiple models. However, it fails to identify whether there is a difference between any two specific models. Hence, the Nemenyi test ought to be conducted in the subsequent step. The values of q_α and CD are derived through reference to the table and subsequent calculations. When the Average Rank Difference (ARD) of the two algorithms exceeds CD, the performance of the two algorithms is significantly distinct. From the foregoing, it is not arduous to deduce that the proposed method surpasses multiple traditional models in terms of performance (see Table 12).

6. Conclusions

In this study, we present four incremental feature selection algorithms that leverage RV and review rate to address the challenge of real-time feature selection in $D\text{-}MIvFD$. Inspired by machine learning principles of regularization and replay, these algorithms enhance the speed, adaptability, and effectiveness of feature selection processes. Our experimental results indicate that these algorithms can rapidly identify optimal feature subsets, reducing storage needs while sustaining high classification accuracy.

This research offers significant insights for future exploration and practical application in related fields. Nonetheless, this paper presents certain limitations, including the method of data update, the type of data input and output, as well as the adaptability in intricate decision-making contexts, such as multi-label decision making scenarios, all of which require improvement. Future work could delve into deeper integration with machine learning frameworks and other disciplines, aiming to broaden the algorithms' applicability to complex decision-making scenarios.

Given the increasing complexity and diversity of real-world data, developing effective implementation strategies for the proposed methods in practical environments is a critical direction for future research. This will ensure the robustness and scalability of the algorithms in the face of evolving data landscapes.

Table 8
Experimental results of different algorithms on twelve datasets when review rate=0.7.

Algorithms	Datasets	PT(s)	TSS	T_1(s)	$\hat{OS}$	T_2(s)	$\hat{FS}$
ObDI-FS	DTCR	0.69	277	**0.00**	1	**0.63**	C_7
	Toxicity	0.20	128	**0.68**	5	4.77	C_{360}
	PC	0.05	67	**0.49**	5	3.28	C_{261}
	DARWIN	0.21	130	1.82	5	5.66	C_{219}
	GGCMF	12.29	198	**0.00**	1	**0.02**	C_{11}
	PSDAS	57.14	2760	**0.00**	1	**0.00**	C_{21}
	Tunadromd	412.19	3348	2.80	1	1261.22	C_{55}
	TRR	103.41	2843	**0.00**	1	**0.00**	C_{12}
	SBD	96.94	3185	**0.00**	1	**0.00**	C_5
	NATICUSdroid	14 888.47	9672	**0.00**	1	1500.96	C_{42}
	AGMPD	3473.24	5063	1.50	1	5.48	C_3
	SSMCR	25 464.62	16 531	0.00	5	0.00	C_1
ObDD-FS	DTCR	0.08	95	**0.02**	1	**0.17**	C_7
	Toxicity	0.03	43	**0.06**	5	**0.33**	C_{16}
	PC	0.00	22	**0.05**	5	**0.11**	C_{252}
	DARWIN	0.01	43	**0.09**	5	**0.36**	C_{147}
	GGCMF	1.26	86	**0.02**	1	**0.07**	C_{10}
	PSDAS	5.77	1013	**0.09**	1	4.59	C_{21}
	Tunadromd	50.79	1115	245.28	5	**0.17**	C_1
	TRR	10.97	1082	**0.06**	1	**0.44**	C_{12}
	SBD	9.58	1130	**0.03**	1	**0.33**	C_5
	NATICUSdroid	1577.08	7332	4700.42	5	7.42	C_1
	AGMPD	309.21	7997	382.17	5	9.20	C_1
	SSMCR	2696.14	8266	0.00	5	0.00	C_1
AtDI-FS	DTCR	0.41	190	**0.02**	1	**0.73**	C_7
	Toxicity	0.08	85	**0.30**	5	2.33	C_{360}
	PC	0.03	45	**0.16**	5	1.24	C_{252}
	DARWIN	0.08	87	**0.87**	5	5.11	C_{201}
	GGCMF	6.44	221	**0.05**	1	**0.15**	C_{17}
	PSDAS	21.27	2202	**0.27**	1	28.17	C_{30}
	Tunadromd	239.24	2231	1457.15	5	**0.69**	C_1
	TRR	38.47	2645	**0.83**	1	1.47	C_{18}
	SBD	38.69	3126	**0.14**	1	**0.06**	C_7
	NATICUSdroid	6560.44	14 095	7.73	1	32 177.23	C_{65}
	AGMPD	1492.14	15 995	1946.47	5	34.09	C_1
	SSMCR	12 163.07	16 604	**0.30**	4	**0.14**	C_3
AtDD-FS	DTCR	0.22	177	**0.00**	1	**0.20**	C_1
	Toxicity	0.08	85	**0.51**	4	1.18	C_{53}
	PC	0.01	45	**0.09**	5	**0.39**	C_{252}
	DARWIN	0.08	87	**0.41**	4	**0.84**	C_{80}
	GGCMF	3.40	144	**0.00**	1	**0.02**	C_6
	PSDAS	22.91	1142	**0.00**	1	**0.03**	C_{11}
	Tunadromd	142.38	2231	481.88	5	**0.69**	C_1
	TRR	38.47	1244	**0.00**	1	**0.20**	C_6
	SBD	31.83	1709	**0.00**	1	**0.00**	C_2
	NATICUSdroid	5098.04	11 346	**0.00**	1	14.94	C_6
	AGMPD	1062.14	15 995	1214.63	5	42.27	C_1
	SSMCR	6330.35	16 531	0.00	5	0.00	C_1

