Information Fusion 124 (2025) 103361

Contents lists available at ScienceDirect

Information Fusion

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

Full length article

Multi-level correlation information fusion via three-way concept-cognitive learning for multi-label learning

Jiaming Wu [a], Eric C.C. Tsang [a,*], Weihua Xu [b], Chengling Zhang [a], Qianshuo Wang [c]

[a] School of Computer Science and Engineering, Macau University of Science and Technology, Macao Special Administrative Region of China
[b] College of Artificial Intelligence, Southwest University, Chongqing, 400715, China
[c] College of Computer and Cyber Security, Fujian Normal University, Fujian, 350007, China

ARTICLE INFO

Keywords:
Concept-cognitive learning
Multi-label learning
Label correlation
Fuzzy formal context
Three-way decision

ABSTRACT

Multi-label learning tasks typically involve complex correlation between labels, which often span across multiple levels. Accurately capturing and fusing these multi-level correlation information is crucial for improving prediction performance and understanding the potential relationship between labels. The current mainstream label correlation acquisition methods mainly focus on statistical analysis of labels. However, these methods lack exploration of the hierarchical structure of correlation, which may lead to the cognitive bias of labels and the decline in predictive performance. To address this, a multi-label learning model with multi-level correlation information fusion via three-way concept-cognitive learning (MCF-3WCCL) is proposed to capture the hierarchical correlation between labels more comprehensively, improve the prediction performance and enhance the interpretability. In this model, three-way concept-cognitive operators are utilized to structurally represent label concepts, thereby capturing the hierarchical correlations among labels. Additionally, the extent of label concepts are used as clues, which are mapped into feature concepts to form the dependencies between labels and features. On this basis, by fusing these feature concepts, the overall cognition of the label is finally formed. Extensive comparative experiments reflect that the proposed method is superior and versatile.

1. Introduction

Traditional supervised learning only allows each sample to be associated with one of multiple candidate labels. However, in the real world, the data is often more complex, with a single sample typically associated with multiple labels. Traditional supervised learning has been unable to meet this demand. As a result, multi-label learning has garnered extensive attention in various fields such as sentiment analysis, text recognition, data mining and so on [1–5]. The hallmark of multi-label learning is that the label variables show the intrinsic correlation in the label space [6–9]. Therefore, an important research direction is how to mine and utilize useful label correlation from complex multi-label data.

Recently, the study of label correlation in multi-label learning has garnered widespread attention [10–12]. label correlation can significantly improve model performance and enhance model interpretability [7]. Based on this, many excellent methods have been developed [13–15]. Dai et al. [16] introduced a method using fuzzy conditional mutual information to calculate mutual information between labels, effectively capturing potential correlations within the label space. GLFS [17] improves generalization performance by using local label correlations, which are learned through jointly learning of common features among similar labels and the specific features of each individual label. 2SML [18] utilizes highly representative instances to learn implicit correlations. Although these methods have achieved excellent performance to varying degrees, most of them focus on modeling label correlations at a single level, overlooking the potential multi-level and hierarchical structural relationships among labels. In addition, some methods are also affected by the class imbalance problem. Zhao et al. [19,20] attempted to alleviate this limitation by introducing label distribution information and label correlations for data enhancement. Other studies have sought to mine both positive and negative semantic information from labels to improve the understanding of label semantics [21,22], but they also lack in-depth modeling of correlation hierarchies. In fact, when processing complex information, the human cognitive system often follows a knowledge formation process that progresses from global to local, and from abstract to concrete. High-level information provides global guidance for overall direction, while low-level information contributes to detail

* Corresponding author.
E-mail addresses: wjmfuzzy@163.com (J. Wu), cctsang@must.edu.mo (E.C.C. Tsang), chxuwh@gmail.com (W. Xu), chengling_zhang@126.com (C. Zhang), 121152022075@student.fjnu.edu.cn (Q. Wang).

https://doi.org/10.1016/j.inffus.2025.103361
Received 6 December 2024; Received in revised form 17 April 2025; Accepted 22 May 2025
Available online 6 June 2025

refinement and supplementation. In multi-label learning, correlations at different levels influence the target label in different ways. Prioritizing the extraction of high-level guidance information and then gradually integrating local details aligns more closely with the logic of human cognitive development. Therefore, establishing a model that simulates this cognitive mechanism and effectively captures multi-level label correlations is of great significance for enhancing the model's reasoning and decision-making abilities in complex label spaces.

It is worth noting that concept-cognitive learning (CCL) is an effective tool for knowledge mining. By simulating human cognitive process, learning concepts and revealing the hierarchical structure between concepts, complex knowledge can be more easily understood and applied. CCL is a novel paradigm in intelligent learning, which is widely used in the fields of cognitive computing and artificial intelligence [23–26]. In the 1980s, Wille [27] introduced formal concept analysis (FCA), a formal method for defining the structure and relationships of concepts. A classical concept typically comprises two components: extent and intent, each of which can be uniquely determined by the other [28,29]. On this basis, CCL simulates the behavior of the human brain in concept learning through a specific cognitive model, aiming to identify concepts and learn objects from given clues [30]. In this process, concepts play a crucial role in knowledge mining and integrating data into wisdom. Recently, scholars have put forward various CCL models and methods to meet different practical needs [31–33]. Guo et al. [34] proposed a concept recall mechanism, which integrates past experience into the system by recalling relevant knowledge, and realizes the dynamic update of knowledge. To effectively avoids the high dependency of cognitive results on the order of attribute cognition, Liu et al. [35] adopted a random strategy independent of attribute order. However, the above methods are still not sufficient for the completeness of knowledge description. It is particularly noteworthy that the CCL model based on the three-way decision theory [36–38] is a significant representative study. This model introduces the idea of three divisions: the positive, negative, and boundary regions. It has achieved significant results in terms of the completeness of knowledge description and the reduction of cognitive bias [39–42]. Taking into account the limitations of individual cognition and the incompleteness of cognitive environment, Yuan et al. [43] established a progressive fuzzy three-way concept. In order to make the generated concept more flexible, Zhang et al. [44] put forward a variable precision three-way concept induced by objects. This method allows multiple suitable concepts to be generated by setting different thresholds. Nevertheless, these methods focus on the concept generation in the feature space and ignore the complex semantic relationships between labels. This limits the application of CCL in multi-label learning.

In view of the advantages of CCL in knowledge mining, CCL provides an effective approach for the precisely mining multi-level correlation information by comprehending and utilizing the relationship and hierarchical structure between concepts. Additionally, label concepts can help us discover implicit relationships between labels rather than relying solely on statistical co-occurrence. This method avoids cognitive biases that may result from label sparsity and builds more interpretable correlation model. Therefore, researchers have combined CCL with multi-label learning and achieved good results. Wu et al. [45] proposed a multi label classification based on the correlation concepts of positive and negative. Liu et al. [46] proposed a multi-level information fusion approach for stochastic concept clustering to deal with missing labels. However, they have a drawback in that they cannot be directly applied to continuous values. Instead, continuous values must be converted into discrete ones through discretization, which may lead to the loss of some information during this process. Therefore, further exploring the combination of CCL and multi-label learning is promising and worthy of investigation.

Based on the above discussion, we propose a multi-level correlation information fusion method via three-way concept-cognitive learning for multi-label learning (MCF-3WCCL). Utilize the three-way concept-cognitive learning to obtain label concepts and calculate the degree of importance of label concepts to the target label based on their structural information. We use the extent of label concepts as clues to construct the relationship between the label and the feature. Further, by fusing these feature concepts, the overall cognition of the label is finally formed. This approach can better leverage the effects of correlation information at different levels on labels, enhancing the model's predictive performance and interpretability. The main contributions of this paper are as follows:

- The three-way concept-cognitive operator integrates positive and negative information to learn the label concept, which makes the acquired knowledge description more accurate and effectively reduces the cognitive bias of the label.
- The model calculates the degree of importance of label concepts to the target label based on their structural information, effectively avoid the system paying too much attention to the local details and losing the integrity and improves the interpretability of the model.
- Constructing label correlation matrix from both extent and intent perspectives enables the multi-label learning model to more comprehensively understand and utilize the intrinsic structural information of the data, thereby improving the prediction performance and generalization ability of the model.

The structure of this paper is as follows: Section 2 reviews the preliminary knowledge relevant to this paper. Section 3 provides a detailed introduction to multi-level label correlation information mining and fusing mechanism based on three-way concept analysis. In Section 4, the MCF-3WCCL is verified and analyzed by experiments. The final section gives the conclusion.

2. Preliminaries

This section briefly reviews multi-label learning, fuzzy formal concept analysis and fuzzy three-way concept analysis to clarify the knowledge and notations used in this paper. For convenience, the main symbols used in this paper are listed in Table 1.

2.1. Multi-label learning

In multi-label learning framework [1,7], a multi-label dataset $\boldsymbol{D} = \left\{\left(\mathbf{x}_i, \mathbf{y}_i\right)\right\}_{i=1}^{n}$ with n samples is given, where $\mathbf{x}_i = \left[x_{i1}, x_{i2}, \ldots, x_{im}\right]$ is the m-dimensional feature vector of x_i and $\mathbf{y}_i = \left[y_{i1}, y_{i2}, \ldots, y_{iq}\right]$ is the q-dimensional label vector of x_i. The value $y_{ij} \in \{0,1\}$ characterizes the binary relationship between the label variable l_j and the sample x_i, where $y_{ij} = 1$ means that the sample x_i is associated with the label l_j, and $y_{ij} = 0$ means that the sample x_i is not associated with the label l_j. The feature matrix of the multi-label dataset is represented as $\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right]^{\mathrm{T}} \in \mathbb{R}^{n\times m}$, and the label matrix is $\mathbf{Y} = \left[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\right]^{\mathrm{T}} \in \{0,1\}^{n\times q}$.

The goal of multi-label learning is to obtain a mapping relationship from the feature space to the label space, and to predict the labels of samples with known features but unknown labels. Hence, the multi-label learning problem can be formulated as follows:

$$\min_{W} \frac{1}{2}\|\mathbf{XW} - \mathbf{Y}\|_F^2 + \lambda\|\mathbf{W}\|_1,$$

where $\mathbf{W} \in \mathbb{R}^{m\times q}$ is a weight matrix, $\|\cdot\|_F$ represents the matrix F-norm, $\|\mathbf{W}\|_1$ represents a regularization term, λ is a tradeoff parameter.

Table 1
The summary of the main symbols.

Symbol	Description	Symbol	Description
$\mathbf{X}$	The feature matrix, $\mathbf{X} \in \mathbb{R}^{n\times m}$	$\tilde{\mathcal{L}}^{\triangledown}$, $\mathcal{H}^{\triangledown}$	A pair of fuzzy three-way concept cognitive operators
$\mathbf{Y}$	The label matrix, $\mathbf{Y} \in \{0,1\}^{n\times q}$	$\mathcal{L}^{\triangle}$, $\mathcal{H}^{\triangle}$	A pair of three-way concept cognitive operators
$\mathbf{W}$	The weight matrix, $\mathbf{W} \in \mathbb{R}^{m\times q}$	$\mathcal{LCS}_j$	The 3WL-concept space of the label l_j
$\|\cdot\|_1$	The l_1-norm	$\mathcal{WLCS}_j$	The weighted 3WL-concept space of the label l_j
$\|\cdot\|_F^2$	The F-norm	$\mathcal{WFCS}_j$	The weighted fuzzy 3WF-concept space of the label l_j
U	The set of objects	pc_j	The overall cognition of the label l_j
A	The set of features	r_{ex}	The extent relevance
L	The set of labels	r_{in}	The intent relevance
$\tilde{A}(a)$	The fuzzy membership degree of a in $\tilde{A}$	r_{ij}	The label correlation between l_i and l_j
$\tilde{R}_A$	The fuzzy relation $\tilde{R}_A : U \times A \to [0,1]$	$\mathbf{R}$	The label correlation matrix
R_L	The binary relation $R_L : U \times L \to \{0,1\}$		

2.2. Fuzzy formal concept analysis

In this subsection, we describe the fuzzy formal context and the fuzzy concept, which are also discussed in [24,28,31].

A triplet $\left(U, A, \tilde{R}_A\right)$ is called a fuzzy formal context, where $U = \left\{x_1, x_2, \cdots, x_n\right\}$ and $A = \left\{a_1, a_2, \cdots, a_m\right\}$ represent an object set and a feature set, respectively. The power sets of U is represented by $\mathcal{P}(U)$. $\tilde{R}_A$ is a fuzzy relation between U and A. The fuzzy relation $\tilde{R}_A : U \times A \to [0,1]$, $\tilde{R}_A(x,a)$ denotes the membership degree of x with respect a.

Assume A is a universe, for any $a \in A$, the value $\tilde{A}(a) : A \to [0,1]$ is referred to as the fuzzy membership degree of a in $\tilde{A}$. Then $\mathcal{F}(A)$ represents the set of all fuzzy subsets of A. Let $\tilde{B}_1$ and $\tilde{B}_2$ be two fuzzy sets on A. If $\tilde{B}_1(a) \leqslant \tilde{B}_2(a)$, then $\tilde{B}_1$ is a subset of $\tilde{B}_2$, i.e., $\tilde{B}_1 \subseteq \tilde{B}_2$.

Definition 1. Let $\left(U, A, \tilde{R}_A\right)$ be a fuzzy formal context. For any $X \subseteq U$, $B \subseteq A$ and $\tilde{B} \in \mathcal{F}(A)$, the concept cognitive operators $\tilde{\mathcal{L}} : \mathcal{P}(U) \to \mathcal{F}(A)$ and $\mathcal{H} : \mathcal{F}(A) \to \mathcal{P}(U)$ are defined by:

$$\tilde{\mathcal{L}}(X)(a) = \bigwedge_{x\in X} \tilde{R}_A(x,a), a \in A,$$
$$\mathcal{H}\left(\tilde{B}\right) = \left\{x \in U \,|\, \forall a \in B, \tilde{B}(a) \leqslant \tilde{R}_A(x,a)\right\}.$$

Property 1. *For $\forall X, X_1, X_2 \in \mathcal{P}(U)$ and $\tilde{B}, \tilde{B}_1, \tilde{B}_2 \in \mathcal{F}(A)$, the following statements hold:*

(1) $X_1 \subseteq X_2 \Rightarrow \tilde{\mathcal{L}}\left(X_2\right) \subseteq \tilde{\mathcal{L}}\left(X_1\right)$;
(2) $\tilde{B}_1 \subseteq \tilde{B}_2 \Rightarrow \mathcal{H}\left(\tilde{B}_2\right) \subseteq \mathcal{H}\left(\tilde{B}_1\right)$;
(3) $X \subseteq \mathcal{H}\tilde{\mathcal{L}}(X), \tilde{B} \subseteq \tilde{\mathcal{L}}\mathcal{H}\left(\tilde{B}\right)$;
(4) $\tilde{\mathcal{L}}\left(X_1 \cup X_2\right) = \tilde{\mathcal{L}}\left(X_1\right) \cap \tilde{\mathcal{L}}\left(X_2\right)$;
(5) $\mathcal{H}\left(\tilde{B}_1 \cup \tilde{B}_2\right) = \mathcal{H}\left(\tilde{B}_1\right) \cap \mathcal{H}\left(\tilde{B}_2\right)$;
(6) $\tilde{\mathcal{L}}\left(X_1 \cap X_2\right) \supseteq \tilde{\mathcal{L}}\left(X_1\right) \cap \tilde{\mathcal{L}}\left(X_2\right)$;
(7) $\mathcal{H}\left(\tilde{B}_1 \cap \tilde{B}_2\right) \supseteq \mathcal{H}\left(\tilde{B}_1\right) \cup \mathcal{H}\left(\tilde{B}_2\right)$.

