INCOMPLETE MULTI-VIEW PARTIAL-LABEL FEATURE SELECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Partial-label feature selection can simultaneously disambiguate candidate labels and select discriminative features when each training instance is associated with multiple candidate labels, and it has proved effective for conventional single-view partial-label problems. However, data in many real-world applications are naturally described by multiple views and may contain missing views, rendering existing methods inadequate. To address this issue, we propose IMPLFS, an incomplete multi-view partial-label feature selection method based on complementary learning. Three key issues need to be addressed: how to learn reliable sample similarities from incomplete views, how to fully exploit complementary information across views, and how to reduce the adverse effect of candidate-label ambiguity on feature evaluation. First, IMPLFS develops a complementary-learning-guided similarity-matrix model that uses observed within-view relations and cross-view complementary relations to complete and enhance the sample-similarity structure of each view. Second, it dynamically integrates the complementary information supplied by different views into a fused graph through graph learning and adaptive view weighting. Finally, it constructs a joint optimization framework for label-confidence estimation and feature selection, allowing label disambiguation and feature evaluation to be updated collaboratively. Experimental results on multiple benchmark datasets demonstrate that IMPLFS outperforms competing feature selection methods and can effectively address the incomplete multi-view partial-label feature selection problem.

1 INTRODUCTION

Partial-label learning (PLL) Tian et al. (2023) is a representative weakly supervised learning framework Ren et al. (2023) in which each training instance is associated with a set of candidate labels, only one of which is the ground-truth label. Because annotating a candidate-label set is often easier than identifying the ground-truth label directly, PLL is more practical than conventional supervised classification in many applications, including face recognition Guillaumin et al. (2010), natural language processing Wisniewski et al. (2014), and EEG signal analysis Zhang & Etemad (2025).

Most existing approaches solve PLL problems by resolving candidate-label ambiguity in the label space Tian et al. (2023); Wang et al. (2025). Recent studies have shown that exploiting useful information in the feature space can also improve disambiguation Liu et al. (2026a); Fan et al. (2025). Feature selection Guyon & Elisseeff (2003) seeks a discriminative, relevant, and stable subset of the original high-dimensional features, preserving task-relevant information while removing redundant and noisy features. It therefore also constitutes a means of extracting useful information from the feature space. Motivated by this connection, Bao et al. (2022) were the first to investigate partial-label feature selection. They proposed a method based on mutual-information dependency maximization and submodular optimization, thereby improving the generalization performance of existing PLL algorithms. Subsequent work has investigated feature selection under ambiguous supervision from several perspectives Qian et al. (2023); Ding et al. (2024); Zhang et al. (2024); Li et al. (2025).

With the rapid development of information technology and increasingly diverse data-acquisition mechanisms, multi-view data have become ubiquitous Yu et al. (2025). Partial-label data also frequently exhibit multi-view characteristics Zhao et al. (2022); Hai et al. (2025). In practical ap-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

plications, however, sensor failures and related causes often result in missing views Tang et al. (2024). Conventional partial-label feature selection techniques primarily target single-view data and are therefore ill suited to partial-label problems involving multiple, incomplete views Bao et al. (2022); Ding et al. (2024); Zhang et al. (2024). Missing views make sample similarities unreliable and leave cross-view cues underexploited, while ambiguous candidate labels can mislead feature ranking. Effective incomplete multi-view partial-label feature selection must therefore recover reliable sample structure, identify a confident label within each candidate set, and select sparse discriminative features in a unified manner.

To meet this goal, we present IMPLFS, the first incomplete multi-view partial-label feature selection method based on complementary learning. IMPLFS addresses the coupled difficulties as follows. First, a complementary-learning-guided similarity-matrix model uses observed within-view relations and cross-view complementary relations to complete and enhance the sample-similarity structure of each view. Second, fused-graph learning and adaptive view weighting dynamically integrate complementary information from different views. Third, a joint optimization framework for label-confidence estimation and feature selection collaboratively updates partial-label disambiguation and feature evaluation. The main contributions are summarized below.

- We investigate incomplete multi-view partial-label feature selection for the first time and propose IMPLFS, a complementary-learning-based method for this setting.
- We propose a unified framework for incomplete multi-view partial-label feature selection. It learns reliable multi-view sample structures through complementary relation modeling and instance-wise adaptive fusion, and jointly optimizes the learned structures, label confidences, and feature weights, thereby coupling multi-view structure learning, candidate-label disambiguation, and discriminative feature selection.
- Experiments on four benchmarks demonstrate consistent gains across four partial-label classifiers and robustness to varying missing-view and feature-selection ratios.

2 Related Work

Partial-label learning. PLL seeks the ground-truth label from each instance's candidate set, with candidate-label disambiguation as its central challenge Tian et al. (2023). Label-confidence estimation was introduced into partial-label feature selection through mutual-information dependence maximization Bao et al. (2022). Subsequent methods coupled feature weighting with feature dependency and label consistency Qian et al. (2023), neighborhood mutual information Ding et al. (2024), adaptive instance relations Zhang et al. (2024), or fuzzy-neighborhood disambiguation Li et al. (2025). Multi-view PLL methods have explored large-margin learning Zhao et al. (2022), prototype-based disambiguation Sun et al. (2023), and contrastive fusion Hai et al. (2025). Recent work has combined label disambiguation, shared-subspace learning, and feature weighting for multi-view partial multi-label feature selection Han et al. (2026). However, existing PLFS methods are mainly single-view, multi-view PLL methods assume complete views and target prediction, and partial multi-label learning permits multiple valid labels. These settings differ from feature selection with missing views and exactly one true label in each candidate set.

Incomplete multi-view learning. Incomplete multi-view learning extracts reliable representations and cross-view structure when instances lack one or more views Tang et al. (2024). Following early partial multi-view clustering Li et al. (2014), representative approaches have used matrix factorization Liu & Song (2022), kernel tensors Wu et al. (2024), imputation-free graphs and cross-view graph matching Bai et al. (2024); Yang et al. (2023), and deep representations Zhang et al. (2022). COMPLETER and dual contrastive prediction further connected missing-view recovery with cross-view consistency Lin et al. (2021; 2023). Recent work considers insufficient label information Jiang et al. (2024b) or incomplete-view partial multi-label classification Liu et al. (2024); Wen et al. (2026). These methods address clustering, representation learning, or multi-label classification; they are unsupervised or involve missing labels or multiple valid labels rather than a single true label hidden in a candidate set, and feature selection is not their objective.