Table 9
Comparison of accuracy (%) about different algorithms under four classifiers.

Classifiers	Datasets	AERAR	HKCMI	IGUFS	INF-UFS	UM	$\Upsilon = 0.3$	$\Upsilon = 0.5$	$\Upsilon = 0.7$
KNN	Toxicity	–	–	–	–	–	**75.91**	73.02	73.02
	PC	–	–	–	–	–	66.27	**72.62**	69.84
	DARWIN	71.43	61.42	58.58	68.36	50.82	68.53	70.14	**74.86**
	Tunadromd	–	–	–	–	–	94.03	94.03	**94.03**
	NATICUSdroid	–	–	–	–	–	88.56	88.56	**88.56**
	PSDAS	–	–	60.78	60.36	–	70.13	73.58	**73.58**
	TRR	–	–	**83.94**	79.35	74.32	75.79	75.79	75.79
	GGCMF	**75.01**	70.68	71.13	66.52	51.64	61.26	62.06	62.06
	DTCR	88.26	91.54	75.91	74.93	74.91	99.35	99.35	**99.92**
	SBD	–	86.14	81.45	76.34	85.68	90.82	90.82	**90.82**
	AGMPD	–	–	**94.28**	91.08	–	92.86	92.86	92.86
	SSMCR	–	–	–	–	–	92.18	92.18	**92.18**

(*continued on next page*)

Table 9 (*continued*).

