Information Fusion 127 (2026) 103715

Contents lists available at ScienceDirect

Information Fusion

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

Multi-granularity interval-intent fuzzy concept-cognitive learning: An attention-enhanced adaptive clustering framework

Yi Ding, Weihua Xu*

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

ARTICLE INFO

Keywords:
Adaptive concept clustering
Concept-cognitive learning (CCL)
Concept prediction
Multi-granularity learning
Multi-level attention mechanism

ABSTRACT

Cognitive processes lie at the heart of artificial intelligence (AI) research, and the Multi-Granularity Interval-Intent Fuzzy Concept-Cognitive Learning model (MIFCL-A) presented in this paper offers a novel perspective on this domain. MIFCL-A innovatively incorporates multi-level attention mechanism to replicate the intricacies of human cognition, utilizing advanced concept cognitive learning methodologies. This model addresses several limitations inherent in existing concept learning frameworks, such as reliance on manual parameter tuning for concept clustering, the generation of pseudo concepts that compromise cognitive consistency, and an overreliance on attribute-based concept attention that neglects the centrality of objects. Our model introduces a multi-granularity concept structure that captures both global (coarse-granularity) and local (fine-granularity) perspectives, integrating global decision concepts with boundary-derived local concepts. It features a hierarchical attention mechanism that applies global attribute attention at the coarse-granularity level and local concept attention at the fine-granularity level. Moreover, an adaptive concept clustering algorithm is incorporated, which negates the need for manual parameter tuning and ensures the precision and robustness of concept evolution across varying granularities. Comparative evaluations indicate that MIFCL-A outperforms current models in terms of classification accuracy and knowledge representation capabilities, establishing its potential as an effective tool for knowledge discovery and data mining.

1. Introduction

Uncertainty plays a crucial role in shaping our understanding and processing of information across many fields of scientific research and decision-making [1]. Whether in the natural sciences, social systems, or machine learning, the inherent unpredictability and variability of data demand robust methods to dealing with uncertainty.

Granular computing (GrC) introduces the concept of “granules” - clusters of similar objects or data points that can be used to simplify complex systems and reduce uncertainty [2]. Granular computing provides a framework for understanding and organizing information at different levels of abstraction, allowing for more efficient processing and decision-making. Currently, numerous granular computing models are widely used to address uncertainty across various scenarios, including fuzzy sets [3], rough sets [4], three-way decision theory [5], and formal concept analysis (FCA) [6], among others. These models provide powerful frameworks for handling complex, uncertain, and imprecise data by partitioning the information into manageable granules or clusters.

Concept, as the fundamental units of human cognition in philosophy, effectively describe the relationship between the essence of objects and their objective nature [7]. In 1982, Wille [8] introduced the theory of FCA, which analyzes formal concepts from a lattice theory perspective. Formal concept lattices serve as a mathematical bedrock for structured knowledge representation, establishing rigorous frameworks and formal semantics to enable principled data analysis in real-world applications. These structures not only crystallize conceptual hierarchies but also provide a foundation for systematic knowledge discovery. Extensive research by scholars has advanced this foundation [9,10]. Inspired by FCA and cognitive computing [11], concept-cognitive learning (CCL) has emerged as a growing research focus. CCL treats concepts as fundamental informational granules, processing abstract information through structured methodologies to uncover latent object-attribute relationships within data. Recent years have witnessed considerable expansion in CCL theory and models, encompassing various concept such as formal concepts [12,13], fuzzy concepts [14–16], three-way concepts [17–19], two-way concepts [20,21], incomplete concept [22].

Different concept systems are suited to various scenarios, as they are designed to address the specific characteristics and challenges of each domain. Zhang et al. [23] established a granular cognitive representation based on object/attribute sufficiency and necessity through

* Corresponding author.
E-mail addresses: drying20@163.com (Y. Ding), chxuwh@swu.edu.cn (W. Xu).

https://doi.org/10.1016/j.inffus.2025.103715
Received 6 May 2025; Received in revised form 2 August 2025; Accepted 2 September 2025
Available online 11 September 2025

rigorous mathematical modeling. This dual-perspective approach has since been extended to fuzzy-based dynamic concept learning [20], classical concept learning with dynamic shifting perspectives [21], and competence-based concept learning in the context of skill acquisition [24].

In 1965, Zadeh [3] introduced fuzzy sets, providing a refined method for data representation via membership functions. Building on this foundation, the integration of fuzzy systems with machine learning has become widespread, exemplified by models such as fuzzy neural networks [25] and fuzzy support vector machines [26]. This synergy extends to concept cognitive learning (CCL), where integrating fuzzy sets advances concept exploration from a fuzzy perspective. Fuzzifying classical formal concepts enables CCL to effectively process continuous numerical data. Specifically, fuzzy concepts offer the distinct advantage of extracting valuable information representations while minimizing information loss during cognitive processes [27]. Complementing this, Wang et al. [28] proposed a novel multi-view fuzzy CCL model that enhances knowledge representation by learning and integrating fuzzy concepts across multiple data views. Notably, clustering techniques-central to discovering inherent structures in unlabeled data-find wide application in areas such as medical testing [29] and social network analysis [30]. Through a fuzzy concept clustering approach, Mi et al. [14] achieved a simplified representation of the concept space. Subsequently, numerous concept-cognitive learning models based on concept clustering emerged. For example, Deng et al. [15] considered attribute-derived fuzzy concepts and explained the predictive mechanism of incremental concept learning by integrating concept clustering spaces with upper and lower approximation spaces. Guo et al. [31] developed a concept learning model for tumor diagnosis in high-dimensional data, based on fuzzy three-way concepts and concept fusion. Liu et al. [32] adopted a cross-granularity concept clustering strategy to integrate feature information for missing multi-label data, effectively enabling label completion. Additionally, attention is a core attribute underlying all human perceptual and cognitive processes [33]. Xu et al. [34] proposed a multi-attention concept-cognitive learning model that incorporates graph attention mechanisms and graph structures for discrete data. Weighting attributes constitutes another form of attention mechanism, as demonstrated by Belohlavek et al. [35], who selected significant concepts through attribute weighting and differential concept weighting. Further extending this paradigm, Zhang et al. [36] constructed weighted fuzzy concepts and implemented a progressive dynamic learning framework for such concepts.

Guo et al. [37] conducted a comprehensive and systematic analysis of concept-cognitive learning, providing a crucial theoretical framework and valuable insights for further advancements in this field. Although the effectiveness of concept clustering processes has been widely validated, and many scholars have extensively considered attribute attention mechanism to enhance concept learning. The current CCL models still exhibit notable limitations, as outlined below:

(1) Neglect of global cognition: While focusing on localized concept refinement, existing models largely overlook the global perspective, a critical aspect of human decision-making processes.

(2) Absence of multi-level attention mechanism: Existing models typically prioritize obtaining attribute attention (via weighting) before calculating the corresponding concept attention. In other words, the attention to the concept is entirely influenced by the attributes within the concept's intent, often neglecting human focus on the object itself (the extent of the concept).

(3) Concept clustering process of human intervention: In most CCL models that consider concept clustering, the clustering process is determined by predefined parameters. This reliance on manually set parameters results in inefficiency and inflexibility when dealing with complex and dynamic data environments.

(4) Pseudo-concepts and interpretability: The clustering process often generates pseudo-concepts, which lack interpretability and fail to adhere to rigorous concept system definitions, resulting in cognitive inconsistencies between the cognitive processes at different stages.

Human cognition operates on a "global-first"" mechanism, prioritizing the processing of information based on coarse-granularity details before delving into finer-granularity specifics [38]. As a brain-inspired learning paradigm, granular-ball computing (GBC) first constructs multi-grained representations of data through adaptive granular balls - each ball covering a local region while preserving global topological structures. Xie et al. [39] proposed a novel method for constructing weighted granules, which adopts local iteration as an approximation for the calculation of global weights, eliminating the need for parameter adjustment during the update process. Xia et al. [40] developed an efficient and robust method for clustering granules from a multi-granularity perspective. Xia et al. [41] established a three-way approximation framework with granular-ball computing, significantly enhancing interpretability in multi-granularity feature selection through fuzzy entropy fusion. Unlike traditional fine-grained processing that operates directly on raw data points, GBC's coarse-to-fine hierarchy aligns with the "global-first" cognition.

Inspired by the GBC's parameter-free adaptability achieved through autonomous ball generation and hierarchical knowledge discovery from coarse to fine granularity, this paper proposes a novel CCL model-attention-based multi-granularity interval-intent fuzzy concept-cognitive learning-which closely aligns with human cognitive mechanisms. Its primary contributions are as follows:

(1) A definition of multi-granularity interval-intent fuzzy concepts is proposed, which includes two types of concepts learned from different granular perspectives: global decision concepts derived from a coarse-granularity (global) perspective and boundary-derived local concepts formed from a fine-granularity (local) perspective.

(2) A multi-level attention mechanism that does not rely solely on attributes is proposed. For global decision concepts, global attribute attention is defined by the idea of maximizing the inter-class distance. Combining global and local cognitive operators, local concept attention (also referred to as concept independence degree) is defined using the extent ratio, which quantifies the differentiation of concepts based on their decision membership degree.

(3) An adaptive concept clustering process is proposed, fully regulated by local concept attention, eliminating the need for parameter settings to update existing concepts or generate new ones. The framework implements alternating learning between concepts of different granularities. By the end of the process, all concepts strictly adhere to formal conceptual definitions, ensuring cognitive consistency and maintaining conceptual rigor throughout the framework.

To improve the representation of concepts, the definition of basic concepts in this paper follows the approach from [42], specifically using interval-intent fuzzy concepts. The main structure of the paper is organized as follows: Section 2 introduces classical fuzzy concepts and interval-intent fuzzy concepts, and provides a clear distinction of the cognitive operators for the latter across different decisions. Section 3 presents the core content of the MIFCL-A, including multi-granularity concepts, multi-level attention mechanism, adaptive concept clustering, and concept prediction. The following section presents extensive numerical experiments to analyze the effectiveness of attention and the model's classification performance. The final section summarizes the content of the paper and provides an analysis of future research directions.

2. Related work

This section briefly reviews some notions related to fuzzy formal context, classical fuzzy concept and interval-intent fuzzy concepts. It is important to note that the interval fuzzy set discussed in this paper differs from the interval-valued fuzzy set. Specifically, when determining the inclusion relation of two sets, the former (this paper) focuses on the inclusion of the interval itself, whereas the latter emphasizes the inclusion of fuzziness.

2.1. Fuzzy formal context and classical fuzzy concept

A typical two-dimensional numerical data table can often be interpreted as a fuzzy formal context. Unlike the classical formal context, it provides a more detailed representation of the membership relationships between objects and attributes.

A fuzzy formal context is defined as a triplet $(G, M, \tilde{I})$, where $G = \{x_1, x_2, \cdots, x_{|G|}\}$ and $M = \{a_1, a_2, \cdots, a_{|M|}\}$ represent the object set and attribute set, respectively. The term $\tilde{I}$ denotes the fuzzy relation between G and M, formally expressed as $\tilde{I} : G \times M \to [0, 1]$. Here, $\tilde{I}(x, a)$ specifies the degree of membership of an object x to an attribute a.

The classical fuzzy attribute set $\tilde{B}^s$ is represented in the form $\{\tilde{B}^s(a_1), \tilde{B}^s(a_2), \cdots, \tilde{B}^s(a_{|M|})\}$, where $\tilde{B}^s(a)$ denotes the degree to which the attribute a belongs to the fuzzy attribute set $\tilde{B}^s$. For convenience, the set of all crisp object subsets in G is denoted by 2^G, while the set of all fuzzy attribute subsets in M is represented as Γ^M.

Given a fuzzy formal context $(G, M, \tilde{I})$, for any $X \in G$ and $\tilde{B}^s \in \Gamma^M$, two operators $f : 2^G \to \Gamma^M$ and $h : \Gamma^M \to 2^G$ are given as follows [14]:

$$f(X) = \left\{ \bigwedge_{x \in X} \tilde{I}(x, a) | a \in M \right\},$$

$$h(\tilde{B}^s) = \{x \in G | \forall a \in M, \tilde{B}^s(a) \leq \tilde{I}(x, a)\}.$$

The pair $(X, \tilde{B}^s)$ is a classical fuzzy concept if $f(X) = \tilde{B}^s$ and $h(\tilde{B}^s) = X$, where X and $\tilde{B}^s$ are the extent and intent of the classical fuzzy concept.