Multi-view feature selection. Multi-view feature selection exploits consistent and complementary information to remove high-dimensional redundancy. Existing methods model cross-view lo-

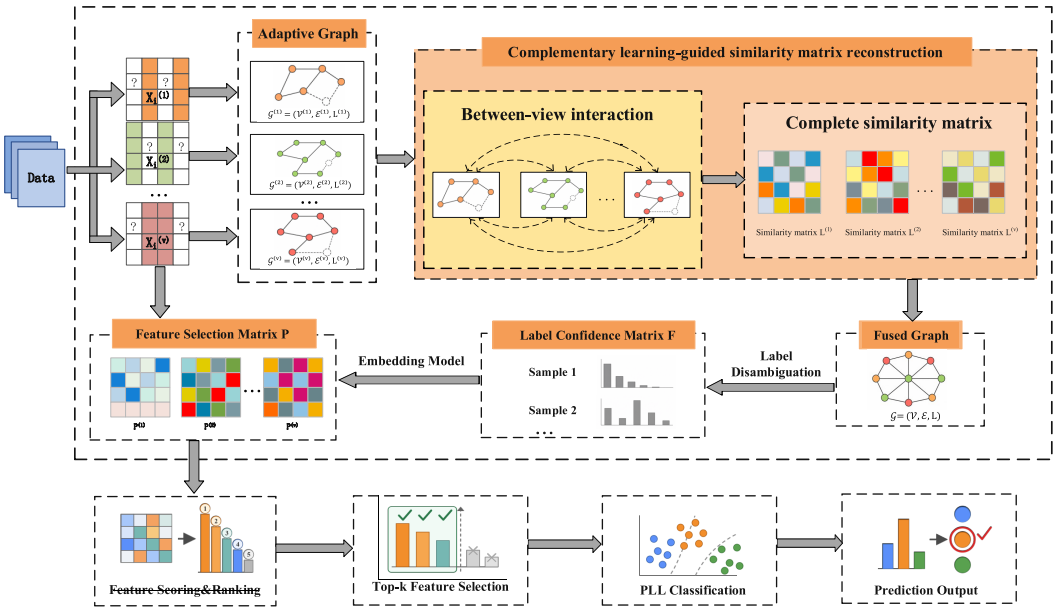


Figure 1: The overall framework of IMPLFS.

cality, diversity, and consensus Tang et al. (2022), complementary features and multi-order similarities Cao & Xie (2024), or adaptive graphs Jiang et al. (2024a). Multi-label extensions use global-view reconstruction Hao et al. (2025b), embedded feature fusion Hao et al. (2025a), and multi-head attention Liu et al. (2026b) to relate views, features, and labels. For missing views, dynamic incomplete-view feature selection Huang et al. (2023) and TIME-FS Huang et al. (2025) jointly consider representation recovery and feature selection, but remain unsupervised and do not resolve candidate-label ambiguity. Thus, prior studies separately advance partial-label feature selection, multi-view PLL, and incomplete multi-view feature selection, while their supervision assumptions, view completeness, and objectives differ. Research on multi-view feature selection under simultaneous view missingness and candidate-label ambiguity remains limited.

3 The IMPLFS Feature Selection Algorithm

3.1 Definitions

Let $\mathcal{X} = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \cdots \times \mathbb{R}^{d_V}$ denote a feature space comprising V views, where $d_v, 1 \le v \le V$, is the dimensionality of the vth view, and let $\mathcal{Y} = \{c_1, c_2, \ldots, c_m\}$ denote the label space containing m class labels. The training set is $\mathcal{D} = \{(\{x_i^{(v)}\}_{v=1}^{V}, C_i) \mid 1 \le i \le n\}$, with view availability indicated by $w_i^{(v)} \in \{0, 1\}$, where n is the number of instances and $C_i \subseteq \mathcal{Y}$ is the candidate-label set of the ith instance. The feature vector of this instance in view v is $\mathbf{x}_i^{(v)} = [x_{i1}^{(v)}, x_{i2}^{(v)}, \ldots, x_{id_v}^{(v)}]^\top \in \mathbb{R}^{d_v}$. We write $X^{(v)} = [\mathbf{x}_1^{(v)}, \mathbf{x}_2^{(v)}, \ldots, \mathbf{x}_n^{(v)}]^\top \in \mathbb{R}^{n \times d_v}$ and $d = \sum_{v=1}^{V} d_v$. Let $Y = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^\top \in \mathbb{R}^{n \times m}$, where $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{im}]^\top \in \{0, 1\}^m$. Here, $y_{ik} = 1$ indicates $c_k \in \mathcal{C}_i$, whereas $y_{ik} = 0$ indicates $c_k \notin \mathcal{C}_i$. Let $F = [f_{ik}] \in \mathbb{R}^{n \times m}$ denote the label-confidence matrix, where f_{ik} is the confidence that c_k is the ground-truth label of $\{\mathbf{x}_i^{(v)}\}_{v=1}^{V}$. The row vector $\mathbf{f}_i = [f_{i1}, f_{i2}, \ldots, f_{im}]$ satisfies $\sum_{k=1}^{m} f_{ik} = 1$ and $0 \le f_{ik} \le y_{ik}$. It can be initialized by setting $f_{ik} = 1/|\mathcal{C}_i|$ if $c_k \in \mathcal{C}_i$, and $f_{ik} = 0$ otherwise. The feature-selection matrix for the vth view is denoted by $P^{(v)} \in \mathbb{R}^{d_v \times m}$, and $\mathbf{p}_j^{(v)} \in \mathbb{R}^m$ denotes its jth row, namely the weights of the jth feature in that view across all labels. Each view is represented by an affinity graph with similarity matrix $L^{(v)} = [l_{ij}^{(v)}] \in \mathbb{R}^{n \times n}$, and $L = [l_{ij}] \in \mathbb{R}^{n \times n}$ denotes the fused similarity matrix. We explicitly write the transposed rows as the column vectors $\boldsymbol{\ell}_i^{(v)} = [l_{i1}^{(v)}, l_{i2}^{(v)}, \ldots, l_{in}^{(v)}]^\top \in \mathbb{R}^n$ and $\boldsymbol{\ell}_i = [l_{i1}, l_{i2}, \ldots, l_{in}]^\top \in \mathbb{R}^n$. Let

$W^{(v)} = \mathrm{diag}(w_1^{(v)}, \ldots, w_n^{(v)})$, where $w_i^{(v)} = 1$ if instance i is observed in view v, and $w_i^{(v)} = 0$ otherwise.