SVM	Toxicity	–	–	–	–	–	59.80	64.61	**64.61**
	PC	–	–	–	–	–	61.90	61.90	**65.48**
	DARWIN	63.40	64.70	64.29	**84.20**	63.26	65.35	68.57	69.44
	Tunadromd	–	–	–	–	–	94.41	94.41	**94.41**
	NATICUSdroid	–	–	–	–	–	87.21	87.21	**87.21**
	PSDAS	–	–	62.37	58.55	–	74.37	74.49	**74.49**
	TRR	–	–	**77.40**	74.95	65.43	71.39	71.39	71.39
	GGCMF	70.15	72.92	**81.70**	81.25	58.48	63.45	67.74	67.74
	DTCR	91.21	93.17	78.86	79.82	79.48	96.56	96.56	**98.13**
	SBD	–	80.45	87.51	87.79	89.72	92.20	92.20	**92.20**
	AGMPD	–	–	91.34	87.27	–	98.34	98.34	**98.34**
	SSMCR	–	–	–	–	–	92.21	92.21	**92.21**
Bayes	Toxicity	–	–	–	–	–	42.77	45.65	**45.65**
	PC	–	–	–	–	–	63.50	69.05	**69.05**
	DARWIN	68.55	74.58	62.89	**86.35**	74.07	53.51	54.86	61.73
	Tunadromd	–	–	–	–	–	67.43	67.43	**67.43**
	NATICUSdroid	–	–	–	–	–	62.59	62.59	**62.59**
	PSDAS	–	–	61.41	57.02	–	**68.88**	68.49	68.49
	TRR	–	–	**71.70**	67.63	56.19	64.94	64.94	64.94
	GGCMF	53.04	70.24	68.30	**81.25**	58.48	54.81	55.61	55.61
	DTCR	85.99	87.97	78.21	78.51	79.48	93.31	93.31	**94.59**
	SBD	–	86.09	80.56	87.74	**88.77**	88.50	88.50	88.50
	AGMPD	–	–	94.35	97.81	–	98.34	98.34	**98.34**
	SSMCR	–	–	–	–	–	92.21	92.21	**92.21**
RF	Toxicity	–	–	–	–	–	64.32	65.28	**65.28**
	PC	–	–	–	–	–	77.38	77.38	**80.95**
	DARWIN	62.16	71.52	70.57	**87.06**	78.48	72.66	73.27	77.14
	Tunadromd	–	–	–	–	–	95.11	95.11	**95.11**
	NATICUSdroid	–	–	–	–	–	89.13	89.13	**89.13**
	PSDAS	–	–	61.72	56.71	–	71.93	72.16	**72.16**
	TRR	–	–	**85.42**	82.90	61.02	78.82	78.82	78.82
	GGCMF	71.26	65.18	**80.90**	80.21	58.48	64.29	69.97	69.97
	DTCR	90.56	90.24	75.25	79.16	76.86	99.35	99.35	**99.92**
	SBD	–	85.28	84.08	89.40	88.93	92.59	92.59	**92.59**
	AGMPD	–	–	88.49	93.55	–	99.07	99.07	**99.07**
	SSMCR	–	–	–	–	–	90.87	90.87	**90.87**

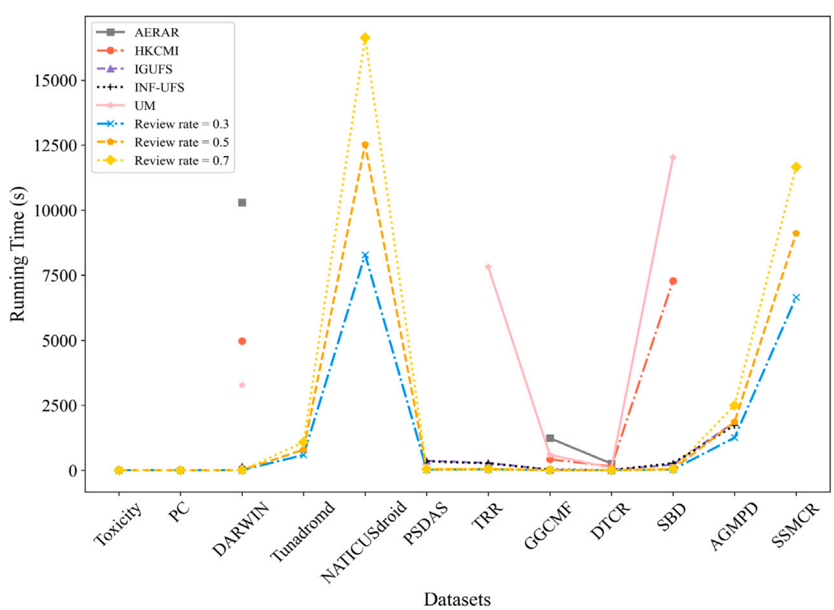

Fig. 8. Comparison of the running time about different algorithms.

