Fuzzy Sets and Systems 529 (2026) 109724

Contents lists available at ScienceDirect

Fuzzy Sets and Systems

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

RF-WFCCL: A random forest-driven weighted fuzzy concept cognitive learning

Weihua Xu *, Chongze Zhang

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

ARTICLE INFO

Keywords:
Conceptual cognitive learning
Random forest
Granular computing
Concept clusters
Object classification

ABSTRACT

Weighted Fuzzy Concept Cognitive Learning (WFCCL) improves concept learning by assigning different weights to attributes. However, existing weighting methods mainly rely on fuzzy entropy and fail to effectively handle the noise, uncertainty, and complexity commonly present in real-world data. Moreover, the key threshold λ for generating Advanced Weighted Fuzzy Concepts (AWFC) has not been thoroughly investigated, which may limit the depth of knowledge discovery and the applicability of the model. To address these issues, this paper proposes a Random Forest-Driven Weighted Fuzzy Concept Cognitive Learning (RF-WFCCL) method. The proposed approach leverages random forests to compute attribute importance, thereby making weight allocation more reflective of data characteristics and reducing the influence of noise and uncertainty. In addition, an adaptive optimization algorithm is designed to automatically determine the approximately optimal threshold $\hat{\lambda}(i)$ for generating AWFC, enhancing the accuracy and applicability of concept cognition. Unlike traditional methods, our model not only improves the adaptability of weight allocation but also introduces a data-driven threshold optimization strategy to facilitate knowledge discovery. Finally, extensive experiments on 12 datasets using 12 classification algorithms demonstrate the feasibility and superiority of the proposed method.

1. Introduction

Cognitive computing [1] is a computer system derived from simulating the human brain, and an integral part of artificial intelligence. Its goal is to address the complexity, uncertainty, and incompleteness inherent in biological systems, thereby enabling processes such as perception, memory, and problem-solving. Over the years, cognitive computing has found widespread applications in fields such as machine learning [2], data science [3], and information processing [4,5].

Information granules, as conceptual forms, are the basic cognitive units in human thought, playing an important role in the process of understanding the world [6]. A concept consists of three parts: extension, intension, and name [7]. The extension refers to a set of objects, while the intension denotes the shared attributes of these objects. The concept name describes both the intension and extension of the concept [8]. Concept learning is one of the essential methods in cognitive learning. In recent years, concept learning has gradually expanded into fields such as granular computing [9,10], rough set theory [11,12], and formal concept analysis [13,14].

Conceptual cognitive learning takes concepts as the fundamental carriers of knowledge, naturally integrating new data into itself by simulating human learning mechanisms, exhibiting good adaptability in dynamic learning environments. With the increase in data types, various conceptual cognitive models have emerged in practical applications, such as formal concepts [15], granular concepts

---

* Corresponding author.
E-mail addresses: chxuwh@gmail.com (W. Xu), a1157260366@163.com (C. Zhang).

https://doi.org/10.1016/j.fss.2025.109724
Received 19 August 2025; Received in revised form 18 November 2025; Accepted 3 December 2025

[16], fuzzy concepts [17], three-way concepts [18] and bidirectional concepts [19–21]. In the foundational domain, conceptual cognitive learning continues to improve. For example, Zhang et al. [16] studied the sufficiency and necessity between objects and attributes, combining granular computing with cognitive learning to simulate human cognitive processes. Yao et al. [22] starting from the perspectives of cognitive informatics and granular computing, explored the learning framework of conceptual cognitive learning. Xu et al. [23] extensively discussed the transformation of arbitrary information granules into sufficient and necessary theories, proposing an effective conversion algorithm, providing new insights into information granule transformation.

The theoretical framework of concept-cognitive learning (CCL) has been gradually improved. To address classification problems, many scholars have proposed various CCL models in recent years. Zhang et al. [24] incorporated weights into fuzzy formal contexts and designed a dynamic update mechanism to enhance classification accuracy in dynamic data environments. Wang et al. [25] proposed a multi-view fuzzy concept recognition method that effectively resolves the challenges of concept representation and integration when data are collected from multiple views in real-world scenarios. Ding et al. [26] improved the flexibility of fuzzy concepts by adopting interval representation and introduced a re-cognition learning mechanism to simulate the secondary cognitive process from fuzziness to precision, making it more aligned with real cognitive patterns. Guo et al. [27] proposed a fuzzy-granular three-way CCL method, which introduces a big concept priority principle and a dynamic update mechanism to achieve concept modeling and dynamic knowledge learning in fuzzy environments. Wu et al. [28] established relationships between feature concepts and multi-label concepts, comprehensively considering both extent relevance and intent relevance, and applied this approach to multi-label classification tasks. Zhou et al. [29] proposed a skill assessment method from the perspective of CCL, converting a specific fuzzy skill function into a fuzzy formal context, effectively addressing the limitations of existing skill assessment techniques in handling noisy response data. Xie et al. [30] proposed a concept-cognitive learning method that integrates causal reasoning and attribute weighting, which constructs a causal concept space based on a partial order structure to effectively capture the causal relationships between attributes and decision classes. Huang et al. [31] proposed a fuzzy concept-cognitive learning method that integrates cognitive logic and knowledge space theory, which establishes cognitive surmise relationships among attributes to effectively enhance the representational ability of fuzzy concepts. Ding et al. [32] proposed a multi-granularity interval-intent fuzzy concept-cognitive learning model that incorporates a hierarchical attention mechanism and an adaptive concept clustering method, effectively improving the accuracy of concept learning.

The field of CCL has significant potential for development. Currently, in the fuzzy formal context, fuzzy entropy is commonly used to measure attribute importance. However, data often contain noise, uncertainty, and complexity, while traditional fuzzy entropy-based weight allocation methods exhibit weak robustness to these factors. This may lead to biased weight calculations, ultimately affecting the learning effectiveness of fuzzy concepts. Moreover, in the process of generating AWFC, the selection of threshold directly influences concept quality and the depth of knowledge discovery. However, existing studies have often overlooked the optimization of threshold, limiting the generalization ability of cognitive learning models. Therefore, designing an adaptive threshold selection mechanism that can dynamically adjust based on data characteristics is a key challenge in enhancing the applicability of CCL.

To overcome the limitations of current methods and promote more accurate AWFC derivation, we have introduced an innovative WFCCL approach that addresses the shortcomings of existing cognitive methods. As shown in Fig. 1, this approach mainly comprises three stages: constructing weighted fuzzy concept space (WFCS), building advanced weighted fuzzy concept space (AWFCS), and concept prediction. For clarity, we assume a three-classification problem here. In the first stage, samples are mapped to WFCs, forming two concept spaces (conditional WFCS and decision concept space), where the conditional WFCS consists of three sub-concept spaces. In the second stage, different threshold values $\lambda(i)$ result in different $\mathcal{G}_*^{\lambda}$, and the optimal $\widehat{\mathcal{G}}_*^{\lambda}$ is obtained by finding the optimal threshold $\widehat{\lambda}(i)$. In the third stage, for any given test object x, we can calculate its similarity to each concept in each sub-concept space, and then derive the final predicted label by synthesizing the maximum similarity in each sub-concept space.

The main contributions of this paper are delineated as follows:

1. **Enhanced Attribute Weighting in Fuzzy contexts:** We introduce a method for calculating attribute importance within a fuzzy setting using RF, which quantifies the significance of each attribute. Subsequently, WFCs are constructed, addressing the influence of noise, uncertainty, and complexity in data and leading to a more scientifically grounded weight allocation.
2. **Advancement to Advanced Weighted Fuzzy Concepts:** Leveraging the foundation of WFCs, we extend our approach to develop AWFCs, incorporating the pivotal role of the threshold $\lambda(i)$. An algorithm is proposed to identify an approximately optimal threshold $\widehat{\lambda}(i)$ for each dataset, thereby enhancing cognitive precision, transcending cognitive barriers characterized by stringent similarity criteria for concept clustering, and markedly improving classification efficiency.
3. **Empirical Validation and Threshold Analysis:** We conducted a comprehensive empirical evaluation using data from 12 datasets sourced from the UCI and KEEL repositories. The performance of RF-WFCCL was compared against 12 state-of-the-art classification algorithms in terms of accuracy and standard deviation. Furthermore, we conducted a detailed analysis of the impact of the threshold $\lambda(i)$ on the classification accuracy of RF-WFCCL, substantiating the rationality of our method for threshold determination. The findings demonstrate that RF-WFCCL achieves a higher average accuracy across the evaluated datasets, outperforming other classification methods.

The remaining sections of this paper are organized as follows: In Section 2, we introduce the relevant knowledge of FFCA, RF, and the motivation behind this study. Section 3 investigates the weighting of attributes based on RF, followed by the construction of WFC, and the generation of AWFC based on selected thresholds. Section 4 discusses how to classify newly added objects and the cognitive learning process based on advanced weighted fuzzy space. In Section 5, the effectiveness of the algorithm is evaluated through comparative experiments with 12 classification algorithms on selected datasets. Finally, we conclude this paper by summarizing our

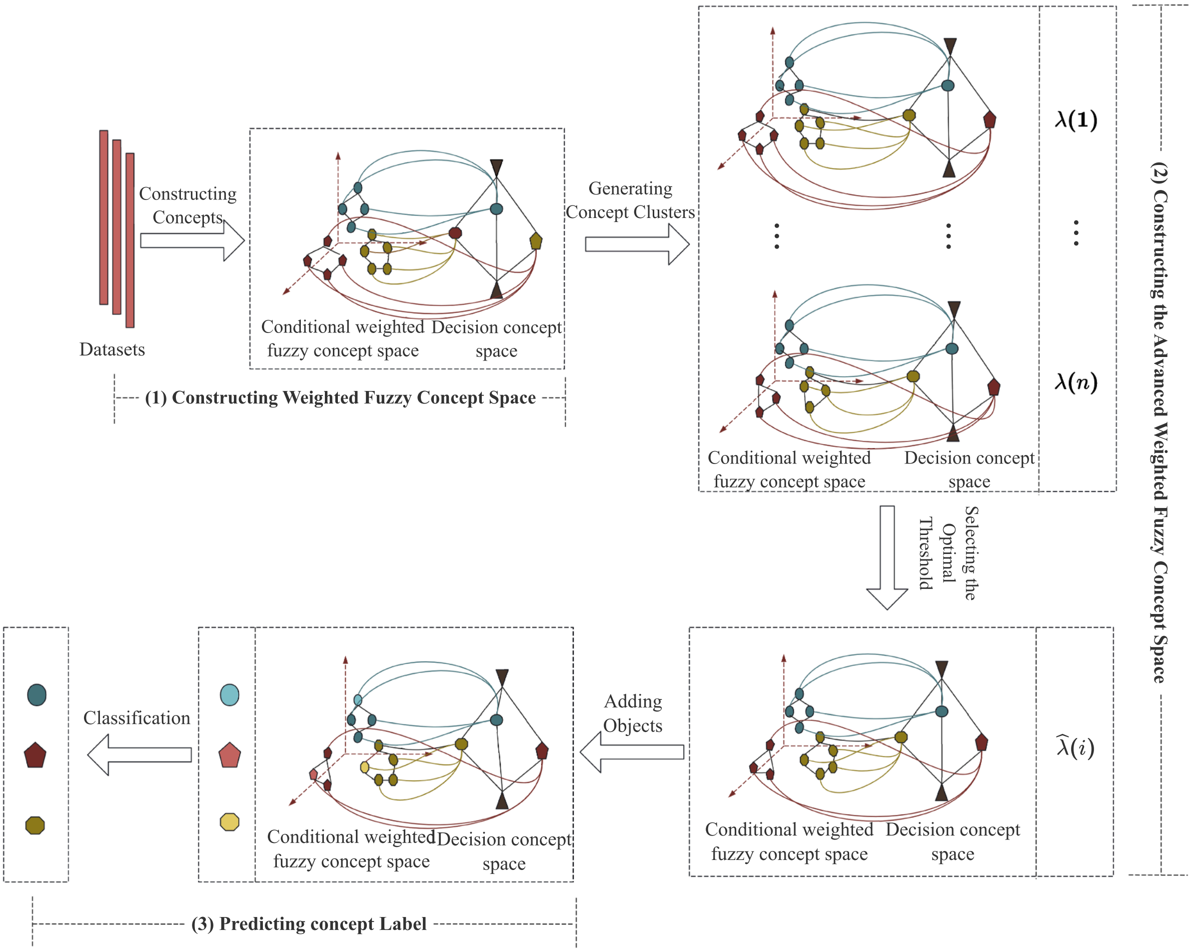


Fig. 1. The entire process of the RF-WFCCL.

work and proposing future research directions. Additionally, to enhance the reader's understanding, Table 1 lists and explains the definitions of the main symbols used in this paper.

2. Preliminaries

In this section, we briefly reviewed some fundamental concepts related to FFCA and the theory of RF.

2.1. Fuzzy formal context analysis (FFCA)

Yahia [33] introduced the FFCA to handle data analysis and knowledge discovery. In practice, it comprises nonempty finite sets of objects and attributes, along with a fuzzy binary relation between the two sets, which are essential for the subsequent discussions.

Presume G is a space, or more accurately a fuzzy set $\widetilde{W}$ of G which is defined as a membership function $\widetilde{W}(\cdot) : G \rightarrow [0,1]$. For any $x \in G$, $\widetilde{W}(x)$ represents the degree of membership of x to $\widetilde{W}$. Then we denote by $\mathcal{F}(G)$ the set of all fuzzy subsets of G.