Thus, a pair $\left(X, \tilde{B}\right)$ is fuzzy concept if $\tilde{\mathcal{L}}(X) = \tilde{B}$ and $\mathcal{H}\left(\tilde{B}\right) = X$. In general, X and $\tilde{B}$ are called extent and intent, respectively, which can be uniquely determined by each other. The fuzzy concept lattice $\tilde{L}\left(U, A, \tilde{R}\right)$ is the referred to all fuzzy concepts in $\left(U, A, \tilde{R}_A\right)$. For $\forall \left(X_1, \tilde{B}_1\right), \left(X_2, \tilde{B}_2\right) \in \tilde{L}\left(U, A, \tilde{R}_A\right)$, the ordered by $\left(X_1, \tilde{B}_1\right) \leqslant \left(X_2, \tilde{B}_2\right) \Leftrightarrow X_1 \subseteq X_2 \Leftrightarrow \tilde{B}_2 \subseteq \tilde{B}_1$.

2.3. Fuzzy three-way concept analysis

Subsequently, the authors [38,40] combines three-way theory with formal concept analysis to explore it through positive and negative information, leading to three-way concept analysis. Correspondingly, a pair of negative concept cognitive operators are represented as follows.

Let $\tilde{R}_A^- = U \times A - \tilde{R}_A$ be the complement of the relation $\tilde{R}_A$, where $\tilde{R}_A^-(x,a) = 1 - \tilde{R}_A(x,a)$ reflects the non-membership degree of object x to attribute a.

Definition 2. Let $\left(U, A, \tilde{R}_A\right)$ be a fuzzy formal context. For any $X \subseteq U$, $B \subseteq A$ and $\tilde{B} \in \mathcal{F}(A)$, the negative concept cognitive operators $\tilde{\mathcal{L}}^- : \mathcal{P}(U) \to \mathcal{F}(A)$ and $\mathcal{H}^- : \mathcal{F}(A) \to \mathcal{P}(U)$ are given as follows:

$$\tilde{\mathcal{L}}^-(X)(a) = \bigwedge_{x\in X} \tilde{R}_A^-(x,a), a \in A,$$
$$\mathcal{H}^-\left(\tilde{B}\right) = \left\{x \in U \,|\, \forall a \in B, \tilde{B}(a) \leqslant \tilde{R}_A^-(x,a)\right\},$$

where $\tilde{B}$ is the fuzzy set on the complement of A.

Furthermore, to represent both positive and negative information concurrently, we combine positive and negative cognitive operators to create three-way cognitive operators.

Definition 3. Let $\left(U, A, \tilde{R}_A\right)$ be a fuzzy formal context. For any $X \subseteq U$ and $\tilde{B}_1, \tilde{B}_2 \in \mathcal{F}(A)$, the fuzzy three-way concept cognitive operators $\tilde{\mathcal{L}}^{\triangledown} : \mathcal{P}(U) \to \mathcal{F}(A) \times \mathcal{F}(A)$ and $\mathcal{H}^{\triangledown} : \mathcal{F}(A) \times \mathcal{F}(A) \to \mathcal{P}(U)$ are defined by:

$$\tilde{\mathcal{L}}^{\triangledown}(X) = \left(\tilde{\mathcal{L}}(X), \tilde{\mathcal{L}}^-(X)\right),$$
$$\mathcal{H}^{\triangledown}\left(\tilde{B}_1, \tilde{B}_2\right) = \mathcal{H}\left(\tilde{B}_1\right) \cap \mathcal{H}^-\left(\tilde{B}_2\right).$$

Then, $\left(X, \left(\tilde{B}_1, \tilde{B}_2\right)\right)$ is a fuzzy three-way concept if $\tilde{\mathcal{L}}^{\triangledown}(X) = \left(\tilde{B}_1, \tilde{B}_2\right)$ and $\mathcal{H}^{\triangledown}\left(\tilde{B}_1, \tilde{B}_2\right) = X$. Obviously, $\left(\mathcal{H}^{\triangledown}\tilde{\mathcal{L}}^{\triangledown}(X), \tilde{\mathcal{L}}^{\triangledown}(X)\right)$ represent the object-induced fuzzy three-way concept.

3. Mining multi-level correlation information based on three-way concept analysis

Generally speaking, humans typically adopt a multi-level cognitive style in the process of understanding labels. High-level information helps us quickly grasp the main characteristics and overarching significance of labels, while low-level information focuses on local features and unique meanings. By processing multi-level information, humans avoid overlooking the global significance while also considering the specific details, so as to make a more accurate and comprehensive understanding of labels. In this section, It should be noted that each label concept contains rich label correlation information. The weights are assigned according to the degree of importance of these multi-level correlation information to the target label. Furthermore, the overall cognition of the target label is obtained by associating and fusing the corresponding feature concepts. Finally, a multi-label learning model, named MCF-3WCCL, is constructed.

3.1. Acquire the multi-level correlation information

In multi-label data, labels are usually scarce and uncertain, making it challenging to accurately identify and learn them. Additionally, complex correlations exist between labels, and relying solely on limited positive label information may result in insufficient model learning or poor generalization. Therefore, to better model and utilize the limited label information, it is crucial to fully utilize both the positive and negative information in the labels, as well as the structural relationships between label semantics. In this subsection, we will use three-way concept analysis to uncover potential multi-level label correlations and enhance the model's understanding of labels.

Table 2
A multi-label context about animals.

U	a_1	a_2	a_3	a_4	l_1	l_2	l_3	l_4	l_5
x_1	0.8	0.4	0.6	0.3	1	1	0	1	0
x_2	0.5	0.7	0.2	0.1	1	0	0	0	0
x_3	0.1	0.9	0.8	0.7	0	0	1	1	1
x_4	0.6	0.2	0.4	0.3	1	0	0	1	0
x_5	0.3	0.5	0.7	0.9	0	0	1	0	1
x_6	0.4	0.8	0.3	0.6	0	0	1	1	0
x_7	0.9	0.6	0.4	0.2	1	0	0	0	1

Definition 4. A quintuple $\langle U, A \cup L, \tilde{R}_A, R_L \rangle$ is called a multi-label context, where $U = \{x_1, x_2, \dots, x_n\}$ represents a non-empty finite sample set, $A = \{a_1, a_2, \dots, a_m\}$ represents a non-empty finite feature set, and $L = \{l_1, l_2, \dots, l_q\}$ represents a non-empty finite label set, which contains q possible label variables. $\tilde{R}_A : U \times A \to [0,1]$ is a fuzzy relation between U and A, $\tilde{R}_A(x, a)$ denotes the membership degree of x with respect a. $R_L : U \times L \to \{0,1\}$ is a binary relation between U and L, $xR_L l = 1$ represents object x with label l.

Definition 5. Let $\langle U, A \cup L, \tilde{R}_A, R_L \rangle$ be a multi-label context. For any $X \subseteq U$ and $L_1, L_2 \subseteq L$, the three-way concept-cognitive operators $\mathcal{L}^{\triangle} : \mathcal{P}(U) \to \mathcal{P}(L) \times \mathcal{P}(L)$ and $\mathcal{H}^{\triangle} : \mathcal{P}(L) \times \mathcal{P}(L) \to \mathcal{P}(U)$ are defined by:

$$\mathcal{L}^{\triangle}(X) = (\mathcal{L}(X), \mathcal{L}^{-}(X)),$$
$$\mathcal{H}^{\triangle}(L_1, L_2) = \mathcal{H}(L_1) \cap \mathcal{H}^{-}(L_2),$$

where, the positive operators $\mathcal{L}(X)$ and $\mathcal{H}(L_1)$ are as follows:

$$\mathcal{L}(X) = \{l \in L | \forall x \in X, xR_L l = 1\},$$
$$\mathcal{H}(L_1) = \{x \in U | \forall l \in L_1, xR_L l = 1\}.$$

The negative operators $\mathcal{L}^{-}(X)$ and $\mathcal{H}^{-}(L_2)$ are as follows:

$$\mathcal{L}^{-}(X) = \{l \in L | \forall x \in X, xR_L l = 0\},$$
$$\mathcal{H}^{-}(L_2) = \{x \in U | \forall l \in L_2, xR_L l = 0\}.$$

Then, $(X, (L_1, L_2))$ is a three-way label concept (3WL-concept) if $\mathcal{L}^{\triangle}(X) = (L_1, L_2)$ and $\mathcal{H}^{\triangle}(L_1, L_2) = X$. Obviously, $\left(\mathcal{H}^{\triangle}\mathcal{L}^{\triangle}(X), \mathcal{L}^{\triangle}(X)\right)$ and $\left(\mathcal{H}^{\triangle}(L_1, L_2), \mathcal{L}^{\triangle}\left(\mathcal{H}^{\triangle}(L_1, L_2)\right)\right)$ are 3WL-concepts. Moreover, if $X \subseteq X'$ or $(L_1', L_2') \subseteq (L_1, L_2)$, then $(X, (L_1, L_2))$ is referred to as a sub-concept of $(X', (L_1', L_2'))$, denoted by $(X, (L_1, L_2)) \leqslant (X', (L_1', L_2'))$.

Definition 6. Let $\langle U, A \cup L, \tilde{R}_A, R_L \rangle$ be a multi-label context, for any $l \in L$ and $x \in U_j \subseteq U$, where U_j is the set of all samples with label l_j, then the 3WL-concept space of the label l_j is defined by :

$$\begin{aligned} \mathcal{LCS}_j = &\{(\mathcal{HL}(x) \cap \mathcal{H}^{-}\mathcal{L}^{-}(x), (\mathcal{L}(x), \mathcal{L}^{-}(x))) | x \in U_j\} \\ &\cup \{(\mathcal{H}(l) \cap \mathcal{H}^{-}(l^{\neg}), (\mathcal{LH}(l), \mathcal{L}^{-}\mathcal{H}^{-}(l^{\neg}))) | l \in L\}, \end{aligned}$$

where $l^{\neg}$ denotes any label except l. $\mathcal{LCS}_j$ denotes the set of all 3WL-concepts under the label l_j. For convenience, we briefly note the 3WL-concept as $(X_k, (L_k^+, L_k^-))$.

In reality, each 3WL-concept has rich label correlation information, L_k^+ contains labels that are common to X_k, and L_k^- contains labels that are not common to X_k. In addition, multi-level relationships can be formed between 3WL-concepts. The following we through a simple example to illustrate.

Example 1. Table 2 is a multi-label context, where $U = \{x_1, x_2, \dots, x_7\}$ correspond to Lion, Elephant, Eagle, Wolf, Swallow, Owl, and Cow, respectively. $A = \{a_1, a_2, a_3, a_4\}$ is a feature set, and $L = \{l_1, l_2, \dots, l_5\}$ is a label set where $l_1, l_2, \dots, l_5$ correspond to Mammal, Felid, Bird, Predator, and Migratory Animal, respectively.

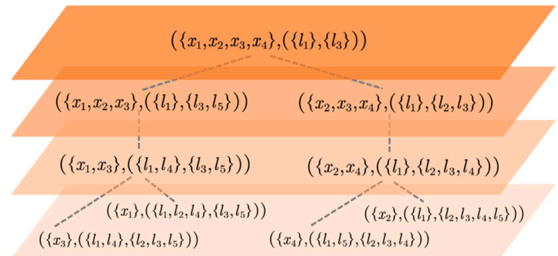

Fig. 1. Multi-level relationships between 3WL-concepts.

Taking the label l_1 as an example, we can calculate the 3WL-concept space of the label l_1 :

$$\mathcal{LCS}_1 = \left\{ \begin{array}{c} (\{x_1, x_2, x_4, x_7\}, (\{l_1\}, \{l_3\})), (\{x_1, x_2, x_4\}, (\{l_1\}, \{l_3, l_5\})), \\ (\{x_2, x_4, x_7\}, (\{l_1\}, \{l_2, l_3\})), \\ (\{x_1, x_4\}, (\{l_1, l_4\}, \{l_3, l_5\})), (\{x_2, x_7\}, (\{l_1\}, \{l_2, l_3, l_4\})), \\ (\{x_1\}, (\{l_1, l_2, l_4\}, \{l_3, l_5\})), \\ (\{x_4\}, (\{l_1, l_4\}, \{l_2, l_3, l_5\})), (\{x_2\}, (\{l_1\}, \{l_2, l_3, l_4, l_5\})), \\ (\{x_7\}, (\{l_1, l_5\}, \{l_2, l_3, l_4\})) \end{array} \right\}.$$

For $(\{x_1, x_2, x_4, x_7\}, (\{l_1\}, \{l_3\}))$, l_1 is the co-occurrence label of $\{x_1, x_2, x_4, x_7\}$, and l_3 is a label that $\{x_1, x_2, x_4, x_7\}$ does not have in common. It shows that l_1 (Mammal) and l_3 (Bird) may be mutually exclusive. In fact, l_1 and l_3 do exist mutually exclusive from the actual semantics. In addition, for $(\{x_1, x_2, x_4\}, (\{l_1\}, \{l_3, l_5\}))$, negative labels $\{l_3, l_5\}$ may exhibit a correlation, because bird is typically migratory animal. Therefore, 3WL-concepts contain rich correlation information, including both positive and negative correlations.

Through the extent of 3WL-concepts, we can get a multi-level relationship as shown in Fig. 1. We can observe that as the hierarchy level increases, the number of included samples gradually increases, and the information covered becomes more extensive. For example, at the top level of the hierarchy, $(\{x_1, x_2, x_4, x_7\}, (\{l_1\}, \{l_3\}))$ has the label l_1 as the common label of its samples. l_1 (Mammal) is a highly generalized label that can typically cover more animal samples. As the hierarchy extends downward, the number of samples decreases, and the information becomes more specific. For example, at the bottom level, $(\{x_1\}, (\{l_1, l_2, l_4\}, \{l_3, l_5\}))$ represents individual lion, which is a more specific subclass of the mammal, which reflects the characteristics of local information. Therefore, constructing multi-level correlation information helps us better understand labels.

3.2. The overall cognition of the label

The human cognitive system tends to establish a knowledge system gradually, progressing from global to local. Based on this, high-level information, serving as the core of global guidance, should be prioritized to ensure that the model or system follows the correct direction. To simulate the process of human recognition and learning labels, we assign weights to 3WL-concepts based on the hierarchical structure and follow a progressive pattern of decreasing weights layer by layer. This approach helps balance the conflict between global objectives and specific details, preventing the system from overly focusing on local details and losing the generality of the objective. Additionally, in real-world data, samples with similar features often share similar labels. Therefore, we maintain a similar topological structure between 3WF-concepts and 3WL-concepts through the extent of 3WL-concepts, thereby modeling the relationship between features and labels more effectively.

Definition 7. Let $\langle U, A \cup L, \tilde{R}_A, R_L \rangle$ be a multi-label context. For any $(X_k, (L_k^+, L_k^-)) \in \mathcal{LCS}_j$, $U_j \subseteq U$ is the set of all samples with label l_j.

The measures of extent X_k, positive intent L_k^+ and negative intent L_k^- are defined as follows:

$$H_{ex}\left(X_k\right) = -\frac{1}{\ln\left|U_j\right|}\ln\frac{\left|X_k\right|}{\left|U_j\right|}, \tag{1}$$

$$H_{in}^{+}\left(L_k^+\right) = -\frac{1}{\ln|L|}\ln\frac{|L| - \left|L_k^+\right| + 1}{|L|}, \tag{2}$$

$$H_{in}^{-}\left(L_k^-\right) = -\frac{1}{\ln|L|}\ln\frac{|L| - \left|L_k^-\right|}{|L|}, \tag{3}$$

where $|\cdot|$ denotes the cardinality.