2.2. Interval-intent fuzzy concept

In many concept-cognitive learning models, the same cognitive operators are defined differently across articles, with two common interpretations: full-decision mapping (the whole decision context) and single-decision mapping (a specific decision context) . This inconsistency in definitions can lead to confusion and variations in the model's application. This section distinguishes by defining different cognitive operators.

Let U denote the unit closed interval [0,1], and $[U]$ represent the set of all closed intervals within U. For a non-empty attribute set M, $\tilde{B} : M \to [U]$ is defined as an interval fuzzy attribute set over M. Similarly, the set of all interval fuzzy attribute subsets on M is denoted as $[\Gamma^M]$.

For any $\tilde{B} \in [\Gamma^M]$, the interval fuzzy attribute set $\tilde{B}$ is represented in the form $\{\tilde{B}(a_1), \tilde{B}(a_2), \cdots, \tilde{B}(a_{|M|})\}$, where $\tilde{B}(a) = [a^-, a^+]$ represents the membership interval of attribute a on $\tilde{B}$. For convenience, it is denoted by $[\![\tilde{B}(a)]\!]^- = a^-$ and $[\![\tilde{B}(a)]\!]^+ = a^+$.

Definition 1. Given a fuzzy formal context $(G, M, \tilde{I})$, for any $X \in G$ and $\tilde{B} \in [\Gamma^M]$, the two full-decision interval-intent cognitive operators $F : 2^G \to [\Gamma^M]$ and $H : [\Gamma^M] \to 2^G$ are defined as follows:

$$F(X) = \left\{ \left[\bigwedge_{x \in X} \tilde{I}(x, a), \bigvee_{x \in X} \tilde{I}(x, a) \right] | a \in M \right\},$$

$$H(\tilde{B}) = \{x \in G | \forall a \in M, [\![\tilde{B}(a)]\!]^- \leq \tilde{I}(x, a) \leq [\![\tilde{B}(a)]\!]^+\}.$$

The definition of an interval-intent fuzzy concept is similar to that of a classical fuzzy concept, namely, $(X, \tilde{B})$ is a full-decision interval-intent fuzzy concept if $F(X) = \tilde{B}$ and $H(\tilde{B}) = X$.

Property 1. *Let $X, X_1, X_2 \subseteq G$ and $\tilde{B}, \tilde{B}_1, \tilde{B}_2 \subseteq [\Gamma^M]$, then we have:*
(1) $F(X_1) \subseteq F(X_2)$ if $X_1 \subseteq X_2$, $H(\tilde{B}_1) \subseteq H(\tilde{B}_2)$ if $\tilde{B}_1 \subseteq \tilde{B}_2$;
(2) $X \subseteq HF(X)$, $\tilde{B} \supseteq FH(\tilde{B})$;
(3) $F(X) = FHF(X)$ and $H(\tilde{B}) = HFH(\tilde{B})$.

Proof. The detailed proof can be found in Ref. [42]. □

According to Definition 1, both full-decision interval-intent cognitive operators apply to the entire set of objects G, and the learned concept extent may include objects with different decisions. However, in some cases, it may be preferable to learn concepts separately for each decision. The quintuple $(G, M, \tilde{I}, D, J)$ is known as a fuzzy formal decision context, where $\tilde{I} : G \times M \to [0, 1]$ and $J : G \times D \to \{0, 1\}$.

Definition 2. Let $(G, M, \tilde{I}, D, J)$ be a fuzzy formal decision context and $G/D = \{G^{d_1}, G^{d_2}, \cdots, G^{d_{|D|}}\}$ be a decision division of object sets G. For any $X \in G^{d_k}$ and $\tilde{B} \in [\Gamma^M]$, the two single-decision interval-intent cognitive operators $F^{d_k} : 2^G \to [\Gamma^M]$ and $H^{d_k} : [\Gamma^M] \to 2^G$ under decision d_k are given as follows:

$$F^{d_k}(X) = \left\{ \left[\bigwedge_{x \in X} \tilde{I}(x, a), \bigvee_{x \in X} \tilde{I}(x, a) \right] | a \in M \right\},$$

$$H^{d_k}(\tilde{B}) = \{x \in G^{d_k} | \forall a \in M, [\![\tilde{B}(a)]\!]^- \leq \tilde{I}(x, a) \leq [\![\tilde{B}(a)]\!]^+\}.$$

The single-decision interval-intent cognitive operators defined in Definition 2 also adhere to the properties of full-decision interval-intent cognitive operators. The concept is defined similarly: $(X, \tilde{B})$ is a local interval-intent fuzzy concept if $F^{d_k}(X) = \tilde{B}$ and $H^{d_k}(\tilde{B}) = X$. And Property 1 is also applicable to single-decision interval-intent cognitive operators.

Property 2. *Let $(G, M, \tilde{I}, D, J)$ be a fuzzy formal decision context and $G/D = \{G^{d_1}, G^{d_2}, \cdots, G^{d_{|D|}}\}$ be a decision division of object sets G. For any $X \in G^{d_k}$ and $\tilde{B} \in [\Gamma^M]$, $(H^{d_k}F^{d_k}(X), F^{d_k}(X))$ and $(H^{d_k}(\tilde{B}), F^{d_k}H^{d_k}(\tilde{B}))$ are both single-decision interval-intent fuzzy concept.*

Proof. According to Property 1, the proof is straightforward. □

Obviously, when the concept learning clue consists of only a single object, the upper and lower bounds of the learned concept intent interval coincide, reducing it to a fuzzy single-valued form. For any single object x, $F^{d_k}(\{x\}) = \{\tilde{B}^s(a) | a \in M\}$ in which $\tilde{B}^s(a) = [\![\tilde{B}(a)]\!]^- = [\![\tilde{B}(a)]\!]^+$. And if there is no ambiguity, $F^{d_k}(\{x\})$ can be written as $F^{d_k}(x)$. Unless otherwise specified, the concepts mentioned in the following paragraphs refer to single-decision interval-intent fuzzy concepts.

3. MIFCL-A

Human cognition involves a dynamic interplay between global and local information processing, with cognitive systems shifting between these levels. From a global perspective, coarse-grained cognition facilitates the comprehension of a system's overall structure and consistency, enabling broad understanding of complex environments. However, when processing heterogeneous information for complex decision-making, humans prioritize boundary information through localized comparisons between entities.

Inspired by this dual cognitive capacity-simultaneously grasping holistic contexts while attending to local distinctions-we propose an attention-based multi-granularity concept-cognitive learning model which introduces global decision concepts (coarse-grained perspective) and boundary-derived local concepts (fine-grained perspective). We construct a multi-level attention system aligned with these conceptual hierarchies. Additionally, we develop an adaptive concept cluster generation method that dynamically restructures concept clusters through alternating coarse-to-fine granularity learning-an iterative mechanism that enhances learning performance by progressively refining conceptual representations.

3.1. Multi-granularity interval-intent fuzzy concept

This subsection introduces two single-decision interval-intent fuzzy concepts from both global and local perspectives: the global decision concept and the boundary-derived local concept. It is important to note that the latter is learned based on the existence of the former, aligning with the principle of cognitive relevance.

3.1.1. Global decision concepts

Definition 3. Let $(G, M, \tilde{I}, D, J)$ be a fuzzy formal decision context and $G/D = \{G^{d_1}, G^{d_2}, \cdots, G^{d_{|D|}}\}$ be a decision division of object sets G. For the set $X = \{x | x \in G^{d_k}\}$, $C_g^{d_k} = (X_g^{d_k}, \tilde{B}_g^{d_k})$ is called the global decision concept under decision d_k, where $X_g^{d_k} = X$ and $\tilde{B}_g^{d_k} = F^{d_k}(X)$.

According to Properties 1 and 2, it is easy to prove that the global decision concept is a single-decision interval-intent fuzzy concept. The number of global decision concepts corresponds to the number of decision categories, meaning there is exactly one global decision concept for each decision space. The essence of the concept primarily reflects the range of attribute values for all objects within the current decision. And we call the set of all global decision concepts the global decision concept space, represented by $CS_g = \{C_g^{d_k} | d_k \in D\}$.

As a holistic abstraction of the decision space, the global decision concept offers a comprehensive and coherent perspective for identifying the macroscopic structure of the data system from a coarse-granularity viewpoint.

3.1.2. Boundary-derived local concepts

The real world is inherently complex and unpredictable, with different global decision concepts often overlapping. As a result, relying solely on coarse-granularity cognition is insufficient for fully distinguishing and learning about entities; recognizing boundary information is essential for a more nuanced understanding.

Definition 4. Given a fuzzy formal sub-context $(G^{d_k}, M, \tilde{I})$, and the global decision concept $(X_g^{d_k}, \tilde{B}_g^{d_k})$ under the decision d_k, the set of boundary objects under decision d_k is defined as $O_g^{d_k} = \{x | \exists a \in M, F^{d_k}(x)(a) \in \{[\![\tilde{B}_g^{d_k}(a)]\!]^-, [\![\tilde{B}_g^{d_k}(a)]\!]^+\}\}$.

Definition 5. Given a fuzzy formal sub-context $(G^{d_k}, M, \tilde{I})$ and the set of boundary objects $O_g^{d_k}$ under decision d_k, $C_b^{d_k} = (X_b^{d_k}, \tilde{B}_b^{d_k}) = (H^{d_k}(\tilde{B}_b^*), F^{d_k} H^{d_k}(\tilde{B}_b^*))$ is called the boundary-derived local concept under decision d_k, where $\tilde{B}_b^* = \{[low(x,a), up(x,a)] | a \in M\}$ for $\forall x \in O_g^{d_k}$.

The lower bound of the attribute serving as a learning cue can be categorized into the following three cases:

(1) Boundary object attribute value: $low(x,a) = F^{d_k}(x)(a)$ if $F^{d_k}(x)(a) - [\![\tilde{B}_g^{d_k}(a)]\!]^- > [\![\tilde{B}_g^{d_k}(a)]\!]^+ - F^{d_k}(x)(a) \neq 0$;

(2) Global concept attribute value: $low(x,a) = [\![\tilde{B}_g^{d_k}(a)]\!]^-$ if $F^{d_k}(x)(a) - [\![\tilde{B}_g^{d_k}(a)]\!]^- \leq [\![\tilde{B}_g^{d_k}(a)]\!]^+ - F^{d_k}(x)(a)$;

(3) Global upper and lower bound average: $low(x,a) = \frac{1}{2}([\![\tilde{B}_g^{d_k}(a)]\!]^- + [\![\tilde{B}_g^{d_k}(a)]\!]^+)$ if $F^{d_k}(x)(a) = [\![\tilde{B}_g^{d_k}(a)]\!]^+$.

The upper bound of the attribute serving as a learning cue can be categorized into the following three cases:

(1) Global concept attribute value: $up(x,a) = [\![\tilde{B}_g^{d_k}(a)]\!]^+$ if $F^{d_k}(x)(a) - [\![\tilde{B}_g^{d_k}(a)]\!]^- > [\![\tilde{B}_g^{d_k}(a)]\!]^+ - F^{d_k}(x)(a)$;

(2) Boundary object attribute value: $up(x,a) = F^{d_k}(x)(a)$ if $0 \neq F^{d_k}(x)(a) - [\![\tilde{B}_g^{d_k}(a)]\!]^- \leq [\![\tilde{B}_g^{d_k}(a)]\!]^+ - F^{d_k}(x)(a)$;

(3) Global upper and lower bound average: $up(x,a) = \frac{1}{2}([\![\tilde{B}_g^{d_k}(a)]\!]^- + [\![\tilde{B}_g^{d_k}(a)]\!]^+)$ if $F^{d_k}(x)(a) = [\![\tilde{B}_g^{d_k}(a)]\!]^-$.

As established by Property 2, the boundary-derived local concept can be formally demonstrated as a single-decision interval-intent fuzzy concept. This concept is constructed through a dual-cue learning mechanism: (1) Object cues (Definition 4): Objects constituting the decision boundary under each decision class; (2) Attribute cues (Definition 5): Synthesized by integrating the attribute values of these boundary objects with the intent of the global decision concept. Critically, attribute cues preserve a hybrid representation of global and local information, thereby capturing a novel fine-grained conceptual perspective.

The set of all boundary-derived local concepts under decision d_k is represented as $CS_b^{d_k} = \{C_{b_1}^{d_k}, C_{b_2}^{d_k}, \cdots, C_{b_n}^{d_k}\}$ and $CS_b^{d_k}$ is referred to boundary-derived local concept space under decision d_k.

Property 3. *For all boundary-derived local concepts under a decision d_k, the following property holds:*

$$2 \leq |CS_b^{d_k}| \leq |O_g^{d_k}| \leq |G^{d_k}|,$$

where $|\cdot|$ indicates the total number of elements in the set.