Let W and V be fuzzy subsets of G. If $\forall\, x \in G$, $\widetilde{W}(x) \leq \widetilde{V}(x)$, then W is called a subset of V, denoted by $W \subseteq V$.

A triplet $(U, M, \widetilde{I})$ is referred to as a fuzzy formal context, where $U = \{x_1, x_2, \ldots, x_n\}$ represents the set of objects and $M = \{b_1, b_2, \ldots, b_m\}$ represents the set of attributes. $\widetilde{I}$ is a fuzzy relation between U and M (i.e. $\widetilde{I} : U \times M \rightarrow [0,1]$) and each $\widetilde{I}(x, b)$ reflects the degree of membership of object x to attribute b.

Definition 1. Yahia et al. [33] Let $(U, M, \widetilde{I})$ be a fuzzy formal context. For $X \subseteq U$ and $\widetilde{A} \in \mathcal{F}(M)$, two operators $F : \mathcal{P}(U) \rightarrow \mathcal{F}(M)$ and $H : \mathcal{F}(M) \rightarrow \mathcal{P}(U)$ are given as follows:

$$F(X, b) = \bigwedge_{x \in X} \tilde{I}(x, b), b \in M, \tag{1}$$

$$H(\tilde{A}) = \{x \in U : \forall b \in M, \tilde{A}(b) \leq \tilde{I}(x, b)\}. \tag{2}$$

Table 1
Meanings of main symbols.

Symbol	Meaning
U	Set of objects.
M	Set of conditional attributes.
$\tilde{I}$	Fuzzy binary relation on $U \times M$.
D	Set of decision attributes.
J	Definite binary relation on $U \times D$.
x	An object.
b	A conditional attribute.
d	A decision attribute.
F, H	A pair of cognitive operators.
N	The number of samples.
$p_{i,k}$	The probability that sample point i belongs to class k.
$\mathcal{K}$	Total number of classes in the sample set.
X	The extension of a fuzzy concept.
$\tilde{A}$	The intension of a fuzzy concept.
w	The weight of a weighted fuzzy concept.
W	The weights of the conditional attributes.
$\theta_{i,j}$	The similarity between weighted fuzzy concepts i and j.
$\mathcal{G}_*$	Weighted fuzzy concept space.
λ	The threshold for weighted fuzzy concept clustering.
$\mathcal{G}_*^{\lambda(i)}$	The advanced weighted fuzzy concept space generated under $\lambda(i)$.
$\mathcal{G}^{\lambda(i),j}$	The j-th concept cluster of $\mathcal{G}_*^{\lambda(i)}$.

Table 2
A fuzzy formal decision context.

U	b_1	b_2	b_3	d
x_1	0.2	0.4	0.7	0
x_2	0.8	0.4	0.6	1
x_3	0.3	0.7	0.8	0
x_4	0.8	0.5	0.8	1

Table 3
Fuzzy concepts of Table 2.

Fuzzy concepts	Terminology
$(U, (0.2, 0.4, 0.6))$	FC_1
$(\{x_2, x_3, x_4\}, (0.3, 0.4, 0.6))$	FC_2
$(\{x_1, x_3, x_4\}, (0.2, 0.4, 0.7))$	FC_3
$(\{x_2, x_4\}, (0.8, 0.4, 0.6))$	FC_4
$(\{x_3, x_4\}, (0.3, 0.5, 0.8))$	FC_5
$(\{x_3\}, (0.3, 0.7, 0.8))$	FC_6
$(\{x_4\}, (0.8, 0.5, 0.8))$	FC_7
$(\emptyset, (1, 1, 1))$	FC_8

Where a pair $(X, \widetilde{A})$ is fuzzy concept satisfying $F(X) = \widetilde{A}$ and $H(\widetilde{A}) = X$. Generally, X is called the extension, and $\widetilde{A}$ is called the intension.

Consider $(U, M, \widetilde{I})$ and (U, D, J) as two fuzzy formal contexts, $\widetilde{I} : U \times M \to [0, 1]$ and $J : U \times D \to \{0, 1\}$. In that case, a fuzzy formal decision context is referred to as $(U, M, \widetilde{I}, D, J)$ where $M \cap D = \emptyset$. M represents the conditional attribute set, while D represents the decision attribute set.

Example 1. A fuzzy formal decision context $(U, M, \widetilde{I}, D, J)$ is shown in Table 2. U and M represent a group of students and a set of indicators used to assess students, respectively. b_1, b_2, b_3 represent the moral education, physical education, intellectual education,respectively. $\widetilde{I}(x_1, b_1)$ refers to the fuzzy membership value of student x_1 in relation to aspect Moral Education. Then $\{d\}$ is the decision attribute set that can partition U into two decision classes $D_1 = \{x_1, x_3\}$ and $D_2 = \{x_2, x_4\}$. $d = 0$ indicates that the overall quality of the student is considered good, while $d = 1$ signifies excellence. The fuzzy concepts generated from Table 2 are presented in Table 3.

2.2. RF

RF fulfills tasks such as classification, regression, and feature importance evaluation by constructing multiple decision trees, making it an ensemble learning method. Next, we will introduce the origin of RF and related concepts such as information gain and *Gini* coefficient.

RF was initially proposed by the American statistician Leo Breiman in 2001 [34]. It originated from the improvement of decision trees using ensemble learning, specifically the 'bagging' (Bootstrap Aggregating) method. Bagging involves creating multiple datasets through bootstrap sampling and training independent classifiers on each dataset. Ultimately, the ensemble model's predictions are obtained through voting or averaging. The introduction of RF aimed primarily to address the overfitting issues associated with decision trees. While traditional decision trees perform well on training sets, they are prone to overfitting noise and subtle variations in training data, leading to poor performance on new data.

In the process of constructing decision trees in random forests, there is a consideration of how to select the optimal splitting attribute. Generally, as the partitioning progresses, we aim for the samples contained in the branching nodes of the decision tree to belong to the same category as much as possible, increasing the 'purity' of the nodes. Information gain and *Gini* index are two commonly used metrics. However, for continuous data types in random forests, the *Gini* index can be directly applied to split continuous features without special processing steps, and its calculation speed is usually faster than that of information gain. Additionally, the *Gini* index performs more robustly when handling class imbalance situations because it focuses more on the purity of nodes rather than the distribution of samples. Considering the characteristics of random forests and the applicability of the *Gini* index, it is more suitable for feature selection and node splitting of continuous data types in random forests. Next, we will briefly introduce the relevant definition of the *Gini* index [35].

In classification problems, let there be $\mathcal{K}$ classes, and let p_k denote the probability that a sample point belongs to the k-th class. The *Gini* index of this probability distribution is defined as:

$$Gini(p) = \sum_{k=1}^{\mathcal{K}} p_k(1-p_k) = 1 - \sum_{k=1}^{\mathcal{K}} p_k^2. \tag{3}$$

For binary classification problems, if the probability that a sample point belongs to the first class is p, the *Gini* index of the distribution is defined as:

$$Gini(p) = 2p(1-p). \tag{4}$$

For a given dataset D, its Gini index is:

$$Gini(D) = 1 - \sum_{k=1}^{\mathcal{K}} \left(\frac{|C_k|}{|D|}\right)^2. \tag{5}$$

Here, c_k represents the subset of samples in D that belong to the k-th class, and $\mathcal{K}$ is the total number of classes.

If the dataset D is split into two parts, D_1 and D_2, based on whether feature A takes a certain possible value a, namely

$$D_1 = \{(x,y) \in D | A(x) = a\}, D_2 = D - D_1.$$

Then, given feature A, the *Gini* index of dataset D is defined as:

$$Gini(D,A) = \frac{|D_1|}{|D|}Gini(D_1) + \frac{|D_2|}{|D|}Gini(D_2). \tag{6}$$

$Gini(D)$ represents the uncertainty of set D, while $Gini(D,A)$ represents the uncertainty of set D after being split by $A = a$. The higher the *Gini* index value, the greater the disorderliness of the sample set.

2.3. Motivation

In the realm of fuzzy sets, the traditional approach to constructing fuzzy concept lattices assigns equal importance to each attribute, leading to uniform WFCs. However, in practical scenarios, attributes are not universally prioritized; instead, selective learning based on preferences is common. This discrepancy has prompted the need for a more nuanced weighting strategy before fuzzy concept construction and subsequent cognitive learning. Existing methods for attribute weighting, which rely on rudimentary mathematical formulas, often result in suboptimal weight allocations that do not reflect the complexities of real-world decision-making. A pertinent example is the evaluation process for university scholarships, where intellectual, moral, and physical criteria are considered, with intellectual capacity typically taking precedence. An imbalanced weighting can lead to inequitable outcomes, such as unfair scholarship distribution. It is thus imperative to develop a more rational weighting mechanism that aligns with human cognitive processes.

During the WFCCL process, similar WFCs are often clustered to form advanced concepts. However, current clustering algorithms do not account for threshold influences, leading to indiscriminate aggregation based solely on similarity. This lack of precision can hinder the cognitive learning process. For instance, while cats, dogs, and tigers are all animals, cats and tigers share a more specific familial relationship within the Felidae family. The absence of threshold considerations can obscure such distinctions. Therefore, establishing reasonable thresholds for concept clustering in practical applications is essential for enhancing the precision of human cognition.

To address these challenges, we consider an algorithm that computes attribute weights using the RF model, a state-of-the-art ensemble learning technique. RF assesses the importance of each attribute across multiple decision trees, effectively reducing the impact of noise and uncertainty in the data. This approach results in a more rational distribution of attribute weights. Furthermore, we have developed an algorithm to determine an approximately optimal threshold for each application scenario when constructing AWFCs. This dual-pronged strategy aims to refine the cognitive learning process and achieve a higher fidelity in knowledge representation.

3. The learning process of the WFCS cognition

In this section, we propose a cognitive learning approach for WFC aimed at aligning with actual cognition.

3.1. The WFC

The steps for computing attribute importance in random forests can be summarized as follows:
(1) Employ the Bootstrap method to randomly select n samples from the dataset for training.
(2) Create a decision tree using the sampled data. At each node:

- Randomly select d features without replacement.
- Use these d features to split the sample set and find the best splitting feature.

(3) Repeat steps (1) to (2) k times, where k is the number of decision trees in the RF.

(4) Calculate the importance of each attribute in each decision tree, and then synthesize these importances to determine the overall importance of the attributes.

Considering a fuzzy formal decision context $(U, M, \widetilde{I}, D, J)$, where $U = \{x_1, x_2, \dots, x_n\}$ represents the set of objects, $M = \{b_1, b_2, \dots, b_m\}$ represents the set of attributes, and $D = \{d_1, d_2, \dots, d_l\}$ denotes the set of decision attributes used for object classification. Furthermore, $U/D = \{D_1, D_2, \dots, D_r\}$ is perceived as a decision partition from U to D.

Definition 2. Let $(U, M, \widetilde{I}, D, J)$ be a fuzzy formal decision context. RF constructs n decision trees based on it, each of which consists of branch nodes and leaf nodes. Branch nodes partition and decide the dataset based on attribute values, while leaf nodes represent the final classification results. The Gini index calculation formula for node i is:

$$Gini_i = \sum_{k=1}^{\mathcal{K}} p_{ik}(1 - p_{ik}). \tag{7}$$

and $\mathcal{K}$ represents the number of categories in the sample set, and p_{ik} represents the estimated probability that the samples in node i belong to the k-th class.

The importance of attribute b_j in node i is represented by the change in the Gini index before and after branching at node i:

$$w'_{ji} = \frac{N_f}{N}(Gini_t - \frac{N_{left}}{N_f} \cdot Gini_{left} - \frac{N_{right}}{N_f} \cdot Gini_{right}). \tag{8}$$

Here, N represents the total number of samples, N_f is the number of samples at the current node, $Gini_f$ is the Gini index of the current node, N_{left} is the number of samples in the left child node, $Gini_{left}$ is the Gini index of the left child node, N_{right} is the number of samples in the right child node, and $Gini_{right}$ is the Gini index of the right child node.

If attribute b_j appears M times in the t-th tree, then the importance of attribute b_j in the t-th tree and in the RF are defined as follows:

$$w'_{tj} = \sum_{i=1}^{M} w'_{ji}, w'_j = \frac{1}{n}\sum_{t=1}^{n} w'_{tj}. \tag{9}$$

Afterward, the weight of attribute b_j is normalized as follows:

$$w_j = \frac{w_{j'}}{\sum_{j=1}^{m} w'_j}. \tag{10}$$

For convenience, we will represent w_j as $w(b_j)$. For an attribute $b \in M$, the weight vector of the attribute is represented as: $W = (w(b_1), w(b_2), \cdots, w(b_m))$, where $w(b_j) \in M$ is the weight vector of each attribute described by the $Gini$ coefficient in the RF.

Definition 3. Let $(U, M, \widetilde{I}, D, J)$ be a fuzzy formal decision context. W is the weight vector of attributes in set M. For $X \subseteq U$ and $\widetilde{A} \in \mathcal{F}(M)$, where a pair $(X, \widetilde{A}, w)$ is a WFC satisfying $F(X) = \widetilde{A}$ and $H(\widetilde{A}) = X$. Where X and $\widetilde{A}$ represent respectively the extension and intension of the concept, w represents the weight of the concept, and its definition is as follows:

$$w = \frac{1}{|M|}\sum_{b_j \in M} \widetilde{A}(b_j)w(b_j). \tag{11}$$

where $\widetilde{A} = (\widetilde{A}(b_1), \widetilde{A}(b_2), \cdots, \widetilde{A}(b_m))$ for all $b_j \in M$ and m is the number of attributes.