3.2 Overall Framework

Figure 1 illustrates the overall framework of IMPLFS. It first completes and enhances view-specific similarity structures through complementary learning and then fuses them using sample-specific view weights. The fused graph guides label-confidence estimation, and the estimated confidences supervise feature selection. These components are coupled in the unified objective and detailed below.

3.2.1 Completing and Enhancing View-Specific Similarities

To mitigate unreliable similarities caused by missing observations and noise, we learn each view-specific graph from both observable within-view relations and complementary structures supplied by the other views:

$$\begin{aligned} \min_{L^{(v)},S} \quad & \sum_{v=1}^{V}\sum_{i,j=1}^{n} \|\mathbf{x}_i^{(v)} - \mathbf{x}_j^{(v)}\|_2^2 w_i^{(v)} w_j^{(v)} l_{ij}^{(v)} + \delta \sum_{v=1}^{V} \|L^{(v)}\|_F^2 + \varphi \sum_{v=1}^{V} \left\| L^{(v)} - \sum_{t\neq v} L^{(t)} S_{tv} \right\|_F^2 + \eta \|S\|_F^2, \\ \text{s.t.} \quad & \sum_{j=1}^{n} l_{ij}^{(v)} = 1,\ l_{ii}^{(v)} = 0,\ l_{ij}^{(v)} \geq 0; \qquad \sum_{t\neq v} S_{tv} = 1,\ S_{vv} = 0,\ S_{tv} \geq 0. \end{aligned} \tag{1}$$

Here, $S \in \mathbb{R}^{V\times V}$ contains the complementary reconstruction coefficients, where S_{tv} measures the contribution of view t to reconstructing view v. The parameters δ, φ, and η control similarity regularization, cross-view reconstruction, and coefficient regularization, respectively. Thus, Eq. (1) combines observable within-view relations with complementary cross-view structures.

3.2.2 Adaptive Fused-Graph Construction

Based on the enhanced view-specific graphs, we construct L using sample-specific adaptive view weights:

$$\begin{aligned} \min_{L,H} \quad & \sum_{i=1}^{n} \left\| \boldsymbol{\ell}_i - \sum_{v=1}^{V} h_i^{(v)} \boldsymbol{\ell}_i^{(v)} \right\|_2^2 \\ \text{s.t.} \quad & \sum_{j=1}^{n} l_{ij} = 1,\ l_{ii} = 0,\ l_{ij} \geq 0;\ \sum_{v=1}^{V} h_i^{(v)} = 1,\ h_i^{(v)} \geq 0. \end{aligned} \tag{2}$$

Let $\boldsymbol{h}_i = (h_i^{(1)}, \ldots, h_i^{(V)})^\top$ and $H = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n)$. Here, $h_i^{(v)}$ measures the contribution of view v to instance i, enabling L to integrate view-specific structures at the instance level.

3.2.3 Label Disambiguation

To resolve candidate-label ambiguity, we estimate F under graph smoothness by encouraging neighboring samples in L to have similar label confidences:

$$\begin{aligned} \min_{F} \quad & \mathrm{Tr}\left(F^\top L_G F\right) \\ \text{s.t.} \quad & \sum_{k=1}^{m} f_{ik} = 1, \quad 0 \leq f_{ik} \leq y_{ik}. \end{aligned} \tag{3}$$

Here, $L_G = D_L - \frac{1}{2}(L + L^\top)$ is the graph Laplacian of L, and the ith diagonal element of D_L is $\sum_j (l_{ij} + l_{ji})/2$.

3.2.4 FEATURE-SELECTION MATRIX LEARNING

Our embedded model selects features by using a feature-weight matrix to reconstruct the label-confidence matrix. Let $\boldsymbol{\alpha} = [\alpha^{(1)}, \ldots, \alpha^{(V)}]^\top$ collect the view weights:

$$\begin{aligned} \min_{P^{(v)}, \boldsymbol{\alpha}} \ & \sum_{v=1}^{V} (\alpha^{(v)})^r \left[\left\| (W^{(v)})^\top (X^{(v)} P^{(v)} - F) \right\|_F^2 + \gamma \left\| P^{(v)} \right\|_{2,1} \right] \\ & \text{s.t.} \quad \sum_{v=1}^{V} \alpha^{(v)} = 1, \quad 0 \le \alpha^{(v)} \le 1. \end{aligned} \tag{4}$$

The $\ell_{2,1}$-norm induces joint row sparsity in $P^{(v)}$, while γ controls its strength and $\alpha^{(v)}$ accounts for unequal view contributions.

3.2.5 UNIFIED OBJECTIVE

Collecting all optimization variables into $\mathcal{Q} = (L, F, H, S, \boldsymbol{\alpha}, (L^{(v)}, P^{(v)})_{v=1}^V)$,

$$\begin{aligned} \min_{\mathcal{Q}} \quad & \sum_{v=1}^{V} \sum_{i,j=1}^{n} \|\mathbf{x}_i^{(v)} - \mathbf{x}_j^{(v)}\|_2^2 w_i^{(v)} w_j^{(v)} l_{ij}^{(v)} + \delta \sum_{v=1}^{V} \|L^{(v)}\|_F^2 \\ & + \varphi \sum_{v=1}^{V} \left\| L^{(v)} - \sum_{t \neq v} L^{(t)} S_{tv} \right\|_F^2 + \eta \|S\|_F^2 + \beta \sum_{i=1}^{n} \left\| \boldsymbol{\ell}_i - \sum_{v=1}^{V} h_i^{(v)} \boldsymbol{\ell}_i^{(v)} \right\|_2^2 \\ & + \mu \operatorname{Tr}(F^\top L_G F) + \theta \sum_{v=1}^{V} (\alpha^{(v)})^r \left[\left\| (W^{(v)})^\top (X^{(v)} P^{(v)} - F) \right\|_F^2 + \gamma \|P^{(v)}\|_{2,1} \right]. \end{aligned} \tag{5}$$

The unified problem is subject to the constraints in Eqs. (1)–(4). The positive parameters balance the corresponding objective terms, and $r > 1$ is the view-weight exponent.