Proof. The above property can be proved in the following three parts:

(1) If an object x satisfies $F^{d_k}(x)(a) < [\![\tilde{B}_g^{d_k}(a)]\!]^+$ or $F^{d_k}(x)(a) > [\![\tilde{B}_g^{d_k}(a)]\!]^-$, then $|O_g^{d_k}| \leq |G^{d_k}|$ follows.

(2) There are numerous factors that can lead to $|CS_b^{d_k}| \leq |O_g^{d_k}|$. As long as the mappings of attribute clues derived from different boundary objects are consistent, the resulting boundary-derived local concept will remain the same.

(3) If there exist two objects such that $F^{d_k}(x_1) \cup F^{d_k}(x_2) = \tilde{B}_g^{d_k}$ and for $\forall a \in M$ there is no object x that makes $F^{d_k}(x)(a) \in \{[\![\tilde{B}_g^{d_k}(a)]\!]^-, [\![\tilde{B}_g^{d_k}(a)]\!]^+\}$, then $|CS_b^{d_k}| = 2$. Otherwise, $|CS_b^{d_k}| > 2$ holds. □

According to this property, the boundary-derived local concept encapsulates structured information about global boundaries. Distinctions among various decisions can be achieved using fewer boundary-derived local concepts. These local concepts delineate the fuzzy transition regions between different decisions or categories, effectively capturing the inherent complexity of fuzzy cognition. We refer to global decision concepts and boundary-derived local concepts collectively as multi-granularity interval-intent fuzzy concepts (or simply multi-granularity concepts). Algorithm 1 outlines the construction process for the concept spaces of these multi-granularity concepts.

Algorithm 1: Construction of multi-granularity concept space.

Input: The fuzzy formal decision context $(G, M, \tilde{I}, D, J)$ and the decision division $G/D = \{G^{d_1}, G^{d_2}, \cdots, G^{d_{|D|}}\}$.

Output: The global decision concept space CS_g and the set of all boundary-derived local concept space CS_b.

```
for G^{d_k} ∈ G/D do
    X = G^{d_k}, O_g^{d_k} = ∅;
    C_g^{d_k} = (X_g^{d_k}, B̃_g^{d_k}) = (X, F^{d_k}(X));
    CS_g ← C_g^{d_k};
    for each j = 1 to |G^{d_k}| do
        for each a ∈ M do
            if F^{d_k}(x)(a) ∈ {[[B̃_g^{d_k}(a)]]^-, [[B̃_g^{d_k}(a)]]^+} then
                O_g^{d_k} ← x_j;
            end
        end
    end
    for each x ∈ O_g^{d_k} do
        Compute B̃_b^* by Definition 5;
        C_b^{d_k} = (H^{d_k}(B̃_b^*), F^{d_k} H^{d_k}(B̃_b^*));
        CS_b^{d_k} ← C_b^{d_k};
    end
    CS_b ← CS_b^{d_k};
end
return CS_g and CS_b
```

3.2. Multi-level attention mechanism

Building on the foundation of global decision concepts and boundary-derived local concepts established earlier, we propose two attention mechanisms to enhance differentiation between distinct information types: global attribute attention and local concept attention.

Global attribute attention aligns with global decision concepts, emphasizing the identification of personalized attribute variations across different decision spaces. Conversely, local concept attention operates on all non-global decision concepts-including but not limited to boundary-derived local concepts. By using concept intents as guiding signals, and through the separate processing of local and global cognitive operators, this approach enables independent concept representations within distinct decision spaces. The inter-class discriminability

of these concepts is quantified via local concept attention as a concept independence degree.

3.2.1. Global attribute attention

To effectively capture attribute-level distinctions across different decision spaces, global attribute attention is introduced as a mechanism to prioritize attributes that contribute significantly to decision differentiation.

Definition 6. Let $(G, M, \tilde{I}, D, J)$ be a fuzzy formal decision context $(G, M, \tilde{I}, D, J)$ and $G/D = \{G^{d_1}, G^{d_2}, \cdots, G^{d_{|D|}}\}$ be a decision division of object sets G. The global attribute attention space is defined as $AttS_g = \{att(a)|a \in M\}$, where $att(a) = \frac{1}{|D|}\sum_{d_k \in D} \exp\left(\frac{\gamma}{2\cdot|D|-2}\sum_{d_m \neq d_k}\left(max\left([\![\tilde{B}_g^{d_k}(a)]\!]^-, [\![\tilde{B}_g^{d_m}(a)]\!]^-\right) - min\left([\![\tilde{B}_g^{d_k}(a)]\!]^+, [\![\tilde{B}_g^{d_m}(a)]\!]^+\right) + 1\right)\right)$ and γ $(\gamma \geq 0)$ is a global attribute attention parameter.

Global attribute attention operates on the principle of maximizing decision boundary differentiation. When decision boundaries for distinct classes under a given attribute exhibit no overlap, all values in the difference term remain positive-resulting in higher attention weights for that attribute. Conversely, when decision boundaries intersect or fully overlap for the same attribute, the summation within the difference term contains negative components, thereby reducing the attribute's attention weight. Attributes receiving stronger global attention exhibit greater distinguishability within the decision space.

3.2.2. Local concept attention

To determine the level of which the concept's attribute scope is exclusive to the current decision, we define concept independence to represent this degree of "exclusivity". And we also can use the independence degree of concepts to represent local concept attention below.

Definition 7. Let $CS_b^{d_k}$ be fuzzy boundary-derived local concept space under decision d_k. The concept independence degree for $C_b^{d_k} = (X_b^{d_k}, \tilde{B}_b^{d_k})$ is defined as follows:

$$ind(C_b^{d_k}) = \frac{|X_b^{d_k}|}{|H(\tilde{B}_b^{d_k})|},$$

where H is full-decision interval-intent cognitive operator and $|\cdot|$ denotes the cardinality.

The independence degree of a concept reflects its specificity within the cognitive process. A higher independence of a concept under decision d_k signifies that the attribute set $\tilde{B}_b^{d_k}$, serving as a intent of the concept, aligns more closely with the attribute interval of decision d_k.

The purpose of Definition 7 is not only to determine the importance of the current concept but also to provide a critical basis for deciding whether subsequent concepts should be updated during the concept cluster generation.

Property 4. *For all boundary-derived local concepts under a decision d_k, the following properties apply:*

(1) $0 < ind(C_b^{d_k}) \leq 1$;

(2) If $ind(C_b^{d_k}) = 1$, *then* $\forall x \in X_b^{d_m}$, $F^{d_m}(x) \notin \tilde{B}_b^{d_k}$ *where* $d_m \neq d_k$.

When $ind(C_b^{d_k}) = 1$, $C_b^{d_k}$ is called a completely independent concept. Property 4 indicates that the attribute value of an object from no other decision falls within the intent of the completely independent concept, meaning that the completely independent concept can be regarded as the exclusive concept for a specific decision. Furthermore, it is evident that for any object $x \in G^{d_k}$, concept $(\{x\}, F^{d_k}(x))$ is also a completely independent concept.

3.3. Adaptive concept clustering

In the preceding discussion, we introduced the boundary-derived local concept, emphasizing that the learning process within decision spaces remains confined to the local regions near decision boundaries. To enhance the universality and adaptability of the cognitive system, this section delves into the process of concept clustering. This approach aims to achieve a more refined and comprehensive cognition of the decision space, surpassing the granularity offered by the global decision concept.

The process of concept clustering, also known as concept cluster generation, involves re-evaluating all objects that are not involved in the formation of the concept and grouping them into existing boundary-derived local concepts or generating new concept independently. In concept-cognitive learning, concept clustering serves as a method to compress and efficiently represent the concept space through the creation of pseudo-concepts. A pseudo-concept's extent is typically defined as the union of the extent of a group concepts, while its intent is represented by the fuzzy attribute set derived from the weighted average of these concepts [14,36]. As a result, pseudo-concepts often deviate from the strict definition of a concept, leading to inconsistencies in cognition between different stages.

For interval-intent fuzzy concepts, the intent is designed to encompass the attribute information of all object sets within the concept's extent. When a new object is introduced, its inclusion requires evaluating the relationship between its attribute values and the interval-intent. This allows for a selective update of the interval boundary values, thereby enabling the extent to be dynamically adjusted. The key advantage of this approach lies in its ability to incorporate new samples while ensuring that the learned concepts continue to adhere to the strict definition of interval-intent fuzzy concepts. Therefore, the concept clustering of interval-intent fuzzy concepts is to re-learn new concepts from new clues.

3.3.1. Cosine similarity measure

Cosine similarity is employed to measure the alignment between objects and boundary-derived local concepts, as it effectively captures the geometric relationship between attribute distributions in multi-dimensional spaces. This measure emphasizes the angle between vectors rather than their magnitudes, ensuring robustness to scale variations and focusing on the structural patterns of attributes.

Definition 8. Given a fuzzy formal sub-context $(G^{d_k}, M, \tilde{I})$ and the boundary-derived local concept space $CS_b^{d_k}$ under decision d_k. For any $x \in G^{d_k}$ and $C_b^{d_k} = (X_b^{d_k}, \tilde{B}_b^{d_k}) \in CS_b^{d_k}$, the cosine similarity between the object x and the concept $C_b^{d_k}$ is defined as:

$$\cos(x, C_b^{d_k}) = \frac{\sum_{a\in M}\tilde{I}(x,a)\cdot mid(\tilde{B}_b^{d_k}(a))}{\sqrt{\sum_{a\in M}\tilde{I}(x,a)^2}\cdot\sqrt{\sum_{a\in M}mid(\tilde{B}_b^{d_k}(a))^2}},$$

where $mid(\tilde{B}_b^{d_k}(a)) = \frac{1}{2}([\![\tilde{B}_b^{d_k}(a)]\!]^- + [\![\tilde{B}_b^{d_k}(a)]\!]^+)$.

Definition 8 serves as the foundation for identifying concept cluster centers in the subsequent section. At the same time, this definition is applicable to all interval-intent fuzzy concepts.

3.3.2. Concept cluster updating method

A concept cluster is a structure formed by a boundary-derived local concept serving as the cluster center, with objects which needs to be cognition acting as the cluster nodes.

The process of concept clustering can be viewed as updating the boundary-derived local concept space derived from the boundaries by learning from objects in the non-boundary range. Simultaneously, new concept clusters are adaptively generated. For each decision space, the concept cluster generation process can be outlined in the following steps:

• **Step 1** Filter objects to be updated: Let P^{d_k} denote the set of all objects that do not overlap with the extent of any boundary-derived local concepts within the given decision d_k. All elements of P^{d_k} need to participate in cluster generation process. First, we randomly select an object $x \in P^{d_k}$, set $CS^{d_k} = CS_b^{d_k}$ and proceed to Step 2.

• **Step 2** Select concept cluster center: For concept space CS^{d_k}, calculate the cosine similarity of the object x with each concept according to Definition 8. The concept $C^{d_k} = (X^{d_k}, \tilde{B}^{d_k})$ with the highest similarity is selected as the cluster center.

• **Step 3** Renewal concept cluster: Let $X' = X^{d_k} \cup \{x\}$. A new concept $C'^{d_k} = (H^{d_k} F^{d_k}(X'), F^{d_k}(X'))$ is generated by learning from the updated object cue X'. If the independence of concepts satisfies $ind(C^{d_k}) \leq ind(C'^{d_k})$, the old concept is replaced with the new one, i.e., $C^{d_k} = (H^{d_k} F^{d_k}(X'), F^{d_k}(X'))$, and continue to Step 4. Otherwise, return to Step 2, select the concept with cosine similarity that is sub-optimal compared to the previous iteration, and repeat Step 3. If the selected object x causes a decrease in the independence of all concepts, then $(\{x\}, F^{d_k}(x))$ is added to the concept space CS^{d_k} as a new concept.

• **Step 4** Update object state: The objects selected in Step 1, along with the new objects in set P^{d_k} learned in Step 3, are considered as having participated in the concept clustering process. These objects are then excluded from subsequent random selections. The above steps are repeated iteratively until $P^{d_k} = \emptyset$, at which point the clustering process concludes.

During concept clustering, the decision to add an object to a new concept cluster's extent hinges on whether it significantly alters the cluster's independence degree. The primary objectives are to maintain cluster centers with high differentiation and minimize the influence of other clustering decisions on the current one. Ultimately, concept clustering aims to adaptively identify concepts with the maximally achievable independence degree (indicating high inter-class discrimination). This ensures that different types of concepts can be efficiently distinguished in subsequent concept prediction tasks.