For two WFC $(X_1, \widetilde{A}_1, w_1)$ and $(X_2, \widetilde{A}_2, w_2)$, the hierarchical order relation is delineated through the subconcept-superconcept relationship:

$$\begin{aligned} &(X_1, \widetilde{A}_1, w_1) \leq (X_2, \widetilde{A}_2, w_2) \Leftrightarrow X_1 \subseteq X_2 \Leftrightarrow \\ &\widetilde{A}_2 \leq \widetilde{A}_1(\text{ or } w_2 \leq w_1). \end{aligned} \tag{12}$$

Additionally, the WFC collectively construct a lattice structure known as the Weighted Fuzzy Concept Lattice (WFCL), abbreviated as $L_w(U, M, \widetilde{I}, D, J)$, under the relation "≤". Furthermore, the pair $(H(F(x)), F(x), w)$ is denoted as a weighted fuzzy granular concept.

Table 4
WFCs in Table 2.

WFC	Terminology
$(U, (0.2, 0.4, 0.6), 0.1371)$	WFC_1
$(\{x_2, x_3, x_4\}, (0.3, 0.4, 0.6), 0.1519)$	WFC_2
$(\{x_1, x_3, x_4\}, (0.2, 0.4, 0.7), 0.1537)$	WFC_3
$(\{x_2, x_4\}, (0.8, 0.4, 0.6), 0.2259)$	WFC_4
$(\{x_3, x_4\}, (0.3, 0.5, 0.8), 0.1871)$	WFC_5
$(\{x_3\}, (0.3, 0.7, 0.8), 0.1908)$	WFC_6
$(\{x_4\}, (0.8, 0.5, 0.8), 0.2611)$	WFC_7
$(\emptyset, (1, 1, 1), 0.3333)$	WFC_8

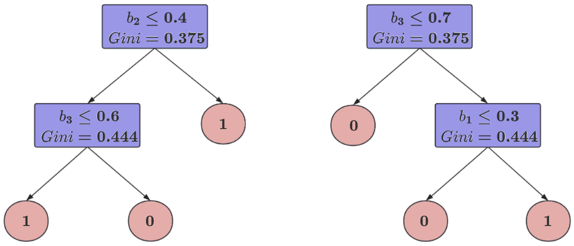


Fig. 2. Two decision trees generated by RF.

Proposition 1. *Let $(U, M, \widetilde{I}, D, J)$ be a fuzzy formal decision context. W represents a vector of weights assigned to attributes within M. For an arbitrary $X_1 \subseteq U$, then $(H(F(X_1)), F(X_1), w_1)$ is a WFC.*

Proof. It can be directly proven by Definition 3. Hence, the effectiveness of this weighting approach can be illustrated with the following example. □

Example 2. (Continued with Example 1) To understand the computation process of WFC, we utilized random forests to generate two decision trees to assess the importance of attributes. The first decision tree was trained with examples and attributes (x_1, x_2, x_3, x_4), corresponding to attributes (b_2, b_3). The second decision tree was trained with examples and attributes (x_1, x_3, x_4), corresponding to attributes (b_1, b_3). Please refer to Fig. 2 for specific details. Additionally, we can utilize the connotation and attribute weights of fuzzy concepts to calculate the weight of each concept using (11), with specific results detailed in Table 4.

From Fig. 2, we can calculate the attribute weights obtained from the first decision tree based on the Gini coefficient of the split nodes as follows: $w(b_2) = 0.042$, $w(b_3) = 0.333$. Similarly, for the second decision tree, the attribute weights are: $w(b_1) = 0.333$, $w(b_3) = 0.042$. Next, by combining the importance of each attribute in each tree and normalizing them, we obtain the final attribute weights as $W = \{0.444, 0.056, 0.5\}$. In other words, when evaluating students' overall performance, academic achievement holds the highest weight, followed by moral education and physical education. This aligns with human cognition. On the other hand, within cognitive science, assigning different weights to attributes can lead to varying degrees of importance of concepts within the conceptual space, consequently affecting human judgments of things. For instance, when assessing a student's overall performance, we often place more emphasis on academic achievements. Therefore, academic performance should logically receive the highest weight. However, if the weight allocation is not sufficiently reasonable, it can ultimately result in less accurate assessments of students.

Based on the preceding discussion, we now have obtained the required WFC. Obviously, in Example 1, we learn fuzzy concepts by exhaustively enumerating information granules. Additionally, considering that information granules are fundamental concepts in granular computing theory, they play a foundational role in human cognition. To reduce time complexity, we integrate information granules into cognitive learning. Therefore, introducing the notion of granular computing (called GrC) is essential in the process of learning fuzzy concepts to decrease computational complexity. Furthermore, constructing the concept space from the given weighted fuzzy granular concepts is the key issue in cognitive learning.

3.2. Construction of the WFCS

In this section, we introduced a method based on GrC to construct the initial WFCS.

Let $(U, M, \widetilde{I}, D, J)$ be a fuzzy formal decision context. $U/D = \{D_1, D_2, \ldots, D_r\}$ is perceived as a decision partition from U to D. For any D_i, the WFCS $\mathcal{G}$ under D_i is represented as follows:

$$\mathcal{G} = \{(H(F(x)), F(x), w) \mid x \in D_i\}. \tag{13}$$

Additionally, we denote the WFCS as $\mathcal{G}_* = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_r\}$, where $\mathcal{G}_i$ is referred to as the WFC subspace of $\mathcal{G}_*$. It is worth noting that each object should undergo sufficient learning to enhance its classification efficacy. Considering the discussion above, we propose the process of constructing the WFCS in Algorithm 1.

Algorithm 1: Construction of the WFCS.

```
Input: A fuzzy formal decision context $(U, M, \widetilde{I}, D, J)$.
Output: WFCS $\mathcal{G}_*$.
Compute the decision partition $U/D = \{D_1, D_2, \ldots, D_r\}$ and the weight W;
for each $D_i \in U/D$ do
    $\mathcal{G}_i \leftarrow \emptyset$;
    for each $x \in D_i$ do
        Calculate the weight value for each attribute;
        Get the fuzzy granular concept $(H(F(x)), F(x))$;
        Calculate the weight $w$ of the concept using Eq (11);
        $\mathcal{G}_i \leftarrow (H(F(x)), F(x), w)$;
    end
end
return $\mathcal{G}_* = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_r\}$.
```

Table 5
A fuzzy formal decision context.

U	b_1	b_2	b_3	d
x_1	0.31	0.59	0.12	0
x_2	0.23	0.48	0.12	0
x_3	0.14	0.63	0.10	0
x_4	0.10	0.47	0.11	0
x_5	0.27	0.74	0.10	0
x_6	0.44	0.60	0.19	0
x_7	0.90	0.31	0.88	1
x_8	0.56	0.31	0.85	1
x_9	0.82	0.58	0.90	1
x_{10}	0.23	0.52	0.57	1

Example 3. Table 5 describes a fuzzy decision formal context $(U, M, \widetilde{I}, D, J)$, in which ten objects are classified into two categories based on the decision attribute d, namely $D_1 = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and $D_2 = \{x_7, x_8, x_9, x_{10}\}$. Here, the weights of attributes are implemented by directly calling the sklearn source code, with n-estimators and random-state both set to 100. The calculated weights of attributes are $W = (0.2641, 0.2397, 0.4962)$. Therefore, the WFCS $\mathcal{G}_1$ generated by class D_1 is:

$$
\begin{aligned}
\mathcal{G}_1 = \{&(\{x_1, x_6\}, (0.31, 0.59, 0.12), 0.0943),\\
&(\{x_1, x_2, x_6\}, (0.23, 0.48, 0.21), 0.0784),\\
&(\{x_3, x_5\}, (0.14, 0.63, 0.1), 0.0792),\\
&(\{x_1, x_2, x_4, x_6\}, (0.1, 0.47, 0.11), 0.0646),\\
&(\{x_5\}, (0.27, 0.74, 0.1), 0.0994),\\
&(\{x_6\}, (0.44, 0.6, 0.19), 0.1181)\}.
\end{aligned}
$$

Similarly, the WFCS $\mathcal{G}_2$ generated by class D_2 is:

$$
\begin{aligned}
\mathcal{G}_2 = \{&(\{x_7\}, (0.9, 0.31, 0.88), 0.2496),\\
&(\{x_7, x_8, x_9\}, (0.56, 0.31, 0.85), 0.2147),\\
&(\{x_9\}, (0.82, 0.58, 0.9), 0.2674),\\
&(\{x_9, x_{10}\}, (0.23, 0.52, 0.57), 0.1561)\}.
\end{aligned}
$$

3.3. Constructing advanced weighted fuzzy conceptual space (AWFCS)

In the previous section, we constructed a WFCS. In fact, there is redundant information among WFCs, and they interact with each other. Therefore, in this section, we propose how to construct an AWFCS based on the WFCS, in order to overcome the limitations of individual cognition and the incompleteness of cognition. Firstly, let us provide the definition of conceptual similarity.

Definition 4. Let $(U, M, \widetilde{I}, D, J)$ be a fuzzy formal decision context. In a weighted conceptual space $\mathcal{G}_i$, if $(X_1, \widetilde{A}_1, w_1)$ is a weighted fuzzy concept in $\mathcal{G}_i$ and $(X_2, \widetilde{A}_2, w_2)$ is its sub-concept, then the similarity between WFCs is expressed as:

$$\theta_{1,2}^{\mathcal{G}_i} = \frac{|X_1 \bigcap X_2|}{|X_1 \bigcap X_2| + 2(\mu|X_1 - X_2| + (1-\mu)|X_2 - X_1|)}. \tag{14}$$

where $\mu = |w_2 - w_1|$, $|X_1 - X_2|$ represents the number of elements existing in X_1 but not in X_2. In fact, since X_2 is a sub-concept of X_1, $|X_2 - X_1| = 0$. Therefore, this formula can be rewritten as:

$$\theta^{\mathcal{G}_i}_{1,2} = \frac{|X_1 \bigcap X_2|}{|X_1 \bigcap X_2| + 2\mu |X_1 - X_2|}. \tag{15}$$

where $\theta^{\mathcal{G}_i}_{1,2}$ reflects the similarity between $(X_1, \widetilde{A}_1, w_1)$ and $(X_2, \widetilde{A}_2, w_2)$. The larger $\theta^{\mathcal{G}_i}_{1,2}$ is, the higher the similarity between $(X_1, \widetilde{A}_1, w_1)$ and $(X_2, \widetilde{A}_2, w_2)$; the smaller $\theta^{\mathcal{G}_i}_{1,2}$ is, the weaker the similarity between $(X_1, \widetilde{A}_1, w_1)$ and $(X_2, \widetilde{A}_2, w_2)$.

Next, to deepen understanding, we define AWFC based on concept similarity.

Algorithm 2: Constructing the AWFC.

```
Input: A WFCS $\mathcal{G}_* = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_r\}$ and a threshold $\lambda$.
Output: The AWFCS $\mathcal{G}^\lambda_* = \{\mathcal{G}^\lambda_1, \mathcal{G}^\lambda_2, \cdots, \mathcal{G}^\lambda_r\}$.
for each $\mathcal{G}_i \in \mathcal{G}_*$ do
    Set $P_i = \emptyset$, $L_i = \emptyset$ and $\widehat{L}_i = \emptyset$;
    // Finding the upper bound concept and sub-concepts.
    for each $(H(F(x_j)), F(x_j), w_j) \in \mathcal{G}_i$ do
        for each $(H(F(x_k)), F(x_k), w_k) \in \mathcal{G}_i$ do
            Set $L_{i,j} = \emptyset$;
            if $(H(F(x_k)), F(x_k), w_k) \subseteq (H(F(x_j)), F(x_j), w_j)$ and $\theta^{\mathcal{G}_i}_{i,j} \geq \lambda$ then
                $L_{i,j} \leftarrow (H(F(x_k)), F(x_k), w_k)$, and $P_i = P_i \cup \{(H(F(x_j)), F(x_j), w_j)\}$;
            end
        end
        $L_i \leftarrow L_{i,j}$;
    end
    // Determining whether the upper bound concept is unique.
    for each $L_{i,m} \in L_i$ do
        if there exists only one concept $(H(F(x_t)), F(x_t), w_t)$ in $P_i$ such that each concept of $L_{i,m}$ is a sub-concept of
          $(H(F(x_t)), F(x_t), w_t)$ then
            $\widehat{L}_i \leftarrow L_{i,m}$;
        end
    end
    // Concept clustering.
    for each $\widehat{L}_{i,m} \in \widehat{L}_i$ do
        Calculate the AWFCS $(X^\lambda_{i,m}, \widetilde{A}^\lambda_{i,m}, w^\lambda_{i,m})$ according to Definition 5;
        $\mathcal{G}^\lambda_i \leftarrow (X^\lambda_{i,m}, \widetilde{A}^\lambda_{i,m}, w^\lambda_{i,m})$;
    end
end
return $\mathcal{G}^\lambda_* = \{\mathcal{G}^\lambda_1, \mathcal{G}^\lambda_2, \cdots, \mathcal{G}^\lambda_r\}$.
```

Definition 5. Let $(U, M, \widetilde{I}, D, J)$ represent a fuzzy formal decision context. For a WFCS $\mathcal{G}_i$, if there exist WFCs $(X_1, \widetilde{A}_1, w_1)$, $(X_2, \widetilde{A}_2, w_2)$, $\cdots$, $(X_n, \widetilde{A}_n, w_n)$ that satisfy $X_1 \subseteq X_2 \subseteq \cdots \subseteq X_n$ (where $(X_n, \widetilde{A}_n, w_n)$ is defined as the upper bound concept), and for any concept $(X_i, \widetilde{A}_i, w_i)$, $\theta^{\mathcal{G}_i}_{i,n} > \lambda$, then this group of concepts can generate an AWFC. Here, we refer to this collection as a concept cluster:

$$X_{i,j} = X_1 \cup X_2 \cup \cdots X_n, \widetilde{A}_{i,j} = \frac{1}{2^{n-1}}(\widetilde{A}_1 + \widetilde{A}_2 + 2\widetilde{A}_3 + 4\widetilde{A}_4 + \cdots + 2^{n-2}\widetilde{A}_n). \tag{16}$$

Then $(X_{i,j}, \widetilde{A}_{i,j}, w_{i,j})$ is an AWFC, where $w_{i,j} = \frac{1}{|M|}\sum_{b_i \in M} \widetilde{A}_{i,j}(b_i) w(b_i)$.

