

# Hypergraph-Based Multi-View Multi-Label Classification via Adaptive High-Order Semantic Fusion

Yi Shan<sup>1</sup>, Liyang Gao<sup>1</sup>, Yuena Lin<sup>1</sup>, Zhen Yang<sup>1</sup>, Gengyu Lyu<sup>1\*</sup>, Honggui Han<sup>1</sup>

<sup>1</sup>College of Computer Science, Beijing University of Technology

yi.shan.bjut@gmail.com, oldgaoglycine@gmail.com, yuenalin@126.com, yangzhen@bjut.edu.cn, lyugengyu@gmail.com, rechardhan@bjut.edu.cn

## Abstract

In multi-view multi-label (MVML) classification, each sample is represented by multiple heterogeneous views and annotated with multiple labels. Existing methods typically exploit pairwise semantic relationships to mine intra-view correlations and align inter-view features for generating structural representations. However, these methods ignore the direct expression of high-order semantic similarities and alignments from a group perspective, which necessitates multi-step aggregation for subsequent feature fusion, leading to the inefficient and incomplete integration of key semantic information. To overcome this limitation, we propose a novel hypergraph-based MVML method with Adaptive High-Order Semantic Fusion (HyperAHSF), which leverages hypergraphs to adaptively model group-level semantic similarities within each view and group-level semantic alignments across different views, enabling more effective feature fusion. Specifically, we first construct view-specific hyperedges by selecting multiple groups of node representations exhibiting high semantic similarity, which captures the group-level semantic similarities within each view, forming view-specific hypergraphs. Furthermore, we establish cross-view hyperedges to connect the multi-view node representations of each sample, which characterizes the group-level semantic alignments across different views, accordingly forming a unified multi-view hypergraph. Afterwards, we employ hypergraph neural networks to efficiently aggregate view-specific information and consensus information from their corresponding hypergraphs via group-level message passing. During the passing process, we impose a label-driven contrastive loss on the consensus information to encourage these representations to cluster toward their corresponding class prototypes, enhancing their discriminability. Finally, the consensus information together with the view-specific information is jointly integrated for multi-label classification. Extensive experiments demonstrate that HyperAHSF outperforms other state-of-the-art methods.

## Introduction

Multi-View Multi-Label (MVML) classification aims to learn from training data, where each sample is described by several heterogeneous view representations and associated with multiple relevant labels (Wang et al. 2025; Hao, Zhang,



Figure 1: An example of MVML in website classification.

and Zhang 2025; Cui et al. 2024). For example, in the commodity website classification task (Figure 1), a single commodity can be described by *picture*, *video*, and *text*. These diverse information sources are used to predict multiple relevant labels for this commodity (i.e., *kitchen*, *knife*, *black*). MVML provides an effective framework to learn a desired multi-label classifier, which bridges these distinct sources of commodity information (views) with different descriptive tags or categories (labels).

The key to learn from MVML data lies in how to effectively integrate intra-view relevant features while comprehensively fuse inter-view heterogeneous features, thereby obtaining structural representations conducive to multi-label classification. For example, (Ruan, Yang, and Li 2022) proposed an MVML method called GCNR, which models view-specific sample correlations based on feature similarity among the corresponding view representations, and then extracts consensus information and view-specific information from these correlations for multi-label classification. (Wei et al. 2025) proposed an MVML method named VAMS, which establishes adjacency relationships among intra-view similar representations and inter-view consistent representations, thus incorporates individual-view specificity and cross-view consensus into a unified framework to derive multi-label prediction scores. However, these methods typically leverage pairwise relationships at the feature level to model intra-view correlations and cross-view alignments, limited by failing to directly express the high-order semantic interactions, which results in two consequences: (1) Message passing relying on pairwise relationships cannot simul-

\*indicates corresponding author.

taneously aggregate information from multiple view representations, leading to incomplete feature fusion. (2) Sample correlation modeling based on feature-space neighborhood relationships fails to utilize labels to further select reliable neighbors, resulting in sub-optimal semantic integration.

To address the above limitation, we propose a novel hypergraph-based MVML method with Adaptive High-Order Semantic Fusion, named HyperAHSF. This method adaptively models high-order semantic similarities within each view and high-order semantic alignments across different views, then employs hypergraph neural networks (HGNNs) to effectively fuse feature from a group-level perspective. Specifically, we first establish a view-specific hypergraph for each view. For each node representation, we adaptively construct view-specific hyperedges by identifying its reliable neighbors with high semantic similarity of labels, which capture group-level semantic similarities within the specific view. Afterwards, we introduce cross-view hyperedges to connect the multi-view node representations of the same sample, which model group-level semantic alignments across different views. Then, the specific hypergraphs are integrated to form a unified multi-view hypergraph and HGNNs are leveraged to perform hyperedge-based message passing on these two types of hypergraphs, where the nodes in view-specific hypergraphs encode view-specific information, while the cross-view hyperedges in unified multi-view hypergraph align consensus information. During this passing process, a label-driven contrastive loss is applied to the representations of cross-view hyperedges to further exploit the high-level semantic information of labels, which encourages these representations to cluster around the corresponding class prototypes and stay away from other class prototypes, thereby enhancing their discriminability. Finally, the consensus information and view-specific information are jointly used for multi-label classification. In summary, the contributions of this paper are as follows:

- We propose a novel MVML method named HyperAHSF, which adaptively captures high-order semantic similarities and alignments of MVML data from a group perspective, enabling effective semantic information fusion.
- To the best of our knowledge, HyperAHSF is the first MVML method to integrate HGNNs with contrastive learning, which leverages hyperedge-based message passing to aggregate more discriminative representations of both cross-view consensus and individual-view specificity for accurate multi-label classification.
- Extensive experimental results and comprehensive experimental analysis have demonstrated that our proposed HyperAHSF can achieve superior performance against other state-of-the-art methods.

## Related Work

### Multi-View Multi-Label Learning

Multi-View Multi-Label Learning is a combination of Multi-View Learning (MVL) and Multi-Label Learning (MLL), which means that MVML must address the challenges of both MVL and MLL (Cui et al. 2024; Wang et al. 2024; Li

et al. 2025c; Zhong, Lyu, and Yang 2025; Ou et al. 2025; Fu et al. 2025; Zou et al. 2025). To learn from such complicated data, (Wu et al. 2019) proposed SIMM, which extracts shared information among views by minimizing adversarial loss, while imposing an orthogonal constraint to preserve the view-specific information of each view. (Lyu et al. 2022) proposed a DeepGCN-based view-specific MVML method, named D-VSM, which bypasses the shared subspace, while each individual view fuses with the complementary information of other views to directly contribute to the final discriminative model. In addition, (Li et al. 2024) proposed DIMvSML to address incomplete MVML scenarios, which constructs view-specific graphs reflecting similarity relations between samples with features. It then builds transfer graphs by aggregating similarity relations from other views, leveraging cross-view consistency for multi-label classification. However, these methods fail to capture complex high-order correlations in the MVML data, resulting in limited expressive ability.