Initially, the concept space comprises boundary-derived local concepts. As clustering progresses, nodes within each cluster are incrementally assigned to their respective cluster centers. Concurrently, objects that weaken the independence of all existing concept clusters trigger the formation of new conceptual clusters. These new clusters then become potential targets for other nodes to merge with in later steps. The final concept space thus integrates both the updated boundary-derived local concepts and the newly formed non-boundary (interior) concepts. We term this post-clustering concept space the Concept Cluster Generation Space, denoted as CS'^{d_k}. The specifics of the concept clustering process are detailed in Algorithm 2.

Fig. 1 provides a comprehensive visualization of the adaptive concept clustering methodology. The clustering process involves three distinct scenarios for each sample undergoing learning: (1) Successful Concept Integration: When aggregated with its most similar concept, the concept independence degree remains stable or improves. This occurs when: the sample count supporting the current decision exceeds alternative decisions (for concepts with initial independence degree less than 1) or only samples supporting this decision are present (for concepts with initial independence degree equal to 1). This scenario represents successful cognitive learning completion. (2) Degraded Concept Independence: Aggregation with the most similar concept reduces the concept's independence, occurring when alternative decisions receive greater sample support than the current decision. In such cases, the algorithm proceeds to evaluate the next most similar concept. (3) Novel Concept Formation: When aggregation with all existing concepts degrades their independence, the sample initiates a new concept, thereby completing its cognitive processing cycle.

It should be noted that the above three situations are not in a parallel relationship. At a certain point in time, a sample can only undergo exactly one of these three possible state transitions. While Fig. 1 schematically represents all three scenarios, it's not parallel processing paths.

Algorithm 2: Concept cluster generation.

Input: The fuzzy formal decision context $(G, M, \tilde{I}, D, J)$, the decision division $G/D = \{G^{d_1}, G^{d_2}, \cdots, G^{d_{|D|}}\}$, and the set of all boundary-derived local concept space $CS_b = \{CS_b^{d_1}, CS_b^{d_{2D}}, \cdots, CS_b^{d_{|D|}}\}$.

Output: Concept cluster generation sapce CS'.

```
for G^{d_k} ∈ G/D do
    CS^{d_k} = CS_b^{d_k};
    Select the set P^{d_k} to be clustered;
    for a random selected object x ∈ P^{d_k} do
        Find the cluster center concept C^{d_k} ∈ CS^{d_k};
        X' = X^{d_k} ∪ {x};
        C'^{d_k} = (H^{d_k} F^{d_k}(X'), F^{d_k}(X'));
        if ind(C^{d_k}) ≤ ind(C'^{d_k}) then
            Substitute C'^{d_k} for C^{d_k};
            Removes objects that have participated in
              clustering from P^{d_k};
        end
        if ind(C_b^{d_k}) > ind(C_b'^{d_k}) then
            Select the sub-optimal cluster center concept and
              back to step 6;
        end
        if no concept update occurred then
            CS^{d_k} ← ({x}, F^{d_k}(x));
        end
    end
    CS' ← CS^{d_k};
end
return CS'
```

3.4. Concept prediction

Humans are often influenced by attention when distinguishing between objects, focusing more on identifying key features. This section introduces a concept prediction method based on multi-level attention mechanism.

Definition 9. Let x represent a new object, and a new concept $C = (\{x\}, F(x))$ will be formed. For any concept $C'^{d_k} = (X'^{d_k}, \tilde{B}'^{d_k}) \in CS'^{d_k}$, we define the attention-based distance between C and C'^{d_k} as follows:

$$Dis(C, C'^{d_k}) = \frac{\sum_{a \in M} |att(a) \cdot (F(x)(a) - mid(\tilde{B}'^{d_k}(a)))|}{ind(C'^{d_k})},$$

where $att(a) \in Att_{S_g}$, $mid(\tilde{B}'^{d_k}(a)) = \frac{1}{2}([\![\tilde{B}'^{d_k}(a)]\!]^- + [\![\tilde{B}'^{d_k}(a)]\!]^+)$and $ind(C'^{d_k})$ is the local concept attention (concept independence degree) of concept C'^{d_k}.

According to Definition 9, when measuring the distance between concepts, global attribute attention is taken into account. Attributes with higher attention contribute more to the distance under the same distance metric. Simultaneously, the distance of the object to be predicted is either increase or maintained based on the independence degree of the known decision concept (boundary-derived local concept or newly formed non-boundary concept). When the independence degree of the concept is very low, we prefer to reduce its influence in cases of equal distance. Therefore, a concept independence term (local concept attention) is included in the denominator of the distance formula, which increases the distance measure for concepts with low independence degree. Conversely, when the independence of a known decision concept is high, such as 1, the distance measure is almost unaffected.

Humans tend to categorize an object into the class to which it is most similar, so the predicted decision of the object aligns with the decision of

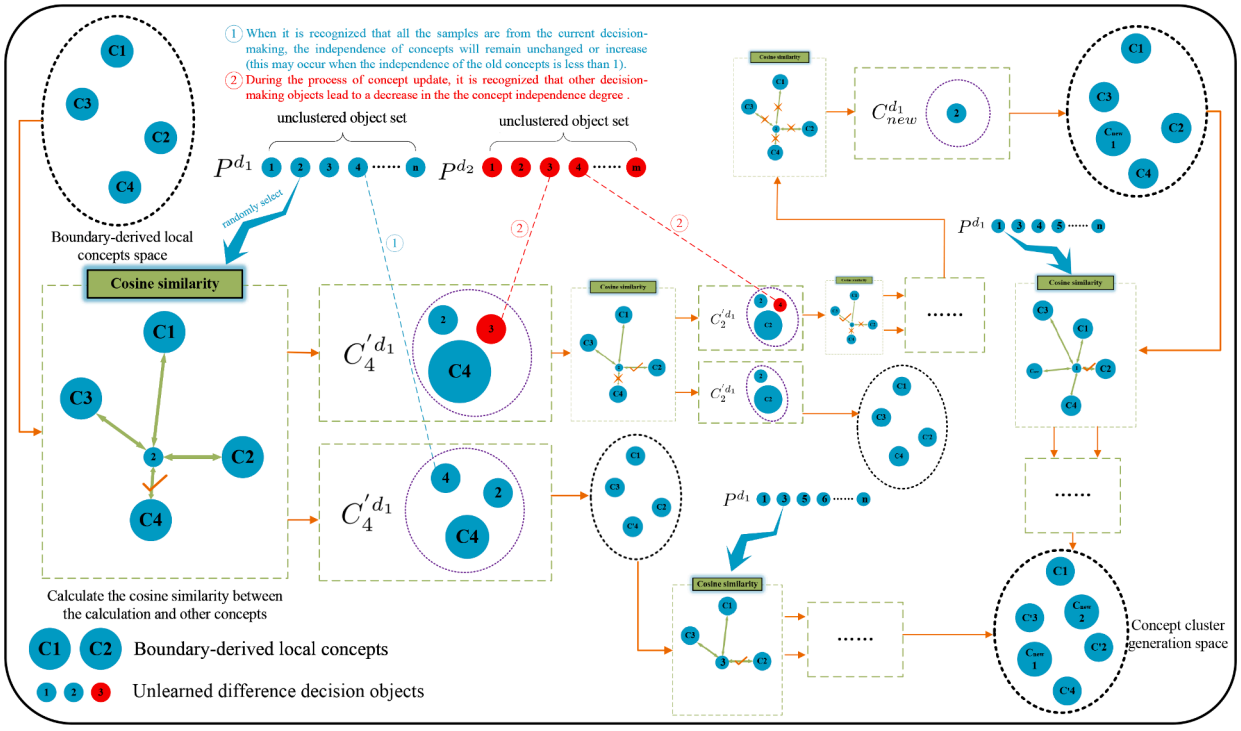

Fig. 1. Concept clustering process diagram.

the concept that is most similar to it. The detailed procedure for concept prediction is delineated in Algorithm 3.

Algorithm 3: Concept prediction.

Input: Concept cluster generation sapce CS', global attribute attention space $AttS_g$ and new object x.
Output: Prediction label d^* of new object x.

Set $Dis\_S = \emptyset$;
for $CS^{d_k} \in CS'$ **do**
  for $C'^{d_k} = (X'^{d_k}, \tilde{B}'^{d_k}) \in CS^{d_k}$ **do**
    Construct new concept $C = (\{x\}, F(x))$;
    Calculate the attention-based distance $Dis(C, C'^{d_k})$ between two concepts by definition 9;
    $Dis\_S \leftarrow Dis(C, C'^{d_k})$;
  end
end
$Dis(C, C'^{d_m}) = \min(Dis\_S)$;
$d^* = d_m$;
return Prediction label d^*

3.5. Overall process and time complexity

The flowchart of this paper is presented in Fig. 2, which outlines the following steps:

(1) Initially, the global decision concept is learned from a global perspective. Based on the boundaries of this global decision concept, all objects contributing to the boundary are identified. These boundary objects are treated as the first learning clue, from which the corresponding attribute value intervals are determined. This leads to the second clue for boundary-derived local concepts from the learning boundaries and further cognitive processing.

(2) The existing boundary-derived local concepts are treated as cluster centers, and knowledge from non-boundary objects is fused sequentially. Objects that do not meet the clustering criteria will adaptively generate new concepts, which can be regarded as updated cluster centers for subsequent iterative processing.

(3) For new objects, they are treated as independent concepts. Using the global decision concept and the boundary-derived local concepts, both global and local attention are calculated. The attention-based distance between the new concepts, the updated old concepts, and the newly generated concepts is computed. The label of the new object is then determined based on the minimal distance.

For the learning of the global decision concept, only a single traversal of the object set and attribute set is required, yielding a time complexity of $O(|G||M|)$. For the boundary-derived local concept, the initial object clue learning involves considering the number of boundary concepts, denoted as N where $N \leq |G|$ according to Property 3. The subsequent acquisition of attribute clues is influenced by the global decision concept, with a time complexity of $O(N|G||M|)$. Considering the overall learning of multi-granularity concepts, the combined time complexity is $O(N|G||M|)$. The time complexity of obtaining global attribute attention and local concept attention is $O(|M||D|)$ and $O(N|G||M|)$, respectively. In the process of concept clustering, which effectively constitutes a re-learning of the entire data space, the time complexities for updating concept clusters and generating new concept clusters are $O(|G|^2|M|)$ and $O(N|G|^2|M|)$, respectively. Finally, the time complexity of concept prediction is $O(|G||M|)$.

4. Experiments

In this section, the effectiveness and feasibility of MIFCL-A under a fuzzy context are validated through experiments. All experiments are conducted on a personal computer equipped with a 12th Gen Intel Core i5-12600KF processor (3.7 GHz) and 32 GB of RAM.