All AWFCs constructed at threshold λ are defined as an AWFCS, represented as $\mathcal{G}^\lambda_* = \{\mathcal{G}^\lambda_1, \mathcal{G}^\lambda_2 \cdots, \mathcal{G}^\lambda_r\}$, where $\mathcal{G}^\lambda_i = \{\mathcal{G}^\lambda_{i,j} \mid j = 1, 2, \cdots, t\} = \{(X^\lambda_{i,j}, \widetilde{A}^\lambda_{i,j}, w^\lambda_{i,j}) \mid j = 1, 2, \cdots, t\}$. Here, t denotes the number of AWFCs in the subspace $\mathcal{G}^\lambda_i$.

The connotations of these subconcepts are assigned different weights based on their importance. In other words, the larger the extension of a subconcept, the greater its importance, and accordingly, the weight of its connotation should also be greater. Algorithm 2 elaborates on how to construct the AWFCS $\mathcal{G}^\lambda_*$.

Example 4. (Continued with Example 3) Here, we set $\lambda = 0.8$. According to Definition 5, the AWFCs are as follows:

$$\mathcal{G}^{\lambda}_{1,1} = (\{x_3, x_5\}, (0.2050, 0.6850, 0.1000), 0.0893);$$
$$\mathcal{G}^{\lambda}_{1,2} = (\{x_1, x_2, x_4, x_6\}, (0.1850, 0.5025, 0.1150), 0.0755);$$
$$\mathcal{G}^{\lambda}_{1,3} = (\{x_6\}, (0.4400, 0.6000, 0.1900), 0.1181);$$
$$\mathcal{G}^{\lambda}_{2,1} = (\{x_7, x_8, x_9\}, (0.7100, 0.3775, 0.8700), 0.2366);$$
$$\mathcal{G}^{\lambda}_{2,2} = (\{x_9, x_{10}\}, (0.5250, 0.5500, 0.7350), 0.2117).$$

The first decision class generated three AWFCs, while the second decision class generated two AWFCs. It can be observed that these AWFCs retain the original information while removing redundant information, thereby enhancing the efficiency of CCL.

3.4. Searching for the approximately optimal parameter $\widehat{\lambda}(i)$

According to Definitions 4 and 5, it is known that the threshold λ has a significant impact on the formation of the AWFCS. Therefore, it is necessary for us to select an approximately optimal threshold λ for each dataset.

Let $\lambda = \lambda(i)$ $(i \in \{1, 2, \cdots, s\})$ and $\lambda(i) \propto i$. For different values of $\lambda(i)$, we can denote all constructed AWFCSs as:

$$\begin{bmatrix} \mathcal{G}^{\lambda(1)}_{*} \\ \mathcal{G}^{\lambda(2)}_{*} \\ \vdots \\ \mathcal{G}^{\lambda(s)}_{*} \end{bmatrix} = \begin{bmatrix} \mathcal{G}^{\lambda(1),1} & \mathcal{G}^{\lambda(1),2} & \cdots & \mathcal{G}^{\lambda(1),m_1} \\ \mathcal{G}^{\lambda(2),1} & \mathcal{G}^{\lambda(2),2} & \cdots & \mathcal{G}^{\lambda(2),m_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{G}^{\lambda(s),1} & \mathcal{G}^{\lambda(s),2} & \cdots & \mathcal{G}^{\lambda(s),m_s} \end{bmatrix}. \tag{17}$$

Here, $\mathcal{G}^{\lambda(i),j}$ $(j \in \{1, 2, \cdots, m_i\})$ represents the concept cluster of $\mathcal{G}^{\lambda(i)}_{*}$, where m_i is the number of concept clusters in $\mathcal{G}^{\lambda(i)}_{*}$, and $\mathcal{G}^{\lambda(i)}_{*}$ is obtained through specific computations using $\lambda(i)$. Moreover, according to Definition 5, any concept cluster can generate an AWFC. Therefore, (17) can be restated as (18).

$$\begin{bmatrix} \mathcal{G}^{\lambda(1)}_{*} \\ \mathcal{G}^{\lambda(2)}_{*} \\ \vdots \\ \mathcal{G}^{\lambda(s)}_{*} \end{bmatrix} = \begin{bmatrix} (X^{\lambda(1)}_1, \widetilde{A}^{\lambda(1)}_1, w^{\lambda(1)}_1) & (X^{\lambda(1)}_2, \widetilde{A}^{\lambda(1)}_2, w^{\lambda(1)}_2) & \cdots & (X^{\lambda(1)}_{m_1}, \widetilde{A}^{\lambda(1)}_{m_1}, w^{\lambda(1)}_{m_1}) \\ (X^{\lambda(2)}_1, \widetilde{A}^{\lambda(2)}_1, w^{\lambda(2)}_1) & (X^{\lambda(2)}_2, \widetilde{A}^{\lambda(2)}_2, w^{\lambda(2)}_2) & \cdots & (X^{\lambda(2)}_{m_2}, \widetilde{A}^{\lambda(2)}_{m_2}, w^{\lambda(2)}_{m_2}) \\ \vdots & \vdots & \ddots & \vdots \\ (X^{\lambda(s)}_1, \widetilde{A}^{\lambda(s)}_1, w^{\lambda(s)}_1) & (X^{\lambda(s)}_2, \widetilde{A}^{\lambda(s)}_2, w^{\lambda(s)}_2) & \cdots & (X^{\lambda(s)}_{m_s}, \widetilde{A}^{\lambda(s)}_{m_s}, w^{\lambda(s)}_{m_s}) \end{bmatrix}. \tag{18}$$

Let $(U, M, \widetilde{I}, D, J)$ is a fuzzy formal context, $U = \{x_1, x_2, \cdots, x_n\}$ is a set of objects, $D = \{1, 2, \cdots, r\}$ represents the decision attribute. and $U/D = \{D_1, D_2, .., D_r\}$ is considered as the decision partition of U on D, where each D_j can generate an advanced weighted fuzzy concept subspace $\mathcal{G}^{\lambda(i)}_{j}$ through $\lambda(i)$. Then, all advanced weighted fuzzy concept subspaces regarding $\lambda(i)$ can be denoted as $\mathcal{G}^{\lambda(i)}_{*}$, namely $\mathcal{G}^{\lambda(i)}_{*} = \{\mathcal{G}^{\lambda(i)}_{1}, \mathcal{G}^{\lambda(i)}_{2}, \cdots, \mathcal{G}^{\lambda(i)}_{r}\}$. Furthermore, facing different $\lambda(i)$, the generated AWFCS can be denoted as:

$$\begin{bmatrix} \mathcal{G}^{\lambda(1)}_{*} & \mathcal{G}^{\lambda(2)}_{*} & \vdots & \mathcal{G}^{\lambda(n)}_{*} \end{bmatrix} = \begin{bmatrix} \mathcal{G}^{\lambda(1)}_{1} & \mathcal{G}^{\lambda(1)}_{2} & \cdots & \mathcal{G}^{\lambda(1)}_{r} \\ \mathcal{G}^{\lambda(2)}_{1} & \mathcal{G}^{\lambda(2)}_{2} & \cdots & \mathcal{G}^{\lambda(2)}_{r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{G}^{\lambda(s)}_{1} & \mathcal{G}^{\lambda(s)}_{2} & \cdots & \mathcal{G}^{\lambda(s)}_{r} \end{bmatrix}. \tag{19}$$

In fact, the objective is to select an approximately optimal $\widehat{\lambda}(i)$ for each dataset within the interval $[0, 1]$. Let $(X_r, \widetilde{A}_r, w_r)$ $(r \in \{1, 2, \cdots, t\})$ be a WFC, then its objective function can be formulated as:

$$\begin{aligned} E(\lambda(i), j) = & \min_{i\in\mathcal{I}, j\in\mathcal{J}, d'} \sum_{r=1}^{t} \left\| (X_r, \widetilde{A}_r, w_r) - \mathcal{G}^{\lambda(i),j}_{d'} \right\|_2^2 - \\ & \max_{i\in\mathcal{I}} \min_{j\in\mathcal{J}} \sum_{d''\in\overline{D}} \sum_{r=1}^{t} \left\| (X_r, \widetilde{A}_r, w_r) - \mathcal{G}^{\lambda(i),j}_{d''} \right\|_2^2 \\ & \text{s.t. } m_i \propto \lambda(i), 0 \le \lambda(i) \le 1. \end{aligned} \tag{20}$$

Where $\mathcal{I} = \{1, 2, \cdots, s\}$, $\mathcal{J} = \{1, 2, \cdots, m_i\}$, $\overline{D} = D \backslash \{d'\}$, $\mathcal{G}^{\lambda(i),j}_{d'}$ represents the concept cluster under decision class d', and d' represents the true decision class of the fuzzy concept $(X_r, \widetilde{A}_r, w_r)$. Therefore, in (20), the first term represents samples being assigned to the true decision class, while the second term represents the opposite scenario.

Let $(X^{\lambda(i)}_{d',j}, \widetilde{A}^{\lambda(i)}_{d',j}, w^{\lambda(i)}_{d',j})$ be the AWFC generated by the concept cluster $\mathcal{G}^{\lambda(i),j}_{d'}$. For any fuzzy concept $(X_r, \widetilde{A}_r, w_r)$, it can be considered as the representation of example x_r with M-dimensional features. Therefore, (20) can be restated as:

$$\begin{aligned} E(\lambda(i), j) = &\min_{i\in\mathcal{I}, j\in\mathcal{J}, d'} \sum_{r=1}^{t} \left\| W(\widetilde{A}_r - \widetilde{A}^{\lambda(i)}_{d',j}) \right\|_2^2 - \\ &\max_{i\in\mathcal{I}} \min_{j\in\mathcal{J}} \sum_{d''\in\overline{D}} \sum_{r=1}^{t} \left\| W(\widetilde{A}_r - \widetilde{A}^{\lambda(i)}_{d'',j}) \right\|_2^2 \\ &\text{s.t. } m_i \propto \lambda(i), 0 \le \lambda(i) \le 1. \end{aligned} \tag{21}$$

Where $W = \{w(b_1), w(b_2), \cdots, w(b_m)\}$ represents the weights of attributes. Since variable j depends on another variable $\lambda(i)$, the objective function can be optimized by updating $\lambda(i)$, namely:

$$\begin{aligned} &\widehat{\lambda}(i) = \arg\min_{i\in\mathcal{I}, j\in\mathcal{J}} E(\lambda(i), j) \\ &\text{s.t. } m_i \propto \lambda(i), 0 \le \lambda(i) \le 1. \end{aligned} \tag{22}$$

In theory, the optimal $\widehat{\lambda}(i)$ can be directly obtained by solving (22). However, due to the difficulty in obtaining the specific functional expression between m_i and $\widehat{\lambda}(i)$, it is challenging to obtain its analytical solution. Therefore, we can use a grid search method to obtain an approximate optimal $\widehat{\lambda}(i)$. Based on the above discussion, Algorithm 3 outlines the process of obtaining an approximate optimal $\widehat{\lambda}(i)$.

Algorithm 3: The process of selecting the optimal $\widehat{\lambda}(i)$ in RF-WFCCL.

```
Input: Training set $\overline{U}$, validation set $V$, and learning rate $\varepsilon$. // $\varepsilon = \{0.0, 0.1, \cdots, 1.0\}$.
Output: Approximately optimal $\widehat{\lambda}(i)$.
Construct the WFCS $\mathcal{G}_*$ based on Algorithm 1;
for $\lambda(i) = 0$ to $1$ do
    Based on Algorithm 2, $\mathcal{G}^{\lambda(i)}_*$ can be obtained;
    for each $x_r \in V$ do
        Construct the weighted fuzzy granular concept $(X_r, \widetilde{A}_r, w_r)$ based on Definition 1 and 3;
        Compute $E(\lambda(i), j)$ based on Eq (21);
    end
    $\lambda(i) = \lambda(i) + \varepsilon$;
end
return $\widehat{\lambda}(i)$.
```

4. Cognitive learning based on AWFC

Given a fuzzy decision formal context $(U, M, \widetilde{I}, D, J)$, the AWFCS demonstrates excellent performance in handling classification problems. When a new object Δx is introduced, determining its category is a question worthy of consideration, and this new object will also alter the original WFCS. In this section, we will explore how to utilize the AWFCS for cognitive learning.

4.1. Prediction of classification labels after the addition of object Δx

In the AWFCS, we can measure the similarity between concepts by computing the Euclidean distance of attributes. The smaller the distance, the greater the similarity between concepts. Therefore, we can calculate the distance between the newly added object Δx and the concepts in $\mathcal{G}^{\lambda}_*$ to determine the category of Δx.