3.3 MODEL OPTIMIZATION

We solve Eq. (5) by block coordinate descent and present the representative updates of $L^{(v)}$, S, and $P^{(v)}$. The remaining updates, detailed derivations, and complete algorithm are provided in Appendix A. Let $\Delta_d = \{\mathbf{z} \in \mathbb{R}^d \mid \mathbf{z} \ge 0,\ \mathbf{1}^\top \mathbf{z} = 1\}$, $\Delta_n^{(i)} = \{\mathbf{z} \in \Delta_n \mid z_i = 0\}$, and let $\Pi_{\mathcal{C}}$ denote Euclidean projection onto $\mathcal{C}$.

Update $L^{(v)}$. After removing constants, the row-wise subproblem is

$$\begin{aligned} & \min_{\boldsymbol{\ell}_i^{(v)} \in \Delta_n^{(i)}} \quad A_i^{(v)} \|\boldsymbol{\ell}_i^{(v)}\|_2^2 + (\mathbf{B}_i^{(v)})^\top \boldsymbol{\ell}_i^{(v)}, \\ & A_i^{(v)} = \delta + \varphi + \varphi \sum_{k \neq v} S_{vk}^2 + \beta (h_i^{(v)})^2, \\ & \mathbf{d}_i^{(v)} = \left(\|\mathbf{x}_i^{(v)} - \mathbf{x}_j^{(v)}\|_2^2 w_i^{(v)} w_j^{(v)} \right)_{j=1}^n, \\ & \mathbf{B}_i^{(v)} = \mathbf{d}_i^{(v)} - 2\varphi \sum_{t \neq v} \boldsymbol{\ell}_i^{(t)} S_{tv} - 2\beta h_i^{(v)} \left(\boldsymbol{\ell}_i - \sum_{t \neq v} h_i^{(t)} \boldsymbol{\ell}_i^{(t)} \right) - 2\varphi \sum_{k \neq v} S_{vk} \left(\boldsymbol{\ell}_i^{(k)} - \sum_{\substack{t \neq k \\ t \neq v}} \boldsymbol{\ell}_i^{(t)} S_{tk} \right). \end{aligned} \tag{6}$$

KKT conditions give the projection update

$$\boldsymbol{\ell}_i^{(v)} \leftarrow \Pi_{\Delta_n^{(i)}} \left(-\frac{\mathbf{B}_i^{(v)}}{2A_i^{(v)}} \right). \tag{7}$$

The fused graph L has the same row-wise simplex-constrained quadratic structure and is updated by the same KKT-based projection method.

Update S. The subproblem is

$$\begin{aligned} \min_{S} \quad & \varphi \sum_{v=1}^{V} \left\| L^{(v)} - \sum_{t \neq v} L^{(t)} S_{tv} \right\|_F^2 + \eta \|S\|_F^2, \\ \text{s.t.} \quad & \sum_{t \neq v} S_{tv} = 1, \qquad S_{vv} = 0, \qquad S_{tv} \geq 0. \end{aligned} \tag{8}$$

Let $\mathbf{m}_v = \mathrm{vec}(L^{(v)})$, and let $\mathcal{M}_{-v}$ contain the vectorized graphs of the other views. For $\mathbf{s} = S_{-v,v}$, the vectorized reconstruction is $\mathrm{vec}(\sum_{t \neq v} L^{(t)} S_{tv}) = \mathcal{M}_{-v}\mathbf{s}$. Expanding the squared norm and removing constants gives a strictly convex simplex-constrained QP. Thus, each column is updated by

$$S_{-v,v} \leftarrow \arg\min_{\mathbf{s} \in \Delta_{V-1}} \quad \mathbf{s}^\top \left(\varphi \mathcal{M}_{-v}^\top \mathcal{M}_{-v} + \eta I \right) \mathbf{s} - 2\varphi \mathbf{m}_v^\top \mathcal{M}_{-v}\mathbf{s}, \qquad S_{vv} = 0. \tag{9}$$

The update of H is a simplex-constrained convex QP for each sample and uses the same optimization method as S.

The label-confidence matrix F is updated by solving a convex QP that combines graph smoothness with view-weighted reconstruction, subject to the constraints in Eq. (3).

Update $P^{(v)}$. The subproblem is

$$\min_{P^{(v)}} \quad \left\| (W^{(v)})^\top (X^{(v)} P^{(v)} - F) \right\|_F^2 + \gamma \|P^{(v)}\|_{2,1}. \tag{10}$$

IRLS introduces the diagonal reweighting matrix $E^{(v,k)}$, with $\varepsilon = 10^{-3}$ to prevent division by zero. Setting the derivative of the resulting quadratic problem to zero gives

$$\begin{aligned} & e_{jj}^{(v,k)} = \frac{1}{2\|\mathbf{p}_j^{(v,k)}\|_2 + \varepsilon}, \qquad E^{(v,k)} = \mathrm{diag}(e_{11}^{(v,k)}, \ldots, e_{d_v d_v}^{(v,k)}), \qquad C^{(v)} = (X^{(v)})^\top W^{(v)} (W^{(v)})^\top, \\ & P^{(v,k+1)} \leftarrow \left[C^{(v)} X^{(v)} + \gamma E^{(v,k)} \right]^{-1} C^{(v)} F. \end{aligned} \tag{11}$$

All variables are alternately updated until convergence; features in each view are then ranked by $\|\mathbf{p}_j^{(v)}\|_2$.

4 Experiments

We evaluate IMPLFS in terms of predictive performance, statistical significance, robustness to missing views and feature-selection ratios, parameter sensitivity, and component effectiveness.

4.1 Experimental Setup

4.1.1 Datasets

We evaluated IMPLFS on four public multi-view partial-label benchmarks: Lostnew, MSRC2new, Yahoo!News50, and Yahoo!News100. Each instance is represented by four views constructed from PCA, LBP, HOG, and ResNet features; the dataset statistics are reported in Table 1. To simulate incomplete views, we randomly masked a prescribed proportion of instances in each view, with missing rates ranging from 0.1 to 0.5. Detailed experimental configurations are left to Appendix B of the supplementary material.

Table 1: Statistics of the experimental datasets.