### Hypergraph Neural Networks

Hypergraph neural networks, due to their ability to effectively mine high-order relationships, possess expressive capabilities that graph neural networks do not have, and thus have attracted increasing research interest. Recently, a series of works have been dedicated to proposing general hypergraph learning frameworks. For example, (Feng et al. 2019) introduced a hypergraph neural networks named HGNN to model complex and high-order data correlations beyond pairwise relationships via hyperedge convolution operations. Based on HGNN, (Gao et al. 2023) proposed the HGNN+, which consists of a systematic hypergraph modeling approach through hyperedge groups and a flexible two-stage message passing scheme in the spatial domain. (Bai, Zhang, and Torr 2021) introduced Hypergraph Convolution and Hypergraph Attention as flexible, end-to-end trainable operators that effectively handle higher-order relationships inherent in many real-world applications. Besides, there are also many works that apply hypergraphs to different downstream tasks, such as identification (Yan et al. 2020; Huang et al. 2025a; Tang et al. 2025; Li et al. 2025b) and prediction (Xu et al. 2022; Huang et al. 2025c; Han et al. 2025), etc.

### The Proposed Method

Formally speaking, let  $\mathcal{X} = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \dots \times \mathbb{R}^{d_T}$  be the feature space with  $T$  views, where  $d_t$  ( $1 \leq t \leq T$ ) is the feature dimension of the  $t$ -th view. Given the MVML training dataset  $\mathcal{D} = \{(\mathbf{X}_i, \mathbf{y}_i) \mid 1 \leq i \leq N\}$  comprising  $N$  samples, where  $\mathbf{X}_i \in \mathcal{X}$  is composed of  $T$  feature vectors  $[\mathbf{x}_i^{(1)}; \mathbf{x}_i^{(2)}; \dots; \mathbf{x}_i^{(T)}]$ , and  $\mathbf{y}_i \in \{0, 1\}^{C \times 1}$  is the label vector with  $C$  classes corresponding to  $\mathbf{X}_i$ , our HyperAHSF aims to learn from these heterogeneous representations of different views to induce an accurate multi-label classifier that predicts appropriate labels for unseen samples. Figure 2 illustrates the overview architecture of HyperAHSF, which consists of four main components: *Multi-View Hypergraph Construction*, *High-Order Semantic Fusion*, *Label-Driven Contrastive Learning* and *Multi-Label Classification*.

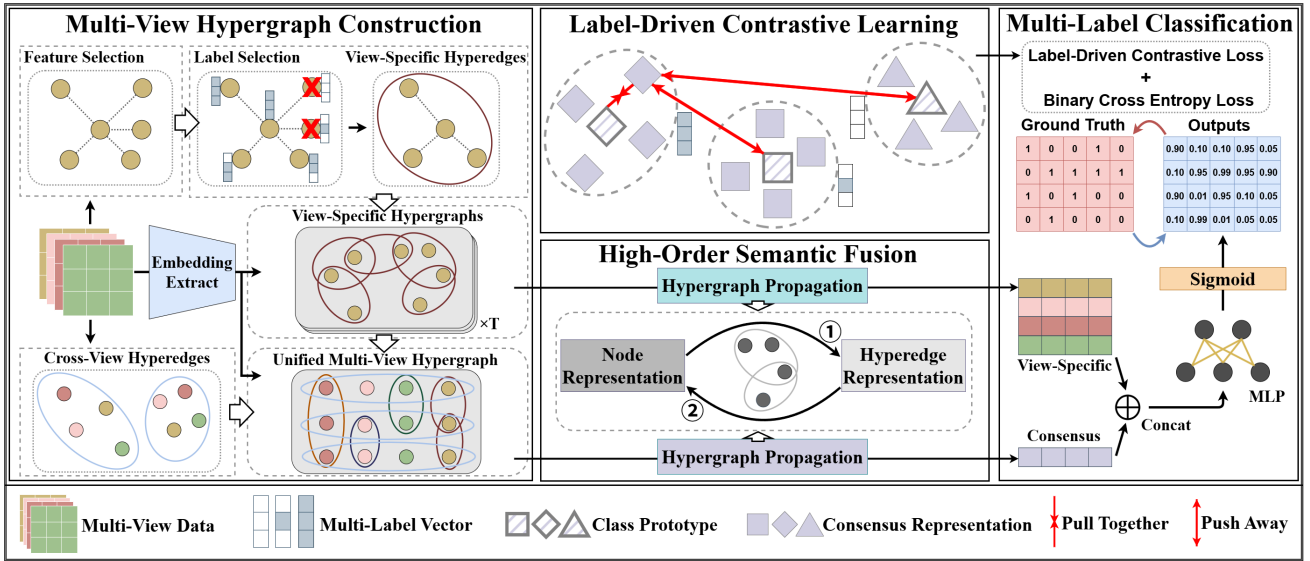


Figure 2: The framework of our proposed HyperAHSF, which consists of four components: (1) *Multi-View Hypergraph Construction*, where view-specific hypergraphs adaptively select multiple groups of node representations with high feature and label similarity, and the unified multi-view hypergraph further connects the multi-view node representations of each sample; (2) *High-Order Semantic Fusion*, where HGNNs are employed to fuse semantic information through message passing between nodes and hyperedges in two types of hypergraphs; (3) *Label-Driven Contrastive Learning*, which encourages the consensus representations to cluster around their corresponding class prototypes and stay away from other class prototypes. (4) *Multi-Label Classification*, where the consensus information and view-specific information is jointly integrated for multi-label prediction.

### Multi-View Hypergraph Construction

A hypergraph can be represented as  $\mathcal{H} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ , which comprises a set of nodes  $\mathcal{V} = \{v_1, v_2, \dots, v_{|\mathcal{V}|}\}$ , a set of hyperedges  $\mathcal{E} = \{e_1, e_2, \dots, e_{|\mathcal{E}|}\}$  and a diagonal matrix of the hyperedge weights  $\mathbf{W} \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{E}|}$ . Each hyperedge  $e_j \subset \mathcal{V}$  is assigned with a positive weight  $\mathbf{W}_{jj}$  to quantify its importance or reliability, where  $1 \leq j \leq |\mathcal{E}|$ . A hypergraph can also be described by its incidence matrix  $\mathbf{H} \in \{0, 1\}^{|\mathcal{V}| \times |\mathcal{E}|}$ , where its entry  $h_{ij} = 1$  indicates that node  $v_i$  is connected by hyperedge  $e_j$ , and  $h_{ij} = 0$  otherwise. For MVML, we first construct  $T$  view-specific hypergraphs  $\mathcal{H}_s^{(t)} = (\mathcal{V}_s^{(t)}, \mathcal{E}_s^{(t)}, \mathbf{W}_s^{(t)})$  with their incidence matrices  $\mathbf{H}_s^{(t)}$  for  $t = 1, 2, \dots, T$ , which capture the group-level semantic similarities within each view. Here, each node  $v_{s,i}^{(t)}$  of  $\mathcal{V}_s^{(t)}$  is described by the corresponding representation  $\mathbf{x}_i^{(t)}$  in the  $t$ -th view. Then, we construct a unified multi-view hypergraph  $\mathcal{H}_c = (\mathcal{V}_c, \mathcal{E}_c, \mathbf{W}_c)$  with its incidence matrix  $\mathbf{H}_c$  based on all view-specific hypergraphs to model the group-level semantic alignments across different views, which means that the node set  $\mathcal{V}_c$  contains multi-view representations of all samples, i.e.,  $\mathcal{V}_c = \bigcup_{t=1}^T \mathcal{V}_s^{(t)}$ .