Property 2. *Let $H_{ex}\left(X_k\right)$, $H_{in}^{+}\left(L_k^+\right)$ and $H_{in}^{-}\left(L_k^-\right)$ be the measures of extent X_k, positive intent L_k^+ and negative intent L_k^-, respectively. Then*

(1) $0 \leqslant H_{ex}\left(X_k\right), H_{in}^{+}\left(L_k^+\right), H_{in}^{-}\left(L_k^-\right) \leqslant 1$.
(2) $H_{ex}\left(X_1\right) \geqslant H_{ex}\left(X_2\right)$, if $X_1 \subseteq X_2$.
(3) $H_{in}^{+}\left(L_1^+\right) \leqslant H_{in}^{+}\left(L_2^+\right)$, if $L_1^+ \subseteq L_2^+$.
(4) $H_{in}^{-}\left(L_1^-\right) \leqslant H_{in}^{-}\left(L_2^-\right)$, if $L_1^- \subseteq L_2^-$.

The first item shows that the value range of $H_{ex}\left(X_k\right)$, $H_{in}^{+}\left(L_k^+\right)$, $H_{in}^{-}\left(L_k^-\right)$ is $[0,1]$. The second to fourth terms indicate that $H_{ex}\left(X_k\right)$ is monotonically decreasing, while $H_{in}^{+}\left(L_k^+\right)$ and $H_{in}^{-}\left(L_k^-\right)$ are monotonically increasing.

Definition 8. Let $\langle U, A \cup L, \tilde{R}_A, R_L\rangle$ be a multi-label context. For any $\left(X_k, \left(L_k^+, L_k^-\right)\right) \in \mathcal{LCS}_j$, H_k is the measure of $\left(X_k, \left(L_k^+, L_k^-\right)\right)$. The weight w_k of $\left(X_k, \left(L_k^+, L_k^-\right)\right)$ is defined as follows :

$$w_k = \frac{1 - H_k}{\sum_{k=1}^{s}\left(1 - H_k\right)}, \tag{4}$$

where $H_k = \frac{H_{ex}\left(X_k\right) + H_{in}^{+}\left(L_k^+\right) + H_{in}^{-}\left(L_k^-\right)}{3}$, $s = \left|\mathcal{LCS}_j\right|$ is the number of 3WL-concepts in the 3WL-concept space for the label l_j. Obviously, $\sum_{k=1}^{s} w_k = 1$.

Property 3. *For any $\left(X_1, \left(L_1^+, L_1^-\right)\right), \left(X_2, \left(L_2^+, L_2^-\right)\right) \in \mathcal{LCS}_j$, if $\left(X_1, \left(L_1^+, L_1^-\right)\right) \leqslant \left(X_2, \left(L_2^+, L_2^-\right)\right)$, then $w_1 \leqslant w_2$.*

Proof. The proof is immediate from Definitions 5, 8 and Property 2.

This property indicates that the big concept will gain greater importance. On the one hand, the big concept has a greater extent, that is, stronger generalization ability. On the other hand, the big concept has a smaller intent, so it is easier to focus attention on the target label.

Definition 9. Let $\langle U, A \cup L, \tilde{R}_A, R_L\rangle$ be a multi-label context. For any $\left(X_k, \left(L_k^+, L_k^-\right)\right) \in \mathcal{LCS}_j$, w_k is its weight. Then $\left(X_k, \left(L_k^+, L_k^-\right), w_k\right)$ is called a weighted 3WL-concept and the weighted 3WL-concept space of label l_j is defined by:

$$\mathcal{WLCS}_j = \left\{\left(X_k, \left(L_k^+, L_k^-\right), w_k\right) | X_k \in \mathcal{LCS}_j\right\}.$$

$\mathcal{WLCS}_j$ denotes the set of all weighted 3WL-concepts under the label l_j. $\mathcal{WLCS} = \left\{\mathcal{WLCS}_1, \mathcal{WLCS}_2, \ldots, \mathcal{WLCS}_q\right\}$ is a set of the weighted 3WL-concept spaces under all labels. On this basis, Algorithm 1 gives the construction process of weighted 3WL-concept space and its time complexity is $O\left(q\left(\left|U_j\right|^2 + 2q^2 + 2k\right)\right)$.

In the above process, we actually determined the degree of importance of different 3WL-concepts when learning label. Next, we use the extent of the weighted 3WL-concepts as clue to obtain the weighted fuzzy three-way feature concepts and fuse these concepts to form an overall cognition of the target label.

Algorithm 1: Constructing the weighted 3WL-concept space.

Input: A multi-label context $\langle U, A \cup L, \tilde{R}_A, R_L\rangle$.
Output: The weighted 3WL-concept space $\mathcal{WLCS} = \left\{\mathcal{WLCS}_1, \mathcal{WLCS}_2, \cdots, \mathcal{WLCS}_q\right\}$.

$\mathcal{WLCS} = \emptyset$;
for $j = 1 : q$ **do**
  $\mathcal{LCS}_j = \emptyset$;
  for *each* $x \in U_j$ *and* $l \in L$ **do**
    Compute 3WL-concept $(\mathcal{HL}(x) \cap \mathcal{H}^{-}\mathcal{L}^{-}(x), (\mathcal{L}(x), \mathcal{L}^{-}(x)))$ and $(\mathcal{H}(l) \cap \mathcal{H}^{-}(l^{\neg}), (\mathcal{LH}(l), \mathcal{L}^{-}\mathcal{H}^{-}(l^{\neg})))$ by Definitions 5 and 6;
    $\mathcal{LCS}_j \leftarrow (\mathcal{HL}(x) \cap \mathcal{H}^{-}\mathcal{L}^{-}(x), (\mathcal{L}(x), \mathcal{L}^{-}(x)))$ and $(\mathcal{H}(l) \cap \mathcal{H}^{-}(l^{\neg}), (\mathcal{LH}(l), \mathcal{L}^{-}\mathcal{H}^{-}(l^{\neg})))$;
  end
  for $\left(X_k, \left(L_k^+, L_k^-\right)\right) \in \mathcal{LCS}_j$ **do**
    Calculate $H_{ex}\left(X_k\right)$ by Eq. (1);
    Calculate $H_{in}^{+}\left(L_k^+\right)$ by Eq. (2);
    Calculate $H_{in}^{-}\left(L_k^-\right)$ by Eq. (3);
  end
  for $k = 1 : \left|\mathcal{LCS}_j\right|$ **do**
    Calculate w_k by Definition 8;
    $\mathcal{WLCS}_j \leftarrow \left(X_k, \left(L_k^+, L_k^-\right), w_k\right)$;
  end
  $\mathcal{WLCS} \leftarrow \mathcal{WLCS}_j$;
end
return $\mathcal{WLCS} = \left\{\mathcal{WLCS}_1, \mathcal{WLCS}_2, \cdots, \mathcal{WLCS}_q\right\}$.

Definition 10. Let $\langle U, A \cup L, \tilde{R}_A, R_L\rangle$ be a multi-label context. For any $\left(X_k, \left(L_k^+, L_k^-\right), w_k\right) \in \mathcal{WLCS}_j$, the weighted fuzzy three-way feature concept space of the label l_j is defined by:

$$\mathcal{WFCS}_j = \left\{\left(\mathcal{H}^{\triangledown}\tilde{\mathcal{L}}^{\triangledown}\left(X_k\right), \tilde{\mathcal{L}}^{\triangledown}\left(X_k\right), w_k\right) | X_k \in \mathcal{WLCS}_j\right\},$$

where $\left(\mathcal{H}^{\triangledown}\tilde{\mathcal{L}}^{\triangledown}\left(X_k\right), \tilde{\mathcal{L}}^{\triangledown}\left(X_k\right), w_k\right)$ is called a weighted fuzzy three-way feature concept (3WF-concept), for short $\left(X_k', \left(\tilde{B}_k'^{+}, \tilde{B}_k'^{-}\right), w_k\right)$.

$\mathcal{WFCS}_j$ denotes the set of all weighted fuzzy 3WF-concepts for the label l_j. The set of weighted fuzzy 3WF-concept spaces of all labels is denoted by $\mathcal{WFCS} = \left\{\mathcal{WFCS}_1, \mathcal{WFCS}_2, \ldots \mathcal{WFCS}_q\right\}$.

Definition 11. Let $\langle U, A \cup L, \tilde{R}_A, R_L\rangle$ be a multi-label context, for all $\left(X_k', \left(\tilde{B}_k'^{+}, \tilde{B}_k'^{-}\right), w_k\right) \in \mathcal{WFCS}_j$, then the pseudo-concept of the label l_j is given by:

$$X_j = X_1' \cup X_2' \cup \cdots \cup X_s';$$
$$\tilde{B}_j^{+} = w_1\tilde{B}_1'^{+} + w_2\tilde{B}_2'^{+} + \cdots + w_s\tilde{B}_s'^{+};$$
$$\tilde{B}_j^{-} = w_1\tilde{B}_1'^{-} + w_2\tilde{B}_2'^{-} + \cdots + w_s\tilde{B}_s'^{-}.$$

After learning all the concepts related to the label, an overall cognition of the label l_j is formed, that is $\left(X_j, \left(\tilde{B}_j^{+}, \tilde{B}_j^{-}\right)\right)$, denoted by pc_j. The set of pseudo-concepts formed by all labels is recorded as $\mathcal{PC} = \left\{pc_1, pc_2, \ldots, pc_q\right\}$.

3.3. The label correlation and multi-label learning model

In the previous subsection, we formed an overall cognition of all labels. This subsection further explores and leverages explicit label correlations. Most existing methods only consider a single perspective, either sample similarity or feature similarity (e.g., cosine similarity), which makes it difficult to fully capture the semantic relationships between labels. However, in multi-label data, label correlations are often jointly determined by both sample-level and feature-level similarities. For example, in music classification, “Music A” and “Music B” may both be labeled as “light music” and “relaxation”, reflecting high label-level consistency and indicating strong sample similarity. Meanwhile, this type of music often has similar audio features such as “low frequency” and “smooth rhythm”, which further promotes the co-occurrence of related labels. Accordingly, we consider both sample-level and feature-level similarities to model label correlations more comprehensively, thereby enhancing the model’s expressive power and generalization performance.

Definition 12. Let $\langle U, A \cup L, \tilde{R}_A, R_L \rangle$ be a multi-label context, $pc_i = \left(X_i, \left(\tilde{B}_i^+, \tilde{B}_i^-\right)\right)$ and $pc_j = \left(X_j, \left(\tilde{B}_j^+, \tilde{B}_j^-\right)\right)$ are the overall cognitions of labels l_i and l_j respectively. The extent relevance and intent relevance between label l_i and l_j are defined as follows:

$$r_{ex}\left(pc_i, pc_j\right) = \frac{\left|X_i \cap X_j\right|}{\left|X_i \cup X_j\right|}, \tag{5}$$

$$r_{in}\left(pc_i, pc_j\right) = \frac{\sum_{k=1}^{m}\left(a_{i,k}^+ \cdot a_{j,k}^+ + a_{i,k}^- \cdot a_{j,k}^-\right)}{\sqrt{\sum_{k=1}^{m}\left({a_{i,k}^+}^2 + {a_{i,k}^-}^2\right)} \cdot \sqrt{\sum_{k=1}^{m}\left({a_{j,k}^+}^2 + {a_{j,k}^-}^2\right)}}, \tag{6}$$

where $a_{i,k}^+ = \tilde{B}_i^+\left(a_k\right)$, $a_{i,k}^- = \tilde{B}_i^-\left(a_k\right)$.

In fact, the extent relevance shows the label correlation from a sample-based perspective, while the intent relevance shows the label correlation from a feature-based perspective. These two aspects are important factors affecting the label correlation. Therefore, combining these two aspects to define label correlation is helpful to enhance the generalization performance.

The label correlation between l_i and l_j is given by:

$$r_{ij} = \gamma \cdot r_{ex}\left(pc_i, pc_j\right) + (1-\gamma) \cdot r_{in}\left(pc_i, pc_j\right). \tag{7}$$

Then, the label correlation matrix can be expressed as:

$$\mathbf{R} = \left(r_{ij}\right)_{q\times q} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1q} \\ r_{21} & r_{22} & \cdots & r_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ r_{q1} & r_{q2} & \cdots & r_{qq} \end{pmatrix}.$$

By considering the label correlation matrix, we can determine the similarity between the prediction coefficients of different label variables in a multi-label model. The MCF-3WCCL algorithm is designed as follows:

$$\min_{\mathbf{W}} \frac{1}{2}\|\mathbf{XW} - \mathbf{Y}\|_F^2 + \frac{\alpha}{2}\|\mathbf{W}^T\mathbf{W} - \mathbf{R}\|_F^2 + \beta\|\mathbf{W}\|_1 \tag{8}$$

where $\alpha, \beta > 0$ are trade-off parameters, and $\mathbf{W} = \left[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q\right] \in \mathbb{R}^{m\times q}$ is the coefficient matrix in the multi-label classification model. For any $l_j \in L$, the sub-classifier is $\mathbf{c}_j = \mathbf{X}\mathbf{w}_j$.

The first term minimizes the gap between the label set predicted by the model and the real label set. By reducing errors, the model can better fit the data and make more accurate predictions. The second term compresses the hypothesis space by introducing label correlation. By considering the label correlation, the objective function optimizes the label prediction by guiding the model to adjust the position of the hyperplane. The last item enhances the sparsity of the model by l_1-norm regularization, simplifies the model and avoids overfitting. In the optimization solution, the accelerated proximal gradient algorithm [12,47] is employed to achieve a fast convergence rate. Building on the above theory, Algorithm 2 is a multi-level correlation information fusion via three-way concept-cognitive learning for multi-label learning.

Next, we will discuss the time complexity of Algorithm 2, the process can be divided into three main parts. (1) The first part is the construction and fusion stage of the weighted fuzzy 3WF-concept (Steps 1–16). Steps 3–7 are used to construct weighted fuzzy 3WF-concepts, where the number of generated concepts does not exceed $n + 2q$, so the time complexity of this part is $O(n + 2q)$. Next, Steps 8–15 fuse up to $n + 2q$ weighted fuzzy 3WF-concepts, where the intent of each concept is an m-dimensional vector. The time complexity of this step is $O(m(n + 2q))$. Since the outer loop iterates over q concept spaces, the total time complexity of Steps 1–16 is $O(qm(n + 2q))$. (2) The second part is the construction of the label correlation matrix (Steps 17–24), which involves calculating the similarities between q pseudo concepts. Since the intent of each pseudo-concept is an m-dimensional vector, the time complexity is $O\left(mq^2\right)$. (3) The third part is the optimization of the objective function and prediction (Steps 25–29), which includes three components: The first term is the reconstruction error, with time complexity $O(nmq)$; The second term is the label correlation constraint, with time complexity $O\left(mq^2\right)$; The third term is the sparse regularization term, with time complexity $O(mq)$. Therefore, the total time complexity of this stage is $O\left(nmq + mq^2 + mq\right)$. In multi-label learning, it is typically assumed that $n \gg m$ and $n \gg q$, so the time complexity of Algorithm 2 can be simplified to $O(nmq)$. Furthermore, considering t iterations of gradient descent, the total time complexity becomes $O(t \cdot nmq)$.

In addition, Fig. 2 presents an overview of the MCF-3WCCL model, which mainly consists of three parts: (1) Data processing stage: The original feature data and label data are processed separately using the three-way cognitive operator, forming the corresponding 3WL-concepts and 3WF-concepts. These concepts provide the semantic-level structural foundation for subsequent modeling. (2) Multi-Level correlation information fusion stage: A hierarchical structure of 3WL-concepts is constructed, and the weights of 3WL-concepts are calculated and gradually decreased layer by layer. Finally, the overall cognition of each label is formed through the information fusion of 3WF-concepts. (3) Prediction stage: The fused overall cognition of the labels is combined with the constructed label correlation matrix to optimize the objective function and generate the final multi-label prediction results.

4. Experimental analysis

In this section, we use real-world multi-label datasets to compare traditional and state-of-the-art multi-label learning algorithms, along with their experimental setups. Through experiments and statistical analysis using five multi-label evaluation metrics, we validate the effectiveness of MCF-3WCCL in multi-label learning. Additionally, the impact of critical parameters on experimental performance is further assessed. All experiments were conducted on a computer with the following configuration: Windows 11 OS, Intel(R) Core(TM) i5-13600KF CPU @ 3.5 GHz, 32 GB RAM, and MATLAB as the programming language.