Dataset	Views	Feature Dim.	Samples	Classes
Lostnew	PCA/LBP/HOG/ResNet	92/59/200/512	1122	16
MSRC2new	PCA/LBP/HOG/ResNet	21/59/200/512	1482	23
Yahoo!News100	PCA/LBP/HOG/ResNet	412/59/200/512	3793	21
Yahoo!News50	PCA/LBP/HOG/ResNet	469/59/200/512	5765	65

4.1.2 COMPARISON METHODS

Following the benchmark protocol of SAUTE Bao et al. (2022), we evaluate four partial-label classifiers in combination with five feature-selection methods:

- PL-AGGD Wang et al. (2022) guides candidate-label disambiguation with an adaptive graph and iteratively alternates between graph updates and label-confidence estimation.
- SURE Feng & An (2019) jointly optimizes classifier training and pseudo-label updates, using maximum infinity-norm regularization to strengthen mutual exclusion among candidate labels.
- IPAL Zhang & Yu (2015) iteratively updates candidate-label confidence through an instance-neighborhood graph and label propagation.
- PL-CL Jia et al. (2023) exploits shared local topology between feature and label spaces to mitigate false-positive candidate labels.
- RDPF Bao et al. (2022) randomly selects the prescribed number of features from the feature space.
- MJE Bao et al. (2022) constructs a feature subset by maximizing its joint entropy.
- MR Bao et al. (2022) selects features according to their relevance to the labels.
- SAUTE Bao et al. (2022) selects feature subsets using mutual-information dependence and submodular optimization.

4.2 COMPARATIVE RESULTS

At a missing rate of 0.3 and a feature-selection ratio of 0.4, Table 2 shows that IMPLFS achieves the highest mean Accuracy in all 16 classifier–dataset combinations. Its consistent advantages across all four partial-label classifiers indicate that the selected features generalize across different downstream learners. Appendix C reports the macro-level Precision, Recall, and F1 results, where IMPLFS leads in most settings and remains competitive in the rest.

Table 2: Classification accuracy on the four datasets (mean±std).

Method	Lostnew	MSRC2new	Yahoo!News100	Yahoo!News50
PL-AGGD-SAUTE	0.5252 ± 0.0283	0.3363 ± 0.0170	0.8304 ± 0.0058	0.7195 ± 0.0173
PL-AGGD-RDPF	0.5025 ± 0.0312	0.3281 ± 0.0200	0.7320 ± 0.0227	0.6576 ± 0.0079
PL-AGGD-MR	0.4560 ± 0.0315	0.3521 ± 0.0169	0.8087 ± 0.0048	0.6748 ± 0.0123
PL-AGGD-MJE	0.5193 ± 0.0157	0.3071 ± 0.0290	0.7950 ± 0.0082	0.6834 ± 0.0069
PL-AGGD-IMPLFS	$\mathbf{0.5697 \pm 0.0185}$	$\mathbf{0.3798 \pm 0.0068}$	$\mathbf{0.8453 \pm 0.0070}$	$\mathbf{0.7484 \pm 0.0073}$
SURE-SAUTE	0.5183 ± 0.0149	0.3303 ± 0.0039	0.8307 ± 0.0093	0.7224 ± 0.0166
SURE-RDPF	0.5153 ± 0.0152	0.2831 ± 0.0068	0.7393 ± 0.0268	0.6557 ± 0.0058
SURE-MR	0.4629 ± 0.0232	0.3311 ± 0.0047	0.8202 ± 0.0104	0.6807 ± 0.0100
SURE-MJE	0.5064 ± 0.0146	0.2869 ± 0.0187	0.7991 ± 0.0111	0.6884 ± 0.0046
SURE-IMPLFS	$\mathbf{0.5460 \pm 0.0051}$	$\mathbf{0.3333 \pm 0.0094}$	$\mathbf{0.8448 \pm 0.0040}$	$\mathbf{0.7497 \pm 0.0076}$
IPAL-SAUTE	0.4065 ± 0.0051	0.3318 ± 0.0047	0.6626 ± 0.0046	0.5724 ± 0.0190
IPAL-RDPF	0.4224 ± 0.0069	0.2959 ± 0.0104	0.5987 ± 0.0054	0.5145 ± 0.0051
IPAL-MR	0.3383 ± 0.0206	0.3236 ± 0.0059	0.6172 ± 0.0206	0.5308 ± 0.0026
IPAL-MJE	0.4333 ± 0.0051	0.3154 ± 0.0191	0.5776 ± 0.0137	0.4894 ± 0.0157
IPAL-IMPLFS	$\mathbf{0.4609 \pm 0.0181}$	$\mathbf{0.3566 \pm 0.0057}$	$\mathbf{0.6705 \pm 0.0058}$	$\mathbf{0.5794 \pm 0.0035}$
PL-CL-SAUTE	0.4115 ± 0.0317	0.3061 ± 0.0343	0.7692 ± 0.0048	0.6349 ± 0.0186
PL-CL-RDPF	0.4500 ± 0.0163	0.2975 ± 0.0201	0.6596 ± 0.0303	0.5697 ± 0.0043
PL-CL-MR	0.3432 ± 0.0223	0.2966 ± 0.0428	0.7440 ± 0.0075	0.5996 ± 0.0192
PL-CL-MJE	0.4580 ± 0.0280	0.2903 ± 0.0366	0.7270 ± 0.0027	0.6100 ± 0.0153
PL-CL-IMPLFS	$\mathbf{0.4777 \pm 0.0178}$	$\mathbf{0.3266 \pm 0.0218}$	$\mathbf{0.7891 \pm 0.0049}$	$\mathbf{0.6740 \pm 0.0110}$

4.3 Two-Sided Paired t-Test

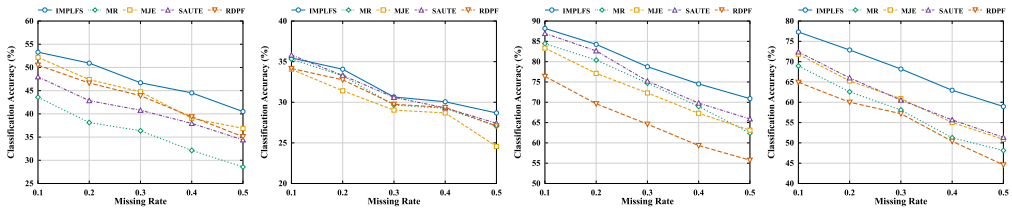
We conducted two-sided paired t-tests using the five matched runs for each comparison. At $p < 0.05$, Table 3 reports 56 wins, 8 ties, and no losses for IMPLFS against the four baselines over all classifier–dataset combinations. These results indicate that the improvements in mean Accuracy are statistically reliable in most comparisons rather than being driven by a small number of runs.