**View-Specific Hypergraph.** To adaptively model the intricate high-order semantic similarities in each view, we adopt a dual-selection strategy to construct view-specific hyperedges, thereby forming the view-specific hypergraphs. Specifically, we first measure the semantic similarity in the feature space to select  $k$  nearest neighbor nodes for each node. Then, we further select reliable neighbor nodes within

them based on the semantic similarity in the label space, constructing view-specific hyperedges  $\mathcal{E}_{vs}^{(t)}$ :

$$e_{vs,i}^{(t)} = \left\{ \mathbf{x}_j^{(t)} \mid \mathbf{x}_j^{(t)} \in \mathcal{K}(\mathbf{x}_i^{(t)}), \text{Sim}(\mathbf{y}_i, \mathbf{y}_j) \geq \text{thr} \right\}, \quad (1)$$

$$\mathcal{E}_{vs}^{(t)} = \left\{ e_{vs,i}^{(t)} \right\}_{i=1}^N, \quad (2)$$

where  $\mathcal{K}(\mathbf{x}_i^{(t)})$  denotes the  $\mathbf{x}_i^{(t)}$  and its  $k$  nearest neighbors in the Euclidean distance,  $\text{Sim}(\cdot, \cdot)$  represents the label similarity function according to the Jaccard index and  $\text{thr}$  is the threshold of label similarity selection. Please note that the label similarity selection can effectively avoid interference of false-positive noise, which refers to nodes that are similar in the feature space but irrelevant in the label space, meaning that this dual-selection strategy can adaptively capture reliable group-level semantic interaction specific to each view.

Additionally, we add a self-loop hyperedge for each node to preserve its inherent features during the message passing (Huang et al. 2025b). Then, we obtain the hyperedge set  $\mathcal{E}_s^{(t)}$  of the view-specific hypergraph  $\mathcal{H}_s^{(t)}$  by integrating view-specific hyperedges  $\mathcal{E}_{vs}^{(t)}$  and self-loop hyperedges  $\mathcal{E}_{sl}^{(t)}$ :

$$\mathcal{E}_s^{(t)} = \mathcal{E}_{vs}^{(t)} \cup \mathcal{E}_{sl}^{(t)}. \quad (3)$$

**Unified Multi-View Hypergraph.** A key assumption of MVML is that representations of the same sample across different views have potential semantic consistency (Liu et al. 2022, 2025; Zhong et al. 2025; Li et al. 2025a; Jiang et al.

2025). To model such important high-order relationship of the group-level semantic alignments, we introduce cross-view hyperedges  $\mathcal{E}_{cv}$ :

$$e_{cv,i} = \left\{ \mathbf{x}_i^{(t)} \mid t = 1, 2, \dots, T \right\}, \quad (4)$$

$$\mathcal{E}_{cv} = \{e_{cv,i}\}_{i=1}^N, \quad (5)$$

where the cross-view hyperedge  $e_{cv,i}$  connects the  $T$  view representations of the  $i$ -th sample. Then, we integrate cross-view hyperedges  $\mathcal{E}_{cv}$  with view-specific hyperedges  $\mathcal{E}_s^{(t)}$  into a unified multi-view hypergraph  $\mathcal{H}_c$ , aiming to jointly exploit inter-view and intra-view high-order correlations:

$$\mathcal{E}_c = \mathcal{E}_{cv} \cup \mathcal{E}_s^{(1)} \cup \mathcal{E}_s^{(2)} \dots \cup \mathcal{E}_s^{(T)}. \quad (6)$$

### High-Order Semantic Fusion

To effectively mine semantic information from high-order semantic similarities and alignments, we employ HGNNs to fuse features from a group perspective in the corresponding  $T$  view-specific hypergraphs  $\mathcal{H}_s^{(t)}$  for  $t = 1, \dots, T$  and a unified multi-view hypergraph  $\mathcal{H}_c$ , respectively. Formally, we first extract view-specific embeddings  $\mathbf{S}^{(t)} \in \mathbb{R}^{N \times h_v}$  and consensus embeddings  $\mathbf{C}^{(t)} \in \mathbb{R}^{N \times h}$  from input  $\mathbf{X}^{(t)}$  via MLPs. Then,  $\mathbf{S}^{(t)}$  is set as the feature matrix of node set  $\mathcal{V}_s^{(t)}$  and  $\mathbf{C} = [\mathbf{C}^{(1)}; \mathbf{C}^{(2)}; \dots; \mathbf{C}^{(T)}]$  is set as the feature matrix of node set  $\mathcal{V}_c$ , respectively. Furthermore, we leverage independent HGNNs to perform message passing in each hypergraph  $\mathcal{H}_s^{(t)}$  and  $\mathcal{H}_c$ , where once propagation of each HGNN can be formulated as follows:

$$\tilde{\mathbf{Z}} = \sigma \left( \mathbf{D}_V^{-1} \mathbf{H} \mathbf{W} \mathbf{D}_E^{-1} \mathbf{H}^\top \mathbf{Z} \Theta \right), \quad (7)$$

$$\mathbf{Q} = \sigma \left( \mathbf{W} \mathbf{H}^\top \mathbf{Z} \right), \quad (8)$$

where  $\mathbf{D}_V$  is the diagonal matrix of node degree,  $\mathbf{D}_E$  is the diagonal matrix of hyperedge degree and  $\mathbf{Z}$  is the feature matrix of nodes for the corresponding hypergraph. Similarly,  $\mathbf{W}$  is the hyperedge weight matrix and  $\mathbf{H}$  is the incidence matrix depend on the corresponding hypergraph. For simplicity, we set all hyperedge weight matrices as the identity matrix, i.e.,  $\mathbf{W}_s^{(1)} = \dots = \mathbf{W}_s^{(T)} = \mathbf{W}_c = \mathbf{I}$ . In addition,  $\Theta$  is the learnable weight matrix and  $\sigma(\cdot)$  is the activation function of the corresponding HGNN.  $\tilde{\mathbf{Z}}$  is the updated feature matrix of nodes and  $\mathbf{Q}$  is the aggregated feature matrix of hyperedges after once propagation of HGNN.