Definition 6. Let $(U, M, \widetilde{I}, D, J)$ be a fuzzy formal decision context. For a newly added object Δx, its membership value relative to $\widetilde{I}$ is $\widetilde{A}$, and the Euclidean distance between Δx and the jth AWFCS $(X^{\lambda}_{i,j}, \widetilde{A}^{\lambda}_{i,j}, w^{\lambda}_{i,j})$ in $\mathcal{G}^{\lambda}_i$ is:

$$ED(\Delta x, X^{\lambda}_{i,j}) = \sqrt{\sum_{b\in M} (w(b)(\widetilde{A}(b) - \widetilde{A}^{\lambda}_{i,j}(b))^2}. \tag{23}$$

The $ED(\Delta x, X^{\lambda}_{i,j})$ reflects the similarity between Δx and $(X^{\lambda}_{i,j}, \widetilde{A}^{\lambda}_{i,j}, w^{\lambda}_{i,j})$. A smaller $ED(\Delta x, X^{\lambda}_{i,j})$ indicates a greater similarity, while a larger $ED(\Delta x, X^{\lambda}_{i,j})$ indicates a lesser similarity. According to the minimum distance principle, Δx should be classified. Algorithm 4 demonstrates the prediction of class labels when adding object Δx.

Algorithm 4: Predict the Category Label of Δx.

```
Input: The AWFCS $\mathcal{G}_*^\lambda = \{\mathcal{G}_1^\lambda, \mathcal{G}_2^\lambda, \cdots, \mathcal{G}_r^\lambda\}$ and the newly added object $\Delta x$.
Output: The category label $d$ of $\Delta x$.
for each $\mathcal{G}_i^\lambda \in \mathcal{G}_*^\lambda$ do
    for each $\mathcal{G}_{i,j}^\lambda \in \mathcal{G}_i^\lambda$ do
        Calculate $ED(\Delta x, \mathcal{G}_{i,j}^\lambda)$ from Definition 6;
    end
    The shortest distance is $s_i = \min(ED(\Delta x, \mathcal{G}_{i,j}^\lambda))$, where $\mathcal{G}_{i,j}^\lambda \in \mathcal{G}_i^\lambda$;
end
Calculate $d = \arg\min_{\{i=1,2,\cdots,r\}} s_i$ in $\mathcal{G}_*^\lambda$;
return The category label $d$ of $\Delta x$.
```

Algorithm 5: The process of concept cognition.

```
Input: The WFCS $\mathcal{G}_* = \{\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_r\}$, the AWFCS $\mathcal{G}_*^\lambda = \{\mathcal{G}_1^\lambda, \mathcal{G}_2^\lambda, \cdots, \mathcal{G}_r^\lambda\}$, decision partition $U/D = \{D_1, D_2, \cdots, D_r\}$, newly
       added data $\Delta x$, threshold $\lambda$
Output: The updated AWFCS $\widehat{\mathcal{G}}_*^\lambda = \{\widehat{\mathcal{G}}_1^\lambda, \widehat{\mathcal{G}}_2^\lambda, \cdots, \widehat{\mathcal{G}}_r^\lambda\}$
The membership degree of $\Delta x$ with respect to $\widetilde{I}$ is $\widetilde{A}$, and update the weights $W$;
Calculate the class label of $\Delta x$ according to Algorithm 4, assuming its class label is $j$;
Set $\widehat{\mathcal{G}}_j = \emptyset$;
if $j$ predicts correctly then
    // Update the concept space of the corresponding class.
    for each $x_k \in D_j$ do
        if $\widetilde{I}(\Delta x, b) \geq \widetilde{I}(x_k, b)$ for each $b \in M$ then
            $X_k \leftarrow \Delta x$ and $\widehat{\mathcal{G}}_j \leftarrow (X_k, \widetilde{A}_k, w_k)$;
        end
    end
end
// Integrate the newly generated granular concept into the corresponding concept space.
Calculate $(\Delta X, \widetilde{A}, w)$ and update $\widehat{\mathcal{G}}_j \leftarrow (\Delta X, \widetilde{A}, w)$;
Update the AWFCS $\widehat{\mathcal{G}}_j^\lambda$ according to Definition 5;
// Update the weight values of all concept representations.
for each $\mathcal{G}_i^\lambda \in \mathcal{G}_*^\lambda (i \neq j)$ do
    for each $\mathcal{G}_{i,k}^\lambda \in \mathcal{G}_i^\lambda$ do
        Recalculate the weight $w_{i,k}$ of $\mathcal{G}_{i,k}^\lambda$ based on $W$;
    end
end
return $\widehat{\mathcal{G}}_*^\lambda = \{\widehat{\mathcal{G}}_1^\lambda, \widehat{\mathcal{G}}_2^\lambda, \cdots, \widehat{\mathcal{G}}_r^\lambda\}$.
```

4.2. *The cognitive process based on advanced weighted fuzzy space concepts*

In general, when a newly added object Δx is given, we first need to determine its class label, assuming the class label is j. At this point, we not only need to update the weights of the concepts but also need to update the concepts in $\mathcal{G}_j$ and $\mathcal{G}_j^\lambda$. Algorithm 5 provides the complete process of concept cognition.

4.3. *Complexity analysis*

We further analyze the time complexity of RF-WFCCL. In the first stage, assuming the time complexity of assigning attribute importance through RF is $O(t_1)$. According to Algorithm 1, we can derive its time complexity as $O(t_1|U|)$. In the second stage, the time complexity of constructing the AWFCS based on Algorithm 2 is $O(|\mathcal{G}_*|^2)$. Assuming the time complexity of step 4 in Algorithm 3 is $O(t_2)$, then based on Algorithms 2 and 3, the time complexity of the second stage is $O(t_2|\mathcal{G}_*||V|)$. In the third stage, we do not need to calculate the similarity with all concepts in the WFCS, only the similarity with concepts in the AWFCS. Therefore, its time complexity is $O(|\widehat{\mathcal{G}}_*^\lambda|)$.

Table 6
Details of the 12 selected datasets.

ID	Dataset	Sample	Attribute	Class	Imbalanced ratio
1	Prostate-ge	102	5966	2	52 : 50
2	Cll-sub-111	111	11,340	3	51 : 49 : 11
3	Lung	203	3312	5	139 : 21 : 20 : 17 : 6
4	Glass	214	9	6	70 : 76 : 17 : 13 : 9 : 29
5	Wdbc	569	30	2	357 : 212
6	Messidor-features	1151	19	2	520 : 611
7	Wifi-localization	2000	7	4	500 : 500 : 500 : 500
8	Student	4424	36	3	2209 : 1421 : 794
9	Twonorm	7400	20	2	3703 : 3697
10	EGS	10,000	12	2	6380 : 3620
11	Magic	19,020	10	2	12,332 : 6688
12	Occupancy	20,560	7	2	15,810 : 4750

5. Experiments

In this section, we will investigate the rationality and effectiveness of the RF-WFCCL algorithm proposed in a fuzzy formal background. Specifically, we will compare it with 12 other classification algorithms and verify its effectiveness through Wilcoxon test. Additionally, we will further explore the impact of different parameters $\lambda(i)$ on classification accuracy to demonstrate the rationality of our designed algorithm for seeking the approximately optimal threshold $\widehat{\lambda}(i)$. We also conducted noise experiments to verify the algorithm's stability and adaptability when facing data perturbations. The datasets used in this experiment are primarily from UCI, KEEL, and publicly available tumor gene datasets, with detailed information provided in Table 6.

5.1. *Experimental setup*

In order to adapt to the fuzzy formal background of this paper, in the data preprocessing stage, all attribute values are first normalized to ensure that the membership values range from 0 to 1. The specific formula is as follows:

$$\widetilde{I}(x_i, b_j) = \frac{f(x_i, b_j) - \min(f(b_j))}{\max(f(b_j)) - \min(f(b_j))}. \tag{24}$$

Here, $f(x_i, b_j)$ represents the value of x_i under attribute b_j in the dataset, while $\min(f(b_j))$ and $\max(f(b_j))$ respectively denote the minimum and maximum values of attribute b_j in the dataset. The normalized value $\widetilde{I}(x_i, b_j)$ is then used as the membership degree of object x_i under attribute b_j.

This paper compares RF-WFCCL with 6 classical machine learning classification algorithms (DT [36], NB [37], RF [34], LR [38], KNN [39], SVM [40]) and 6 fuzzy-based classification algorithms. The latter includes three fuzzy KNN-based algorithms (IFKNN [41], PFKNN [42], FRNN [43]) and three advanced concept cognitive learning models based on fuzzy formal context (DMPWFC [24], IFCRL [26], F3WG-CCL [27]). Each dataset was randomly divided into 80 % for training and 20 % for testing. To reduce statistical errors and improve the reliability of the results, each experiment was independently repeated ten times, and the mean and standard deviation of the outcomes were calculated to evaluate the overall performance and stability of different classification methods. For model parameters, the Random Forest was configured with 100 decision trees, using the Gini index as the splitting criterion, while all other parameters followed the default settings of the RandomForestClassifier in the scikit-learn library. To ensure experimental fairness, all

Table 7
Comparison of sccuracy (including Mean % ± Standard Deviation %) between RF-WFCCL and 6 classical machine learning classification algorithms.

ID	RF-WFCCL	KNN	SVM	DT	NB	RF	LR
1	**90.27 ± 6.1420**	80.00 ± 6.6667	89.52 ± 5.1287	77.14 ± 6.9985	63.81 ± 9.8054	83.33 ± 6.6667	88.48 ± 7.6886
2	**77.50 ± 6.2368**	56.52 ± 7.2975	73.91 ± 5.2494	64.35 ± 8.0145	59.13 ± 7.3271	59.13 ± 7.2975	74.35 ± 6.4125
3	**96.81 ± 2.4065**	94.15 ± 3.6504	95.12 ± 1.5426	85.37 ± 4.0690	81.95 ± 2.2354	90.24 ± 3.6504	94.88 ± 2.5464
4	**98.37 ± 1.4657**	86.97 ± 6.9144	79.30 ± 3.5192	97.91 ± 1.8524	83.02 ± 5.4092	96.27 ± 2.5896	74.41 ± 5.6007
5	**96.84 ± 1.8128**	96.14 ± 1.7189	96.47 ± 1.8996	91.84 ± 2.5438	94.03 ± 1.8316	95.61 ± 1.6643	96.49 ± 1.4678
6	**66.02 ± 1.7519**	60.08 ± 2.9284	62.68 ± 2.4164	62.51 ± 2.7665	55.54 ± 1.9852	65.54 ± 3.6435	64.71 ± 2.8862
7	97.10 ± 0.6782	**97.40 ± 0.8000**	96.97 ± 0.6368	96.62 ± 0.7093	97.02 ± 0.7454	97.15 ± 0.3571	96.75 ± 0.7499
8	70.61 ± 1.6196	65.50 ± 1.0145	75.19 ± 1.0615	67.98 ± 1.5779	67.24 ± 1.5397	74.53 ± 1.6707	**75.61 ± 1.0242**
9	**97.81 ± 0.5268**	96.71 ± 0.1939	97.72 ± 0.3097	84.53 ± 1.1512	97.75 ± 0.2815	96.95 ± 0.4636	97.71 ± 0.3293
10	95.50 ± 0.3821	89.61 ± 0.7453	95.39 ± 0.4170	94.62 ± 0.4355	**96.10 ± 0.3435**	95.14 ± 0.7453	95.31 ± 0.3113
11	**84.44 ± 0.6943**	82.89 ± 0.4655	79.04 ± 0.5165	82.06 ± 0.8623	72.50 ± 0.4552	83.68 ± 0.4655	78.95 ± 0.5302
12	**99.10 ± 0.1812**	98.93 ± 0.1832	98.74 ± 0.1760	99.04 ± 0.1872	97.01 ± 0.2038	99.08 ± 0.1832	98.08 ± 0.2399
Average	**89.1975**	83.7417	86.6708	83.6642	80.4250	86.3875	86.3108

Table 8
Comparison of accuracy (including Mean % ± Standard Deviation %) between RF-WFCCL and 6 fuzzy-based classification algorithms.