Table 3: Win/tie/loss counts for IMPLFS against the baselines over all datasets.

Classifier	MR	MJE	SAUTE	RDPF
PL-AGGD	4/0/0	4/0/0	4/0/0	4/0/0
SURE	3/1/0	4/0/0	3/1/0	4/0/0
IPAL	4/0/0	4/0/0	3/1/0	4/0/0
PL-CL	3/1/0	2/2/0	3/1/0	3/1/0
Total	14/2/0	14/2/0	13/3/0	15/1/0

4.4 Effect of Missing Views

Figure 2 reports the Accuracy of IMPLFS and the four baseline feature selectors as the missing rate increases from 0.1 to 0.5, using PL-CL with the feature-selection ratio fixed at 0.4. Although all methods deteriorate as fewer views are observed, IMPLFS exhibits the smallest Accuracy drop between missing rates 0.1 and 0.5 on all four datasets. This consistently slower degradation demonstrates its robustness to increasing view incompleteness.

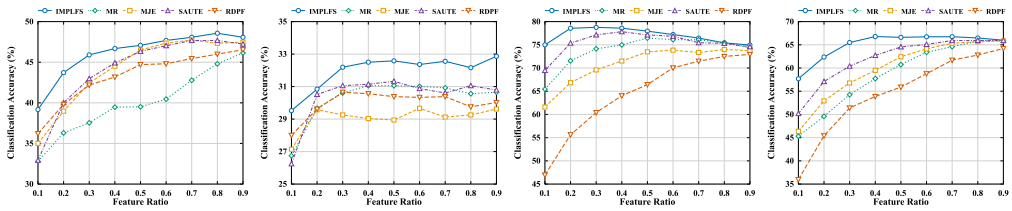


(a) Lostnew (b) MSRC2new (c) Yahoo!News100 (d) Yahoo!News50

Figure 2: Classification accuracy of feature-selection methods under different missing rates.

4.5 Effect of the Feature-Selection Ratio

Figure 3 reports accuracy of IMPLFS and the four baseline feature selectors as the retained-feature ratio varied from 0.1 to 0.9. Across the four datasets, performance improved from small ratios and then remained stable or changed moderately; a ratio of 0.4 consistently lay within a stable, near-optimal region. We therefore used 0.4 as a default that balances performance and compactness.

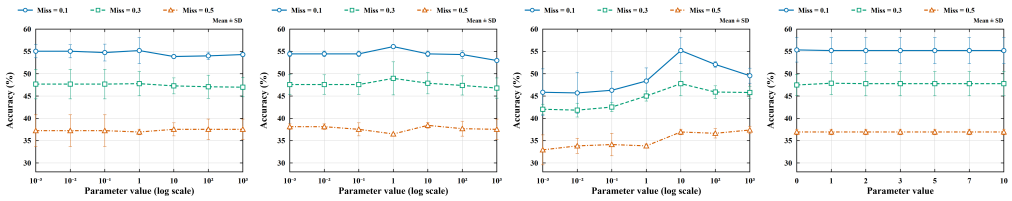


(a) Lostnew (b) MSRC2new (c) Yahoo!News100 (d) Yahoo!News50

Figure 3: Accuracy of IMPLFS across feature-selection ratios on four datasets (mean $\pm$ std).

4.6 Parameter Sensitivity

Figure 4 reports the effects of φ, β, γ, and r on Lostnew under missing rates of 0.1, 0.3, and 0.5. Accuracy varies smoothly over the tested ranges of φ, β, and r, whereas γ has a more pronounced effect. Complete sensitivity results are provided in Appendix D.



(a) Varying φ (b) Varying β (c) Varying γ (d) Varying r

Figure 4: Parameter sensitivity on Lostnew under different missing rates.

4.7 Ablation Study

Table 4 evaluates complementary reconstruction (CR) and partial-label disambiguation (PLD) over five missing rates using PL-CL. The w/o CR variant removes the complementary-reconstruction component, whereas w/o PLD fixes the initialized label-confidence matrix F throughout optimization. Removing CR or PLD reduces Accuracy in all 20 settings, with average drops of 1.69 and 1.87 percentage points, respectively. For w/o PLD, the largest dataset-level average drop is 3.64 points on Lostnew. These consistent degradations show that CR improves relation learning under incomplete views, while PLD reduces candidate-label ambiguity during feature evaluation. Additional results with PL-AGGD, SURE, and IPAL are provided in Appendix E.

Table 4: Ablation results under PL-CL (Accuracy, %). −CR and −PLD denote the variants without the corresponding components.

MR	Lostnew			MSRC2new			YahooNews50			YahooNews100		
	Full	−CR	−PLD	Full	−CR	−PLD	Full	−CR	−PLD	Full	−CR	−PLD
0.1	**53.29**	47.87	48.37	**35.06**	34.91	35.01	**77.31**	76.49	76.29	**88.19**	87.52	87.42
0.2	**49.14**	45.99	46.11	**34.07**	33.11	31.06	**71.87**	71.58	71.20	**84.25**	81.96	82.13
0.3	**46.71**	43.22	42.20	**30.65**	29.51	30.56	**67.19**	65.82	65.25	**78.76**	77.89	78.01
0.4	**43.52**	39.17	38.99	**29.06**	28.14	28.13	**61.06**	59.94	59.87	**73.52**	72.88	73.37
0.5	**38.18**	36.30	36.97	**28.69**	27.87	25.89	**56.96**	55.72	55.25	**69.80**	67.60	67.77

5 Conclusion

This paper studies incomplete multi-view partial-label feature selection and proposes IMPLFS, a unified framework based on complementary graph learning. The method reconstructs view-specific relations from cross-view complementary information, integrates them through adaptive fusion, and jointly learns label confidences and feature-selection matrices. Comprehensive experiments on four benchmarks paired competing feature selectors with four partial-label classifiers and further examined multiple evaluation metrics, missing rates, retained-feature ratios, model variants, and hyperparameter settings. The results consistently demonstrate the effectiveness of IMPLFS, its robustness to increasing view incompleteness and different feature budgets, and the contributions of complementary reconstruction and partial-label disambiguation. Future work will consider structured or non-random missing mechanisms and more efficient optimization.