Furthermore, we exploit the outputs  $\tilde{\mathbf{Z}}$  of Eq. (7) as its inputs and repeat the above propagation operation, then obtaining the desired representations  $\hat{\mathbf{S}}^{(t)} \in \mathbb{R}^{N \times h'_v}$  of the nodes in the view-specific hypergraph  $\mathcal{H}_s^{(t)}$  that encode view-specific information, and representation  $\hat{\mathbf{Q}} \in \mathbb{R}^{N \times h'}$  of cross-view hyperedges in the unified multi-view hypergraph  $\mathcal{H}_c$  that aligns consensus information.

---

### Algorithm 1: The training process of HyperAHSF

---

**Input:** MVML dataset  $\mathcal{D} = \{(\mathbf{X}_i, \mathbf{y}_i) \mid 1 \leq i \leq N\}$ ; The trade-off coefficient  $\alpha$ ; The number of epochs  $I_m$ .

**Output:** The classification model HyperAHSF.

**Process:**

- 1: Construct view-specific hypergraphs  $\mathcal{H}_s^{(t)}$  ( $1 \leq t \leq T$ ) by Eq. (1)-(3).
  - 2: Construct unified multi-view hypergraph  $\mathcal{H}_c$  based on the  $\mathcal{H}_s^{(t)}$  by Eq. (4)-(6).
  - 3: **for** epoch = 1 to  $I_m$  **do**
  - 4:   **//Forward Propagation**
  - 5:   Extract the node features  $\mathbf{S}^{(t)}$  ( $1 \leq t \leq T$ ) and  $\mathbf{C}$  from multi-view data  $\mathbf{X}^{(t)}$  via MLPs.
  - 6:   Conduct the propagation of HGNNs in the hypergraphs  $\mathcal{H}_s^{(t)}$  and  $\mathcal{H}_c$  by Eq. (7)-(8), respectively.
  - 7:   Compute the label-driven contrastive loss  $\mathcal{L}_{lc}$  applied to cross-view hyperedges by Eq. (9).
  - 8:   Compute the cross-entropy loss  $\mathcal{L}_{ce}$  for multi-label classification by Eq. (10)-(11).
  - 9:   **//Backward Propagation**
  - 10:   Update the model parameters by minimizing overall loss  $\mathcal{L}$  in Eq. (12).
  - 11: **end for**
- 

### Label-Driven Contrastive Learning

To further exploit the high-level semantic information of the labels, we apply a label-driven contrastive loss to the representation  $\hat{\mathbf{Q}} = [\hat{\mathbf{q}}_1, \hat{\mathbf{q}}_2, \dots, \hat{\mathbf{q}}_N]^T$  of cross-view hyperedges. Specifically, we first divide the  $N$  cross-view hyperedges into  $G$  valid label groups, where a label group is valid if it contains more than one hyperedge. Hyperedges  $\hat{\mathbf{q}}_i$  and  $\hat{\mathbf{q}}_j$  are assigned to the same group if their label vectors are identical ( $\mathbf{y}_i = \mathbf{y}_j$ ). Let  $\mathcal{I}_g$  denote the set of cross-view hyperedge indices in the  $g$ -th valid label group ( $|\mathcal{I}_g| > 1$ ) and  $\bar{\mathbf{q}}_g = \frac{1}{|\mathcal{I}_g|} \sum_{i \in \mathcal{I}_g} \hat{\mathbf{q}}_i$  denote the class prototype of the  $g$ -th valid label group. Accordingly, the label-driven contrastive loss can be formulated as follows:

$$\mathcal{L}_{lc} = \frac{1}{G} \sum_{g=1}^G \left( \frac{1}{N - |\mathcal{I}_g|} \sum_{i \notin \mathcal{I}_g} \mathcal{S}_{gi} - \frac{1}{|\mathcal{I}_g|} \sum_{j \in \mathcal{I}_g} \mathcal{S}_{gj} \right), \quad (9)$$

where  $\mathcal{S}_{gi}$  denotes the cosine similarity between the class prototype  $\bar{\mathbf{q}}_g$  and the cross-view hyperedge  $\hat{\mathbf{q}}_i$ . This loss aims to pull the cross-view hyperedges together to their class prototype while pushing away from other class prototypes, enhancing the discriminability of consensus information.

### Multi-Label Classification

For the  $i$ -th sample, we concatenate its consensus information  $\hat{\mathbf{q}}_i$  and its view-specific information  $\hat{\mathbf{s}}_i^{(t)}$  to obtain its multi-label prediction scores  $\mathbf{p}_i$ :

$$\mathbf{p}_i = \text{Classifier} \left( \hat{\mathbf{q}}_i \parallel \hat{\mathbf{s}}_i^{(1)} \parallel \hat{\mathbf{s}}_i^{(2)} \parallel \dots \parallel \hat{\mathbf{s}}_i^{(T)} \right), \quad (10)$$

where  $\mathbf{p}_i \in [0, 1]^{C \times 1}$  is the multi-label prediction score vector for the  $i$ -th sample,  $\parallel$  is the concatenation operation.

Then, the multi-label classification loss is computed by the binary cross-entropy loss:

$$\mathcal{L}_{ce} = - \sum_{i=1}^N \sum_{j=1}^C (y_{ij} \log(p_{ij}) + (1 - y_{ij}) \log(1 - p_{ij})). \quad (11)$$

Finally, we train the model by minimizing the overall loss:

$$\mathcal{L} = \mathcal{L}_{ce} + \alpha \mathcal{L}_{lc}, \quad (12)$$

where  $\alpha$  is a hyperparameter that balances the contribution of  $\mathcal{L}_{ce}$  and  $\mathcal{L}_{lc}$ . Algorithm 1 illustrates the whole training process of our proposed HyperAHSF.

## Experiments

### Experimental Settings

We evaluate our proposed HyperAHSF on six widely used MVML datasets: *Emotions*, *Yeast*, *Scene*, *Corel5k*, *Pascal* and *Espgame*, which can be downloaded from Mulan website (<http://mulan.sourceforge.net/datasets-mlc.html>). Then, we employ six state-of-the-art methods for comparative analysis: **FIMAN** (Wu et al. 2020), **D-VSM** (Lyu et al. 2022), **IMvMLC** (Wen et al. 2024), **ML-BVAE** (Fu et al. 2024), **VAMS** (Wei et al. 2025) and **TMvML** (Zhong et al. 2025). The parameter configurations for all comparative methods follow the suggestions from the corresponding literature. Furthermore, six widely-used multi-label metrics are employed: Hamming Loss (HL), Average Precision (AP), One Error (OE), Ranking Loss (RL), Coverage (Cov) and Micro-F1 (Mic), whose formal definitions can be found in (Sun and Zong 2021). Finally, we perform five-fold cross-validation on each dataset to conduct experimental comparisons, which are implemented using the PyTorch framework and executed on a server equipped with a 16 vCPU AMD EPYC 9K84 96-Core Processor, an H20-NVLink GPU and 150 GB of RAM.