To evaluate the performance of different models, nine classic classification datasets (see https://archive.ics.uci.edu/andhttps://sci2s.ugr.es/keel) and three high-dimensional tumor datasets (see https://jundongl.github.io/scikit-feature/datasets.html) are randomly selected for experimentation. The details of these datasets, including the number of objects, attributes, and decisions, are provided in Table 1. Typically, the

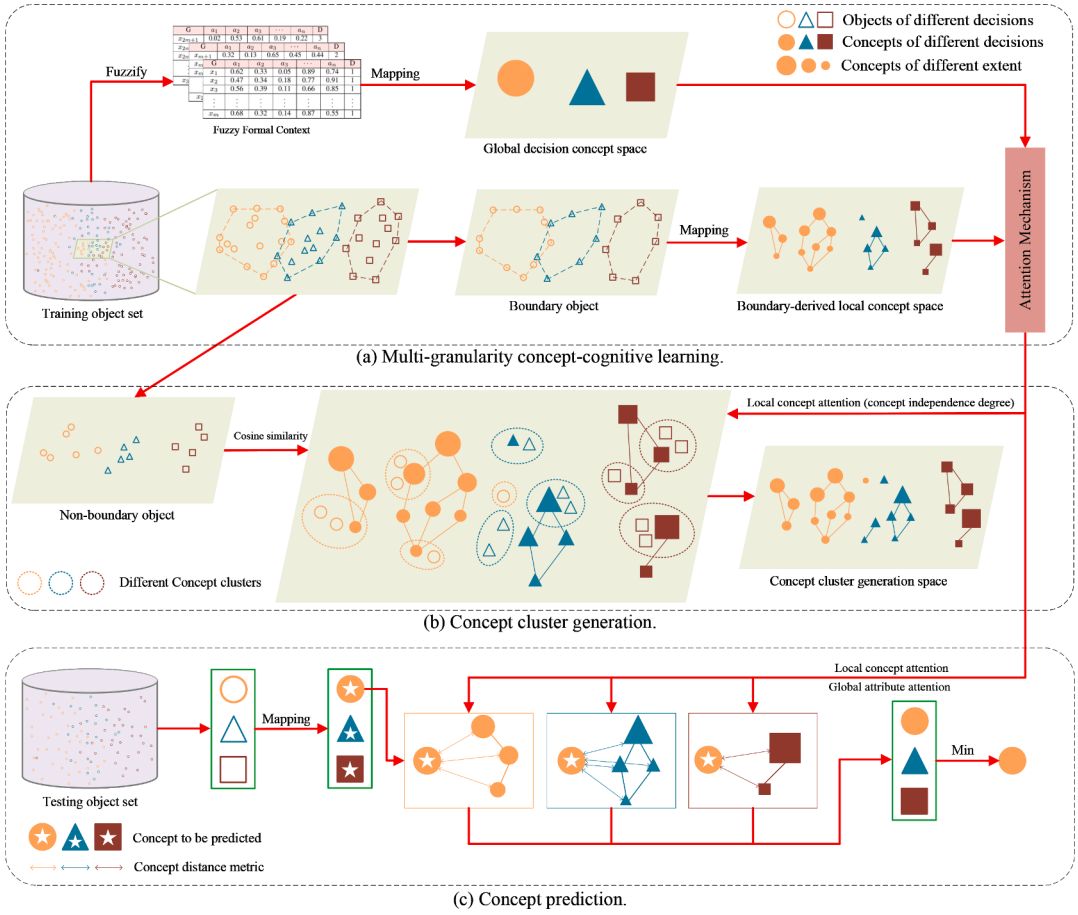

Fig. 2. Overall flow chart of MIFCL-A (the selection of boundary objects in process (a) serves as a simplified representation, with the actual acquisition method detailed in Definition 4).

Table 1
Details of the experimental datasets.

ID	Dataset	Object	Attribute	Class
1	Wine	178	13	3
2	Glass	214	9	6
3	Newthyroid	215	5	3
4	Balance	625	4	3
5	Australian	690	14	2
6	Vehicle	846	18	4
7	Titanic	2201	3	2
8	Mushroom	8124	22	2
9	ESSG	10,000	13	2
10	GLIOMA	50	4434	4
11	TOX_171	171	5748	4
12	Lung	203	3313	4

max-min function is used to transform the attribute values in the dataset into fuzzy form, mapping them to the range [0, 1]. The specific formula is as follows:

$$\tilde{I}(x,a) = \frac{l(x,a) - min(l(a))}{max(l(a)) - min(l(a))},$$

where the $l(x,a)$ denotes the value of object x under attribute a, and the $max(l(a))$ and $min(l(a))$ are the maximum and minimum values of all objects on attribute a, respectively.

All datasets are randomly divided into two subsets in each experiment: a training set comprising 80 % of the data and a test set comprising the remaining 20 %. The results of all classification experiments are derived from ten independent runs, with the final outcomes representing the average of these runs.

4.1. Performance of attention mechanism

This section evaluates the impact of multi-level attention mechanism and concept clustering on classification performance through ablation experiments. Additionally, the influence of the sole hyperparameter γ on the performance of the model is further analyzed.

4.1.1. Ablation experiments

To take into account various influences, the model is mainly divided into three parts for self-comparison analysis. Specifically, it includes whether there is a concept clustering process, whether there is global attribute attention (if global attribute attention is used, set $\gamma = 4$), and whether there is local concept attention. Since local concept attention as a clustering indicator cannot be discarded in the concept clustering process, only the three contrast models considering the concept clustering process are adjusted to the prediction formula in Definition 9. They are MIFCL-NGA (removing the global attribute attention in the numerator), MIFCL-NLA (removing the local concept attention in the denominator), and MIFCL-NA (removing both the numerator and denominator attentions). On this basis, the boundary-derived local concept spaces are exported as the final concept spaces, and the four contrast models of the concept clustering process are removed. They are MIFCL-NCA (only removing the clustering process, considering both attentions), MIFCL-NCNGA (removing the clustering process and the global attribute attention in the numerator of Definition 9), MIFCL-NCNLA (removing the clustering process and the local concept attention in the denominator of Definition 9), and MIFCL-NCNA (removing the clustering process and both attentions in the numerator and denominator of Definition 9). To ensure fairness, 10 experiments are conducted on the same

Table 2
Attention ablation experiments on twelve datasets.

ID	MIFCL-A	MIFCL-NGA	MIFCL-NLA	MIFCL-NA	MIFCL-NCA	MIFCL-NCNGA	MIFCL-NCNLA	MIFCL-NCNA
1	**98.05 ± 2.16**	**98.05 ± 2.16**	96.94 ± 2.90	96.94 ± 2.90	92.78 ± 3.56	92.22 ± 3.23	92.78 ± 3.56	92.22 ± 3.23
2	75.35 ± 6.25	**75.81 ± 6.26**	75.34 ± 6.25	**75.81 ± 6.26**	75.58 ± 6.00	75.58 ± 6.11	75.58 ± 5.23	75.35 ± 5.90
3	**95.58 ± 3.67**	95.35 ± 4.88	**95.58 ± 3.67**	95.35 ± 4.88	90.47 ± 3.36	91.40 ± 2.76	90.70 ± 4.65	91.86 ± 3.92
4	**86.00 ± 1.65**	**86.00 ± 1.65**	70.39 ± 3.89	70.39 ± 3.89	**86.00 ± 1.65**	**86.00 ± 1.65**	68.16 ± 3.67	68.16 ± 3.67
5	**85.22 ± 2.61**	83.47 ± 2.45	83.57 ± 3.16	81.30 ± 3.26	**85.22 ± 2.61**	83.47 ± 2.45	83.99 ± 2.49	81.92 ± 3.30
6	**71.00 ± 2.76**	70.50 ± 2.12	**71.00 ± 2.76**	70.50 ± 2.12	68.98 ± 2.98	67.80 ± 2.64	68.98 ± 3.03	67.89 ± 2.60
7	**77.42 ± 1.39**	**77.42 ± 1.39**	33.85 ± 2.17	33.85 ± 2.17	**77.42 ± 1.39**	**77.42 ± 1.39**	68.29 ± 2.07	68.29 ± 2.07
8	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**
9	**94.75 ± 0.43**	94.38 ± 0.46	88.72 ± 0.53	87.44 ± 0.45	93.40 ± 0.35	86.25 ± 0.29	93.23 ± 0.38	86.26 ± 0.30
10	**80.00 ± 9.42**	77.00 ± 8.16	**80.00 ± 9.42**	77.00 ± 8.16	**80.00 ± 9.42**	77.00 ± 8.16	**80.00 ± 9.42**	77.00 ± 8.16
11	**85.14 ± 6.10**	82.86 ± 6.26	**85.14 ± 6.10**	82.86 ± 6.26	**85.14 ± 6.10**	82.86 ± 6.26	**85.14 ± 6.10**	82.86 ± 6.26
12	**95.12 ± 2.19**	**95.12 ± 2.19**	94.15 ± 1.19	94.15 ± 1.19	**95.12 ± 2.19**	**95.12 ± 2.19**	94.15 ± 1.19	94.15 ± 1.19

randomly split training and test sets, and the average accuracy and standard deviation are presented in Table 2 (optimal values are shown in bold).

The results show that the model incorporating both types of attention and concept clustering, MIFCL-A, achieves the best performance across eleven datasets. Furthermore, MIFCL-A outperforms (with consistent performance in two datasets) MIFCL-NCNA which removes concept clustering and two types of attention in 10 datasets. It is noted that the results of datasets 4, 7, 8 and 12 are the same in the four sub-experiments (MIFCL-A, MIFCL-NGA, MIFCL-NCA, MIFCL-NCNGA), indicating that the global attribute attention and concept clustering have not played a role at this time and the local concept attention in the multi-level attention mechanism has a greater impact on the model which conforms to the human tendency to pay more attention to local details when recognizing things. The model MIFCL-NCA without considering concept clustering has the same effect as the one considering concept clustering in seven datasets. The reason is that the recognized boundary concepts cover most of the information, and only a few or no new objects have participated in the process of concept clustering. To some extent, this demonstrates the effectiveness of boundary-derived local concepts. The situation where the accuracy increases when the global attention is removed as seen in Dataset 2 is caused by the selection of parameter γ. And global attribute attention is more effective for datasets with clearly distinguishable decision boundaries, such as datasets 3, 6, 10 and 11. Therefore, even without the global attribute attention mechanism, the model (MIFCL-NLA) can still achieve the same performance as MIFCL-A. It is worth noting that in the three high-dimensional tumor datasets 10, 11 and 12, the models with attention mechanisms outperformed those without (regardless of whether concept clustering is considered or not), which further demonstrates the effectiveness and feasibility of the attention mechanism in high-dimensional data.

4.1.2. Parameter analysis of global attribute attention

According to Definition 6, the parameter γ plays a crucial role in influencing global attribute attention. To investigate the impact of attention parameter γ on classification performance in detail, we set the parameter within the range of [0, 9] with a step size of 1 and conducted experiments on twelve datasets. When $\gamma = 0$, the attention for each attribute is set to 1, which makes the model equivalent to the MIFCL-NGA in ablation experiments. The experimental results are shown in Fig. 3. In datasets 1, 2, 3, 5, 6, and 9, the classification performance initially shows a consistent upward trend as the attention parameter increases, although some may exhibit a decline at later stages. However, for datasets 4, 7, 8 and 10, global attribute attention has no impact on classification performance which is further confirmed in Table 2, as the classification process for these datasets relies more heavily on local concept attention. The accuracy of both the high-dimensional tumor datasets 11 and 12 show a trend of flat and then decreasing. Overall, global attribute attention provides the model with improved classification performance. The values of each dataset parameter γ in experiments below are obtained at the maximum accuracy in Fig. 3.

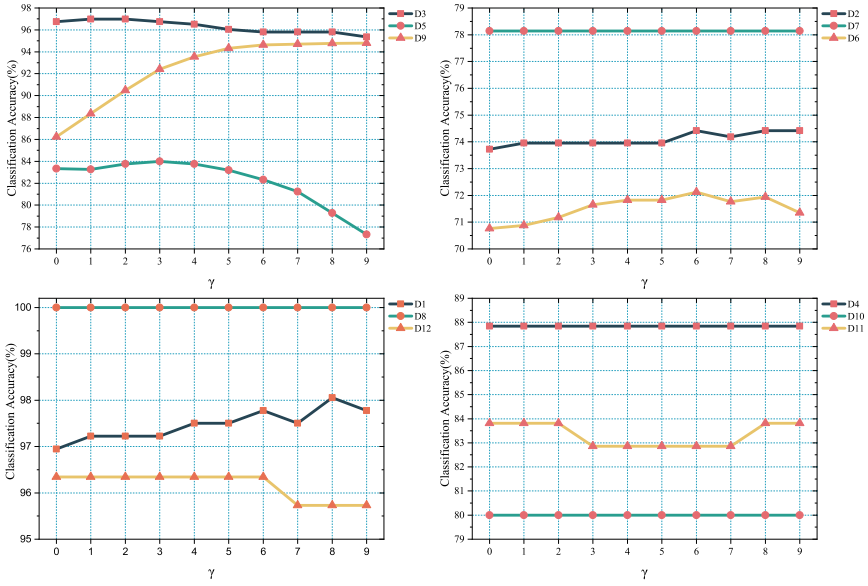

Fig. 3. Classification performance of different global attribute attention parameter γ on twelve datasets.

Table 3
Comparison of accuracy (Mean ± Standard Deviation%) among MIFCL-A and eight classic classification algorithms.