AI USE STATEMENT

Generative AI tools were used primarily to assist with literature search and discovery for the Related Work section. All cited papers were independently retrieved, read, selected, and verified by the authors, who were solely responsible for their interpretation and use in the manuscript. Generative AI tools were not used to develop the scientific ideas, methodology, mathematical derivations, experimental results, or conclusions. The authors take full responsibility for the accuracy and originality of the work.

REFERENCES

Shunshun Bai, Qinghai Zheng, Xiaojin Ren, and Jihua Zhu. Graph-guided imputation-free incomplete multi-view clustering. *Expert Systems with Applications*, 258:125165, Dec 2024. ISSN 0957-4174. doi: 10.1016/j.eswa.2024.125165. URL http://dx.doi.org/10.1016/j.eswa.2024.125165.

Wei-Xuan Bao, Jun-Yi Hang, and Min-Ling Zhang. Submodular feature selection for partial label learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, pp. 26–34. ACM, 2022. doi: 10.1145/3534678.3539292. URL https://doi.org/10.1145/3534678.3539292.

Zhiwen Cao and Xijiong Xie. Multi-view unsupervised complementary feature selection with multi-order similarity learning. *Knowledge-Based Systems*, 283:111172, 2024. doi: 10.1016/j.knosys.2023.111172. URL https://doi.org/10.1016/j.knosys.2023.111172.

Jinfei Ding, Wenbin Qian, Yihui Li, Wenji Yang, and Jintao Huang. Partial label feature selection via label disambiguation and neighborhood mutual information. *Information Sciences*, 680:121163, Oct 2024. doi: 10.1016/j.ins.2024.121163. URL https://doi.org/10.1016/j.ins.2024.121163.

Jinfu Fan, Linqing Huang, Chaoyu Gong, Yang You, Min Gan, and Zhongjie Wang. Kmt-pll: K-means cross-attention transformer for partial label learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2):2789–2800, Feb 2025. doi: 10.1109/TNNLS.2023.3347792. URL https://doi.org/10.1109/TNNLS.2023.3347792.

Lei Feng and Bo An. Partial label learning with self-guided retraining. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1):3542–3549, 2019. doi: 10.1609/aaai.v33i01.33013542.

Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Multiple instance metric learning from automatically labeled bags of faces. In *Computer Vision – ECCV 2010*, volume 6311 of *Lecture Notes in Computer Science*, pp. 634–647. Springer, 2010. doi: 10.1007/978-3-642-15549-9_46. URL https://doi.org/10.1007/978-3-642-15549-9_46.

Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003. URL https://www.jmlr.org/papers/v3/guyon03a.html.

Qiuru Hai, Yongjian Deng, Yuena Lin, Zheng Li, Zhen Yang, and Gengyu Lyu. Cfdm: Contrastive fusion and disambiguation for multi-view partial-label learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(16):17005–17013, Apr 2025. doi: 10.1609/aaai.v39i16.33869. URL https://doi.org/10.1609/aaai.v39i16.33869.

Yemin Han, Jing Chai, Fa Zhu, Xingchi Chen, and David Camacho. Multi-view partial multi-label feature selection based on label disambiguation and shared subspace. *Applied Intelligence*, 56:205, 2026. doi: 10.1007/s10489-026-07206-0.

Pingting Hao, Wanfu Gao, and Liang Hu. Embedded feature fusion for multi-view multi-label feature selection. *Pattern Recognition*, 157:110888, 2025a. doi: 10.1016/j.patcog.2024.110888. URL https://doi.org/10.1016/j.patcog.2024.110888.

Pingting Hao, Kunpeng Liu, and Wanfu Gao. Uncertainty-aware global-view reconstruction for multi-view multi-label feature selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17068–17076, 2025b. doi: 10.1609/aaai.v39i16.33876. URL https://doi.org/10.1609/aaai.v39i16.33876.

Yanyong Huang, Kejun Guo, Xiuwen Yi, Zhong Li, and Tianrui Li. Incremental unsupervised feature selection for dynamic incomplete multi-view data. *Information Fusion*, 96:312–327, 2023. doi: 10.1016/j.inffus.2023.03.018. URL https://doi.org/10.1016/j.inffus.2023.03.018.

Yanyong Huang, Minghui Lu, Wei Huang, Xiuwen Yi, and Tianrui Li. Time-fs: Joint learning of tensorial incomplete multi-view unsupervised feature selection and missing-view imputation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(16):17503–17510, 2025. doi: 10.1609/aaai.v39i16.33924. URL https://doi.org/10.1609/aaai.v39i16.33924.

Yuheng Jia, Chongjie Si, and Min-Ling Zhang. Complementary classifier induced partial label learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 974–983, 2023.

Bingbing Jiang, Xingyu Wu, Xiren Zhou, Yi Liu, Anthony G. Cohn, Weiguo Sheng, and Huanhuan Chen. Semi-supervised multiview feature selection with adaptive graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(3):3615–3629, 2024a. doi: 10.1109/tnnls.2022.3194957. URL https://doi.org/10.1109/tnnls.2022.3194957.

Zhangqi Jiang, Tingjin Luo, and Xinyan Liang. Deep incomplete multi-view learning network with insufficient label information. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(11):12919–12927, Mar 2024b. ISSN 2159-5399. doi: 10.1609/aaai.v38i11.29189. URL http://dx.doi.org/10.1609/aaai.v38i11.29189.

Junqi Li, Wenbin Qian, Wenji Yang, Suxuan Liu, and Jintao Huang. Fuzzy neighborhood-based partial label feature selection via label iterative disambiguation. *International Journal of Approximate Reasoning*, 179:109358, Apr 2025. doi: 10.1016/j.ijar.2024.109358. URL https://doi.org/10.1016/j.ijar.2024.109358.

Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28(1), June 2014. ISSN 2159-5399. doi: 10.1609/aaai.v28i1.8973. URL http://dx.doi.org/10.1609/aaai.v28i1.8973.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11169–11178. IEEE, June 2021. doi: 10.1109/cvpr46437.2021.01102. URL http://dx.doi.org/10.1109/cvpr46437.2021.01102.

Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2023. ISSN 1939-3539. doi: 10.1109/tpami.2022.3197238. URL http://dx.doi.org/10.1109/tpami.2022.3197238.