4.1. Datasets and experiment settings

To assess the effectiveness of MCF-3WCCL in multi-label learning, 16 multi-label datasets were selected from Mulan.[1] Table 3 provides a detailed description of each dataset, where “Sample” denotes the number of samples, “Dim” refers to the feature dimension, “Label” indicates the number of labels, “Cardinality” represents the average number of labels per sample, and “Type” describes the dataset type.

To describe the performance of MCF-3WCCL, the following 10 multi-label algorithms are chosen as comparison methods:

• ML-KNN [14]: A widely-used benchmark algorithm for multi-label classification, which does not consider label correlations or label-specific feature selection. Parameter $k = 10$ and the smoothing parameter s is set to 1.

• MRDM [3]: This is a method based on manifold regularization and dependency maximization, which evaluates the dependency between

[1] https://mulan.sourceforge.net/datasets-mlc.html.

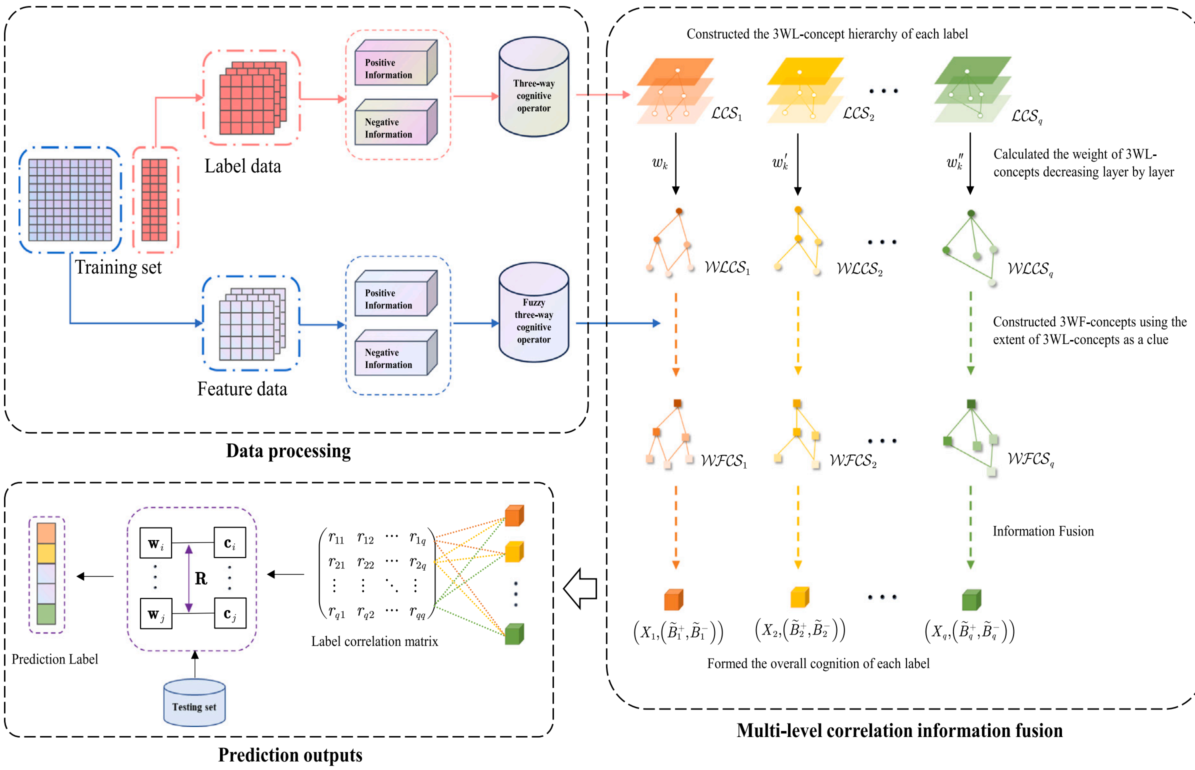

Fig. 2. Overview of MCF-3WCCL model.

manifold space and label space by introducing HSIC-based measurements. The parameters $\alpha = 1$ and $\beta, \gamma \in \{10^{-2}, 10^{-1}, \dots, 10^{2}\}$.

• LPLC [22]: Using local positive and negative pairwise label correlations, an effective multi-label classification Bayesian model is established. The parameters $k = 10$ and grid search for $\alpha \in \{0.1, 0.2, \dots, 1\}$.

• GLFS [17]: The algorithm improves the generalization performance through local label correlations, specially, both label-group and instance-group correlations are used together to enhance model training. The parameters $\alpha, \lambda \in \{0.1, 0.2, \dots, 1\}$ and $\beta, \gamma \in \{10^{-3}, 10^{-2}, \dots, 10^{3}\}$.

• 2SML [18]: The feature manifold and label manifold share a weight matrix, leveraging prior knowledge of correlations, while implicit correlations are learned by selecting highly representative instances. The parameters $\lambda_1 = 10^{-3}, \lambda_2 = 10^{-3}, \lambda_3 = 10^{-4}$ and $\alpha = 0.6, \beta = 1 - \alpha$.

• LLSF-DL [12]: The algorithm with l_1-norm regularization used to learn label-specific features. This algorithm incorporates both second-order and higher-order label correlations. The parameters α, β, γ are searched in $\{4^{-5}, 4^{-4}, \dots, 4^{5}\}$, and ρ is searched in $\{0.1, 1, 10\}$.

• JLCLS [10]: This method enhances the learning process of label-specific features in similar labels by label correlations. The parameters $\alpha, \beta, \theta \in \{2^{-10}, 2^{-9}, \dots, 2^{10}\}$.

• MSWL [4]: This algorithm employs sparse representation to utilize global label correlations in a one-vs-all manner and combines this with local correlations to enhance generalization performance. The parameters $\alpha, \beta \in \{10^{-3}, 10^{-2}, \dots, 10^{3}\}$, $\gamma \in \{10^{-6}, 10^{-5}, \dots, 10^{6}\}$ and $\lambda = 0.1$.

• GLMAM [11]: This algorithm uses a mechanism that uses attention to simultaneously exploit label and instance information to capture both global and local label correlations. The parameters $\lambda_1, \lambda_2, \lambda_3 \in [10^{-6}, 10^{6}]$ and $g \in [2^{0}, 2^{7}]$.

• MDFS [2]: Manifold regularization is used to mine label correlations. The parameters $\alpha = 1$ and $\beta, \lambda \in \{10^{-3}, 10^{-2}, \dots, 10^{3}\}$.

The parameters for each of the above comparison methods are set as the corresponding literature suggested. In MCF-3WCCL algorithm, the parameters $\alpha, \beta \in \{4^{-5}, 4^{-4}, \dots, 4^{5}\}$ and $\gamma \in \{0, 0.1, 0.2, \dots, 1\}$. We uploaded our code on github, specifically at https://github.com/Jimmy4629/MCF-3WCCL.git.

Table 4 summarizes the correlation strategies employed by different methods.

4.2. Experimental results

This experiment employed five-fold cross-validation to evaluate the performance of the MCF-3WCCL and 10 other comparison methods based on five evaluation metrics: Average Precision (↑), Coverage (↓), One-error (↓), Ranking Loss (↓) and Hamming Loss (↓), where "↑" denotes the bigger the better, and "↓" indicates the smaller the better. The data was randomly split into training (80%) and testing (20%) sets, and the experiments were repeated five times to ensure fairness. The final performance is presented as the mean of the classification results across each fold in cross-validation, given as 'mean ± variance'. The last two rows present the average (Ave.) performance across all datasets and the Win/Tie/Loss record indicates the number of datasets where MCF-3WCCL performs better, equally, or worse than the other comparison methods, respectively. Tables 5–9 present the experimental results comparing MCF-3WCCL with other methods. The best results for each dataset are emphasized in bold.

The experimental results presented in Tables 5–9 lead to the following conclusions:

• The performance of MCF-3WCCL on most datasets is better than the other 10 multi-label methods, which significantly reflects its advantages. Whether in Ave. or Win/Tie/Loss statistics, MCF-3WCCL performs best, showing the superiority and effectiveness of its classification performance.

• Combined with Table 4, it can be observed that most methods considering label correlation (e.g., MCF-3WCCL, LLSF-DL, MSWL, GLMAM, MDFS, JLCLS) perform better than methods that do not consider label correlation (e.g., ML-KNN, MRDM). This result indicates that label correlation plays a key role in multi-label learning tasks.

Algorithm 2: Multi-level correlation information fusion via three-way concept-cognitive learning for multi-label learning (MCF-3WCCL).

Input: The weighted 3WF-concept space
$\mathcal{WLCS} = \{\mathcal{WLCS}_1, \mathcal{WLCS}_2, \cdots, \mathcal{WLCS}_q\}$.
Output: The outputs $\hat{\mathbf{Y}}$.

$\mathcal{WLCS} = \{\mathcal{WLCS}_1, \mathcal{WLCS}_2, \cdots, \mathcal{WLCS}_q\}$, $\mathcal{PC} = \emptyset$;
for $j = 1 : q$ **do**
  $\mathcal{WFCS}_j = \emptyset$;
  for $X_k, w_k \in \mathcal{WLCS}_j$ **do**
    Construct weighted fuzzy 3WF-concept $\left(\mathcal{H}^{\nabla}\tilde{\mathcal{L}}^{\nabla}(X_k), \tilde{\mathcal{L}}^{\nabla}(X_k), w_k\right)$ by Definitions 3 and 10;
    $\mathcal{WFCS}_j \leftarrow \left(\mathcal{H}^{\nabla}\tilde{\mathcal{L}}^{\nabla}(X_k), \tilde{\mathcal{L}}^{\nabla}(X_k), w_k\right)$;
  end
  $X_j = \emptyset$; $\tilde{B}_j^+ = \emptyset$; $\tilde{B}_j^- = \emptyset$;
  for $\left(X'_k, \left(\tilde{B}_k'^+, \tilde{B}_k'^-\right), w_k\right) \in \mathcal{WFCS}_j$ **do**
    $X_j = X_j \cup X'_k$;
    $\tilde{B}_j^+ = \tilde{B}_j^+ + w_k \tilde{B}_k'^+$;
    $\tilde{B}_j^- = \tilde{B}_j^- + w_k \tilde{B}_k'^-$;
  end
  $pc_j \leftarrow \left(X_j, \left(\tilde{B}_j^+, \tilde{B}_j^-\right)\right)$;
  $\mathcal{PC} \leftarrow pc_j$;
end
for $i = 1 : q$ **do**
  for $j = 1 : q$ **do**
    Calculate $r_{ex}(pc_i, pc_j)$ by Eq. (5);
    Calculate $r_{in}(pc_i, pc_j)$ by Eq. (6);
    Calculate r_{ij} by Eq. (7);
  end
end
Acquire the label correlation matrix $\mathbf{R} = (r_{ij})_{q\times q}$;
Construct the multi-label learning model in Eq. (8);
Get the optimal solution $\mathbf{W}^*$;
Predict the outputs $\hat{\mathbf{Y}} = \mathbf{X}_{test}\mathbf{W}^*$;
return $\hat{\mathbf{Y}}$.

Table 3
Description of the multi-label datasets.

ID	Dataset	Sample	Dim	Label	Cardinality	Type
1	Flags	194	19	7	3.392	Image
2	Cal500	502	68	174	26.044	Music
3	CHD49	555	49	6	2.580	Medicine
4	Emotion	593	72	6	1.869	Music
5	Birds	645	260	19	1.014	Audio
6	Genbase	662	1186	27	1.252	Biology
7	Medical	978	1449	45	1.245	Text
8	Enron	1702	1001	53	3.378	Text
9	Image	2000	294	5	1.236	Image
10	Scene	2407	294	6	1.074	Image
11	Social	5000	1047	39	1.283	Text
12	Computers	5000	681	33	1.508	Text
13	Health	5000	612	32	1.663	Text
14	Business	5000	438	30	1.588	Text
15	20NG	19 300	1006	20	1.029	Text
16	Tmc2007	28 596	500	22	2.22	Text

• Among all the methods considering label correlation, the methods using global correlation (e.g., MCF-3WCCL, LLSF-DL, JLCLS, MSWL, GLMAM, MDFS) are generally superior to the methods using only local correlation (e.g., LPLC, GLFS, 2SML). The possible reason is that global correlation can provide more holistic information and help to capture a wider connection between labels. In contrast, methods that only focus on local correlation may overly rely on specific features or samples, which can lead to overfitting and adversely affect the generalization ability of the model.

• Although MSWL, GLMAM and MDFS consider both local and global correlations in design, they do not explore the multi-level relationship between local and global correlations. This limitation may lead to the lack of effective strategies in balancing local and global correlations, preventing them from fully leveraging the complementary advantages of both. As a result, these methods still underperform compared to MCF-3WCCL.

In view of the above results, it is necessary to balance the global correlation and local correlation by exploring and combining multi-level correlation information. Global correlation provides holistic guidance, while local correlation can mine micro-level of specific features or samples. Through the effective fusion of multi-level correlation information, the performance and adaptability of the model can be improved while avoiding over-fitting, which provides a direction for further research on multi-label learning. In summary, MCF-3WCCL has achieved significant success in multi-label learning tasks by virtue of its fusion of multi-level correlation information.

4.3. Statistical analysis

To analyze the relative performance of the 11 multi-label methods in greater depth, we will conduct the Friedman test [48] and the Bonferroni-Dunn test [49] to investigate the statistical significance of the experimental results.

The Friedman test [48] is a commonly used statistical test for repeated measures designs. For any evaluation metric, the average ranking of the ith algorithm can be calculated as $R_i = \frac{1}{N}\sum_{j=1}^{N} r_{ij}$, where N is the number of multi-label datasets, and r_{ij} denotes the ranking of the ith algorithm on the jth dataset. The Friedman statistic F_F follows the F-distribution with numerator degrees of freedom $(k-1)$ and denominator degrees of freedom $(k-1) * (N-1)$, and can be calculated as follows:

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

where $\chi_F^2 = \frac{12N}{k(k+1)}\left(\sum_{i=1}^{k} R_i^2 - \frac{k(k+1)^2}{4}\right)$, k is the number of algorithms.

Table 10 shows the Friedman statistics for different metrics of MCF-3WCCL, with $k = 11$, $N = 16$, and the significance level $\theta = 0.05$, the Friedman statistics of the five evaluation metrics are greater than the critical value 1.8943. As a result, the null hypothesis that the methods perform equally is rejected, indicating significant differences in the performance of the algorithms.

To further validate these findings, we employ a post hoc analysis using the Bonferroni-Dunn test [49] to clearly demonstrate the performance gap between MCF-3WCCL and the other comparison algorithms. In general, if the average ranks differ by at least the critical difference ($CD = q_\theta\sqrt{\frac{k(k+1)}{6N}}$), the performance of the two algorithms is significantly different. For the Bonferroni-Dunn test, when $k = 11$, $\theta = 0.05$, then $q_\theta = 2.807$. Therefore, with $N = 16$, we can calculate the critical difference $CD = 3.2915$. If the average ranks of our algorithm and a comparison algorithm are within one CD, their relative performance is considered statistically similar. To visually represent this, Fig. 3 presents the CD plot for each evaluation metric. In each sub-figure, algorithms falling outside the red line are deemed to have significantly different performance compared to MCF-3WCCL. From Fig. 3, it is evident that MCF-3WCCL consistently achieves the highest rankings across all evaluation metrics. Overall, the performance of MCF-3WCCL is significantly superior to most other algorithms.

Table 4
Correlation strategies of different methods.

Method	No correlation	Local correlation	Global correlation	Multi-level correlation
ML-KNN	✓			
MRDM	✓			
LPLC		✓		
GLFS		✓		
2SML		✓		
LLSF-DL			✓	
JLCLS			✓	
MSWL		✓	✓	
GLMAM		✓	✓	
MDFS		✓	✓	
MCF-3WCCL		✓	✓	✓

Table 5
Comparison results of MCF-3WCCL with other comparison methods on Average Precision (↑) metric.