ID	MIFCL-A	KNN	DT	NB	Bagging	AdaBoost	ELM	RBFN	IBK
1	**97.78 ± 1.94**	96.11 ± 3.09	90.55 ± 3.33	95.83 ± 2.84	93.89 ± 2.72	93.51 ± 3.81	70.00 ± 6.67	97.22 ± 2.78	95.56 ± 2.55
2	**74.41 ± 7.29**	72.56 ± 3.57	71.39 ± 6.16	50.69 ± 6.72	24.18 ± 4.18	41.08 ± 5.59	38.37 ± 6.35	47.21 ± 5.51	67.91 ± 5.68
3	**96.98 ± 2.83**	95.11 ± 2.83	94.19 ± 2.60	95.34 ± 2.75	80.69 ± 10.61	79.23 ± 5.21	81.39 ± 5.20	87.44 ± 3.77	94.65 ± 2.09
4	86.84 ± 3.01	78.32 ± 1.90	77.04 ± 3.47	**91.52 ± 1.64**	83.75 ± 2.51	90.80 ± 2.41	48.08 ± 3.45	68.88 ± 4.24	82.32 ± 3.22
5	**86.08 ± 3.05**	83.11 ± 3.13	81.08 ± 2.11	80.50 ± 3.03	79.20 ± 2.24	85.02 ± 2.51	85.21 ± 2.45	80.87 ± 9.64	83.91 ± 2.35
6	**71.94 ± 2.92**	69.05 ± 2.86	69.41 ± 2.98	44.35 ± 3.58	51.00 ± 2.63	58.03 ± 4.56	41.00 ± 3.24	40.53 ± 2.62	69.47 ± 3.40
7	**78.46 ± 1.59**	76.37 ± 4.84	77.37 ± 1.45	76.62 ± 1.31	77.39 ± 2.05	77.14 ± 0.99	77.80 ± 1.55	77.73 ± 1.69	77.05 ± 5.10
8	**100.00 ± 0.00**	99.99 ± 0.02	**100.00 ± 0.00**	92.32 ± 0.50	93.48 ± 0.34	**100.00 ± 0.00**	96.29 ± 0.89	82.21 ± 9.17	**100.00 ± 0.00**
9	94.70 ± 0.24	89.58 ± 0.70	94.38 ± 0.66	**95.98 ± 0.25**	94.72 ± 0.26	95.84 ± 0.21	94.35 ± 0.64	88.73 ± 0.79	90.57 ± 0.52
10	**80.00 ± 10.95**	75.00 ± 17.80	63.00 ± 14.94	30.00 ± 12.47	31.00 ± 13.70	54.67 ± 15.96	21.00 ± 13.75	18.00 ±8.71	76.00 ± 12.00
11	**83.57 ± 8.06**	76.86 ± 7.06	60.00 ± 5.22	77.71 ± 5.99	50.28 ± 7.28	47.69 ± 7.73	27.43 ± 7.25	18.29 ± 6.41	73.14 ± 6.41
12	**95.12 ± 1.09**	94.86 ± 2.00	84.15 ± 4.77	80.49 ± 3.81	89.27 ± 3.29	81.31 ± 4.25	47.80 ± 8.94	67.07 ± 4.53	94.87 ± 2.77
Average	**87.1567**	83.9100	80.2133	75.9458	70.7375	75.3600	60.7267	64.5150	83.7875

Table 4
Comparison of accuracy (Mean ± Standard Deviation%) among MIFCL-A and eight fuzzy-based algorithms.

ID	MIFCL-A	IF-KNN	FRNN	PFKNN	CFKNN	GAFuzzyKNN	FuzzyKNN	EF-KNN-IVF	FuzzyNPC
1	**97.78 ± 1.94**	96.11 ± 3.09	95.56 ± 3.56	96.67 ± 2.42	94.17 ± 2.90	94.72 ± 1.94	95.28 ± 3.52	95.56 ± 2.83	77.00 ± 14.33
2	**74.41 ± 7.29**	67.44 ± 7.50	52.56 ± 9.02	44.88 ± 7.06	60.00 ± 7.04	64.19 ± 7.22	60.47 ± 7.99	66.98 ± 6.72	37.20 ± 9.64
3	**96.98 ± 2.83**	94.88 ± 2.50	68.37 ± 3.63	93.95 ± 4.05	93.02 ± 5.30	93.95 ± 4.05	92.09 ± 4.90	93.02 ± 4.53	75.11± 4.77
4	86.84 ± 3.01	79.52 ± 4.11	**88.32 ± 2.09**	72.00 ± 4.71	75.44 ± 3.73	84.32 ± 2.35	80.00 ± 1.75	81.84 ± 3.33	62.00 ± 10.49
5	**86.08 ± 3.05**	83.19 ± 1.80	83.11 ± 4.62	85.72 ± 2.05	74.20 ± 2.05	85.14 ± 2.10	84.64 ± 2.33	83.33 ± 3.15	75.07 ± 6.88
6	**71.94 ± 2.92**	71.05 ± 2.84	53.53 ± 3.10	44.53 ± 2.82	65.29 ± 3.13	58.18 ± 3.52	71.47 ± 1.61	68.76 ± 3.46	32.58 ± 4.63
7	**78.46 ± 1.59**	75.56 ± 5.03	75.92 ± 1.56	74.01 ± 1.34	69.75 ± 11.77	68.48 ± 10.13	73.72 ± 8.95	78.34 ± 1.75	73.67 ± 4.58
8	**100.00 ± 0.00**	99.99 ± 0.02	94.14 ± 0.49	87.78 ± 0.56	99.94 ± 0.08	**100.00 ± 0.00**	99.94 ± 0.08	99.99 ± 0.02	68.64 ± 8.73
9	**94.70 ± 0.24**	89.19 ± 0.74	63.29 ± 0.88	87.42 ± 0.54	83.05 ± 0.51	91.11 ± 0.63	89.38 ± 0.65	90.29 ± 0.77	68.09 ± 4.81
10	**80.00 ± 10.95**	73.00 ± 15.52	71.00 ± 11.36	75.00 ± 12.04	76.00 ± 13.56	76.00 ± 11.14	72.00 ± 11.66	73.00 ± 17.34	41.00 ± 16.40
11	**83.57 ± 8.06**	83.14 ± 6.81	82.29 ± 5.54	68.57 ± 6.13	70.00 ± 4.82	73.71 ± 4.92	79.14 ± 4.95	59.71 ± 9.68	33.14 ± 8.87
12	**95.12 ± 1.09**	93.41 ± 2.89	94.39 ± 3.78	94.15 ± 3.96	94.39 ± 2.89	**95.12 ± 2.89**	94.63 ± 3.74	93.41 ± 2.19	71.95 ± 9.33
Average	**87.1567**	83.8733	76.8733	77.0567	79.6042	82.0767	82.7300	82.0192	59.6208

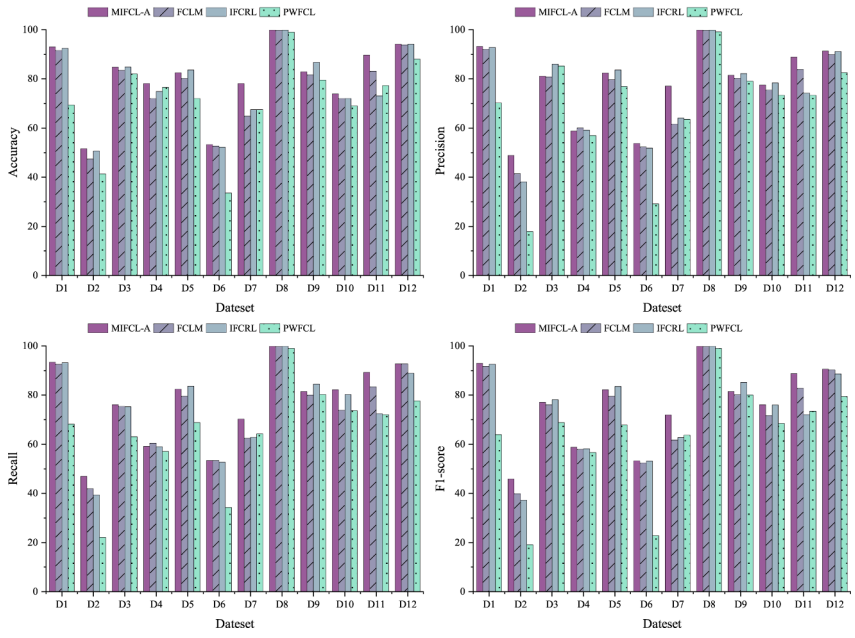

Fig. 4. Performance comparison under 10 % Gaussian noise contamination (Accuracy/Precision/Recall/F1-score).

4.2. Classification performance analysis

In this section, we compare a total of nineteen classification methods, including eight classical machine learning classifiers-K-NearestNeighbor (KNN) [43], Decision Tree (DT) [44], Naive Bayes (NB) [45], Bootstrap Aggregating (Bagging) [46], Adaptive Boosting (AdaBoost) [47], Extreme Learning Machine (ELM) [48], Radial Basis Function Network (RBFN) [49], and Instance-base KNN (IBK) [50]-along with eight fuzzy-based classifiers: Intuitionistic Fuzzy KNN (IF-KNN) [51], Fuzzy-rough KNN (FRNN) [52], Pruned Fuzzy KNN (PFKNN) [53], Condensed Fuzzy KNN (CFKNN) [54], Genetic Algorithm Fuzzy KNN (GAFuzzyKNN) [55], Fuzzy KNN (FuzzyKNN) [56], Evolutionary FKNN-Interval-Valued Fuzzy Sets (EF-KNN-IVFS) [57], and Fuzzy Nearest Prototype Classifier (FuzzyNPC) [56]. Additionally, we examine three advanced CCL classifiers: Fuzzy Concept-cognitive Learning Model (FCLM) [14], Interval-intent Fuzzy Concept-cognitive Learning Model

Table 5
Average classification performance (mean ± standard deviation %) of four CCL algorithms.

Dataset	Mechanism	Accuracy	Precision	Recall	F1-score
Wine	MIFCL-A	**97.50 ± 1.94**	**97.39 ± 2.25**	**97.87 ± 1.57**	**97.48 ± 2.03**
	FCLM	94.44 ± 3.93	94.39 ± 4.03	95.24 ± 3.32	94.23 ± 4.16
	IFCRL	96.39 ± 2.17	95.94 ± 2.22	97.07 ± 1.59	96.27 ± 2.00
	PWFCL	86.11 ± 11.11	85.20 ± 14.73	88.02 ± 9.67	85.58 ± 12.96
Glass	MIFCL-A	**75.58 ± 5.01**	70.34 ± 6.05	**71.68 ± 11.73**	**69.07 ± 8.89**
	FCLM	65.35 ± 525	63.99 ± 8.82	64.00 ± 6.88	61.38 ± 6.12
	IFCRL	73.72 ± 6.98	**72.41 ± 11.73**	71.34 ± 7.05	67.91 ± 7.54
	PWFCL	33.95 ± 7.29	5.79 ± 1.37	17.00 ± 1.00	8.56 ± 1.60
Newthyroid	MIFCLL-A	**95.58 ± 3.81**	94.19 ± 4.62	**95.28 ± 4.13**	**94.27 ± 4.51**
	FCLM	95.11 ± 3.36	**97.54 ± 2.11**	89.24 ± 7.48	92.45 ± 5.21
	IFCRL	93.48 ± 2.90	93.82 ± 4.58	90.46 ± 4.58	91.34 ± 3.38
	PWFCL	94.65 ± 3.13	96.09 ± 3.76	89.91 ± 6.15	92.07 ± 4.63
Balance	MIFCL-A	86.56 ± 5.03	**87.76 ± 10.06**	68.08 ± 4.96	**69.70 ± 6.70**
	FCLM	**88.16 ± 3.21**	80.15 ± 13.59	**68.57 ± 3.80**	69.66 ± 5.29
	IFCRL	72.72 ± 3.91	57.00 ± 2.44	56.02 ± 2.75	56.13 ± 2.43
	PWFCL	60.39 ± 12.16	59.43 ± 3.95	55.21 ± 6.19	51.25 ± 7.23
Australian	MIFCL-A	**85.94 ± 1.86**	**85.76 ± 1.88**	**85.50 ± 1.86**	**85.59 ± 1.84**
	FCLM	82.02 ± 2.58	82.28 ± 3.03	80.82 ± 3.04	81.16 ± 2.96
	IFCRL	84.85 ± 2.48	84.63 ± 2.42	84.89 ± 2.61	84.63 ± 2.47
	PWFCL	84.20 ± 3.02	84.33 ± 2.91	83.54 ± 3.13	83.73 ± 3.07
Vehicle	MIFCL-A	**72.29 ± 2.89**	**71.48 ± 3.05**	**72.53 ± 2.78**	**71.70 ± 2.90**
	FCLM	69.00 ± 3.69	68.97 ± 2.76	69.63 ± 2.51	69.17 ± 2.67
	IFCRL	71.58 ± 1.68	70.64 ± 1.87	71.42 ± 1.35	70.15 ± 1.88
	PWFCL	36.29 ± 5.67	31.33 ± 11.93	35.82 ± 5.19	30.33 ± 7.72
Titanic	MIFCL-A	**78.69 ± 1.62**	**77.85 ± 2.05**	**70.57 ± 1.84**	**71.12 ± 2.03**
	FCLM	62.43 ± 16.41	65.75 ± 7.28	62.35 ± 6.06	57.87 ± 13.72
	IFCRL	46.64 ± 1.72	59.73 ± 1.82	57.41 ± 1.49	45.85 ± 1.91
	PWFCL	40.99 ± 14.61	30.44 ± 27.17	50.58 ± 1.23	29.47 ± 8.59
Mushroom	MIFCL-A	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**
	FCLM	99.99 ± 0.02	99.99 ± 0.02	99.99 ± 0.03	99.99 ± 0.02
	IFCRL	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**
	PWFCL	99.99 ± 0.02	99.99 ± 0.02	99.99 ± 0.02	99.99 ± 0.02
EGSS	MIFCL-A	**94.33 ± 0.04**	**94.19 ± 0.05**	**92.94 ± 0.19**	**93.51 ± 0.10**
	FCLM	86.02 ± 0.13	85.16 ± 0.07	84.62 ± 0.13	84.87 ± 0.07
	IFCRL	91.88 ± 0.56	91.85 ± 0.49	90.27 ± 0.81	90.97 ± 0.69
	PWFCL	80.13 ± 0.74	79.22 ± 0.48	81.47 ± 0.64	79.44 ± 0.63
GLIOMA	MIFCL-A	80.00 ± 10.95	**81.04 ± 12.58**	**80.17 ± 16.03**	**78.21 ± 14.36**
	FCLM	69.00 ± 9.43	69.17 ± 13.42	69.21 ± 17.42	65.72 ± 14.29
	IFCRL	**83.00 ± 11.87**	79.70 ± 17.93	78.13 ± 20.57	76.50 ± 20.34
	PWFCL	75.00 ± 9.22	73.31 ± 13.73	71.79 ± 11.29	68.95 ± 11.03
TOX_171	MIFCL-A	84.00 ± 3.18	83.98 ± 4.03	83.15 ± 3.92	83.93 ± 4.19
	FCLM	**86.29 ± 5.54**	**86.61 ± 7.07**	**88.72 ± 4.25**	**86.26 ± 6.82**
	IFCRL	80.85 ± 5.43	81.95 ± 5.12	81.80 ± 6.09	80.63 ± 6.15
	PWFCL	76.00 ± 6.89	77.97 ± 5.41	76.95 ± 6.11	76.20 ± 6.17
Lung	MIFCL-A	**95.12 ± 1.09**	**92.71 ± 4.96**	**92.63 ± 4.82**	**92.18 ± 4.61**
	FCLM	94.15 ± 1.62	92.45 ± 6.28	91.78 ± 3.92	90.83 ± 3.81
	IFCRL	94.15 ± 4.11	93.69 ± 6.82	90.74 ± 8.85	91.28 ± 7.86
	PWFCL	88.44 ± 4.28	83.29 ± 14.97	79.63 ± 10.32	79.84 ± 11.66