ID	RF-FWCCL	DMPWFC	IFCRL	F3WG-CCL	IF-KNN	PFKNN	FRNN
1	**90.27 ± 6.1420**	81.82 ± 9.0909	80.95 ± 5.6344	83.64 ± 7.0605	84.76 ± 6.3174	62.38 ± 8.0952	84.76 ± 6.9985
2	**77.50 ± 6.2368**	67.50 ± 8.4564	60.87 ± 7.4503	63.33 ± 8.5297	63.48 ± 8.6988	61.30 ± 7.3913	63.04 ± 8.5751
3	**96.81 ± 2.4065**	89.52 ± 5.5533	95.61 ± 2.6269	93.33 ± 4.8562	92.44 ± 4.8105	93.41 ± 4.5039	94.88 ± 2.3010
4	**98.37 ± 1.4657**	33.95 ± 7.2205	68.60 ± 4.4133	86.82 ± 5.9091	88.60 ± 3.8142	82.55 ± 5.7201	71.16 ± 8.3978
5	**96.84 ± 1.8128**	90.88 ± 1.5789	95.26 ± 1.6736	96.67 ± 2.2807	96.75 ± 1.4171	92.28 ± 2.6548	82.01 ± 4.4771
6	**66.02 ± 1.7519**	60.82 ± 2.9945	63.03 ± 2.1048	62.16 ± 2.5413	60.73 ± 1.8574	56.02 ± 2.5257	60.08 ± 3.1563
7	97.10 ± 0.6782	94.38 ± 0.9437	88.20 ± 2.7221	94.90 ± 2.2450	**97.17 ± 0.7668**	94.67 ± 0.6712	94.44 ± 0.6964
8	70.61 ± 1.6196	62.11 ± 1.4382	**71.08 ± 0.7365**	65.30 ± 1.3878	65.26 ± 1.6958	63.67 ± 1.4331	53.58 ± 1.5351
9	**97.81 ± 0.5268**	80.97 ± 1.2455	97.68 ± 0.3310	95.47 ± 0.6817	96.57 ± 0.4400	97.70 ± 0.2642	97.52 ± 0.4033
10	**95.50 ± 0.3821**	80.68 ± 1.8143	92.02 ± 0.6442	87.38 ± 1.2504	89.45 ± 0.6301	87.53 ± 0.7050	64.10 ± 0.6465
11	**84.44 ± 0.6943**	77.24 ± 0.6514	79.66 ± 0.8201	82.23 ± 0.7426	82.91 ± 0.4493	76.35 ± 0.4815	64.85 ± 0.7493
12	99.10 ± 0.1812	96.59 ± 0.4421	98.91 ± 0.2521	**99.14 ± 0.2871**	98.27 ± 0.1351	89.93 ± 0.3575	99.02 ± 0.1436
Average	**89.1975**	76.3717	82.6558	84.1975	84.6992	79.8158	77.4533

Table 9
Comparison of F1-score (including Mean % ± Standard Deviation %) between RF-WFCCL and 6 classical machine learning classification algorithms.

ID	RF-WFCCL	KNN	SVM	DT	NB	RF	LR
1	**89.77 ± 5.9368**	81.21 ± 4.5885	89.70 ± 3.3974	76.40 ± 8.7113	58.49 ± 7.7627	76.80 ± 8.4511	89.26 ± 6.3755
2	**79.93 ± 6.8100**	74.53 ± 5.1306	71.33 ± 7.7672	72.21 ± 8.4376	56.70 ± 8.8097	77.45 ± 8.1058	73.39 ± 8.6390
3	**81.60 ± 5.5916**	61.05 ± 8.1311	79.95 ± 6.0363	58.30 ± 6.4267	69.01 ± 6.566	57.59 ± 6.9268	80.77 ± 7.5810
4	**98.15 ± 5.5556**	81.01 ± 8.7201	54.95 ± 9.4868	92.42 ± 7.2491	82.92 ± 4.5118	84.55 ± 9.6002	43.60 ± 5.8457
5	96.73 ± 2.1506	96.86 ± 0.8471	96.65 ± 1.3259	92.60 ± 2.6866	93.28 ± 3.2764	95.34 ± 1.5722	**97.28 ± 2.2258**
6	**65.66 ± 2.0264**	61.16 ± 1.4642	65.14 ± 2.5124	62.67 ± 2.6655	62.07 ± 2.0450	62.07 ± 1.6468	65.43 ± 2.4804
7	97.22 ± 0.8464	**98.47 ± 0.6182**	98.24 ± 0.7454	96.50 ± 1.0895	98.38 ± 0.5376	97.21 ± 0.6469	97.12 ± 1.0699
8	65.83 ± 2.1269	56.12 ± 1.0999	**67.22 ± 1.2202**	61.26 ± 2.0644	58.38 ± 1.8513	64.86 ± 1.5124	66.53 ± 1.6112
9	**97.81 ± 0.7044**	96.48 ± 0.4958	97.64 ± 0.3679	84.25 ± 1.3023	97.75 ± 0.3282	92.39 ± 0.3821	96.97 ± 0.3233
10	**95.64 ± 0.3738**	88.37 ± 0.6702	94.18 ± 0.3622	94.29 ± 0.4914	95.60 ± 0.4936	95.24 ± 0.2888	95.56 ± 0.3238
11	**83.79 ± 1.0299**	80.20 ± 0.7406	76.27 ± 0.4043	80.00 ± 0.7042	65.32 ± 0.6517	82.96 ± 0.6336	75.95 ± 0.5897
12	**98.89 ± 0.1353**	98.81 ± 0.1376	98.44 ± 0.1985	98.85 ± 0.1493	95.74 ± 0.3276	98.65 ± 0.1192	98.56 ± 0.1763
Average	**87.5850**	81.1892	82.4758	80.8125	77.8033	82.0925	81.7017

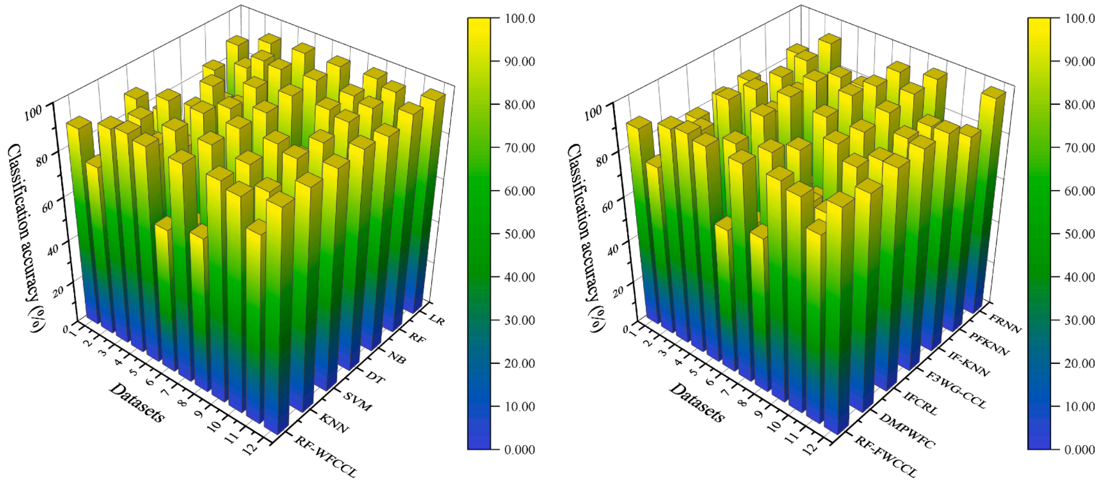
Table 10
Comparison of F1-score (including Mean % ± Standard Deviation %) between RF-WFCCL and 6 fuzzy-based classification algorithms.

ID	RF-FWCCL	DMPWFC	IFCRL	F3WG-CCL	IF-KNN	PFKNN	FRNN
1	**89.77 ± 5.9368**	79.22 ± 9.8865	80.76 ± 5.6099	82.83 ± 7.4410	82.32 ± 5.3571	57.65 ± 9.7442	82.48 ± 2.3389
2	**79.93 ± 6.8100**	74.11 ± 9.2273	67.66 ± 8.4504	75.24 ± 8.3498	72.67 ± 4.8697	76.19 ± 8.3640	78.26 ± 7.9073
3	81.60 ± 5.5916	82.64 ± 7.0921	80.17 ± 7.4437	**83.41 ± 7.3606**	69.11± **7.4420**	60.08 ± 8.0850	74.74 ± 6.3694
4	**98.15 ± 5.5556**	28.49 ± 6.8396	59.13 ± 4.8074	73.60 ± 4.0135	81.50 ± 6.5741	68.62 ± 9.3668	46.12 ± 5.4721
5	**96.73 ± 2.1506**	85.18 ± 2.5845	94.71 ± 1.8984	96.32 ± 2.5629	96.63 ± 2.1101	94.65 ± 1.9835	82.92 ± 3.6181
6	**65.66 ± 2.0264**	58.41 ± 4.1643	62.84 ± 2.2044	61.86 ± 2.6441	60.79 ± 1.3831	57.28 ± 3.4988	60.24 ± 2.7154
7	97.22 ± 0.8464	95.79 ± 1.6070	87.89 ± 2.5930	94.50 ± 2.3493	**98.25 ± 0.5971**	95.83 ± 0.9768	95.23 ± 1.6006
8	**65.83 ± 2.1269**	59.64 ± 1.8222	62.23 ± 0.8534	61.62 ± 1.2196	57.72 ± 1.3449	60.22 ± 1.162	35.88 ± 0.8977
9	97.81 ± 0.7044	77.49 ± 2.6658	95.67 ± 0.3308	95.47 ± 0.6823	96.18 ± 0.5124	97.83 ± 0.3491	**97.90 ± 0.4299**
10	**95.64 ± 0.3738**	80.20 ± 1.7747	91.20 ± 0.7186	86.37 ± 1.3946	88.39 ± 0.4454	86.23 ± 0.6139	39.02 ± 0.3777
11	**83.79 ± 1.0299**	75.63 ± 0.7395	77.40 ± 1.4196	80.34 ± 0.9412	80.38 ± 0.5018	74.04 ± 0.5152	39.32 ± 0.2671
12	98.89 ± 0.1353	97.71 ± 0.5281	98.79 ± 0.3102	98.24 ± 0.3681	**99.16 ± 0.1517**	85.1 ± 0.4103	43.45 ± 0.1149
Average	**87.5850**	74.5425	79.8758	82.4833	81.9250	76.1433	64.6300

algorithms were implemented in Python 3.10 and executed under the same hardware environment, specifically an Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz processor with 8 GB of memory.

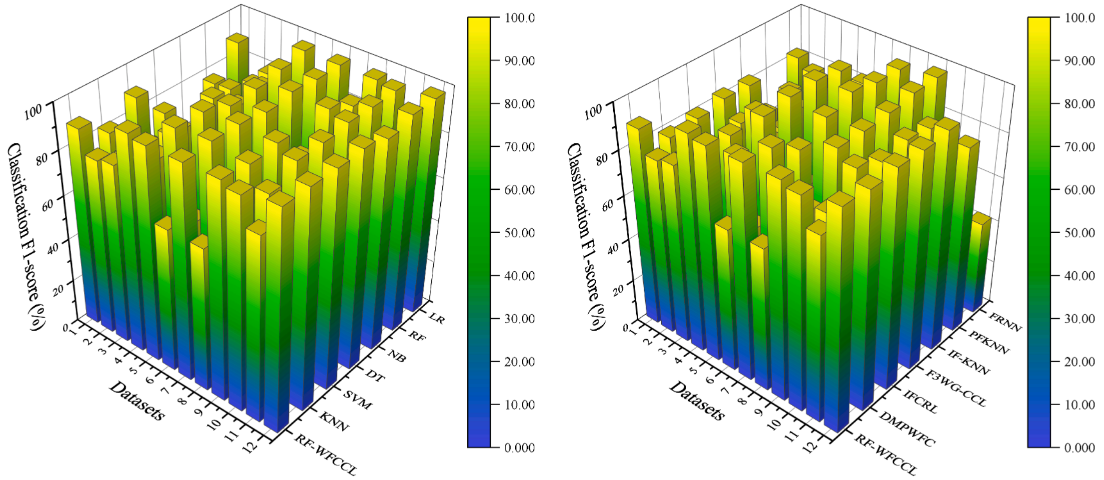
5.2. Comparative experimental analysis

We used classification accuracy and F1-score as evaluation metrics to assess the performance of different algorithms. The detailed results of RF-WFCCL compared with classical machine learning classification algorithms on the selected 12 datasets are shown in Tables 7 and 8. Overall, RF-WFCCL achieved an average accuracy of 89.1975 % and an average F1-score of 87.5850 % across the 12 datasets, obtaining the highest accuracy and F1-score on multiple datasets. Tables 9 and 10 present the accuracy and F1-score



(a) Classic classification (b) Fuzzy-based classification

Fig. 3. Comparison of accuracy bar graphs between RF-WFCCL, classical classification algorithms, and fuzzy-based classification algorithms.



(a) Classic classification (b) Fuzzy-based classification

Fig. 4. Comparison of F1-score bar graphs between RF-WFCCL, classical classification algorithms, and fuzzy-based classification algorithms.

comparisons between RF-WFCCL and fuzzy-based classification algorithms. Compared with DMPWFC, IFCRL, and F3WG-CCL, RF-WFCCL improved the average accuracy by 12.8258 %, 6.5417 %, and 5 %, respectively, and the average F1-score by 13.0425 %, 7.7092 %, and 5.1017 %, respectively. This demonstrates that RF-WFCCL effectively enhances the knowledge representation capability of concepts and improves the depth of knowledge cognition. Furthermore, in the comparative experiment involving 12 different classification algorithms, RF-WFCCL not only exhibited superior accuracy and F1-score but also showed lower standard deviation, indicating strong stability. This performance suggests that RF-WFCCL possesses excellent classification capability, enabling it to adapt to different data characteristics. To provide a more intuitive presentation of the experimental results, we created bar charts (Figs. 3 and 4), which illustrate the accuracy and F1-score comparisons of different classifiers across datasets, further highlighting the substantial advantages of RF-WFCCL in classification performance.

To examine the significant differences between RF-WFCCL and other classification algorithms, we conducted a Wilcoxon paired test with a p-value threshold of 0.05. Table 11 presents the average rankings and Wilcoxon test results for RF-WFCCL and other classification algorithms across 12 datasets. As shown in Table 11, RF-WFCCL achieved an average ranking of 1.7500 in accuracy and 2.0833 in F1-score across the 12 datasets. Moreover, the p-values from the Wilcoxon test between RF-WFCCL and other classification algorithms were all below 0.05, indicating that RF-WFCCL is statistically effective. In conclusion, RF-WFCCL is an excellent cognitive learning method.

Table 11
Average rankings of 13 classification algorithms and Wilcoxon test results.