### Experimental Results

Table 2 illustrates the comparison between HyperAHSF and the other six methods on six evaluation metrics, where mean results and standard deviations are recorded. According to the 252 statistical comparisons, we can observe that:

- In terms of the six comparative methods, HyperAHSF is superior to **FIMAN**, **TMvML**, **VAMS** in all cases, and also outperforms **ML-BVAE**, **IMvMLC** and **D-VSM** in 97.2% cases, respectively.

Evaluation Metrics	$\tau_F$	Critical Value
Hamming Loss	17.105	2.421 Methods:7, Datasets:6
Average Precision	37.712	
One Error	25.545	
Ranking Loss	17.652	
Coverage	18.333	
Micro-F1	25.361	

Table 1: Friedman statistics  $\tau_F$  of each evaluation metric (at 0.05 significance level).

- In terms of the six evaluation metrics, HyperAHSF achieves the best performance on Average Precision, One Error and Micro-F1, and is also superior to other methods over 97.2% cases on Ranking Loss, Coverage and Hamming Loss, respectively.
- In terms of the six employed datasets, HyperAHSF achieves the best performance in all datasets except *Pascal*. And in the dataset *Pascal*, it is also superior to other comparative methods in 91.7% cases.
- Notably, HyperAHSF significantly outperforms graph-based methods **VAMS** and **D-VSM** in most cases, demonstrating the effectiveness of leveraging hypergraph to mine high-order semantic correlations in MVML data.

To comprehensively evaluate the superiority of our proposed HyperAHSF, the Friedman test (Demšar 2006) is conducted as the statistical test to analyze the relative performance among all comparative methods. As shown in Table 1, the null hypothesis of distinguishable performance among all comparative algorithms is rejected at 0.05 significance level. Furthermore, we employ the post hoc Bonferroni-Dunn test (Demšar 2006) to compare the relative performance among the comparative algorithms. Figure 3 illustrates the CD diagrams on each evaluation metric, where the average rank of each algorithm is marked along the axis. According to Figure 3, it is observed that HyperAHSF consistently ranks 1st across all evaluation metrics.

## Further Analysis

### Ablation Study

To evaluate the effectiveness of each component in our proposed HyperAHSF, we perform an ablation study to compare HyperAHSF with its four degenerated algorithms: HyperAHSF-w/o ConLos, HyperAHSF-w/o LabSel, HyperAHSF-w/o SemSim and HyperAHSF-w/o SemAli. Table 3 records the experimental results on *Emotions*, *Yeast* and *Corel5k* datasets, where the best performance is shown in bold across all evaluation metrics. Specifically, HyperAHSF is significantly superior to the HyperAHSF-w/o SemAli and HyperAHSF-w/o SemSim, indicating that the explicit mining of high-order semantic similarities and alignments in MVML data is crucial to effective fusion of key semantic information, therefore obtaining high-quality representations for multi-label classification. Additionally, compared with ignoring label-driven contrastive loss, model performance drops more significantly when ignoring adaptive label selection, which means it is important to capture more reliable semantic interactions, and contrastive learning exploiting label semantics can further enhance the discriminability of learned representations. In general, HyperAHSF significantly outperforms its four degenerated algorithms, which also demonstrates the superiority of leveraging hypergraph to adaptively model the group-level semantic similarities and alignments and the effectiveness of our designed label-driven contrastive learning.

### Sensitivity Analysis

We conduct sensitivity analysis of our HyperAHSF with respect to its two key parameters: the trade-off coefficient  $\alpha$  of