(IFCRL) [42], and Dynamic Updating Mechanism Progressive Weighted Fuzzy Concept-cognitive Learning Model (DMPWFC) [36]. Since this paper does not deal with the dynamic update of objects, the algorithm in Reference 13 is de-dynamicized and simply referred to as PWFCL.

4.2.1. Comparison with classical machine learning method

We compare the performance of MIFCL-A with eight classic machine learning classifiers using accuracy as the evaluation metric. The results, as shown in Table 3 (optimal values are shown in bold), demonstrate that MIFCL-A outperforms all other models across the most datasets.

Notably, MIFCL-A, DT, AdaBoost, and IBK achieve 100 % classification accuracy on dataset 8, demonstrating perfect classification capability for specific datasets. Crucially, MIFCL-A outperforms all competitors by achieving the highest accuracy in ten datasets, while NB leads in only two. With an average accuracy across twelve datasets that exceeds second-place KNN by 3.25 % and third-place IBK by 3.37 %, MIFCL-A demonstrates superior classification performance. Furthermore, our model excels across all three high-dimensional tumor datasets, confirming its robustness in handling complex, high-dimensional data.

4.2.2. Comparison with fuzzy-based method

Since MIFCL-A leverages interval-intent fuzzy concepts, we compare it against eight fuzzy-based KNN variants. As shown in Table 4 (optimal values bolded), the MIFCL-A model demonstrates superior classification performance across twelve diverse datasets. It achieves the highest accuracy in dataset 8, with particularly significant margins in dataset 2 (10.22 % higher than GAFuzzyKNN) and dataset 3 (28.61 % higher than FRNN). Notably, MIFCL-A maintains robust performance with consistently low standard deviations (e.g., 0.24 in dataset 9), indicating exceptional stability. With an average accuracy of 87.1567 %, it outperforms the second-best algorithm (IF-KNN at 83.8733 %) by 3.28 percentage points and the weakest comparator (FuzzyNPC at 59.6208 %) by 27.54

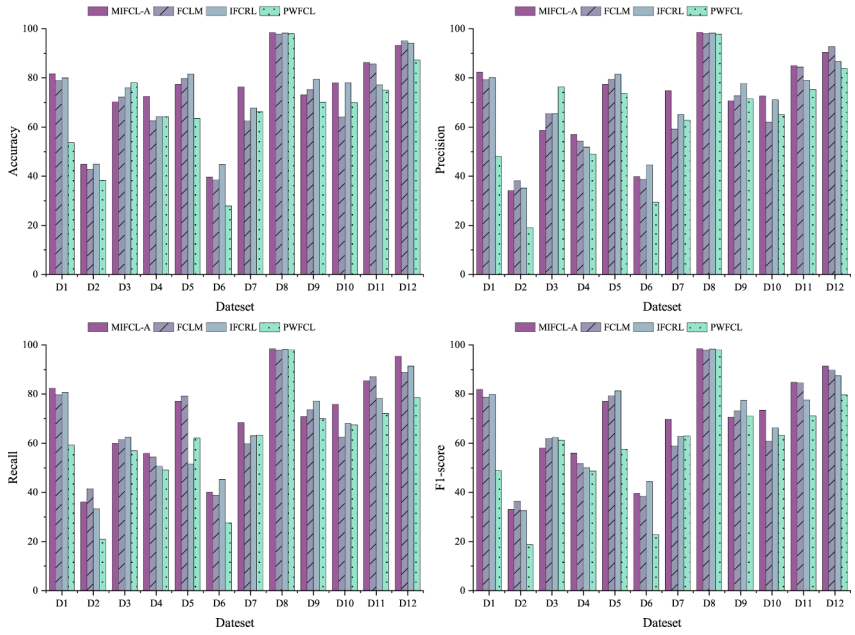

Fig. 5. Performance comparison under 20 % Gaussian noise contamination (Accuracy/Precision/Recall/F1-score).

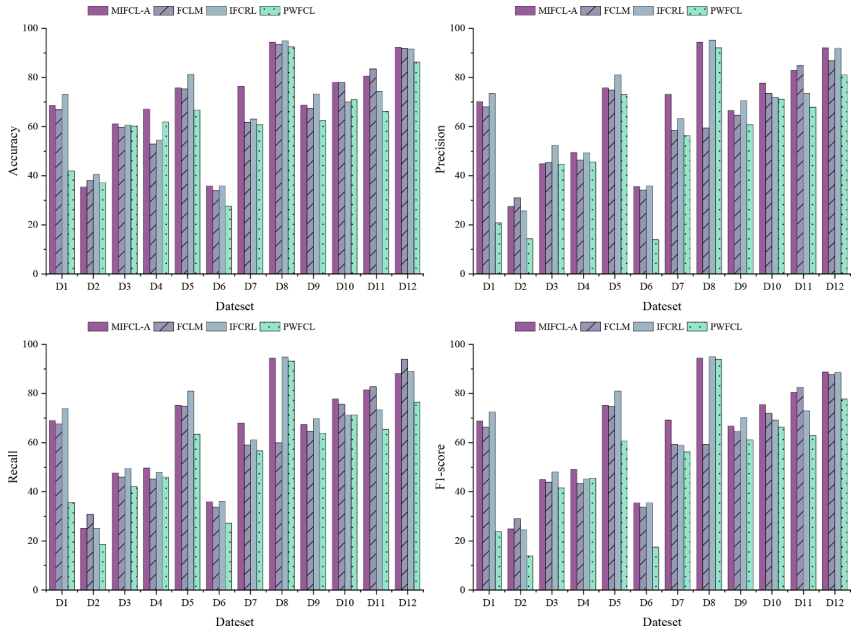

Fig. 6. Performance comparison under 30 % Gaussian noise contamination (Accuracy/Precision/Recall/F1-score).

percentage points. While FRNN and GAFuzzyKNN attain optimal results on one and two datasets respectively, MIFCL-A demonstrates superior overall accuracy - exceeding these models by 10.28 % and 5.08 % respectively. Furthermore, MIFCL-A achieves the best average rank (1.08) in fuzzy classification comparisons. This performance advantage extends to high-dimensional tumor datasets, where MIFCL-A maintains exceptional results against other fuzzy methods.

4.2.3. Comparison with CCL method

To further validate the effectiveness of the proposed model, we compare MIFCL-A with three advanced CCL models across four metrics: accuracy, precision, recall, and F1-score. The results are shown in Table 5, with the best values highlighted in bold. MIFCL-A achieved the best results in accuracy, precision, recall and F1 score 9, 9, 10 and 11 times, respectively, across nine datasets. This demonstrates that MIFCL-A outperforms existing CCL models, offering superior classification performance and better stability.

To evaluate the robustness of the models, we introduce Gaussian noise at varying intensities (10 %, 20 %, and 30 %) into the dataset to analyze the performance of the four CCL models. The models' performance metrics, including accuracy, precision, recall, and F1-score, are presented in Figs. 4–6. Comprehensive experimental results demonstrate that MIFCL-A achieved superior performance, attaining the highest scores in 25, 20, 19, and 20 instances across the four evaluation metrics under the three different noise conditions, respectively. In certain cases, adding moderate noise appears to improve model performance.

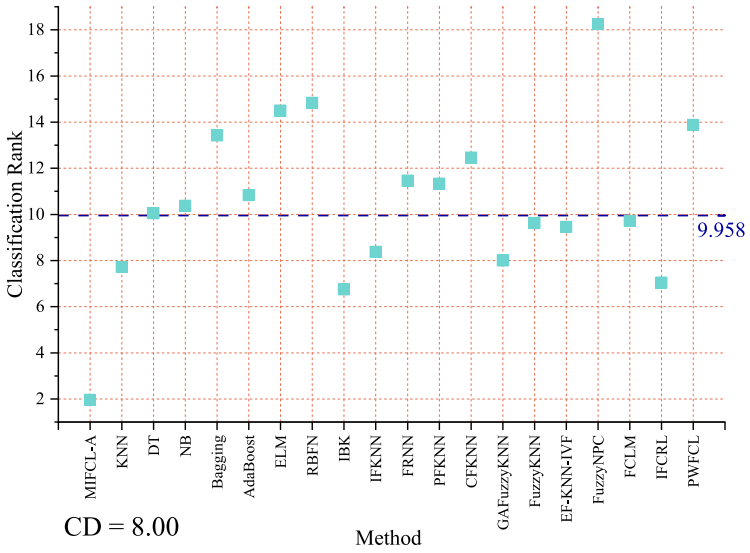

Fig. 7. The comprehensive ranking chart of MIFCL-A compared with nineteen other algorithms (MIFCL-A demonstrates significant differences compared to the algorithms positioned above the purple dashed line).

This effect may occur because: the noise helps smooth decision boundaries, acting as an implicit regularizer or in high-dimensional spaces with limited samples, the noise effectively expands the data coverage.

4.2.4. Significance analysis

We conduct the Friedman test to assess whether there are significant differences among the performance of the various algorithms. The formula for the F-statistic in Friedman's test is as follows:

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} \sim F(k-1,(k-1)(N-1)),$$

where $\chi_F^2 = \frac{12}{Nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3N(k+1)$, k and N are the number of different algorithms and datasets, respectively. $R_i = \sum_{j=1}^{N} r_j^i$ indicates the sum rank of i-th algorithm on all the datasets, and r_j^i indicates the rank of i-th algorithm on j-th dataset.

We propose the null hypothesis (H_0) that there is no significant difference in the performance of the different algorithms at a significance level of 0.05. Based on the test results, the p-value (0.0008) is smaller than the significance level of 0.05, leading to the rejection of the null hypothesis which indicates that there are significant differences in the overall performance of the algorithms. Additionally, the F-value $F_F = 6.65 > F(19, 209) = 1.63$, further supporting this conclusion.