Dataset	ML-KNN	MRDM	LPLC	GLFS	2SML	LLSF-DL	JLCLS	MSWL	GLMAM	MDFS	MCF-3WCCL
Flags	0.8033$_{\pm0.0251}$	0.8058$_{\pm0.0200}$	0.8032$_{\pm0.0290}$	0.7840$_{\pm0.0076}$	0.6840$_{\pm0.0339}$	0.7883$_{\pm0.0338}$	0.8083$_{\pm0.0249}$	0.7998$_{\pm0.0002}$	0.6569$_{\pm0.0339}$	0.8020$_{\pm0.0127}$	**0.8148**$_{\pm0.0349}$
Cal500	0.4937$_{\pm0.0088}$	0.4899$_{\pm0.0125}$	0.4967$_{\pm0.0095}$	0.4840$_{\pm0.0036}$	0.4135$_{\pm0.0088}$	0.4993$_{\pm0.0025}$	0.4991$_{\pm0.0069}$	0.4868$_{\pm0.0000}$	0.3399$_{\pm0.0062}$	0.4959$_{\pm0.0021}$	**0.5054**$_{\pm0.0081}$
CHD49	0.7715$_{\pm0.0159}$	0.7938$_{\pm0.0127}$	0.7848$_{\pm0.0195}$	0.7733$_{\pm0.0095}$	0.7932$_{\pm0.0155}$	0.7723$_{\pm0.0228}$	0.7919$_{\pm0.0301}$	0.7964$_{\pm0.0002}$	0.7848$_{\pm0.0243}$	0.7655$_{\pm0.0074}$	**0.8067**$_{\pm0.0272}$
Emotion	0.7889$_{\pm0.0187}$	0.8011$_{\pm0.0046}$	0.8005$_{\pm0.0281}$	0.7705$_{\pm0.0318}$	0.7961$_{\pm0.0225}$	0.7969$_{\pm0.0278}$	0.7654$_{\pm0.0184}$	0.7863$_{\pm0.0039}$	0.7880$_{\pm0.0311}$	0.7998$_{\pm0.0259}$	**0.8109**$_{\pm0.0187}$
Birds	0.7168$_{\pm0.0261}$	0.7210$_{\pm0.0187}$	0.6902$_{\pm0.0168}$	0.7011$_{\pm0.0378}$	0.7686$_{\pm0.0132}$	0.6992$_{\pm0.0266}$	0.7070$_{\pm0.0466}$	0.7640$_{\pm0.0000}$	0.7539$_{\pm0.0352}$	0.7074$_{\pm0.0181}$	**0.7763**$_{\pm0.0286}$
Genbase	0.3189$_{\pm0.0330}$	0.9868$_{\pm0.0043}$	0.9899$_{\pm0.0013}$	0.9501$_{\pm0.1025}$	0.6230$_{\pm0.4563}$	0.9964$_{\pm0.0023}$	0.9943$_{\pm0.0044}$	0.9793$_{\pm0.0357}$	0.9819$_{\pm0.0128}$	0.9844$_{\pm0.0046}$	**0.9974**$_{\pm0.0011}$
Medical	0.8024$_{\pm0.0100}$	0.8618$_{\pm0.0104}$	0.7658$_{\pm0.0099}$	0.7849$_{\pm0.0977}$	0.4275$_{\pm0.3733}$	0.8662$_{\pm0.0244}$	0.8466$_{\pm0.0213}$	0.7028$_{\pm0.0130}$	0.9076$_{\pm0.0113}$	0.7881$_{\pm0.0185}$	**0.9077**$_{\pm0.0099}$
Enron	0.6239$_{\pm0.0082}$	0.6341$_{\pm0.0078}$	0.5209$_{\pm0.0141}$	0.6506$_{\pm0.0303}$	0.6899$_{\pm0.0146}$	0.6846$_{\pm0.0114}$	0.6961$_{\pm0.0078}$	0.4484$_{\pm0.0683}$	0.5774$_{\pm0.0171}$	0.6253$_{\pm0.0202}$	**0.7054**$_{\pm0.0106}$
Image	0.7867$_{\pm0.0112}$	0.7512$_{\pm0.0080}$	0.7884$_{\pm0.0103}$	0.7435$_{\pm0.0364}$	0.7467$_{\pm0.0135}$	0.7841$_{\pm0.0186}$	0.7259$_{\pm0.0064}$	0.7801$_{\pm0.0004}$	0.7735$_{\pm0.0245}$	0.7816$_{\pm0.0179}$	**0.7917**$_{\pm0.0073}$
Scene	**0.8579**$_{\pm0.0130}$	0.7544$_{\pm0.0276}$	0.8383$_{\pm0.0097}$	0.7771$_{\pm0.0770}$	0.8394$_{\pm0.0103}$	0.7793$_{\pm0.0119}$	0.8432$_{\pm0.0119}$	0.8281$_{\pm0.0002}$	0.8134$_{\pm0.0159}$	0.8486$_{\pm0.0212}$	0.8521$_{\pm0.0077}$
Social	0.7338$_{\pm0.0085}$	0.7225$_{\pm0.0145}$	0.7360$_{\pm0.0101}$	0.7329$_{\pm0.0124}$	0.7781$_{\pm0.0074}$	0.7424$_{\pm0.0119}$	0.7765$_{\pm0.0121}$	0.6751$_{\pm0.0330}$	0.7713$_{\pm0.0110}$	0.7399$_{\pm0.0102}$	**0.7782**$_{\pm0.0091}$
Computers	0.6388$_{\pm0.0112}$	0.6095$_{\pm0.0045}$	0.5780$_{\pm0.0058}$	0.6345$_{\pm0.0092}$	0.7077$_{\pm0.0123}$	0.6311$_{\pm0.0036}$	0.7047$_{\pm0.0049}$	0.7008$_{\pm0.0000}$	0.6963$_{\pm0.0150}$	0.6617$_{\pm0.0120}$	**0.7086**$_{\pm0.0084}$
Health	0.6711$_{\pm0.0096}$	0.7285$_{\pm0.0147}$	0.4810$_{\pm0.0101}$	0.7266$_{\pm0.0215}$	**0.7899**$_{\pm0.0056}$	0.6533$_{\pm0.0113}$	0.7849$_{\pm0.0092}$	0.7797$_{\pm0.0000}$	0.7788$_{\pm0.0090}$	0.7196$_{\pm0.0129}$	0.7850$_{\pm0.0060}$
Business	0.8790$_{\pm0.0099}$	0.8647$_{\pm0.0080}$	0.8708$_{\pm0.0016}$	0.8788$_{\pm0.0035}$	0.8871$_{\pm0.0108}$	0.8774$_{\pm0.0041}$	0.8805$_{\pm0.0041}$	0.8807$_{\pm0.0000}$	0.8769$_{\pm0.0076}$	0.8793$_{\pm0.0045}$	**0.8872**$_{\pm0.0069}$
20NG	0.6030$_{\pm0.0111}$	0.5722$_{\pm0.0035}$	0.4581$_{\pm0.0101}$	0.4619$_{\pm0.1195}$	0.8170$_{\pm0.0031}$	0.7319$_{\pm0.0022}$	0.8346$_{\pm0.0045}$	0.8201$_{\pm0.0000}$	0.8208$_{\pm0.0023}$	0.6205$_{\pm0.0257}$	**0.8352**$_{\pm0.0031}$
Tmc2007	0.7926$_{\pm0.0026}$	–	0.7538$_{\pm0.0033}$	0.6766$_{\pm0.0436}$	0.8198$_{\pm0.0016}$	0.7465$_{\pm0.0039}$	0.8335$_{\pm0.0032}$	0.8253$_{\pm0.0000}$	0.8343$_{\pm0.0023}$	0.7920$_{\pm0.0124}$	**0.8394**$_{\pm0.0012}$
Ave.	0.7051	0.7398	0.7098	0.7207	0.7238	0.7531	0.7804	0.7534	0.7597	0.7507	**0.8001**
Win/Tie/Loss	15/0/1	15/0/0	16/0/0	16/0/0	15/0/1	16/0/0	16/0/0	16/0/0	16/0/0	16/0/0	

Table 6
Comparison results of MCF-3WCCL with other comparison methods on Coverage (↓) metric.

Dataset	ML-KNN	MRDM	LPLC	GLFS	2SML	LLSF-DL	JLCLS	MSWL	GLMAM	MDFS	MCF-3WCCL
Flags	0.5491$_{\pm0.0228}$	0.5472$_{\pm0.0151}$	0.5604$_{\pm0.0420}$	0.5492$_{\pm0.0131}$	0.6661$_{\pm0.0350}$	0.5530$_{\pm0.0440}$	0.5494$_{\pm0.0155}$	0.5483$_{\pm0.0003}$	0.7046$_{\pm0.0588}$	0.5718$_{\pm0.0141}$	**0.5460**$_{\pm0.0368}$
Cal500	0.7513$_{\pm0.0204}$	0.7504$_{\pm0.0184}$	0.8174$_{\pm0.0043}$	0.7642$_{\pm0.0043}$	0.9186$_{\pm0.0069}$	0.7462$_{\pm0.0139}$	0.7460$_{\pm0.0089}$	0.7985$_{\pm0.0000}$	0.9071$_{\pm0.0042}$	0.7499$_{\pm0.0021}$	**0.7457**$_{\pm0.0221}$
CHD49	0.4817$_{\pm0.0201}$	0.4522$_{\pm0.0165}$	0.4571$_{\pm0.0075}$	0.4638$_{\pm0.0164}$	0.4605$_{\pm0.0053}$	0.4856$_{\pm0.0093}$	0.4538$_{\pm0.0254}$	0.4541$_{\pm0.0002}$	0.4540$_{\pm0.0184}$	0.4844$_{\pm0.0076}$	**0.4345**$_{\pm0.0281}$
Emotion	0.3067$_{\pm0.0205}$	0.2987$_{\pm0.0177}$	0.3249$_{\pm0.0178}$	0.3130$_{\pm0.0223}$	0.3100$_{\pm0.0126}$	0.3010$_{\pm0.0170}$	0.3220$_{\pm0.0150}$	0.3110$_{\pm0.0033}$	0.3047$_{\pm0.0112}$	0.3255$_{\pm0.0169}$	**0.2971**$_{\pm0.0178}$
Birds	0.1528$_{\pm0.0199}$	0.1571$_{\pm0.0189}$	0.1820$_{\pm0.0137}$	0.1731$_{\pm0.0289}$	0.1520$_{\pm0.0256}$	0.1653$_{\pm0.0144}$	0.1677$_{\pm0.0162}$	0.1523$_{\pm0.0001}$	0.1598$_{\pm0.0302}$	0.1498$_{\pm0.0105}$	**0.1420**$_{\pm0.0150}$
Genbase	0.2115$_{\pm0.0195}$	0.0201$_{\pm0.0063}$	0.0186$_{\pm0.0019}$	0.0406$_{\pm0.0284}$	0.1024$_{\pm0.1100}$	0.0138$_{\pm0.0023}$	0.0129$_{\pm0.0052}$	0.0192$_{\pm0.0152}$	0.0313$_{\pm0.0129}$	0.0105$_{\pm0.0016}$	**0.0104**$_{\pm0.0021}$
Medical	0.0610$_{\pm0.0067}$	0.0498$_{\pm0.0076}$	0.0811$_{\pm0.0033}$	0.0779$_{\pm0.0199}$	0.3258$_{\pm0.2404}$	0.0571$_{\pm0.0084}$	0.0350$_{\pm0.0134}$	0.1335$_{\pm0.0053}$	**0.0320**$_{\pm0.0085}$	0.0632$_{\pm0.0073}$	0.0380$_{\pm0.0075}$
Enron	0.2578$_{\pm0.0147}$	0.2427$_{\pm0.0053}$	0.4029$_{\pm0.0251}$	0.2676$_{\pm0.0097}$	0.2801$_{\pm0.0103}$	0.2317$_{\pm0.0113}$	0.2609$_{\pm0.0160}$	0.5126$_{\pm0.0637}$	0.4049$_{\pm0.0171}$	**0.2566**$_{\pm0.0060}$	0.2572$_{\pm0.0129}$
Image	0.1971$_{\pm0.0129}$	0.2182$_{\pm0.0084}$	0.2078$_{\pm0.0121}$	0.2392$_{\pm0.0254}$	0.2172$_{\pm0.0105}$	0.1942$_{\pm0.0132}$	0.2288$_{\pm0.0087}$	0.2023$_{\pm0.0003}$	0.2083$_{\pm0.0230}$	0.2063$_{\pm0.0122}$	**0.1938**$_{\pm0.0039}$
Scene	0.0885$_{\pm0.0032}$	0.1451$_{\pm0.0180}$	0.0893$_{\pm0.0047}$	0.1342$_{\pm0.0510}$	0.0909$_{\pm0.0057}$	0.1420$_{\pm0.0070}$	0.0884$_{\pm0.0047}$	0.1013$_{\pm0.0001}$	0.1104$_{\pm0.0082}$	0.0883$_{\pm0.0136}$	**0.0868**$_{\pm0.0069}$
Social	0.0830$_{\pm0.0036}$	0.0839$_{\pm0.0070}$	0.1183$_{\pm0.0045}$	0.0897$_{\pm0.0042}$	0.0827$_{\pm0.0055}$	0.0841$_{\pm0.0069}$	0.0828$_{\pm0.0073}$	0.1692$_{\pm0.0159}$	0.0997$_{\pm0.0128}$	0.0836$_{\pm0.0046}$	**0.0824**$_{\pm0.0061}$
Computers	0.1286$_{\pm0.0077}$	0.1373$_{\pm0.0055}$	0.2141$_{\pm0.0281}$	0.1331$_{\pm0.0042}$	0.1182$_{\pm0.0079}$	**0.1134**$_{\pm0.0052}$	0.1190$_{\pm0.0059}$	0.1313$_{\pm0.0000}$	0.1359$_{\pm0.0081}$	0.1192$_{\pm0.0035}$	0.1409$_{\pm0.0057}$
Health	0.1072$_{\pm0.0029}$	0.0949$_{\pm0.0066}$	0.1696$_{\pm0.0053}$	**0.0983**$_{\pm0.0070}$	0.1068$_{\pm0.0055}$	0.0991$_{\pm0.0017}$	0.1034$_{\pm0.0041}$	0.1068$_{\pm0.0000}$	0.1209$_{\pm0.0069}$	0.0990$_{\pm0.0031}$	0.1268$_{\pm0.0046}$
Business	0.0955$_{\pm0.0055}$	0.0825$_{\pm0.0034}$	0.1040$_{\pm0.0040}$	0.0780$_{\pm0.0024}$	0.0794$_{\pm0.0074}$	**0.0657**$_{\pm0.0036}$	0.0776$_{\pm0.0058}$	0.0735$_{\pm0.0000}$	0.0846$_{\pm0.0059}$	0.0781$_{\pm0.0028}$	0.0712$_{\pm0.0058}$
20NG	0.1709$_{\pm0.0058}$	0.2159$_{\pm0.0046}$	0.2628$_{\pm0.0073}$	0.2804$_{\pm0.0731}$	0.0550$_{\pm0.0017}$	0.1218$_{\pm0.0016}$	0.0470$_{\pm0.0007}$	0.0546$_{\pm0.0000}$	0.0554$_{\pm0.0011}$	0.1555$_{\pm0.0141}$	**0.0464**$_{\pm0.0009}$
Tmc2007	0.1470$_{\pm0.0010}$	–	0.1977$_{\pm0.0044}$	0.2334$_{\pm0.0260}$	0.1285$_{\pm0.0013}$	0.2566$_{\pm0.0037}$	0.1239$_{\pm0.0018}$	0.1239$_{\pm0.0000}$	0.1292$_{\pm0.0032}$	0.1503$_{\pm0.0066}$	**0.1238**$_{\pm0.0016}$
Ave.	0.2369	0.2331	0.2630	0.2460	0.2559	0.2269	0.2137	0.2433	0.2464	0.2245	**0.2089**
Win/Tie/Loss	14/0/2	12/0/3	16/0/0	14/0/2	14/0/2	12/0/4	13/0/3	14/0/2	13/0/3	13/0/3	

4.4. Parameter sensitivity analysis

The MCF-3WCCL algorithm's objective function involves three important parameters: α, β and γ. The trade-off parameters α and β are coefficients of the regularization terms, primarily aimed at reducing model complexity. When constructing label correlations, γ can balance extent relevance and intent relevance. A series of experiments were conducted using five evaluation metrics to analyze the impact of α, β, and γ on the performance of MCF-3WCCL. Given the limited space, we selected six datasets from various domains to demonstrate the influence of these parameters on performance. All parameters except the one being analyzed are fixed at their optimal values.