Table 10
Comparison of accuracy (%) about different algorithms under four classifiers.

Classifiers	Datasets	GDGRN	UHDFK	$\Upsilon = 0.3$	$\Upsilon = 0.5$	$\Upsilon = 0.7$
KNN	Toxicity	64.95	68.27	**75.91**	73.02	73.02
	PC	**73.24**	62.38	66.27	72.62	69.84
	DARWIN	71.36	66.70	68.53	70.14	**74.86**
	Tunadromd	**95.56**	90.72	94.03	94.03	94.03
	NATICUSdroid	–	–	**88.56**	88.56	88.56
	PSDAS	58.54	64.39	70.13	**73.58**	73.58
	TRR	**87.29**	79.43	75.79	75.79	75.79
	GGCMF	**75.37**	67.36	61.26	62.06	62.06
	DTCR	94.78	83.10	99.35	99.35	**99.92**
	SBD	**98.95**	80.19	90.82	90.82	90.82
	AGMPD	**96.37**	88.72	92.86	92.86	92.86
	SSMCR	–	–	**92.18**	92.18	92.18
SVM	Toxicity	**69.52**	63.29	59.80	64.61	64.61
	PC	**68.93**	65.49	61.90	61.90	65.48
	DARWIN	53.69	**82.34**	65.35	68.57	69.44
	Tunadromd	91.23	86.93	**94.41**	94.41	94.41
	NATICUSdroid	–	–	**87.21**	87.21	87.21
	PSDAS	58.98	63.11	74.37	**74.49**	74.49
	TRR	**84.92**	70.86	71.39	71.39	71.39
	GGCMF	67.22	61.60	63.45	**67.74**	67.74
	DTCR	92.16	78.86	96.56	96.56	**98.13**
	SBD	**95.29**	84.39	92.20	92.20	92.20
	AGMPD	97.74	83.38	**98.34**	98.34	98.34
	SSMCR	–	–	**92.21**	92.21	92.21
Bayes	Toxicity	**47.12**	42.55	42.77	45.65	45.65
	PC	63.36	68.92	63.5	**69.05**	69.05
	DARWIN	**66.34**	60.76	53.51	54.86	61.73
	Tunadromd	24.61	53.77	**67.43**	67.43	67.43
	NATICUSdroid	–	–	**62.59**	62.59	62.59
	PSDAS	42.91	50.19	**68.88**	68.49	68.49
	TRR	**75.53**	68.87	64.94	64.94	64.94
	GGCMF	52.27	**63.85**	54.81	55.61	55.61
	DTCR	79.96	87.33	93.31	93.31	**94.59**
	SBD	**91.55**	83.51	88.50	88.50	88.50
	AGMPD	94.92	91.71	**98.34**	98.34	98.34
	SSMCR	–	–	**92.21**	92.21	92.21
RF	Toxicity	56.69	59.26	64.32	**65.28**	65.28
	PC	63.12	73.64	77.38	77.38	**80.95**
	DARWIN	**77.30**	74.39	72.66	73.27	77.14
	Tunadromd	**96.91**	91.64	95.11	95.11	95.11
	NATICUSdroid	–	–	**89.13**	89.13	89.13
	PSDAS	60.74	68.44	71.93	**72.16**	72.16
	TRR	72.53	71.67	**78.82**	78.82	78.82
	GGCMF	65.85	64.39	64.29	**69.97**	69.97
	DTCR	93.55	96.19	99.35	99.35	**99.92**
	SBD	90.74	86.20	**92.59**	92.59	92.59
	AGMPD	97.51	94.44	**99.07**	99.07	99.07
	SSMCR	–	–	**90.87**	90.87	90.87

Table 11
Comparison of running times about different algorithms on twelve datasets.