To further explore these differences, we performed the Nemenyi post-hoc test, which is commonly used following the Friedman test when there are multiple comparisons. The Nemenyi test is designed to identify which specific pairs of algorithms show statistically significant differences in their performance. The critical range critical difference (CD) formula for the difference between the average order values is calculated from the Nemenyi test as follows:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}},$$

where q_α is the critical value corresponding to the significance level, k and N are the number of different algorithms and datasets, respectively. In this paper, q_α is determined to be 3.314 based on the degrees of freedom and significance level of the test and $CD = 8$. The CD diagram, shown in Fig. 7, provides a clear visual representation of the differences between MIFCL-A and other algorithms. The results indicate that MIFCL-A exhibits significant differences compared to eleven algorithms: DT, NB, Bagging, AdaBoost, ELM, RBFN, FRNN, PFKNN, CFKNN, FuzzyNPC and PWFCL.

5. Conclusion

This paper presents a multi-granularity interval-intent fuzzy concept-cognitive learning model with multi-level attention mechanism, which simulates the human cognitive process of grasping the big picture while focusing on local details. Multi-granularity concepts are first acquired separately by considering a global perspective, which captures overall contextual information, and a local perspective, which focuses on finer boundary differences. Building on this, a multi-level attention mechanism is introduced, where global attribute attention captures the overall context and large-scale patterns, while local concept attention filters local information, emphasizing key boundary concepts and ignoring irrelevant details. Futhermore, adaptive concept clustering serves as a pivotal step in simulating the human cognitive process of alternating between coarse-granularity and fine-granularity understanding, which allows the model to dynamically adjust the level of granularity. Meanwhile, the new concepts in the clustering process strictly adhere to concept definitions, ensuring consistency throughout the cognitive process. Finally, extensive experiments validate the model's feasibility and effectiveness.

Multi-granularity concepts are used to represent knowledge at different levels. MIFCL-A, while performing multi-granularity knowledge representation, introduces a multi-level attention mechanism to achieve an adaptive concept updating and generation process. However, it is important to note that when dealing with datasets with numerous decisions, the high overlap of attributes between decisions often leads to lower effectiveness of global attribute attention. Additionally, the inter-cluster fusion in the adaptive concept cluster generation process requires further consideration. In the future, we will continue to think about and promote the research of relevant work.

CRediT authorship contribution statement

Yi Ding: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation; **Weihua Xu:** Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Data availability

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

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.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NOs. 62376229, 12371465), the Natural Science Foundation of Chongqing (NOs. CSTB2023NSCQ-LZX0027, CSTB2023NSCQ-MSX1063) and Southwest University Graduate Research Innovation Project (NO. SWUB25086).

References

[1] R. Hastie, R.M. Dawes, Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making, Sage, 2010.
[2] A. Bargiela, W. Pedrycz, Granular computing, in: Handbook on Computer Learning and Intelligence: Volume 2: Deep Learning, Intelligent Control and Evolutionary Computation, World Scientific, 2022, pp. 97–132.
[3] L.A. Zadeh, Fuzzy sets, Inf. Contr. 8 (3) (1965) 338–353.
[4] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[5] Y. Yao, Three-way decision and granular computing, Int. J. Approx. Reason. 103 (2018) 107–123.
[6] R. Wille, Concept lattices and conceptual knowledge systems, Comput. Math. Appl. 23 (6–9) (1992) 493–515.
[7] Y. Wang, On concept algebra: a denotational mathematical structure for knowledge and software modeling, Int. J. Cognit. Inf. Natural Intell. 2 (2) (2008) 1–19.
[8] R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, in: Formal Concept Analysis: 7th International Conference, ICFCA 2009 Darmstadt, Germany, May 21–24, 2009 Proceedings 7, Springer, 2009, pp. 314–339.
[9] T. Zhang, M. Rong, H. Shan, M. Liu, Stability analysis of incremental concept tree for concept cognitive learning, Int. J. Mach. Learn. Cybern. 13 (1) (2022) 11–28.
[10] F. Hao, D. Park, G. Min, Y. Jeong, J. Park, k-Cliques mining in dynamic social networks based on triadic formal concept analysis, Neurocomputing 209 (2016) 57–66.
[11] D.H. Jonassen, T.C. Reeves, N. Hong, D. Harvey, K. Peters, Concept mapping as cognitive learning and assessment tools, J. Interact. Learn. Res. 8 (3) (1997) 289.
[12] R. Belohlavek, Introduction to Formal Concept Analysis, Palacky University, Department of Computer Science, Olomouc 47 (2008).
[13] H. Zhi, J. Li, Granule description based on formal concept analysis, Knowl. Based Syst. 104 (2016) 62–73.
[14] Y. Mi, Y. Shi, J. Li, W. Liu, M. Yan, Fuzzy-based concept learning method: exploiting data with fuzzy conceptual clustering, IEEE Trans. Cybern. 52 (1) (2020) 582–593.
[15] X. Deng, J. Li, Y. Qian, J. Liu, An emerging incremental fuzzy concept-cognitive learning model based on granular computing and conceptual knowledge clustering, IEEE Trans. Emerg. Top. Comput. Intell. 8 (3) (2024) 2417–2432.
[16] D. Guo, W. Xu, Y. Qian, W. Ding, M-FCCL: memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[17] K. Yuan, W. Xu, W. Li, W. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inf. Sci. 584 (2022) 127–147.
[18] J. Li, C. Huang, J. Qi, Y. Qian, W. Liu, Three-way cognitive concept learning via multi-granularity, Inf. Sci. 378 (2017) 244–263.
[19] L. Wei, L. Liu, J. Qi, T. Qian, Rules acquisition of formal decision contexts based on three-way concept lattices, Inf. Sci. 516 (2020) 529–544.
[20] W. Xu, D. Guo, Y. Qian, W. Ding, Two-way concept-cognitive learning method: a fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. 31 (6) (2022) 1885–1899.
[21] W. Xu, D. Guo, J. Mi, Y. Qian, K. Zheng, W. Ding, Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural Netw. Learn. Syst. 34 (10) (2023) 6798–6812.
[22] Z. Liu, J. Li, X. Zhang, X. Wang, Incremental incomplete concept-cognitive learning model: a stochastic strategy, IEEE Trans. Neural Netw. Learn. Syst. (2023).
[23] W.X. Zhang, W.H. Xu, Cognitive model based on granular computing, Chin. J. Eng. Math. 24 (6) (2007) 957–971.
[24] X. Xie, W. Xu, J. Li, A novel concept-cognitive learning method: a perspective from competences, Knowl. Based Syst. 265 (2023) 110382.
[25] R. Rafieisangari, N. Shiri, A multi-adaptive neuro-fuzzy inference system with variable thresholds for heartbeat classification, Artif. Intell. Health 1 (4) (2024) 43–60.
[26] C. Lin, S. Wang, Fuzzy support vector machines, IEEE Trans. Neural Netw. 13 (2) (2002) 464–471.
[27] C. Zhang, E.C.C. Tsang, W. Xu, Y. Lin, L. Yang, J. Wu, Dynamic updating variable precision three-way concept method based on two-way concept-cognitive learning in fuzzy formal contexts, Inf. Sci. 655 (2024) 119818.
[28] J. Wang, W. Xu, W. Ding, Y. Qian, Multi-view fuzzy concept-cognitive learning with high-order information fusion of fuzzy attributes, IEEE Trans. Fuzzy Syst. 32 (12) (2024) 6965–6978.
[29] C. Thomas, R.R. Prasad, Health-care app detection using optimized clustering, Artif. Intell. Health 1 (4) (2024) 16–29.
[30] B. Amiri, R. Karimianghadim, A novel text clustering model based on topic modelling and social network analysis, Chaos Solitons Fract. 181 (2024) 114633.
[31] D. Guo, W. Xu, Fuzzy-based concept-cognitive learning: an investigation of novel approach to tumor diagnosis analysis, Inf. Sci. 639 (2023) 118998.
[32] Z. Liu, J. Li, X. Zhang, X. Wang, Multi-level information fusion for missing multi-label learning based on stochastic concept clustering, Inf. Fusion 115 (2025) 102775.
[33] M.I. Posner, S.E. Petersen, et al., The attention system of the human brain, Annu. Rev. Neurosci. 13 (1) (1990) 25–42.
[34] W. Xu, Y. Chen, Multi-attention concept-cognitive learning model: a perspective from conceptual clustering, Knowl. Based Syst. 252 (2022) 109472.
[35] R. Belohlavek, J. Macko, Selecting important concepts using weights, in: International Conference on Formal Concept Analysis, Springer, 2011, pp. 65–80.
[36] C. Zhang, E.C.C. Tsang, W. Xu, Y. Lin, L. Yang, Incremental concept-cognitive learning approach for concept classification oriented to weighted fuzzy concepts, Knowl. Based Syst. 260 (2023) 110093.
[37] D. Guo, W. Xu, W. Ding, Y. Yao, X. Wang, W. Pedrycz, Y. Qian, Concept-cognitive learning survey: mining and fusing knowledge from data, Inf. Fusion 109 (2024) 102426.
[38] S. Xia, G. Wang, X. Gao, X. Lian, Granular-ball computing: an efficient, robust, and interpretable adaptive multi-granularity representation and computation method, arXiv preprint arXiv:2304.11171. (2023).
[39] J. Xie, C. Hua, S. Xia, Y. Cheng, G. Wang, X. Gao, W-GBC: an adaptive weighted clustering method based on granular-ball structure, in: 2024 IEEE 40th International Conference on Data Engineering (ICDE), IEEE, 2024, pp. 914–925.
[40] S. Xia, B. Shi, Y. Wang, J. Xie, G. Wang, X. Gao, GBCT: efficient and adaptive clustering via granular-ball computing for complex data, IEEE Trans. Neural Netw. Learn. Syst. 36 (7) (2025) 12159–12172.
[41] D. Xia, G. Wang, Q. Zhang, J. Yang, S. Xia, Three-way approximations fusion with granular-ball computing to guide multi-granularity fuzzy entropy for feature selection, IEEE Trans. Fuzzy Syst. 32 (10) (2024) 5963–5977.
[42] Y. Ding, W. Xu, W. Ding, Y. Qian, IFCRL: interval-intent fuzzy concept re-cognition learning model, IEEE Trans. Fuzzy Syst. 32 (6) (2024) 3581–3593.
[43] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN model-based approach in classification, in: On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3–7, 2003. Proceedings, Springer, 2003, pp. 986–996.
[44] Y. Song, L.U. Ying, Decision tree methods: applications for classification and prediction, Shanghai Arch. Psychiatry 27 (2) (2015) 130.
[45] K.P. Murphy, Naive bayes classifiers, Univers. Br. Columb. 18 (60) (2006) 1–8.
[46] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140.
[47] T. Hastie, S. Rosset, J. Zhu, H. Zou, Multi-class adaboost, Stat. Interface 2 (3) (2009) 349–360.
[48] G. Huang, Q. Zhu, C. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.
[49] D. Lowe, D. Broomhead, Multivariable functional interpolation and adaptive networks, Complex Syst. 2 (3) (1988) 321–355.
[50] D.W. Aha, D. Kibler, M.K. Albert, Instance-based learning algorithms, Mach. Learn. 6 (1991) 37–66.
[51] L. Kuncheva, An intuitionistic fuzzy k-nearest neighbors rule, Notes Intuit Fuzzy Sets 1 (1995).
[52] M. Sarkar, Fuzzy-rough nearest neighbor algorithms in classification, Fuzzy Sets Syst. 158 (19) (2007) 2134–2152.
[53] M. Arif, M.U. Akram, F.A. Afsar, Pruned fuzzy K-nearest neighbor classifier for beat classification, J. Biomed. Sci. Eng. 3 (04) (2010) 380.
[54] J. Zhai, N. Li, M. Zhai, The condensed fuzzy k-nearest neighbor rule based on sample fuzzy entropy, in: 2011 International Conference on Machine Learning and Cybernetics, 1, IEEE, 2011, pp. 282–286.
[55] X. Hu, C. Xie, Improving fuzzy k-NN by using genetic algorithm, J. Comput. Inf. Syst. 1 (2) (2005) 203–213.
[56] J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nearest neighbor algorithm, IEEE Trans. Syst. Man Cybern. (4) (1985) 580–585.
[57] J. Derrac, F. Chiclana, S. García, F. Herrera, Evolutionary fuzzy k-nearest neighbors algorithm using interval-valued fuzzy sets, Inf. Sci. 329 (2016) 144–163.