Fig. 4 shows the impact of γ on the performance of MCF-3WCCL. Each row corresponds to a different dataset and shows the influences of γ on the five evaluation metrics. The results show that: For the Genbase dataset, the performance is best when $\gamma = 1$, indicating that Genbase can be completely dependent on the extent relevance. In most datasets, the parameter γ achieves high performance on some intermediate values, indicating that the multi-label classification performance is affected by both the extent relevance and intent relevance. Parameter γ can well integrate the relationship between the two to achieve the best performance. It is worth noting that when $\gamma \in [0.3, 0.5]$, MCF-3WCCL is more likely to achieve optimal performance across most datasets. Overall, This analysis highlights the importance of γ in

Table 7
Comparison results of MCF-3WCCL with other comparison methods on One-error (↓) metric.

Dataset	ML-KNN	MRDM	LPLC	GLFS	2SML	LLSF-DL	JLCLS	MSWL	GLMAM	MDFS	MCF-3WCCL
Flags	$0.2293_{\pm0.0869}$	$0.2293_{\pm0.0388}$	$0.2320_{\pm0.0430}$	$0.2395_{\pm0.0230}$	$0.4111_{\pm0.1016}$	$0.3025_{\pm0.0664}$	$0.2185_{\pm0.0735}$	$0.2341_{\pm0.0000}$	$0.4284_{\pm0.1057}$	$0.2207_{\pm0.0457}$	$\mathbf{0.1987}_{\pm0.0604}$
Cal500	$0.1183_{\pm0.0297}$	$0.1176_{\pm0.0325}$	$0.1360_{\pm0.0233}$	$\mathbf{0.0904}_{\pm0.0028}$	$0.5149_{\pm0.0220}$	$0.1235_{\pm0.0230}$	$0.1184_{\pm0.0207}$	$0.2129_{\pm0.0000}$	$0.3555_{\pm0.0313}$	$0.1002_{\pm0.0012}$	$0.1175_{\pm0.0227}$
CHD49	$\mathbf{0.2342}_{\pm0.0332}$	$0.2354_{\pm0.0205}$	$0.2378_{\pm0.0389}$	$0.2781_{\pm0.0122}$	$0.2590_{\pm0.0423}$	$0.3027_{\pm0.0226}$	$0.2498_{\pm0.0505}$	$0.2348_{\pm0.0010}$	$0.2681_{\pm0.0691}$	$0.2424_{\pm0.0136}$	$0.2378_{\pm0.0468}$
Emotion	$0.2919_{\pm0.0329}$	$0.2648_{\pm0.0038}$	$0.2450_{\pm0.0548}$	$0.3183_{\pm0.0558}$	$0.2632_{\pm0.0410}$	$0.2851_{\pm0.0610}$	$0.3338_{\pm0.0249}$	$0.2979_{\pm0.0102}$	$0.3035_{\pm0.0647}$	$0.2635_{\pm0.0463}$	$\mathbf{0.2445}_{\pm0.0351}$
Birds	$0.3550_{\pm0.0445}$	$0.3457_{\pm0.0305}$	$0.3938_{\pm0.0210}$	$0.3575_{\pm0.0434}$	$0.2778_{\pm0.0223}$	$0.4124_{\pm0.0499}$	$0.3519_{\pm0.0670}$	$0.2744_{\pm0.0000}$	$0.2940_{\pm0.0479}$	$0.3687_{\pm0.0301}$	$\mathbf{0.2713}_{\pm0.0514}$
Genbase	$0.8906_{\pm0.0749}$	$0.0136_{\pm0.0056}$	$0.0014_{\pm0.0013}$	$0.0539_{\pm0.1349}$	$0.0016_{\pm0.0033}$	$\mathbf{0.0000}_{\pm0.0000}$	$0.0030_{\pm0.0060}$	$0.0223_{\pm0.0439}$	$0.0061_{\pm0.0121}$	$0.0284_{\pm0.0087}$	$0.0015_{\pm0.0030}$
Medical	$0.2587_{\pm0.0121}$	$0.1779_{\pm0.0151}$	$0.2870_{\pm0.0111}$	$0.2508_{\pm0.1230}$	$\mathbf{0.0658}_{\pm0.0808}$	$0.1668_{\pm0.0312}$	$0.2035_{\pm0.0184}$	$0.3386_{\pm0.0281}$	$0.1270_{\pm0.0198}$	$0.2756_{\pm0.0214}$	$0.1227_{\pm0.0103}$
Enron	$0.3137_{\pm0.0161}$	$0.3184_{\pm0.0335}$	$0.4929_{\pm0.0136}$	$0.2637_{\pm0.0390}$	$0.2392_{\pm0.0277}$	$0.2262_{\pm0.0191}$	$0.2261_{\pm0.0147}$	$0.5116_{\pm0.0781}$	$0.3349_{\pm0.0213}$	$0.3187_{\pm0.0328}$	$\mathbf{0.2186}_{\pm0.0166}$
Image	$0.3270_{\pm0.0180}$	$0.3900_{\pm0.0157}$	$0.3235_{\pm0.0322}$	$0.3860_{\pm0.0586}$	$0.4005_{\pm0.0220}$	$0.3350_{\pm0.0286}$	$0.4355_{\pm0.0123}$	$0.3383_{\pm0.0008}$	$0.3439_{\pm0.0269}$	$0.3242_{\pm0.0290}$	$\mathbf{0.3185}_{\pm0.0145}$
Scene	$0.2506_{\pm0.0227}$	$0.3976_{\pm0.0388}$	$0.2611_{\pm0.0158}$	$0.3551_{\pm0.1101}$	$0.2699_{\pm0.0196}$	$0.3365_{\pm0.0188}$	$0.2643_{\pm0.0227}$	$0.2813_{\pm0.0003}$	$0.3037_{\pm0.0281}$	$0.2528_{\pm0.0319}$	$\mathbf{0.2455}_{\pm0.0101}$
Social	$0.3474_{\pm0.0129}$	$0.3614_{\pm0.0198}$	$0.3130_{\pm0.0175}$	$0.3370_{\pm0.0171}$	$0.2820_{\pm0.0084}$	$0.3492_{\pm0.0167}$	$0.2874_{\pm0.0189}$	$0.3835_{\pm0.0330}$	$0.2833_{\pm0.0097}$	$0.3305_{\pm0.0152}$	$\mathbf{0.2776}_{\pm0.0077}$
Computers	$0.4378_{\pm0.0124}$	$0.4748_{\pm0.0090}$	$0.4624_{\pm0.0145}$	$0.4427_{\pm0.0106}$	$0.3551_{\pm0.0170}$	$0.4758_{\pm0.0053}$	$0.3642_{\pm0.0064}$	$0.3488_{\pm0.0000}$	$0.3510_{\pm0.0152}$	$0.4095_{\pm0.0165}$	$\mathbf{0.3472}_{\pm0.0110}$
Health	$0.4272_{\pm0.0111}$	$0.3436_{\pm0.0167}$	$0.6578_{\pm0.0165}$	$0.3390_{\pm0.0263}$	$0.2590_{\pm0.0131}$	$0.4934_{\pm0.0180}$	$0.2610_{\pm0.0205}$	$0.2655_{\pm0.0002}$	$0.2580_{\pm0.0192}$	$0.3581_{\pm0.0161}$	$\mathbf{0.2488}_{\pm0.0068}$
Business	$0.1194_{\pm0.0116}$	$0.1338_{\pm0.0092}$	$0.1138_{\pm0.0064}$	$0.1194_{\pm0.0049}$	$0.1193_{\pm0.0115}$	$0.1348_{\pm0.0052}$	$0.1202_{\pm0.0092}$	$0.1196_{\pm0.0000}$	$0.1241_{\pm0.0101}$	$0.1908_{\pm0.0042}$	$\mathbf{0.1190}_{\pm0.0059}$
20NG	$0.5154_{\pm0.0145}$	$0.5270_{\pm0.0030}$	$0.6173_{\pm0.0066}$	$0.6438_{\pm0.1307}$	$0.2708_{\pm0.0052}$	$0.3617_{\pm0.0038}$	$0.2530_{\pm0.0075}$	$0.2652_{\pm0.0000}$	$0.2632_{\pm0.0039}$	$0.4998_{\pm0.0317}$	$\mathbf{0.2481}_{\pm0.0050}$
Tmc2007	$0.2337_{\pm0.0033}$	–	$0.2660_{\pm0.0040}$	$0.3185_{\pm0.0409}$	$0.2115_{\pm0.0041}$	$0.2620_{\pm0.0034}$	$0.1922_{\pm0.0065}$	$0.2097_{\pm0.0000}$	$0.1860_{\pm0.0019}$	$0.2336_{\pm0.0132}$	$\mathbf{0.1824}_{\pm0.0011}$
Ave.	0.3344	0.2887	0.3818	0.2996	0.2625	0.2855	0.2427	0.2710	0.2644	0.2716	**0.2125**
Win/Tie/Loss	15/0/1	14/0/1	13/0/3	15/0/1	15/0/1	15/0/1	16/0/0	15/0/1	16/0/0	15/0/1	

Table 8
Comparison results of MCF-3WCCL with other comparison methods on Ranking Loss (↓) metric.

Dataset	ML-KNN	MRDM	LPLC	GLFS	2SML	LLSF-DL	JLCLS	MSWL	GLMAM	MDFS	MCF-3WCCL
Flags	$0.2256_{\pm0.0320}$	$0.2252_{\pm0.0142}$	$0.2234_{\pm0.0353}$	$0.2535_{\pm0.0106}$	$0.3989_{\pm0.0134}$	$0.2359_{\pm0.0339}$	$0.2189_{\pm0.0273}$	$0.2304_{\pm0.0005}$	$0.4336_{\pm0.0644}$	$0.2350_{\pm0.0107}$	$\mathbf{0.2139}_{\pm0.0430}$
Cal500	$0.1845_{\pm0.0068}$	$0.1846_{\pm0.0025}$	$0.1977_{\pm0.0054}$	$0.1866_{\pm0.0013}$	$0.2346_{\pm0.0059}$	$0.1847_{\pm0.0017}$	$0.1802_{\pm0.0044}$	$0.2019_{\pm0.0000}$	$0.3191_{\pm0.0050}$	$0.1857_{\pm0.0008}$	$\mathbf{0.1789}_{\pm0.0036}$
CHD49	$0.2310_{\pm0.0153}$	$0.2089_{\pm0.0136}$	$0.2060_{\pm0.0136}$	$0.2272_{\pm0.0143}$	$0.2161_{\pm0.0127}$	$0.2551_{\pm0.0191}$	$0.2077_{\pm0.0299}$	$0.2064_{\pm0.0003}$	$0.2154_{\pm0.0135}$	$0.2534_{\pm0.0088}$	$\mathbf{0.1920}_{\pm0.0309}$
Emotion	$0.1718_{\pm0.0174}$	$0.1643_{\pm0.0065}$	$0.1748_{\pm0.0241}$	$0.1948_{\pm0.0283}$	$0.1709_{\pm0.0155}$	$0.1630_{\pm0.0150}$	$0.1871_{\pm0.0141}$	$0.1779_{\pm0.0043}$	$0.1709_{\pm0.0207}$	$0.1758_{\pm0.0219}$	$\mathbf{0.1578}_{\pm0.0191}$
Birds	$0.1059_{\pm0.0137}$	$0.1121_{\pm0.0128}$	$0.1412_{\pm0.0148}$	$0.1310_{\pm0.0259}$	$0.0987_{\pm0.0131}$	$0.1150_{\pm0.0076}$	$0.1157_{\pm0.0189}$	$0.1004_{\pm0.0001}$	$0.1050_{\pm0.0214}$	$0.1085_{\pm0.0102}$	$\mathbf{0.0917}_{\pm0.0170}$
Genbase	$0.1958_{\pm0.0185}$	$0.0060_{\pm0.0035}$	$0.0054_{\pm0.0011}$	$0.0194_{\pm0.0261}$	$0.4012_{\pm0.4889}$	$0.0011_{\pm0.0008}$	$0.0023_{\pm0.0016}$	$0.0084_{\pm0.0145}$	$0.0118_{\pm0.0066}$	$0.0014_{\pm0.0011}$	$\mathbf{0.0006}_{\pm0.0004}$
Medical	$0.0431_{\pm0.0070}$	$0.0335_{\pm0.0070}$	$0.0639_{\pm0.0023}$	$0.0543_{\pm0.0199}$	$0.6080_{\pm0.4801}$	$0.0429_{\pm0.0078}$	$0.0232_{\pm0.0107}$	$0.1034_{\pm0.0040}$	$\mathbf{0.0204}_{\pm0.0053}$	$0.0453_{\pm0.0064}$	$0.0244_{\pm0.0066}$
Enron	$0.0938_{\pm0.0059}$	$0.0892_{\pm0.0035}$	$0.1756_{\pm0.0129}$	$0.0932_{\pm0.0050}$	$0.0950_{\pm0.0032}$	$0.0900_{\pm0.0043}$	$\mathbf{0.0782}_{\pm0.0058}$	$0.2793_{\pm0.0591}$	$0.1754_{\pm0.0102}$	$0.0926_{\pm0.0036}$	$0.0900_{\pm0.0052}$
Image	$0.1795_{\pm0.0113}$	$0.2067_{\pm0.0099}$	$0.1907_{\pm0.0151}$	$0.2271_{\pm0.0334}$	$0.2028_{\pm0.0105}$	$0.1749_{\pm0.0165}$	$0.2195_{\pm0.0068}$	$0.1876_{\pm0.0004}$	$0.1910_{\pm0.0254}$	$0.1915_{\pm0.0155}$	$\mathbf{0.1746}_{\pm0.0052}$
Scene	$0.0900_{\pm0.0075}$	$0.1567_{\pm0.0224}$	$0.0976_{\pm0.0073}$	$0.1451_{\pm0.0608}$	$0.0921_{\pm0.0060}$	$0.1513_{\pm0.0095}$	$0.0889_{\pm0.0062}$	$0.1037_{\pm0.0001}$	$0.1138_{\pm0.0097}$	$0.0893_{\pm0.0163}$	$\mathbf{0.0871}_{\pm0.0078}$
Social	$0.0595_{\pm0.0032}$	$0.0615_{\pm0.0062}$	$0.0915_{\pm0.0039}$	$0.0641_{\pm0.0033}$	$0.0566_{\pm0.0030}$	$0.0568_{\pm0.0054}$	$0.0562_{\pm0.0041}$	$0.1430_{\pm0.0145}$	$0.0687_{\pm0.0098}$	$0.0606_{\pm0.0038}$	$\mathbf{0.0561}_{\pm0.0051}$
Computers	$0.0897_{\pm0.0054}$	$0.0970_{\pm0.0034}$	$0.1749_{\pm0.0207}$	$0.0915_{\pm0.0029}$	$0.0778_{\pm0.0057}$	$\mathbf{0.0767}_{\pm0.0041}$	$0.0809_{\pm0.0038}$	$0.1037_{\pm0.0000}$	$0.1052_{\pm0.0074}$	$0.0820_{\pm0.0032}$	$0.0966_{\pm0.0053}$
Health	$0.0670_{\pm0.0023}$	$0.0541_{\pm0.0034}$	$0.1128_{\pm0.0028}$	$0.0572_{\pm0.0055}$	$0.0529_{\pm0.0021}$	$0.0597_{\pm0.0011}$	$\mathbf{0.0521}_{\pm0.0017}$	$0.0587_{\pm0.0000}$	$0.0661_{\pm0.0048}$	$0.0664_{\pm0.0027}$	$0.0660_{\pm0.0023}$
Business	$0.0396_{\pm0.0034}$	$0.0452_{\pm0.0033}$	$0.0556_{\pm0.0022}$	$0.0452_{\pm0.0019}$	$0.0384_{\pm0.0045}$	$\mathbf{0.0328}_{\pm0.0016}$	$0.0384_{\pm0.0023}$	$0.0400_{\pm0.0000}$	$0.0450_{\pm0.0037}$	$0.0449_{\pm0.0023}$	$0.0356_{\pm0.0025}$
20NG	$0.1767_{\pm0.0059}$	$0.2240_{\pm0.0047}$	$0.3242_{\pm0.0019}$	$0.2912_{\pm0.0777}$	$0.0551_{\pm0.0014}$	$0.1244_{\pm0.0013}$	$0.0468_{\pm0.0007}$	$0.0555_{\pm0.0000}$	$0.0561_{\pm0.0011}$	$0.1609_{\pm0.0146}$	$\mathbf{0.0464}_{\pm0.0011}$
Tmc2007	$0.0605_{\pm0.0008}$	–	$0.0914_{\pm0.0017}$	$0.1151_{\pm0.0180}$	$0.0459_{\pm0.0005}$	$0.1194_{\pm0.0030}$	$0.0424_{\pm0.0008}$	$0.0439_{\pm0.0000}$	$0.0457_{\pm0.0013}$	$0.0617_{\pm0.0047}$	$\mathbf{0.0423}_{\pm0.0009}$
Ave.	0.1259	0.1246	0.2109	0.1373	0.1778	0.1177	0.1024	0.1278	0.1339	0.1160	**0.0971**
Win/Tie/Loss	15/0/1	13/0/2	16/0/0	14/0/2	14/0/2	13/0/3	12/0/4	15/0/1	15/0/1	15/0/1	