Datasets	AERAR	HKCMI	IGUFS	INF-UFS	UM	$\Upsilon = 0.3$	$\Upsilon = 0.5$	$\Upsilon = 0.7$
Toxicity	–	–	–	–	–	**1.83**	**1.99**	**2.64**
PC	–	–	–	–	–	**0.56**	**1.30**	**1.48**
DARWIN	10 296.71	4971.36	147.29	121.80	3274.29	**1.82**	**2.54**	**3.89**
Tunadromd	–	–	–	–	–	587.17	791.33	1073.62
NATICUSdroid	–	–	–	–	–	8288.97	12 521.12	16 633.18
PSDAS	–	–	360.97	341.71	–	19.12	33.60	35.06
TRR	–	–	286.69	264.45	7815.74	34.54	39.21	48.58
GGCMF	1233.05	426.06	16.43	20.40	594.52	**3.50**	**4.98**	**5.93**
DTCR	254.33	144.43	3.96	3.19	86.47	**0.51**	**0.72**	**0.79**
SBD	–	7285.33	214.76	267.81	12 035.97	**25.30**	**34.92**	**44.40**
AGMPD	–	–	1847.43	1724.19	–	1253.49	1858.40	2493.14
SSMCR	–	–	–	–	–	6647.04	9103.83	11 663.66

Table 12
Results of Friedman tests with different classifiers.

Metrics	KNN	SVM	Bayes	Random Forest
χ^2	52.17	49.43	31.28	56.96
P	0.00	0.00	0.00	0.00

Table 13
Classification and Summary of relevant research literature.

Methods	Categories	Research contents	Disadvantages
Feature selection	Multi-label feature selection [12,23,27] Utilizing the triple nested equivalence class rough set [18] By K-Nearest Neighbor rough set and mutual information [17] Utilizing the sub-tolerance relation class [19] Using all historical data for incremental algorithm [30,31,33] Filtering partial historical data or Utilizing historical results for incremental algorithm [4,16,19] Discarding historical data for incremental algorithm [32,35]	Within the specific context, algorithms have been developed.	Do not fit $D\text{-}MIvFD$ framework. Advanced theoretical ideas or architectures are not combined.
Interval valued data processing	Feature selection [30,31,33,35] Information fusion [28]	These examines effective methods for feature selection, information fusion in $IVDIS$.	These apply primarily to static decision-making scenarios.
Contrast algorithms	Based on α-approximate equal relation ($AERAR$) [31] By fuzzy complementary mutual information ($HKCMI$) [32] Interval valued feature selection based on the graph theory ($IGUFS$) [33] Graph feature filtering ($INF\text{-}UFS$) [34] Based on θ-rough degree (UM) [35] Via granular rectangular neighborhood rough set ($GDGRN$) [24] Using FCM and K-Nearest Neighbor rough set ($UHDFK$) [2]	Based on various metrics, the feature selection under the framework of interval or single valued is realized.	Not applicable to $D\text{-}MIvFD$ framework. Applicability need to be improved.
Our approach	Incremental Feature Selection for $D\text{-}MIvFD$ Datasets	Inspired by the regularization and replay ideas, algorithms are proposed for four update modes.	These update patterns are fixed. Not applicable to multi-label issues.

CRediT authorship contribution statement

Zihan Feng: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation. **Xiaoyan Zhang:** Writing – review & editing, Validation, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgments

The authors would like to thank the Associate Editor and the reviewers for their insightful comments and suggestions.

This work was supported by the National Natural Science Foundation of China (Grant NO.12371465, 62376229) and the Chongqing Natural Science Foundation (Grant NO.CSTB2023NSCQ-MSX1063).

Data availability

No data was used for the research described in the article.