Datasets	Metrics	FIMAN	D-VSM	IMvMLC	ML-BVAE	VAMS	TMvML	HyperAHSF
Emotions	HL	0.231±0.013	0.179±0.017	0.330±0.021	0.317±0.012	0.193±0.019	0.227±0.015	<b>0.178±0.018</b>
	AP	0.806±0.027	0.835±0.027	0.782±0.021	0.572±0.022	0.826±0.012	0.774±0.023	<b>0.842±0.036</b>
	OE	0.258±0.042	0.214±0.046	0.303±0.032	0.568±0.029	0.235±0.027	0.319±0.046	<b>0.202±0.071</b>
	RL	0.161±0.026	0.137±0.015	0.183±0.019	0.423±0.035	0.144±0.013	0.184±0.019	<b>0.127±0.029</b>
	Cov	7.796±0.189	1.624±0.038	1.873±0.037	3.163±0.242	1.659±0.055	1.914±0.066	<b>1.601±0.166</b>
	Mic	0.671±0.014	0.700±0.034	0.482±0.010	0.107±0.091	0.692±0.015	0.635±0.033	<b>0.706±0.032</b>
Yeast	HL	0.216±0.004	0.196±0.003	0.313±0.007	0.232±0.004	0.204±0.007	0.214±0.004	<b>0.189±0.008</b>
	AP	0.740±0.007	0.760±0.005	0.738±0.006	0.712±0.007	0.763±0.009	0.745±0.007	<b>0.779±0.008</b>
	OE	0.257±0.013	0.215±0.010	0.248±0.009	0.253±0.012	0.222±0.008	0.250±0.010	<b>0.214±0.018</b>
	RL	0.187±0.005	0.172±0.003	0.180±0.005	0.204±0.007	0.168±0.004	0.180±0.006	<b>0.157±0.007</b>
	Cov	6.673±0.074	6.562±0.133	6.538±0.094	6.706±0.094	6.376±0.096	6.509±0.099	<b>6.152±0.151</b>
	Mic	0.111±0.008	0.655±0.008	0.468±0.003	0.479±0.008	0.656±0.012	0.577±0.007	<b>0.679±0.011</b>
Scene	HL	0.195±0.005	0.075±0.005	0.197±0.007	0.179±0.001	0.090±0.001	0.101±0.003	<b>0.074±0.009</b>
	AP	0.827±0.010	0.890±0.011	0.844±0.004	0.717±0.023	0.878±0.014	0.795±0.004	<b>0.897±0.014</b>
	OE	0.280±0.018	0.189±0.016	0.263±0.009	0.464±0.032	0.198±0.014	0.207±0.011	<b>0.176±0.025</b>
	RL	0.107±0.006	0.058±0.007	0.086±0.004	0.171±0.017	0.070±0.010	0.078±0.015	<b>0.056±0.009</b>
	Cov	0.628±0.020	0.372±0.051	0.516±0.024	0.942±0.090	0.434±0.058	0.459±0.070	<b>0.363±0.043</b>
	Mic	0.616±0.008	0.777±0.015	0.308±0.002	0.304±0.095	0.748±0.016	0.713±0.023	<b>0.793±0.027</b>
Core15k	HL	0.018±0.000	0.012±0.000	0.158±0.008	0.013±0.000	0.012±0.008	0.017±0.004	<b>0.011±0.000</b>
	AP	0.430±0.007	0.475±0.008	0.333±0.008	0.286±0.000	0.452±0.009	0.369±0.006	<b>0.540±0.007</b>
	OE	0.489±0.017	0.410±0.018	0.613±0.019	0.626±0.019	0.432±0.016	0.575±0.014	<b>0.354±0.012</b>
	RL	0.085±0.000	0.084±0.004	0.114±0.003	0.188±0.008	0.082±0.004	0.152±0.002	<b>0.065±0.002</b>
	Cov	53.94±0.790	53.21±1.947	70.80±2.256	108.7±4.792	53.20±2.218	84.87±3.437	<b>43.03±0.813</b>
	Mic	0.361±0.009	0.399±0.004	0.029±0.000	0.025±0.001	0.372±0.005	0.414±0.008	<b>0.456±0.012</b>
Pascal	HL	0.116±0.002	<b>0.048±0.000</b>	0.074±0.000	0.062±0.001	0.110±0.013	0.080±0.035	<b>0.048±0.001</b>
	AP	0.721±0.003	0.767±0.003	0.695±0.008	0.659±0.002	0.759±0.008	0.706±0.006	<b>0.781±0.005</b>
	OE	0.313±0.011	0.274±0.008	0.438±0.011	0.426±0.007	0.283±0.006	0.283±0.010	<b>0.262±0.005</b>
	RL	0.118±0.003	0.077±0.001	<b>0.063±0.002</b>	0.106±0.001	0.101±0.002	0.111±0.006	0.069±0.003
	Cov	3.486±0.081	2.342±0.044	<b>1.908±0.065</b>	2.966±0.028	2.231±0.038	2.482±0.022	2.205±0.076
	Mic	0.008±0.002	0.636±0.004	0.136±0.001	0.385±0.031	0.426±0.026	0.394±0.014	<b>0.643±0.008</b>
Espgame	HL	0.028±0.000	0.017±0.000	0.021±0.002	0.017±0.000	0.019±0.001	0.022±0.001	<b>0.016±0.000</b>
	AP	0.284±0.002	0.364±0.003	0.273±0.001	0.257±0.001	0.266±0.002	0.223±0.012	<b>0.412±0.006</b>
	OE	0.628±0.004	0.464±0.010	0.615±0.013	0.615±0.010	0.470±0.011	0.690±0.023	<b>0.437±0.012</b>
	RL	0.154±0.002	0.133±0.003	0.141±0.000	0.211±0.002	0.144±0.002	0.217±0.014	<b>0.114±0.001</b>
	Cov	102.8±1.183	86.75±1.634	92.21±0.113	136.5±1.681	89.91±1.816	133.4±7.478	<b>80.17±0.824</b>
	Mic	0.242±0.002	0.332±0.003	0.034±0.000	0.008±0.002	0.320±0.000	0.022±0.001	<b>0.335±0.005</b>

Table 2: Performance comparisons between HyperAHSF and other comparative methods, where the best performance is shown in bold. For HL, OE, RL and Cov, the lower the value, the better performance. For AP and Mic, the higher the better.

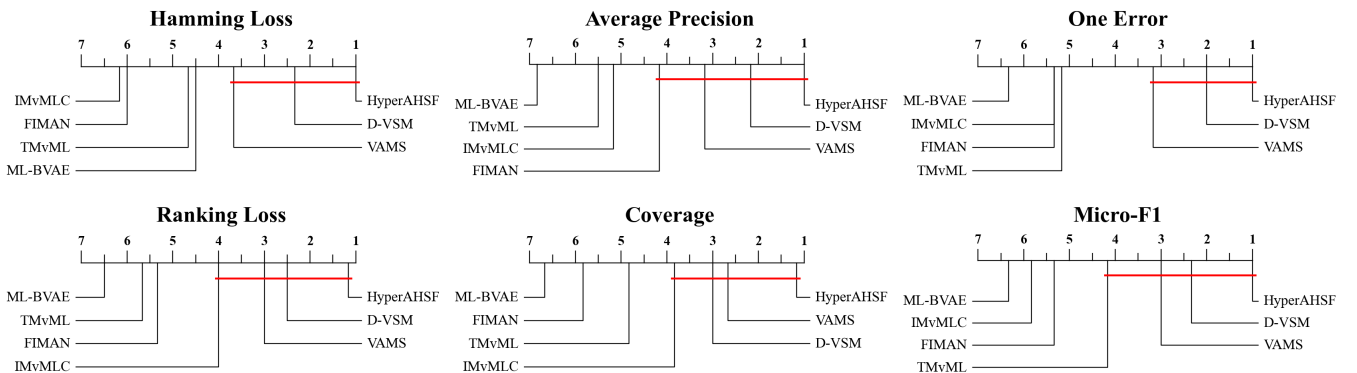


Figure 3: Experimental comparisons between HyperAHSF and all other comparative algorithms with the Bonferroni-Dunn test (CD = 3.213 at 0.05 significance level). Algorithms not connected with HyperAHSF are significantly inferior to HyperAHSF.