	RF-FWCCL	KNN	SVM	DT	NB	RF	LR	DMPWFC	IFCRL	F3WG-CCL	IFKNN	PFKNN	FRNN
Accuracy ranking	1.7500	7.2500	4.7500	7.4167	8.6667	5.3333	5.1667	10.5833	7.4167	6.9167	6.5000	9.9167	9.3333
P-value	-	0.0007	0.0046	0.0002	0.0007	0.0061	0.0061	0.0002	0.0005	0.0005	0.0007	0.0002	0.0002
F1-score ranking	2.0833	6.5000	5.6667	7.6667	8.0000	6.2500	5.3333	9.9167	7.6667	6.8333	6.4167	9.2500	9.4167
P-value	-	0.0024	0.0046	0.0002	0.0012	0.0002	0.0046	0.0005	0.0002	0.0005	0.0017	0.0005	0.0007

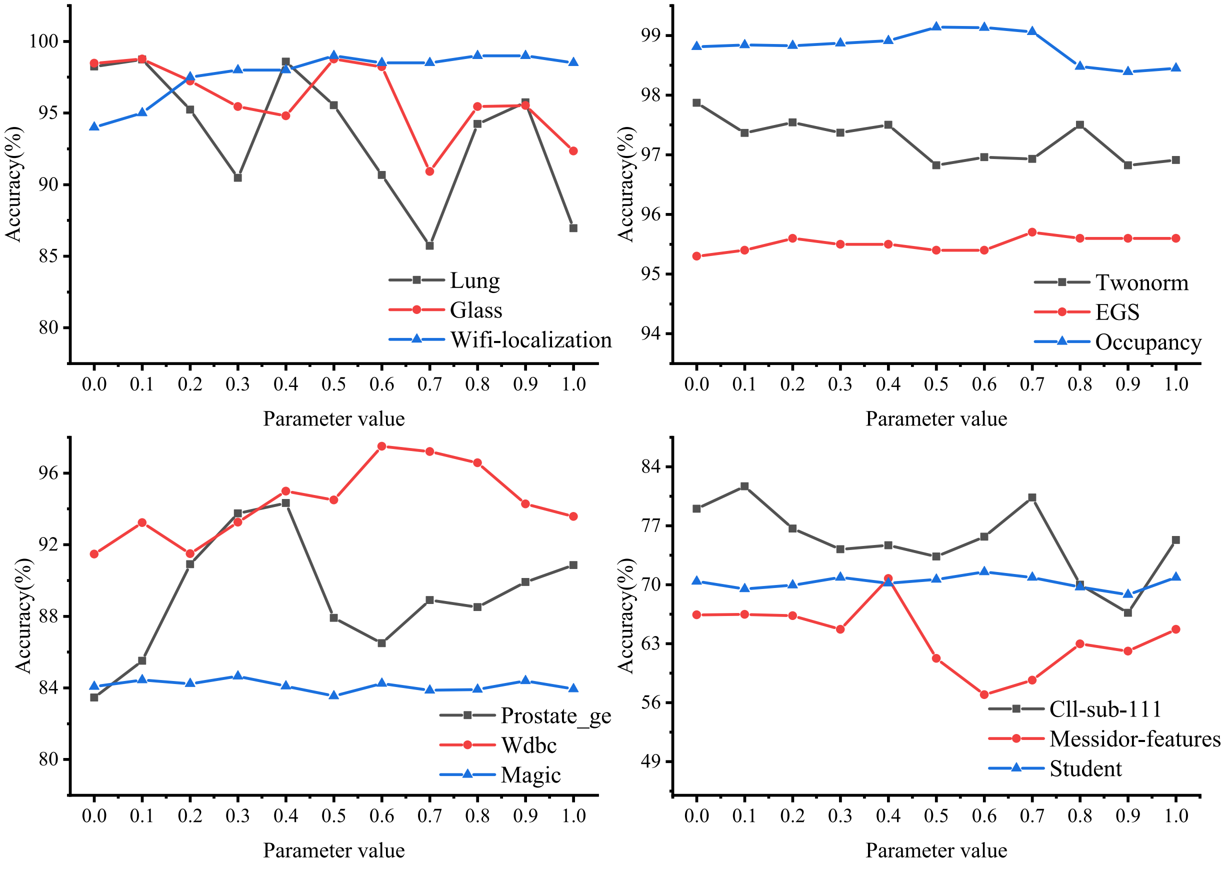


Fig. 5. The relationship between accuracy of RF-WFCCL on 12 datasets and parameter $\lambda(i)$.

Table 12
Comparison of accuracy (Mean % ± Standard Deviation %) between RF-WFCCL and 6 classical machine learning classification algorithms at a noise level of 0.01.

ID	RF-WFCCL	KNN	SVM	DT	NB	RF	LR
1	**90.48 ± 6.0984**	79.29 ± 7.3925	86.48 ± 5.6344	75.52 ± 6.6586	61.05 ± 8.2798	77.90 ± 7.6190	87.90 ± 4.2857
2	**76.00 ± 6.5409**	55.39 ± 6.4348	72.39 ± 5.4307	63.91 ± 7.5355	58.26 ± 6.3958	60.65 ± 7.3712	72.17 ± 6.5361
3	**95.78 ± 2.7192**	90.83 ± 2.4512	94.85 ± 2.4512	83.39 ± 5.3659	79.54 ± 3.5672	88.05 ± 5.2821	92.63 ± 1.4634
4	**98.64 ± 2.0830**	87.21 ± 4.7858	77.33 ± 2.7907	94.65 ± 2.3372	81.16 ± 6.5340	91.30 ± 3.9191	72.98 ± 5.8325
5	95.61 ± 1.5517	95.11 ± 0.9649	95.37 ± 1.5193	91.07 ± 3.6894	92.81 ± 1.7456	**96.05 ± 1.7212**	95.67 ± 1.9536
6	**65.57 ± 2.1741**	59.25 ± 1.4021	60.76 ± 2.0138	59.87 ± 3.3162	56.97 ± 2.7118	62.73 ± 3.2695	61.67 ± 2.0945
7	**97.30 ± 0.4274**	97.21 ± 0.4975	93.95 ± 0.5339	93.03 ± 0.6842	96.95 ± 0.3905	94.85 ± 0.6713	96.65 ± 0.5099
8	69.66 ± 1.0346	63.73 ± 1.5803	**72.05 ± 0.9588**	65.99 ± 1.3317	65.54 ± 0.8291	70.44 ± 0.9375	71.97 ± 1.5674
9	**97.53 ± 0.2306**	96.39 ± 0.3861	97.18 ± 0.3521	84.12 ± 0.9220	97.38 ± 0.3495	92.02 ± 0.4504	96.83 ± 0.2826
10	94.42 ± 0.3528	89.39 ± 0.8076	94.32 ± 0.2250	94.58 ± 0.4326	**95.42 ± 0.1955**	94.32 ± 0.3362	94.74 ± 0.3248
11	**84.81 ± 0.4598**	82.50 ± 0.4657	79.09 ± 0.5769	79.37 ± 0.6013	72.37 ± 0.9462	83.65 ± 0.6618	78.98 ± 0.5935
12	**98.91 ± 0.1134**	98.89 ± 0.1323	98.64 ± 0.0904	98.58 ± 0.1291	96.76 ± 0.2628	98.93 ± 0.1572	98.12 ± 0.1693
Average	**88.7258**	82.9325	85.2008	82.0067	79.5175	84.2408	85.0258

5.3. Sensitivity analysis

According to Algorithm 3, (18) and (19), it is evident that the parameter $\lambda(i)$ significantly influences the entire weighted fuzzy concept cluster $\widehat{\mathcal{G}}_{*}^{\lambda}$. Furthermore, as per (20), the performance of RF-WFCCL is essentially determined by $\widehat{\mathcal{G}}_{*}^{\lambda}$, underscoring the necessity of analyzing how variations in parameter $\lambda(i)$ affect RF-WFCCL's performance. Specifically, for each dataset, we set the learning step of $\lambda(i)$ in Algorithm 3 to 0.1, i.e., $\lambda(i) = \{0, 0.1, \cdots, 1.0\}$, and conducted 10 independent experiments on the same dataset splits for the same $\lambda(i)$ parameters to calculate their average accuracy. Fig. 5 illustrates the trend of RF-WFCCL's average accuracy with varying parameter $\lambda(i)$. We observe significant variations in accuracy for each dataset under different thresholds of $\lambda(i)$, highlighting the importance of selecting an appropriate parameter for each dataset. Additionally, we notice that the highest accuracy achieved for each dataset under different thresholds in Fig. 5 is close to the accuracy obtained through our algorithm for finding the optimal threshold, providing indirect validation of the rationality of our proposed algorithm.

5.4. Anti-noise interference capability

To evaluate the model's resistance to noise interference, we added two different levels of Gaussian white noise (0.01 and 0.05) to the attributes of 12 datasets. By running multiple experiments and calculating the average accuracy, we compared the results with 12 classification algorithms to analyze the model's robustness under different noise conditions. The experimental results are presented in Tables 12–15.

The experimental results show that as the noise level increases, the overall classification accuracy tends to decrease. However, compared to other models, RF-WFCCL demonstrates superior stability, maintaining good classification performance even in high-noise environments. At noise levels of 0.01 and 0.05, RF-WFCCL still achieves the best classification results on most datasets, indicating its stronger resistance to data disturbances and making it more reliable in practical applications.

Table 13
Comparison of accuracy (Mean % ± Standard Deviation %) between RF-WFCCL and 6 fuzzy-based classification algorithms at a noise level of 0.01.

ID	RF-WFCCL	DMPWFC	IFCRL	F3WG-CCL	IF-KNN	PFKNN	FRNN
1	**90.48 ± 6.0984**	79.62 ± 8.4728	78.43 ± 6.1473	81.73 ± 7.5515	80.29 ± 4.2857	59.81 ± 4.8562	77.43 ± 6.5465
2	**76.00 ± 6.5409**	64.53 ± 7.4825	56.52 ± 8.1201	60.33 ± 7.3333	60.26 ± 7.7765	60.09 ± 8.2667	58.72 ± 8.3793
3	**95.78 ± 2.7192**	86.48 ± 4.6342	96.10 ± 2.7160	91.19 ± 2.8571	91.63 ± 3.0463	92.12 ± 3.6177	93.61 ± 1.8252
4	**98.64 ± 2.0830**	35.57 ± 7.6835	71.63 ± 5.9564	85.45 ± 6.3636	85.62 ± 4.0882	79.49 ± 8.0022	72.56 ± 7.1793
5	95.61 ± 1.5517	87.59 ± 1.5374	94.18 ± 1.5379	95.09 ± 2.8070	**95.84 ± 1.8946**	90.04 ± 2.4436	79.04 ± 3.7586
6	**65.57 ± 2.1741**	62.91 ± 2.5439	63.90 ± 1.9185	63.88 ± 3.7626	58.94 ± 3.5019	55.58 ± 2.5257	60.74 ± 4.0288
7	97.30 ± 0.4274	94.52 ± 0.8562	89.18 ± 2.1274	93.50 ± 1.8166	**97.34 ± 0.5937**	95.90 ± 1.1023	94.25 ± 1.1292
8	69.66 ± 1.0346	60.43 ± 1.5432	**69.95 ± 0.7854**	63.24 ± 1.4528	66.81 ± 1.2631	62.85 ± 1.6978	55.49 ± 1.3772
9	**97.53 ± 0.2306**	80.75 ± 1.2642	96.72 ± 0.3954	95.62 ± 0.6412	95.42 ± 0.2652	96.66 ± 0.3299	95.82 ± 0.3550
10	**94.42 ± 0.3528**	78.23 ± 1.4383	90.34 ± 0.5216	88.87 ± 1.4734	88.13 ± 0.4359	87.64 ± 0.5106	63.08 ± 0.9722
11	**84.81 ± 0.4598**	77.06 ± 0.7645	79.42 ± 0.6428	81.64 ± 0.6328	82.39 ± 0.5703	74.73 ± 0.5652	64.56 ± 0.6983
12	98.91 ± 0.1134	93.42 ± 0.3694	97.86 ± 0.3264	**98.95 ± 0.2674**	98.06 ± 0.1072	88.68 ± 0.4270	98.56 ± 0.4694
Average	**88.7258**	75.0925	82.0192	83.2908	83.3942	78.6325	76.1550

Table 14
Comparison of accuracy (Mean % ± Standard Deviation %) between RF-WFCCL and 6 classical machine learning classification algorithms at a noise level of 0.05.