References

[1] Z.Y. Jiang, R.S. Liu, S.Z. Yang, et al., DRNet: Learning a dynamic recursion network for chaotic rain streak removal, Pattern Recognit. (2024) 111004.
[2] W.H. Xu, Y. Zhang, Y.H. Qian, A novel unsupervised feature selection for high-dimensional data based on FCM and k -nearest neighbor rough sets, IEEE Trans. Neural Netw. Learn. Syst. 36 (6) (2025) 10889–10898.
[3] X.W. Xue, J. He, L. Ma, et al., ASF-Net: Robust video deraining via temporal alignment and online adaptive learning, Pattern Recognit. (2024) 110973.
[4] J. Zhao, D.Y. Wu, J.X. Wu, et al., Consistency approximation: Incremental feature selection based on fuzzy rough set theory, Pattern Recognit. (2024) 110652.
[5] Z.H. Liu, J.J. Wang, H. Chen, et al., Relation-specific feature augmentation for unbiased scene graph generation, Pattern Recognit. (2024) 110936.
[6] J.W. Duan, J.Q. Xiong, Y.H. Li, et al., Deep learning based multimodal biomedical data fusion: An overview and comparative review, Inf. Fusion (2024) 102536.
[7] Z.H. Zheng, Y.M. Chen, Q.B. Hou, et al., Zone evaluation: Revealing spatial bias in object detection, IEEE Trans. Pattern Anal. Mach. Intell. 46 (12) (2024) 8636–8651.
[8] R. Ofori-Boateng, M. Aceves-Martins, N. Wiratunga, et al., Towards the automation of systematic reviews using natural language processing, machine learning, and deep learning: a comprehensive review, Artif. Intell. Rev. 57 (8) (2024) 1–60.
[9] D.D. Guo, W.H. Xu, W.P. Ding, et al., Concept-cognitive learning survey: Mining and fusing knowledge from data, Inf. Fusion (2024) 102426.
[10] A. Grimaldi, V. Boutin, S.H. Ieng, et al., A robust event-driven approach to always-on object recognition, Neural Netw. (2024) 106415.
[11] W.H. Xu, Q.Y. Bu, Matrix-based incremental feature selection method using weight-partitioned multigranulation rough set, Inform. Sci. 681 (2024) 121219.
[12] P. Zhang, W.F. Gao, Feature relevance term variation for multi-label feature selection, Appl. Intell. 51 (7) (2021) 1–16.
[13] H.Z. Chu, H.Z. Liu, J.B. Zhuo, et al., Occlusion-guided multi-modal fusion for vehicle-infrastructure cooperative 3D object detection, Pattern Recognit. (2024) 110939.
[14] T. Gao, S.L. Xia, M.K. Liu, et al., MSNet: Multi-scale network for object detection in remote sensing images, Pattern Recognit. (2024) 110983.
[15] C.Z. Wang, C.Y. Wang, Y.H. Qian, et al., Feature selection based on weighted fuzzy rough sets, IEEE Trans. Fuzzy Syst. 32 (7) (2024) 4027–4037.
[16] W.H. Xu, Y.F. Yang, Y. Ding, et al., Incremental feature selection approach to multi-dimensional variation based on matrix dominance conditional entropy for ordered data set, Appl. Intell. 54 (6) (2024) 4890–4910.
[17] W.H. Xu, C.C. Liu, Incremental reduction of imbalanced distributed mixed data based on k-nearest neighbor rough set, Internat. J. Approx. Reason. 172 (2024) 109218.
[18] J. Zhao, D.Y. Wu, Y.X. Zhou, et al., Rough set theory-based group incremental approach to feature selection, Inform. Sci. 675 (2024) 120733.