Emotions	Hamming Loss	Average Precision	One Error	Ranking Loss	Coverage	Micro-F1
HyperAHSF-w/o ConLos	0.182±0.020	0.839±0.031	0.208±0.064	0.129±0.024	1.613±0.145	0.698±0.038
HyperAHSF-w/o LabSel	0.189±0.032	0.829±0.044	0.222±0.067	0.137±0.028	1.634±0.153	0.693±0.037
HyperAHSF-w/o SemAli	0.188±0.030	0.827±0.041	0.221±0.066	0.141±0.027	1.671±0.149	0.694±0.042
HyperAHSF-w/o SemSim	0.183±0.027	0.828±0.046	0.238±0.071	0.136±0.031	1.627±0.142	0.696±0.044
HyperAHSF	<b>0.178±0.018</b>	<b>0.842±0.036</b>	<b>0.202±0.071</b>	<b>0.127±0.029</b>	<b>1.601±0.166</b>	<b>0.706±0.032</b>
Yeast	Hamming Loss	Average Precision	One Error	Ranking Loss	Coverage	Micro-F1
HyperAHSF-w/o ConLos	0.194±0.006	0.773±0.003	0.221±0.014	0.161±0.005	6.230±0.147	0.670±0.006
HyperAHSF-w/o LabSel	0.194±0.005	0.774±0.006	0.219±0.016	0.159±0.006	6.157±0.167	0.667±0.004
HyperAHSF-w/o SemAli	0.198±0.007	0.768±0.010	0.225±0.021	0.164±0.008	6.293±0.135	0.662±0.012
HyperAHSF-w/o SemSim	0.195±0.006	0.775±0.009	0.227±0.022	0.162±0.007	6.284±0.163	0.678±0.006
HyperAHSF	<b>0.189±0.008</b>	<b>0.779±0.008</b>	<b>0.214±0.018</b>	<b>0.157±0.007</b>	<b>6.152±0.151</b>	<b>0.679±0.011</b>
Core5k	Hamming Loss	Average Precision	One Error	Ranking Loss	Coverage	Micro-F1
HyperAHSF-w/o ConLos	<b>0.011±0.000</b>	0.536±0.007	0.361±0.014	0.067±0.002	43.77±0.799	0.453±0.011
HyperAHSF-w/o LabSel	0.012±0.002	0.486±0.009	0.420±0.017	0.067±0.003	43.24±0.932	0.320±0.014
HyperAHSF-w/o SemAli	<b>0.011±0.000</b>	0.538±0.006	0.364±0.008	0.066±0.002	43.55±0.523	0.451±0.010
HyperAHSF-w/o SemSim	0.017±0.014	0.335±0.027	0.570±0.046	0.193±0.036	105.1±2.566	0.074±0.029
HyperAHSF	<b>0.011±0.000</b>	<b>0.540±0.007</b>	<b>0.354±0.012</b>	<b>0.065±0.002</b>	<b>43.03±0.813</b>	<b>0.456±0.012</b>

Table 3: The ablation study of HyperAHSF on *Emotions*, *Yeast* and *Core5k* datasets, where HyperAHSF-w/o ConLos, HyperAHSF-w/o LabSel, HyperAHSF-w/o SemSim and HyperAHSF-w/o SemAli ignore the label-driven contrastive loss, the adaptive label selection, the group-level semantic similarities, and the group-level semantic alignments, respectively.

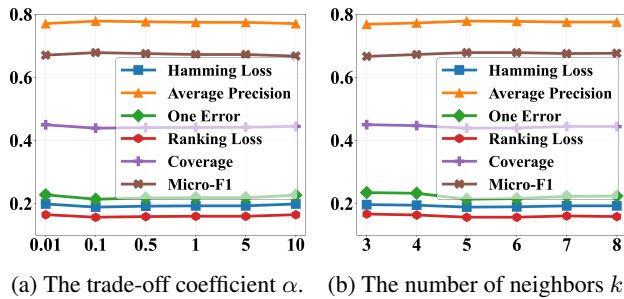


Figure 4: The sensitivity analysis of our proposed HyperAHSF for parameters  $\alpha$  and  $k$ , where the Coverage metric is normalized by the number of class to make all metrics be presented in the unified figure (Figure 5 in the same way).

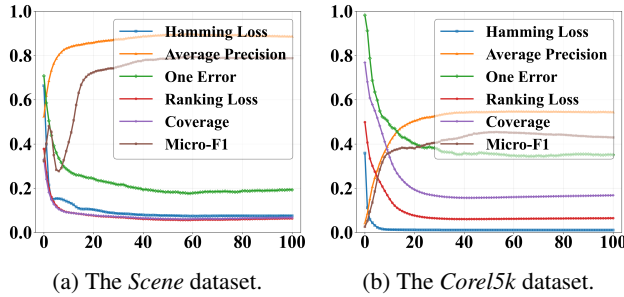


Figure 5: The convergence analysis of our proposed HyperAHSF on *Scene* and *Core5k* datasets.

the label-driven contrastive loss and the number of semantically similar neighbors  $k$ . Figure 4 shows the performance of HyperAHSF under different parameter configuration of  $\alpha$

and  $k$  on *Yeast* dataset. According to Figure 4, we can observe that the model performance first improves gradually, and then stabilizes with a slight downward tendency as  $\alpha$  or  $k$  increases. In our experiments, the optimal values of  $\alpha$  and  $k$  tend to be set to 0.1 and 5, respectively, but vary with appropriate adjustments in different datasets.

### Convergence Analysis

We conduct convergence analysis of our proposed HyperAHSF on *Scene* and *Core5k* datasets, where the experimental results are illustrated in Figure 5. As shown in Figure 5, we can easily find that the performance of HyperAHSF first rises sharply and then immediately becomes stable as the number of epochs increases, which empirically demonstrates the convergence of our proposed HyperAHSF.

### Conclusion

In this paper, we propose a novel hypergraph-based MVML method with Adaptive High-Order Semantic Fusion named HyperAHSF, which leverages hypergraphs to adaptively model group-level semantic similarities within each view and group-level semantic alignments across different views, enabling more effective feature fusion. Compared with existing MVML methods, HyperAHSF can directly express high-order semantic interactions and further exploit label-driven contrastive learning to enhance the discriminability of learned representations, significantly improving the model's performance of multi-label classification.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62306020), the National College

Students' Innovation and Entrepreneurship Training Program of BJUT (No. GJDC2025-01-32), the Beijing Natural Science Foundation (No. QY25368), and the National Key Research and Development Program of China (No. 2023YFB3107100).