Table 9
Comparison results of MCF-3WCCL with other comparison methods on Hamming Loss (↓) metric.

Dataset	ML-KNN	MRDM	LPLC	GLFS	2SML	LLSF-DL	JLCLS	MSWL	GLMAM	MDFS	MCF-3WCCL
Flags	$0.3102_{\pm0.0209}$	$0.3042_{\pm0.0238}$	$0.3436_{\pm0.0218}$	$0.3571_{\pm0.0235}$	$0.4924_{\pm0.0158}$	$0.2944_{\pm0.0164}$	$0.2977_{\pm0.0236}$	$0.2899_{\pm0.0007}$	$0.4572_{\pm0.0404}$	$0.3567_{\pm0.0109}$	$\mathbf{0.2874}_{\pm0.0433}$
Cal500	$0.1391_{\pm0.0011}$	$0.1385_{\pm0.0022}$	$0.1560_{\pm0.0039}$	$0.1378_{\pm0.0007}$	$0.1379_{\pm0.0020}$	$\mathbf{0.1369}_{\pm0.0015}$	$0.1373_{\pm0.0023}$	$0.1465_{\pm0.0000}$	$0.1500_{\pm0.0034}$	$0.1438_{\pm0.0007}$	$0.1370_{\pm0.0008}$
CHD49	$0.3180_{\pm0.0100}$	$0.3114_{\pm0.0152}$	$0.3336_{\pm0.0103}$	$0.2998_{\pm0.0061}$	$0.2910_{\pm0.0097}$	$0.3032_{\pm0.0243}$	$0.3311_{\pm0.0036}$	$0.3195_{\pm0.0004}$	$0.4291_{\pm0.0152}$	$0.3360_{\pm0.0076}$	$\mathbf{0.2874}_{\pm0.0133}$
Emotion	$0.2096_{\pm0.0210}$	$0.2019_{\pm0.0140}$	$0.2162_{\pm0.0113}$	$0.2126_{\pm0.0221}$	$0.2086_{\pm0.0121}$	$0.2193_{\pm0.0145}$	$0.2088_{\pm0.0087}$	$0.2068_{\pm0.0002}$	$0.3109_{\pm0.0113}$	$0.2054_{\pm0.0105}$	$\mathbf{0.1959}_{\pm0.0108}$
Birds	$0.0573_{\pm0.0033}$	$0.0560_{\pm0.0059}$	$0.0619_{\pm0.0024}$	$0.0614_{\pm0.0026}$	$0.0494_{\pm0.0032}$	$0.0578_{\pm0.0014}$	$0.0543_{\pm0.0048}$	$0.0528_{\pm0.0001}$	$0.0735_{\pm0.0023}$	$0.0575_{\pm0.0023}$	$\mathbf{0.0470}_{\pm0.0055}$
Genbase	$0.0756_{\pm0.0024}$	$0.0177_{\pm0.0064}$	$0.0052_{\pm0.0009}$	$0.0067_{\pm0.0081}$	$0.3276_{\pm0.3439}$	$0.0017_{\pm0.0007}$	$0.0069_{\pm0.0011}$	$0.0034_{\pm0.0044}$	$0.0428_{\pm0.0033}$	$0.0042_{\pm0.0007}$	$\mathbf{0.0007}_{\pm0.0007}$
Medical	$0.0158_{\pm0.0007}$	$0.0237_{\pm0.0011}$	$0.0179_{\pm0.0011}$	$0.0143_{\pm0.0028}$	$0.1299_{\pm0.2044}$	$0.0146_{\pm0.0067}$	$0.0154_{\pm0.0006}$	$0.0563_{\pm0.0304}$	$0.0264_{\pm0.0011}$	$0.0161_{\pm0.0010}$	$\mathbf{0.0138}_{\pm0.0016}$
Enron	$0.0527_{\pm0.0012}$	$0.0569_{\pm0.0033}$	$0.0603_{\pm0.0016}$	$0.0527_{\pm0.0034}$	$0.0478_{\pm0.0011}$	$0.0522_{\pm0.0017}$	$0.0581_{\pm0.0024}$	$0.1251_{\pm0.0208}$	$0.0637_{\pm0.0008}$	$0.0536_{\pm0.0013}$	$\mathbf{0.0453}_{\pm0.0008}$
Image	$\mathbf{0.1762}_{\pm0.0079}$	$0.1804_{\pm0.0100}$	$0.1883_{\pm0.0073}$	$0.1981_{\pm0.0178}$	$0.2110_{\pm0.0037}$	$0.2060_{\pm0.0032}$	$0.2056_{\pm0.0030}$	$0.1890_{\pm0.0002}$	$0.2471_{\pm0.0062}$	$0.1771_{\pm0.0079}$	$0.1784_{\pm0.0083}$
Scene	$\mathbf{0.0866}_{\pm0.0074}$	$0.1019_{\pm0.0041}$	$0.0980_{\pm0.0042}$	$0.1183_{\pm0.0191}$	$0.1128_{\pm0.0065}$	$0.1350_{\pm0.0035}$	$0.1757_{\pm0.0012}$	$0.1156_{\pm0.0000}$	$0.1786_{\pm0.0023}$	$0.0972_{\pm0.0081}$	$0.1019_{\pm0.0036}$
Social	$0.0230_{\pm0.0005}$	$0.0261_{\pm0.0006}$	$0.0222_{\pm0.0005}$	$0.0236_{\pm0.0021}$	$0.0216_{\pm0.0003}$	$0.0229_{\pm0.0002}$	$0.0254_{\pm0.0003}$	$0.1713_{\pm0.0163}$	$0.0328_{\pm0.0007}$	$0.0224_{\pm0.0005}$	$\mathbf{0.0205}_{\pm0.0009}$
Computers	$0.0407_{\pm0.0010}$	$0.0394_{\pm0.0013}$	$0.0411_{\pm0.0003}$	$0.0397_{\pm0.0009}$	$0.0343_{\pm0.0005}$	$0.0384_{\pm0.0017}$	$0.0351_{\pm0.0007}$	$0.0344_{\pm0.0000}$	$0.0457_{\pm0.0004}$	$0.0381_{\pm0.0013}$	$\mathbf{0.0339}_{\pm0.0013}$
Health	$0.0474_{\pm0.0012}$	$0.0401_{\pm0.0006}$	$0.0810_{\pm0.0033}$	$0.0383_{\pm0.0024}$	$0.0349_{\pm0.0006}$	$0.0394_{\pm0.0005}$	$0.0440_{\pm0.0010}$	$0.0355_{\pm0.0000}$	$0.0519_{\pm0.0013}$	$0.0418_{\pm0.0026}$	$\mathbf{0.0332}_{\pm0.0004}$
Business	$0.0272_{\pm0.0014}$	$0.0273_{\pm0.0005}$	$0.0275_{\pm0.0006}$	$0.0284_{\pm0.0004}$	$0.0271_{\pm0.0010}$	$\mathbf{0.0261}_{\pm0.0010}$	$0.0309_{\pm0.0002}$	$0.0429_{\pm0.0050}$	$0.0529_{\pm0.0007}$	$0.0274_{\pm0.0005}$	$0.0267_{\pm0.0016}$
20NG	$0.0395_{\pm0.0005}$	$0.0376_{\pm0.0005}$	$0.0490_{\pm0.0002}$	$0.0397_{\pm0.0047}$	$0.6174_{\pm0.0016}$	$0.0357_{\pm0.0007}$	$0.0359_{\pm0.0004}$	$0.0325_{\pm0.0000}$	$0.0510_{\pm0.0003}$	$0.0397_{\pm0.0030}$	$\mathbf{0.0324}_{\pm0.0009}$
Tmc2007	$0.0653_{\pm0.0007}$	–	$0.0711_{\pm0.0006}$	$0.0832_{\pm0.0040}$	$0.6009_{\pm0.0034}$	$0.0594_{\pm0.0005}$	$0.0609_{\pm0.0006}$	$0.0665_{\pm0.0000}$	$0.1001_{\pm0.0007}$	$0.0654_{\pm0.0029}$	$\mathbf{0.0588}_{\pm0.0005}$
Ave.	0.1053	0.1042	0.1108	0.1070	0.2090	0.1027	0.1077	0.1180	0.1446	0.1052	**0.0938**
Win/Tie/Loss	14/0/2	15/0/0	16/0/0	16/0/0	16/0/0	14/0/2	16/0/0	16/0/0	16/0/0	16/0/0	

integrating extent and intent relevances to improve the performance and generalization ability of MCF-3WCCL.

Fig. 5 illustrates the influence of α and β on the performance of MCF-3WCCL. Each row corresponds to a different dataset and shows the influences of α and β on the five evaluation metrics. The results indicate the following: Adjusting parameters α and β in the multi-label learning model significantly affects each evaluation metric. While the optimal parameter settings may vary for different datasets, the overall trend is consistent. These consistent trends suggest that when tuning model parameters, one can refer to these common patterns to optimize

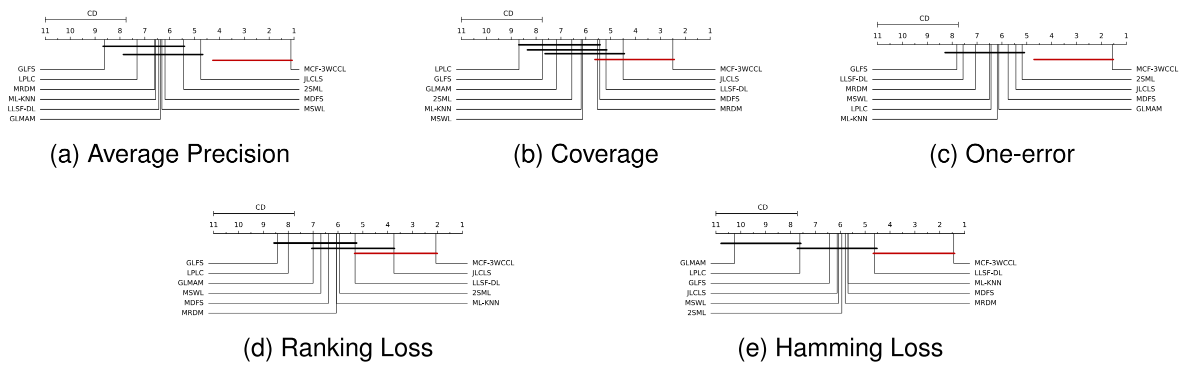

Fig. 3. Bonferroni-Dunn test of MCF-3WCCL is compared with other comparison algorithms.

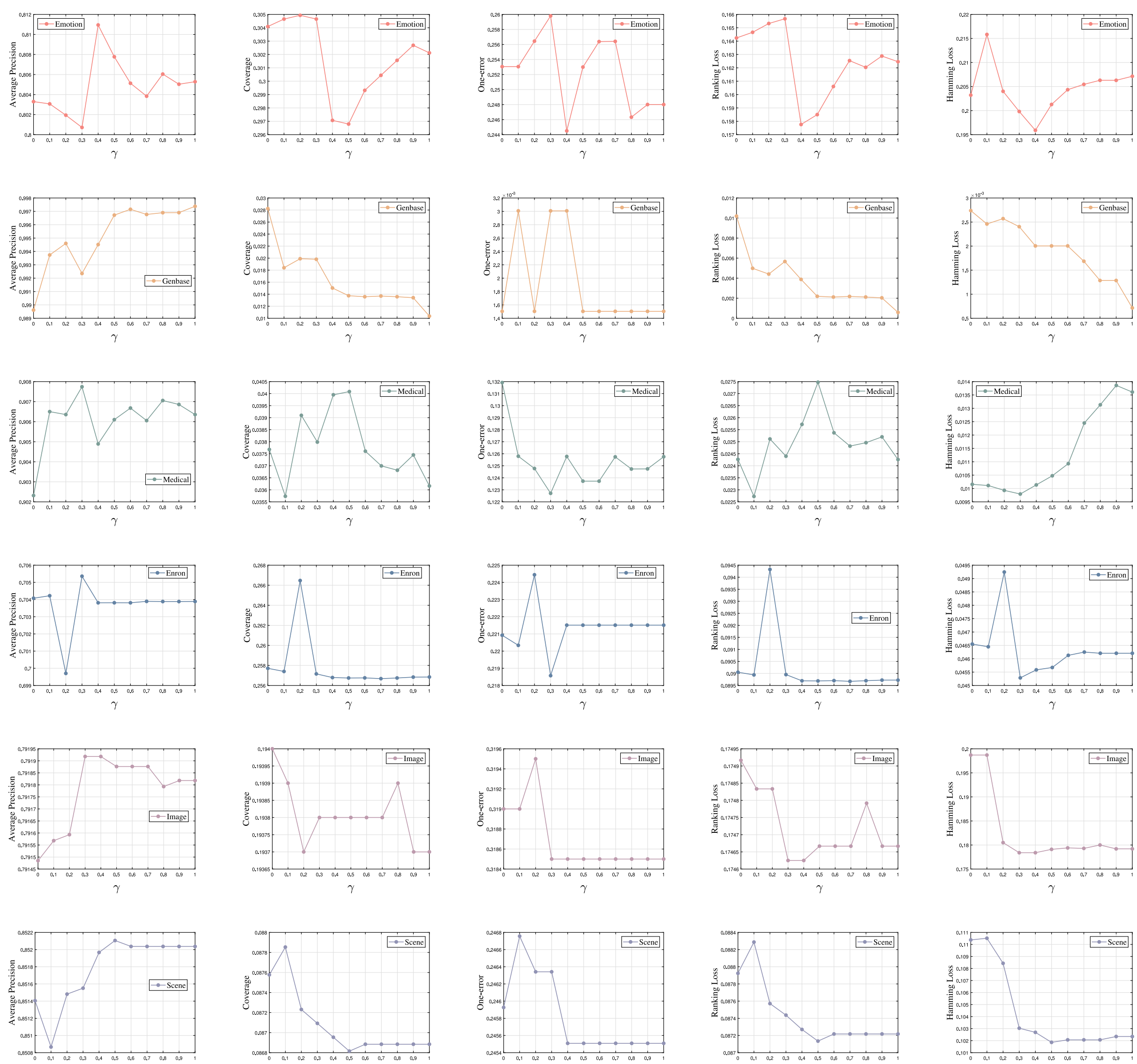

Fig. 4. The influence of parameter γ on the performance of MCF-3WCCL.