[19] J. Zhao, Y. Ling, F.L. Huang, et al., Incremental feature selection for dynamic incomplete data using sub-tolerance relations, Pattern Recognit. 148 (2024) 110125.
[20] K.H. Yuan, W.H. Xu, D.Q. Miao, A local rough set method for feature selection by variable precision composite measure, Appl. Soft Comput. 155 (2024) 111450.
[21] Z. Li, Y.C. Wang, D.D. Xu, et al., Tbnet: A texture and boundary-aware network for small weak object detection in remote-sensing imagery, Pattern Recognit. (2024) 110976.
[22] K.H. Yuan, D.Q. Miao, et al., Multigranularity data analysis with zentropy uncertainty measure for efficient and robust feature selection, IEEE Trans. Cybern. (2024).
[23] Q.Q. Han, L. Hu, W.F. Gao, Integrating label confidence-based feature selection for partial multi-label learning, Pattern Recognit. 161 (2025) 111281.
[24] X.Y. Zhang, X. Shen, Graph-driven feature selection via granular-rectangular neighborhood rough sets for interval-valued data sets, Appl. Soft Comput. 170 (2025) 112716.
[25] B.K. Vo, H.S. Nguyen, An attribute ranking method based on rough sets and interval-valued fuzzy sets, Internat. J. Approx. Reason. 170 (2024) 109215.
[26] J.J. Wang, X.N. Li, An overlap function-based three-way intelligent decision model under interval-valued fuzzy information systems, Expert Syst. Appl. 238 (2024) 122036.
[27] P.T. Hao, W.F. Gao, L. Hu, Embedded feature fusion for multi-view multi-label feature selection, Pattern Recognit. 157 (2025) 110888.
[28] W.H. Xu, K. Cai, D.D. Wang, A novel information fusion method using improved entropy measure in multi-source incomplete interval-valued datasets, Internat. J. Approx. Reason. 164 (2024) 109081.
[29] Z.H. Xie, W.Z. Wu, L.X. Wang, et al., Entropy based optimal scale selection and attribute reduction in multi-scale interval-set decision tables, Int. J. Mach. Learn. Cybern. (2024) 1–22.
[30] X.Y. Zhang, Z.H. Feng, Feature selection based on contradictory state sequence for multi-scale interval valued decision table, Inform. Sci. 677 (2024) 120926.
[31] X. Liu, J. Dai, J. Chen, et al., Unsupervised attribute reduction based on α-approximate equal relation in interval-valued information systems, Int. J. Mach. Learn. Cybern. 11 (9) (2020) 2021–2038.
[32] Z. Yuan, H.M. Chen, X.L. Yang, et al., Fuzzy complementary entropy using hybrid-kernel function and its unsupervised attribute reduction, Knowl.-Based Syst. 231 (2021) 107398.
[33] W.H. Xu, M. Huang, Z. Jiang, et al., Graph-based unsupervised feature selection for interval-valued information system, IEEE Trans. Neural Netw. Learn. Syst. 35 (9) (2024) 12576–12589.
[34] G. Roffo, S. Melzi, U. Castellani, et al., Infinite feature selection: a graph-based feature filtering approach, IEEE Trans. Pattern Anal. Mach. Intell. 43 (12) (2020) 4396–4410.
[35] J.H. Dai, W.T. Wang, J.S. Mi, Uncertainty measurement for interval-valued information systems, Inform. Sci. 251 (2013) 63–78.

Zihan Feng received the B.Sc. degree from the School of Mathematics and Statistics, Xinxiang University, Xinxiang, China, in 2022. She is currently working toward the M.Sc. degree in computer science and technology with the School of Artificial Intelligence, Southwest University, Chongqing, China. His research interests cover feature selection, granular computing, and knowledge discovery.

Xiaoyan Zhang received the B.Sc. degree in mathematics from Yanbei Normal University, Datong, China, the M.Sc. degree in mathematics from Guangxi University, Nanning, China, and the Ph.D. degree in mathematics from Northwest University, Xi'an China, in 2001, 2005 and 2017, respectively. She is currently a Professor of College of Artificial Intelligence, Southwest University, Chongqing, China. She has published over 50 articles in international journals and book chapters. Her current research interests include mathematical foundations of artificial intelligence, fuzzy sets, granular computing, knowledge discovery, etc