## References

- Bai, S.; Zhang, F.; and Torr, P. H. 2021. Hypergraph Convolution and Hypergraph Attention. *Pattern Recognition*, 110: 107637.
- Cui, J.; Xie, Y.; Liu, C.; Huang, Q.; Li, M.; and Wen, J. 2024. Deep dual Incomplete Multi-View Multi-Label Classification Via Label Semantic-Guided Contrastive Learning. *Neural Networks*, 180: 106674.
- Demšar, J. 2006. Statistical Comparisons of Classifiers Over Multiple Data Sets. *The Journal of Machine Learning Research*, 7(1): 1–30.
- Feng, Y.; You, H.; Zhang, Z.; Ji, R.; and Gao, Y. 2019. Hypergraph Neural Networks. In *AAAI Conference on Artificial Intelligence*, 3558–3565.
- Fu, K.; Du, C.; Wang, S.; and He, H. 2024. Multi-View Multi-Label Fine-Grained Emotion Decoding From Human Brain Activity. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7): 9026–9040.
- Fu, L.; Deng, B.; Huang, S.; Liao, T.; Zhang, C.; and Chen, C. 2025. Learn from Global Rather Than Local: Consistent Context-Aware Representation Learning for Multi-View Graph Clustering. In *International Joint Conference on Artificial Intelligence*, 5145–5153.
- Gao, Y.; Feng, Y.; Ji, S.; and Ji, R. 2023. HGNN+: General Hypergraph Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 3181–3199.
- Han, X.; Zhou, H.; Tian, Z.; Du, S.; and Gao, Y. 2025. Inter-Intra Hypergraph Computation for Survival Prediction on Whole Slide Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7): 6006–6021.
- Hao, P.; Zhang, H.; and Zhang, Y. 2025. Tensor-based Opposing yet Complementary Learning for Multi-view Multi-label Feature Selection. In *ACM International Conference on Multimedia*, 1822–1831.
- Huang, C.; Gao, C.; Li, M.; Li, Y.; Wang, X.; Jiang, Y.; and Huang, X. 2025a. Correlation Information Enhanced Graph Anomaly Detection via Hypergraph Transformation. *IEEE Transactions on Cybernetics*, 55(6): 2865–2878.
- Huang, S.; Fu, L.; Zhuang, S.; Qiu, Y.; Huang, B.; Cui, Z.; and Zhang, T. 2025b. Going Beyond Consistency: Target-oriented Multi-view Graph Neural Network. In *International Joint Conference on Artificial Intelligence*, 5426–5434.
- Huang, Z.; Li, T.; Wang, X.; Yang, K.; Deng, C.; Feng, J.; and Li, Y. 2025c. Predicting Mobile App Usage With Context-Aware Dynamic Hypergraphs. *IEEE Transactions on Mobile Computing*, 24(6): 5511–5524.
- Jiang, B.; Liu, J.; Wang, Z.; Zhang, C.; Yang, J.; Wang, Y.; Sheng, W.; and Ding, W. 2025. Semi-supervised multi-view feature selection with adaptive similarity fusion and learning. *Pattern Recognition*, 159: 111159.
- Li, Q.; Luo, T.; Jiang, M.; Jiang, Z.; Hou, C.; and Li, F. 2025a. Semi-Supervised Multi-View Multi-Label Learning with View-Specific Transformer and Enhanced Pseudo-Label. In *AAAI Conference on Artificial Intelligence*, 18430–18438.
- Li, Q.; Luo, T.; Jiang, M.; Liao, J.; and Jiang, Z. 2024. Deep Incomplete Multi-View Network Semi-Supervised Multi-Label Learning with Unbiased Loss. In *ACM International Conference on Multimedia*, 9048–9056.
- Li, Y.; Hu, X.; Li, P.; Wang, L.; and You, Z. 2025b. Hypergraph Representation Learning for Identifying CircRNA-Disease Associations. *Pattern Recognition*, 168: 111835.
- Li, Y.; Zhang, J.; Wu, H.; Du, G.; and Long, J. 2025c. Consistent and Specific Multi-View Multi-Label Learning with Correlation Information. *Information Sciences*, 687: 121395.
- Liu, C.; Wen, J.; Xu, Y.; Zhang, B.; Nie, L.; and Zhang, M. 2025. Reliable Representation Learning for Incomplete Multi-View Missing Multi-Label Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6): 4940–4956.
- Liu, W.; Yuan, J.; Lyu, G.; and Feng, S. 2022. Label driven latent subspace learning for multi-view multi-label classification. *Applied Intelligence*, 53(4): 3850–3863.
- Lyu, G.; Deng, X.; Wu, Y.; and Feng, S. 2022. Beyond Shared Subspace: A View-Specific Fusion for Multi-View Multi-Label Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7647–7654.
- Ou, S.; Xue, Z.; Qin, L.; Li, Y.; Liang, M.; Wu, J.; Zhang, X.; Beheshti, A.; and Qi, Y. 2025. Incomplete Multi-View Multi-Label Classification via Diffusion-Guided Redundancy Removal. In *AAAI Conference on Artificial Intelligence*, 19758–19766.
- Ruan, C.; Yang, L.; and Li, H. 2022. Graph Convolution Network Based Representation for Multi-View Multi-Label Learning. In *IEEE International Conference on Multimedia and Expo*, 1–6.
- Sun, S.; and Zong, D. 2021. LCBM: A Multi-View Probabilistic Model for Multi-Label Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8): 2682–2696.
- Tang, Y.; Ding, J.; Chen, Y.; Gao, Y.; Jiang, A.; and Wang, C. 2025. Anxiety Disorder Identification with Biomarker Detection through Subspace-Enhanced Hypergraph Neural Network. *Neural Networks*, 187: 107293.
- Wang, J.; Feng, S.; Lyu, G.; and Yuan, J. 2024. SURER: Structure-Adaptive Unified Graph Neural Network for Multi-View Clustering. In *AAAI Conference on Artificial Intelligence*, 15520–15527.
- Wang, Y.; Li, Q.; Chang, D.; Wen, J.; Xiao, F.; and Zhao, Y. 2025. A Category-Driven Contrastive Recovery Network for Double Incomplete Multi-View Multi-Label Classification. *IEEE Transactions on Multimedia*, 27: 4008–4017.
- Wei, H.; Deng, Y.; Hai, Q.; Lin, Y.; Yang, Z.; and Lyu, G. 2025. Multi-View Multi-Label Classification via View-Label Matching Selection. In *AAAI Conference on Artificial Intelligence*, 21456–21464.

- Wen, J.; Liu, C.; Deng, S.; Liu, Y.; Fei, L.; Yan, K.; and Xu, Y. 2024. Deep Double Incomplete Multi-View Multi-Label Learning With Incomplete Labels and Missing Views. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8): 11396–11408.
- Wu, J.-H.; Wu, X.; Chen, Q.; Hu, Y.; and Zhang, M.-L. 2020. Feature-Induced Manifold Disambiguation for Multi-View Partial Multi-label Learning. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 557–565.
- Wu, X.; Chen, Q.; Hu, Y.; Wang, D.-B.; Chang, X.; Wang, X.; and Zhang, M.-L. 2019. Multi-View Multi-Label Learning with View-Specific Information Extraction. In *International Joint Conference on Artificial Intelligence*, 3884–3890.
- Xu, C.; Li, M.; Ni, Z.; Zhang, Y.; and Chen, S. 2022. GroupNet: Multiscale Hypergraph Neural Networks for Trajectory Prediction with Relational Reasoning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6488–6497.
- Yan, Y.; Qin, J.; Chen, J.; Liu, L.; Zhu, F.; Tai, Y.; and Shao, L. 2020. Learning Multi-Granular Hypergraphs for Video-Based Person Re-Identification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2896–2905.
- Zhong, Q.; Lyu, G.; and Yang, Z. 2025. Align While Fusion: A Generalized Nonaligned Multiview Multilabel Classification Method. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4): 7627–7636.
- Zhong, Q.; Shan, Y.; Wang, H.; Yang, Z.; and Lyu, G. 2025. Tensorized Multi-View Multi-Label Classification via Laplace Tensor Rank. In *International Conference on Machine Learning*.
- Zou, J.; Chen, Y.; Jiang, B.; Zhou, P.; Du, L.; Duan, L.; and Qian, Y. 2025. Robust Tensor Learning with Graph Diffusion for Scalable Multi-view Graph Clustering. In *ACM International Conference on Multimedia*, 2207–2215.