ID	RF-WFCCL	KNN	SVM	DT	NB	RF	LR
1	**88.18 ± 7.1002**	78.10 ± 7.5714	85.52 ± 4.1513	78.57 ± 6.8180	60.95 ± 8.5628	74.29 ± 6.2575	80.95 ± 4.4924
2	**73.33 ± 6.0554**	50.91 ± 8.2980	67.65 ± 6.9526	60.87 ± 6.7356	53.83 ± 7.3271	55.35 ± 9.8764	67.91 ± 5.8332
3	**94.76 ± 2.8716**	92.59 ± 2.7160	90.59 ± 3.4836	82.34 ± 6.4715	75.39 ± 2.1951	88.02 ± 5.3835	90.61 ± 3.0463
4	**92.64 ± 4.7912**	80.02 ± 4.4572	73.91 ± 4.6801	83.26 ± 5.8649	77.21 ± 5.0738	88.30 ± 6.0822	75.58 ± 3.9262
5	93.16 ± 2.6549	94.46 ± 1.6850	92.75 ± 1.8834	88.93 ± 2.3142	92.46 ± 2.5180	93.77 ± 2.4007	**94.53 ± 2.9963**
6	**63.24 ± 2.2655**	54.16 ± 2.6204	56.41 ± 2.3746	59.57 ± 2.9437	57.22 ± 2.5644	61.04 ± 2.4025	63.42 ± 2.6633
7	95.45 ± 0.7684	**95.70 ± 0.9605**	95.65 ± 0.3881	91.53 ± 0.7537	95.38 ± 1.0323	93.77 ± 0.8906	93.82 ± 0.7422
8	66.41 ± 1.6051	60.87 ± 1.8251	69.27 ± 1.1862	62.07 ± 1.7291	63.67 ± 1.4999	68.05 ± 1.1849	**69.66 ± 1.3751**
9	**93.12 ± 0.4005**	90.45 ± 0.4757	92.61 ± 0.4192	80.82 ± 0.8678	91.94 ± 0.4759	90.42 ± 0.5429	93.85 ± 0.2854
10	**91.50 ± 0.3823**	82.19 ± 0.5603	88.66 ± 0.4358	89.85 ± 0.4819	90.99 ± 0.4129	90.81 ± 0.5027	91.21 ± 0.3097
11	78.32 ± 0.5374	73.68 ± 0.7287	70.15 ± 0.5714	73.92 ± 0.6257	67.55 ± 0.4444	**78.35 ± 0.6712**	74.89 ± 0.4868
12	**98.42 ± 0.1254**	98.36 ± 0.1167	97.39 ± 0.1556	97.04 ± 0.2801	96.26 ± 0.1612	98.22 ± 0.1790	97.14 ± 0.1917
Average	**85.7108**	79.2908	81.7133	79.0642	76.9042	81.6992	82.8039

Table 15
Comparison of accuracy (Mean % ± Standard Deviation %) between RF-WFCCL and 6 fuzzy-based classification algorithms at a noise level of 0.05.

ID	RF-WFCCL	DMPWFC	IFCRL	F3WG-CCL	IF-KNN	PFKNN	FRNN
1	**88.18 ± 7.1002**	77.56 ± 8.4672	77.46 ± 5.3683	80.35 ± 6.3472	81.90 ± 7.6190	58.43 ± 8.8960	79.81 ± 5.7143
2	**73.33 ± 6.0554**	62.46 ± 7.4698	53.97 ± 6.9552	60.24 ± 7.5478	66.09 ± 7.1359	64.35 ± 7.0659	57.96 ± 8.5652
3	**94.76 ± 2.8716**	86.46 ± 6.2363	91.58 ± 2.8426	89.32 ± 5.3573	88.41 ± 2.6829	91.46 ± 4.7857	89.15 ± 2.7160
4	**92.64 ± 4.7912**	38.46 ± 7.6426	65.32 ± 5.6321	79.54 ± 5.3684	80.49 ± 4.4673	76.74 ± 6.0644	66.79 ± 5.2054
5	93.16 ± 2.6549	86.42 ± 2.7832	92.28 ± 1.9754	**93.54 ± 2.0876**	92.12 ± 1.0416	86.33 ± 1.4783	78.98 ± 3.5783
6	**63.24 ± 2.2655**	55.33 ± 2.8594	59.86 ± 2.8426	60.02 ± 2.3685	57.51 ± 1.6564	52.63 ± 2.3652	58.48 ± 4.0888
7	95.45 ± 0.7684	92.46 ± 0.8962	85.36 ± 2.4736	91.97 ± 2.0536	**95.95 ± 0.7649**	93.55 ± 0.8718	93.25 ± 1.7783
8	66.41 ± 1.6051	56.28 ± 1.6284	**68.86 ± 0.6528**	60.36 ± 1.5962	60.15 ± 1.7505	57.84 ± 1.4977	58.56 ± 1.5667
9	**93.12 ± 0.4005**	72.85 ± 1.4236	93.04 ± 0.3567	90.24 ± 0.5247	90.78 ± 0.5922	92.73 ± 0.5130	89.99 ± 0.4068
10	**91.50 ± 0.3823**	74.40 ± 2.0846	84.46 ± 0.7452	82.34 ± 1.3532	83.83 ± 0.6701	83.78 ± 0.6816	70.66 ± 0.9553
11	**78.32 ± 0.5374**	66.46 ± 0.6012	72.46 ± 0.6724	77.34 ± 0.8624	78.08 ± 0.6295	70.14 ± 0.5744	70.92 ± 0.4930
12	98.42 ± 0.1254	96.48 ± 0.2376	98.52 ± 0.2628	**98.87 ± 0.2672**	98.43 ± 0.1973	87.95 ± 0.2403	77.11 ± 0.5467
Average	**85.7108**	72.1350	78.5975	80.3442	81.1450	76.3275	74.3050

6. Conclusion

This paper presents the RF-WFCCL model, a novel cognitive learning approach that leverages the attribute importance derived from RF to enhance the WFCCL framework. The model's innovative methodology involves training on fuzzy background data to ascertain attribute weights, which are then utilized to construct a WFC . By refining this space through the elimination of redundant information, we generate an AWFC that more closely mirrors human cognitive processes.

Our approach introduces an algorithm for selecting an optimal threshold $\widehat{\lambda}(i)$ for each dataset, which is pivotal for the formation of an AWFC . This contributes to a more precise and efficient cognitive learning model. To verify the effectiveness of RF-WFCCL, we conducted extensive comparative experiments on datasets from UCI, KEEL, and publicly available tumor gene datasets. The experimental results indicate that RF-WFCCL is statistically more effective compared to other classification algorithms and demonstrates superior classification performance.

However, the current RF-WFCCL model has its limitations. It primarily addresses changes in the concept space when new objects are introduced and does not yet accommodate the addition of new attributes. To advance the model's capabilities, especially for concept cognitive learning with a limited number of samples, integrating RF-WFCCL with semi-supervised learning techniques will be a critical next step. Additionally, considering the complexity of human cognition where objects often have multiple labels, extending RF-WFCCL to handle multi-label scenarios will be an essential area of future research.

In summary, the RF-WFCCL model offers a significant advancement in cognitive learning by providing a more nuanced and dynamic framework for concept space evolution. The model's success in classification tasks underscores its potential for real-world applications. Future work will focus on expanding the model's flexibility and applicability, ensuring it can meet the diverse and evolving demands of cognitive learning tasks.

CRediT authorship contribution statement

Weihua Xu: Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization; **Chongze Zhang:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Investigation, Formal analysis, Data curation.

Data availability

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

Declaration of competing interest

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

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under grant 62376229 and in part by the Natural Science Foundation of Chongqing under grant CSTB2023NSCQ-LZX0027.

References

[1] Y. Shi, Y. Mi, J. Li, W. Liu, Concept-cognitive learning model for incremental concept learning, IEEE Transact. Syst. Man Cybernet.: Syst. 51 (2) (2018) 809–821.
[2] X. Zhu, W. Pedrycz, Z. Li, Construction and evaluation of information granules: from the perspective of clustering, IEEE Transact. Syst. Man Cybernet.: Syst. 52 (3) (2020) 2024–2037.
[3] Y. Zhang, T. Zhao, D. Miao, W. Pedrycz, Granular multilabel batch active learning with pairwise label correlation, IEEE Transact. Syst. Man Cybernet.: Syst. 52 (5) (2021) 3079–3091.
[4] K. Yuan, D. Miao, W. Pedrycz, H. Zhang, L. Hu, Multigranularity data analysis with zentropy uncertainty measure for efficient and robust feature selection, IEEE Trans. Cybern. 55 (2) (2024).
[5] K. Yuan, D. Miao, W. Pedrycz, W. Ding, H. Zhang, Ze-HFS: zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery, IEEE Trans. Knowl. Data Eng. 36 (11) (2024) 7326–7339.
[6] W. Xu, Y. Chen, Multi-attention concept-cognitive learning model: a perspective from conceptual clustering, Knowl. Base. Syst. 252 (2022) 109472.
[7] K. Yuan, W. Xu, W. Li, W. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inf. Sci. (N.Y.) 584 (2022) 127–147.
[8] D. Guo, W. Xu, W. Ding, Y. Yao, X. Wang, W. Pedrycz, Y. Qian, Concept-cognitive learning survey: mining and fusing knowledge from data, Inform. Fusion 109 (2024) 102426.
[9] C. Ma, L. Zhang, W. Pedrycz, W. Lu, The long-term prediction of time series: a granular computing-based design approach, IEEE Transact. Syst. Man Cybernet.: Syst. 52 (10) (2022) 6326–6338.
[10] P. Tang, Y. Dong, S. Mao, H.-L. Wei, J. Jin, Online classification of network traffic based on granular computing, IEEE Transact. Syst. Man Cybernet.: Syst. 53 (8) (2023) 5199–5211.
[11] Z. Pawlak, L. Polkowski, A. Skowron, Rough set theory, KI 15 (3) (2001) 38–39.
[12] Q. Zhang, C. Yang, G. Wang, A sequential three-way decision model with intuitionistic fuzzy numbers, IEEE Transact. Syst. Man Cybernet.: Syst. 51 (5) (2019) 2640–2652.
[13] J. Ding, C. Zhang, D. Li, J. Zhan, W. Li, Y. Yao, Three-way decisions in generalized intuitionistic fuzzy environments: survey and challenges, Artif. Intell. Rev. 57 (2) (2024) 38.
[14] H. Zhi, J. Li, Y. Li, Multilevel conflict analysis based on fuzzy formal contexts, IEEE Trans. Fuzzy Syst. 30 (12) (2022) 5128–5142.
[15] R. Wille, Concept lattices and conceptual knowledge systems, Comput. Math. Applic. 23 (6–9) (1992) 493–515.
[16] W.X. Zhang, W.H. Xu, Cognitive model based on granular computing, Chin. J. Eng. Math. 24 (6) (2007) 957–971.
[17] X. Kang, D. Li, S. Wang, K. Qu, Formal concept analysis based on fuzzy granularity base for different granulations, Fuzzy Set. Syst. 203 (2012) 33–48.
[18] C. Huang, J. Li, C. Mei, W.-Z. Wu, Three-way concept learning based on cognitive operators: an information fusion viewpoint, Int. J. Approxim.Reason. 83 (2017) 218–242.
[19] W. Xu, D. Guo, Y. Qian, W. Ding, Two-way concept-cognitive learning method: a fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. (2022) 1885–1899.
[20] 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. (2023) 6798–6812.
[21] X. Zhang, D. Guo, W. Xu, Two-way concept-cognitive learning with multi-source fuzzy context, Cognit. Comput. 15 (5) (2023) 1526–1548.
[22] Y. Yao, Interpreting concept learning in cognitive informatics and granular computing, IEEE Transact. Syst. Man Cybernet., Part B (Cybernet.) 39 (4) (2009) 855–866.
[23] W. Xu, J. Pang, S. Luo, A novel cognitive system model and approach to transformation of information granules, Int. J. Approxim. Reason. 55 (3) (2014) 853–866.
[24] 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. Base. Syst. 260 (2023) 110093.
[25] 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) 1–14.
[26] 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.
[27] D. Guo, W. Xu, Y. Qian, W. Ding, Fuzzy-granular concept-cognitive learning via three-way decision: performance evaluation on dynamic knowledge discovery, IEEE Trans. Fuzzy Syst. 32 (3) (2024) 1409–1423.
[28] J. Wu, E.C.C. Tsang, W. Xu, C. Zhang, L. Yang, Correlation concept-cognitive learning model for multi-label classification, Knowl. Base. Syst. 290 (2024) 111566.
[29] Y.-F. Zhou, H.-L. Yang, J.-J. Li, D.-L. Wang, Skill assessment method: a perspective from concept-cognitive learning, Fuzzy Set. Syst. 508 (2025) 109331.
[30] N. Xie, E. Yan, X. Yao, Q. Chen, T. Hao, A concept-cognitive learning approach from the perspective of causal reasoning, Neurocomputing (2025) 131876.
[31] J. Huang, Y. Lin, W. Sun, Fuzzy concept cognitive learning based on knowledge space theory, AIMS Math. 10 (9) (2025) 22127–22149.
[32] Y. Ding, W. Xu, Multi-granularity interval-intent fuzzy concept-cognitive learning: an attention-enhanced adaptive clustering framework, Inform. Fusion (2025) 103715.
[33] S.B. Yahia, K. Arour, A. Slimani, A. Jaoua, Discovery of compact rules in relational databases, Inform. Sci. J. 4 (3) (2000) 497–511.
[34] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[35] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[36] Y.-Y. Song, L.U. Ying, Decision tree methods: applications for classification and prediction, Shanghai Arch. Psychiatry 27 (2) (2015) 130.
[37] K.P. Murphy, et al., Naive bayes classifiers, Univers. Brit. Columb. 18 (60) (2006) 1–8.
[38] D.W. Hosmer, Jr, S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, John Wiley & Sons, 2013.
[39] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27.
[40] S. Huang, N. Cai, P.P. Pacheco, S. Narrandes, Y. Wang, W. Xu, Applications of support vector machine (SVM) learning in cancer genomics, Cancer Genom. Proteom. 15 (1) (2018) 41–51.
[41] L. Kuncheva, An intuitionistic fuzzy k-nearest neighbors rule, Notes Intuit. Fuzzy Set. 1 (1995) 56-60.
[42] M. Arif, M.U. Akram, et al., Pruned fuzzy K-nearest neighbor classifier for beat classification, J. Biomed. Sci. Eng. 3 (04) (2010) 380.
[43] M. Sarkar, Fuzzy-rough nearest neighbor algorithms in classification, Fuzzy Set. Syst. 158 (19) (2007) 2134–2152.