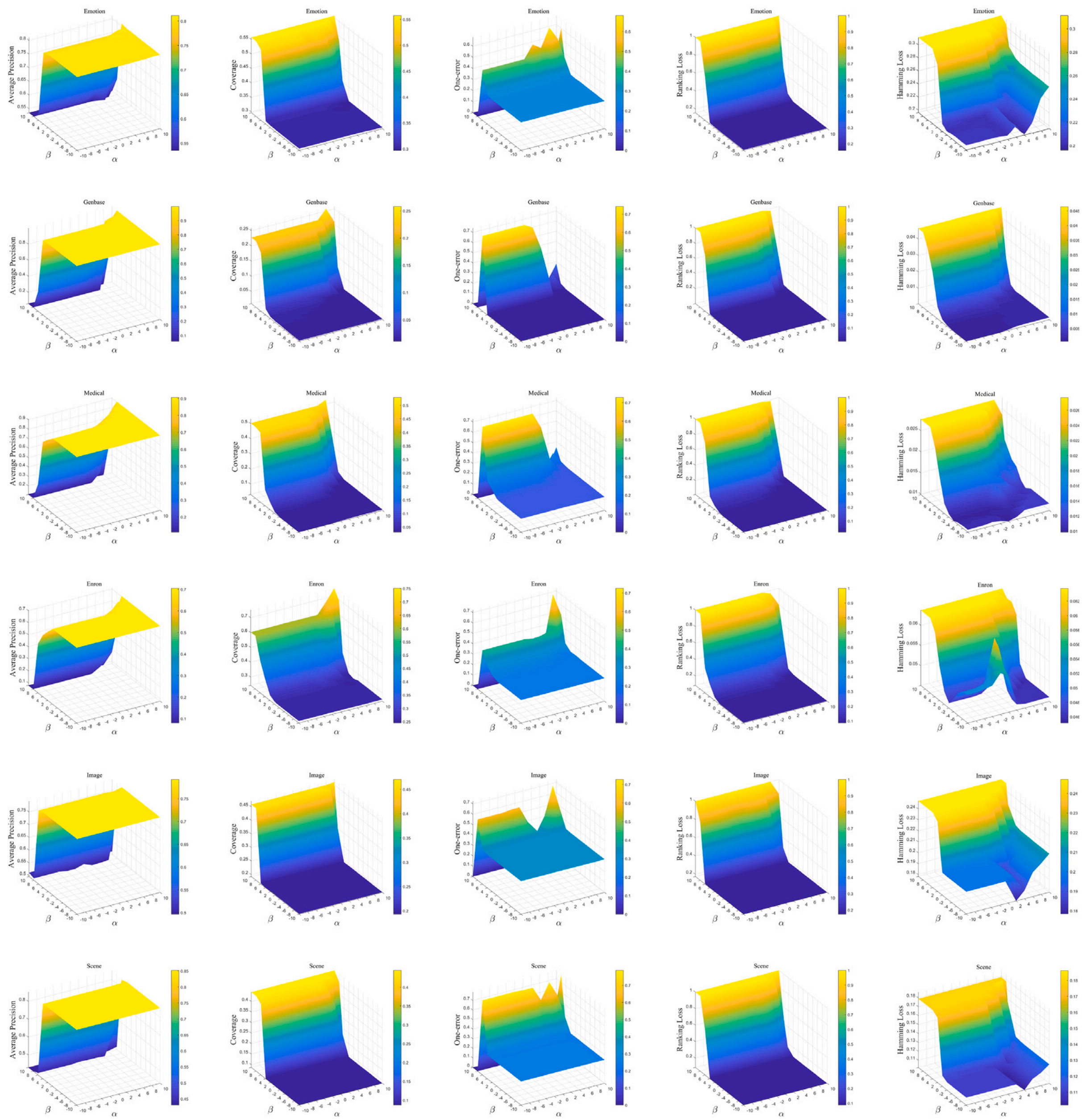

Fig. 5. The influence of parameters α and β on the performance of MCF-3WCCL.

Table 10
Friedman test and the critical value.

Evaluation metric	F_F	Critical value($\theta = 0.05$)
Average precision	6.1807	
Coverage	4.1098	
One-error	4.1054	1.8943
Ranking loss	5.3467	
Hamming loss	8.5581	

performance. This analysis demonstrates the importance of fine-tuning parameters α and β to attain optimal performance across different datasets, guided by the observed trends.

5. Conclusion

A core challenge of multi-label learning tasks is how to accurately capture and fuse complex and multi-level correlations among labels. This paper proposes a multi-level correlation information fusion model via three-way concept-cognitive learning for multi-label learning. The model employs the three-way concept-cognitive operator to structurally represent label concepts, effectively capturing the multi-level correlation information among labels. It calculates the degree of importance of each label concept to the target label based on the structural information of the label concepts. Further, through the mapping of label concept extent, the multi-level correlation information is extended to the feature layer to form a dependency between labels and features. On the basis of feature concept fusion, the overall cognition of the label is formed. A series of experimental results indicate that the model demonstrates superior predictive performance and strong interpretability. This multi-level correlation modeling method based on concept-cognitive learning provides a new solution for multi-label learning tasks, and provides a deeper ability to understand the potential associations between labels.

This method effectively fuses multi-level correlation information, thereby improving both the performance and interpretability of multi-label learning. However, the method is applicable only to complete multi-label data. In real-world scenarios, multi-label data is often incomplete, with some labels missing for certain instances. For such incomplete label data, the 3WL-concepts derived by the method are also incomplete, making it difficult to leverage the extents of 3WL-concepts as cues to obtain corresponding 3WF-concepts. This limitation restricts the method's applicability in broader and more realistic environments. Therefore, our future work will focus on further extending this method to address the challenge of constructing accurate label concepts from incomplete label data, thereby enhancing its applicability and robustness in missing multi-label learning.

CRediT authorship contribution statement

Jiaming Wu: Investigation, Writing – original draft, Software, Methodology, Conceptualization. **Eric C.C. Tsang:** Validation, Supervision, Methodology, Investigation. **Weihua Xu:** Validation, Supervision. **Chengling Zhang:** Visualization, Software. **Qianshuo Wang:** Investigation, Data curation.

Declaration of competing interest

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

Acknowledgments

This paper is supported by the Macao Science and Technology Development Fund (0036/2023/RIA1, 0037/2024/RIA1), Jiangsu Kaite Auto Parts Co., Ltd. (MUST Code: 1330) and the Natural Science Foundation of China (No. 61976245).

Data availability

Data will be made available on request.

References

[1] W.W. Liu, H.B. Wang, X.B. Shen, I.W. Tsang, The emerging trends of multi-label learning, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2022) 7955–7974.
[2] J. Zhang, Z.M. Luo, C.D. Li, et al., Manifold regularized discriminative feature selection for multi-label learning, Pattern Recognit. 95 (2019) 136–150.
[3] R. Huang, Z.J. Wu, Multi-label feature selection via manifold regularization and dependence maximization, Pattern Recognit. 120 (2021) 108149.
[4] J. Zhang, S.Z. Li, M. Jiang, K.C. Tan, Learning from weakly labeled data based on manifold regularized sparse model, IEEE Trans. Cybern. 52 (5) (2022) 3841–3854.
[5] Y. Ma, X.Y. Zou, Q.Z. Pan, M. Yan, G.Q. Li, Target-embedding autoencoder with knowledge distillation for multi-label classification, IEEE Trans. Emerg. Top. Comput. Intell. 8 (3) (2024) 2506–2517.
[6] Y.L. Fan, J.H. Liu, J.N. Tang, P.Z. Liu, et al., Learning correlation information for multi-label feature selection, Pattern Recognit. 145 (2024) 109899.
[7] M.L. Zhang, Z.H. Zhou, A review on multi-label learning algorithms, IEEE Trans. Knowl. Data Eng. 26 (8) (2014) 1819–1837.
[8] C.M. Zhu, D.Q. Miao, Z. Wang, et al., Global and local multi-view multi-label learning, Neurocomputing 371 (2020) 67–77.
[9] Z.B. Yu, M.L. Zhang, Multi-label classification with label-specific feature generation: a wrapped approach, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2022) 5199–5210.
[10] Y.B. Wang, W.J. Zheng, Y.S. Cheng, D.W. Zhao, Joint label completion and label-specific features for multi-label learning algorithm, Soft Comput. 24 (2020) 6553–6569.
[11] Y.S. Chen, Qian K., F. Min, Global and local attention-based multi-label learning with missing labels, Inf. Sci. 594 (2022) 20–42.
[12] J. Huang, G.R. Li, Q.M. Huang, X.D. Wu, Learning label-specific features and class-dependent labels for multi-label classification, IEEE Trans. Knowl. Data Eng. 28 (12) (2016) 3309–3323.
[13] M.L. Zhang, L. Wu, LIFT: Multi-label learning with label-specific features, IEEE Trans. Pattern Anal. Mach. Intell. 37 (1) (2015) 107–120.
[14] M.L. Zhang, Z.H. Zhou, ML-KNN: a lazy learning approach to multi-label learning, Pattern Recognit. 40 (2007) 2038–2048.
[15] J.N. Li, Y.C. Zhang, S.W. Chen, R.F. Xu, Enhancing multi-label classification via dynamic label-order learning, Proc. AAAI Conf. Artif. Intell. 38 (17) (2024) 18527–18535.
[16] J.H. Dai, W.Y. Huang, C.C. Zhang, J. Liu, Multi-label feature selection by strongly relevant label gain and label mutual aid, Pattern Recognit. 145 (2024) 109945.
[17] J. Zhang, H.R. Wu, M. Jiang, et al., Group-preserving label-specific feature selection for multi-label learning, Expert. Syst. Appl. 213 (2023) 118861.
[18] K. Qian, X.Y. Min, Y.S. Chen, F. Min, Weight matrix sharing for multi-label learning, Pattern Recognit. 136 (2023) 109156.
[19] X.Y. Zhao, Y.X. An, N. Xu, X. Geng, Fusion label enhancement for multi-label learning, in: Proceedings of the International Joint Conference on Artificial Intelligence, 2022, pp. 3773–3779.
[20] X.Y. Zhao, Y.X. An, N. Xu, X. Geng, Variational continuous label distribution learning for multi-label text classification, IEEE Trans. Knowl. Data Eng. 36 (6) (2024) 2716–2729.
[21] X.Y. Che, D.G. Chen, J. Deng, J.S. Mi, Exploiting local label correlation from sample perspective for multi-label classification via three-way decision theory, Appl. Soft Comput. 149 (2023) 110950.
[22] J. Huang, G.R. Li, S.H. Wang, et al., Multi-label classification by exploiting local positive and negative pairwise label correlation, Neurocomputing 257 (2017) 164–174.
[23] W.Z. Wu, Y. Leung, J.S. Mi, Granular computing and knowledge reduction in formal contexts, IEEE Trans. Knowl. Data Eng. 21 (10) (2009) 1461–1474.
[24] W.H. Xu, W.T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy dataset, IEEE Trans. Cybern. 46 (2016) 366–379.
[25] Y. Shi, Y.L. Mi, J.H. Li, W.Q. Liu, Concept-cognitive learning model for incremental concept learning, IEEE Trans. Syst. Man Cybern.: Syst. 51 (2) (2021) 809–821.
[26] D.D. Guo, W.H. Xu, W.P. Ding, et al., Concept-cognitive learning survey: mining and fusing knowledge from data, Inf. Fusion 109 (2024) 102426.
[27] R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, in: Ordered Sets, Dordrecht, The Netherlands, 1982, pp. 445–470.
[28] W.X. Zhang, J.M. Ma, S.Q. Fan, Variable threshold concept lattices, Inf. Sci. 177 (2007) 4883–4892.
[29] J.M. M, W.X. Zhang, Axiomatic characterizations of dual concept lattices, Int. J. Approx. Reason. 54 (2013) 690–697.
[30] J.H. Li, C.L. Mei, W.H. Xu, Y.H. Qian, Concept learning via granular computing: a cognitive viewpoint, Inf. Sci. 298 (2015) 447–467.
[31] Y.L. Mi, Y. Shi, J.H. Li, et al., Fuzzy-based concept learning method: exploiting data with fuzzy conceptual clustering, IEEE Trans. Cybern. 52 (1) (2022) 582–593.
[32] W.H. Xu, D.D. Guo, Y.H. Qian, W.P. Ding, Two-way concept-cognitive learning method: a fuzzy-based progressive learning, IEEE Trans. Fuzzy Syst. 31 (6) (2023) 1885–1899.
[33] W.H. Xu, D.D. Guo, J.S. Mi, et al., Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural Netw. Learn. Syst. 34 (10) (2023) 6798–6812.
[34] D.D. Guo, W.H. Xu, Y.H. Qian, W.P. Ding, M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[35] Z.M. Liu, J.H. Li, X. Zhang, X.Z. Wang, Incremental incomplete concept-cognitive learning model: a stochastic strategy, IEEE Trans. Neural Netw. Learn. Syst. 36 (1) (2025) 393–406.
[36] Y.Y. Yao, Three-way decision and granular computing, Int. J. Approx. Reason. 103 (2018) 107–123.
[37] J.M. Wu, D.Y. Liu, Z.H. Huang, J.J. Li, Multi-scale decision systems with test cost and applications to three-way multi-attribute decision-making, Appl. Intell. 54 (2024) 3591–3605.
[38] Y.Y. Yao, Three-way granular computing, rough sets, and formal concept analysis, Int. J. Approx. Reason. 116 (2020) 106–125.
[39] Y.Y. Yao, Interpreting concept learning in cognitive informatics and granular computing, IEEE Trans. Syst. Man Cybern. Part B 39 (4) (2009) 855–866.
[40] J.J. Qi, T. Qian, L. Wei, The connections between three-way and classical concept lattices, Knowl.-Based Syst. 91 (2016) 143–151.
[41] J.H. Li, C.C. Huang, J.J. Qi, et al., Three-way cognitive concept learning via multi-granularity, Inf. Sci. 378 (2017) 244–263.
[42] D.D. Guo, W.H. Xu, Y.H. Qian, W.P. 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.
[43] K.H. Yuan, W.H. Xu, W.T. Li, W.P. Ding, An incremental learning mechanism for object classification based on progressive fuzzy three-way concept, Inf. Sci. 584 (2022) 127–147.
[44] C.L. Zhang, E.C.C. Tsang, W.H. Xu, et al., Dynamic updating variable precision three-way concept method based on two-way concept-cognitive learning in fuzzy formal contexts, Inf. Sci. 655 (2024) 119818.

[45] J.M. Wu, E.C.C. Tsang, W.H. Xu, et al., Correlation concept-cognitive learning model for multi-label classification, Knowl.-Based Syst. 290 (2024) 111566.
[46] Z.M. Liu, J.H. Li, X. Zhang, et al., Multi-level information fusion for missing multi-label learning based on stochastic concept clustering, Inf. Fusion 115 (2025) 102775.
[47] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[48] M. Friedman, A comparison of alternative tests of significance for the problem of m ranking, Ann. Math. Stat. 11 (1940) 86–92.
[49] O.J. Dunn, Multiple comparisons among means, J. Am. Stat. Assoc. 56 (1961) 52–64.