Chengliang Liu, Jinlong Jia, Jie Wen, Yabo Liu, Xiaoling Luo, Chao Huang, and Yong Xu. Attention-induced embedding imputation for incomplete multi-view partial multi-label classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 13864–13872, 2024. doi: 10.1609/aaai.v38i12.29293. URL https://doi.org/10.1609/aaai.v38i12.29293.

Jun-Ying Liu, Jian-Ping Sun, Ya-Hong Zhao, Bin-Bin Jia, and Min-Ling Zhang. Partial label learning with semi-supervised clustering disambiguation. *Pattern Recognition*, 171:112138, Mar 2026a. doi: 10.1016/j.patcog.2025.112138. URL https://doi.org/10.1016/j.patcog.2025.112138.

Xiangyu Liu and Peng Song. Incomplete multi-view clustering via virtual-label guided matrix factorization. *Expert Systems with Applications*, 210:118408, Dec 2022. ISSN 0957-4174. doi: 10.1016/j.eswa.2022.118408. URL http://dx.doi.org/10.1016/j.eswa.2022.118408.

Yuzhou Liu, Jiarui Liu, and Wanfu Gao. Redundancy-optimized multi-head attention networks for multi-view multi-label feature selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 23945–23953, 2026b. doi: 10.1609/aaai.v40i28.39571. URL https://doi.org/10.1609/aaai.v40i28.39571.

Wenbin Qian, Yihui Li, Qianzhi Ye, Weiping Ding, and Wenhao Shu. Disambiguation-based partial label feature selection via feature dependency and label consistency. *Information Fusion*, 94:152–168, Jun 2023. doi: 10.1016/j.inffus.2023.01.019. URL https://doi.org/10.1016/j.inffus.2023.01.019.

Zeyu Ren, Shuihua Wang, and Yudong Zhang. Weakly supervised machine learning. *CAAI Transactions on Intelligence Technology*, 8(3):549–580, Apr 2023. doi: 10.1049/cit2.12216. URL https://doi.org/10.1049/cit2.12216.

Shiding Sun, Xiaotong Yu, and Yingjie Tian. Multi-view prototype-based disambiguation for partial label learning. *Pattern Recognition*, 141:109625, Sept 2023. ISSN 0031-3203. doi: 10.1016/j.patcog.2023.109625. URL http://dx.doi.org/10.1016/j.patcog.2023.109625.

Chang Tang, Xiao Zheng, Xinwang Liu, Wei Zhang, Jing Zhang, Jian Xiong, and Lizhe Wang. Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 34(10):4705–4716, 2022. doi: 10.1109/tkde.2020.3048678. URL https://doi.org/10.1109/tkde.2020.3048678.

Jingjing Tang, Qingqing Yi, Saiji Fu, and Yingjie Tian. Incomplete multi-view learning: Review, analysis, and prospects. *Applied Soft Computing*, 153:111278, Mar 2024. doi: 10.1016/j.asoc.2024.111278. URL https://doi.org/10.1016/j.asoc.2024.111278.

Yingjie Tian, Xiaotong Yu, and Saiji Fu. Partial label learning: Taxonomy, analysis and outlook. *Neural Networks*, 161:708–734, Apr 2023. doi: 10.1016/j.neunet.2023.02.019. URL https://doi.org/10.1016/j.neunet.2023.02.019.

Deng-Bao Wang, Min-Ling Zhang, and Li Li. Adaptive graph guided disambiguation for partial label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):8796–8811, Dec 2022. ISSN 1939-3539. doi: 10.1109/tpami.2021.3120012. URL http://dx.doi.org/10.1109/tpami.2021.3120012.

Guangtai Wang, Jintao Huang, Yiqiang Lai, and Chi-Man Vong. Dealing with partial labels by knowledge distillation. *Pattern Recognition*, 158:110965, Feb 2025. doi: 10.1016/j.patcog.2024.110965. URL https://doi.org/10.1016/j.patcog.2024.110965.

Jie Wen, Jiang Long, Xiaohuan Lu, Chengliang Liu, Xiaozhao Fang, and Yong Xu. Partial multiview incomplete multilabel learning via uncertainty-driven reliable dynamic fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(1):236–250, 2026. doi: 10.1109/tpami.2025.3603677. URL https://doi.org/10.1109/tpami.2025.3603677.

Guillaume Wisniewski, Nicolas Pécheux, Souhir Gahbiche-Braham, and François Yvon. Cross-lingual part-of-speech tagging through ambiguous learning. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1779–1785. Association for Computational Linguistics, 2014. doi: 10.3115/v1/D14-1187. URL https://aclanthology.org/D14-1187/.

Tingting Wu, Songhe Feng, and Jiazheng Yuan. Low-rank kernel tensor learning for incomplete multi-view clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(14):15952–15960, Mar 2024. ISSN 2159-5399. doi: 10.1609/aaai.v38i14.29526. URL http://dx.doi.org/10.1609/aaai.v38i14.29526.

Jing-Hua Yang, Le-Le Fu, Chuan Chen, Hong-Ning Dai, and Zibin Zheng. Cross-view graph matching for incomplete multi-view clustering. *Neurocomputing*, 515:79–88, Jan 2023. ISSN 0925-2312. doi: 10.1016/j.neucom.2022.10.007. URL http://dx.doi.org/10.1016/j.neucom.2022.10.007.

Zhiwen Yu, Ziyang Dong, Chenchen Yu, Kaixiang Yang, Ziwei Fan, and C. L. Philip Chen. A review on multi-view learning. *Frontiers of Computer Science*, 19(7):197334, 2025. doi: 10.1007/s11704-024-40004-w. URL https://doi.org/10.1007/s11704-024-40004-w.

Changqing Zhang, Yajie Cui, Zongbo Han, Joey Tianyi Zhou, Huazhu Fu, and Qinghua Hu. Deep partial multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2402–2415, 2022. doi: 10.1109/tpami.2020.3037734. URL https://doi.org/10.1109/TPAMI.2020.3037734.

Guangyi Zhang and Ali Etemad. Partial label learning for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 16(3):2381–2395, Jul 2025. doi: 10.1109/TAFFC.2025.3562027. URL https://doi.org/10.1109/TAFFC.2025.3562027.

Min-Ling Zhang and Fei Yu. Solving the partial label learning problem: An instance-based approach. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pp. 4048–4054, 2015.

Zan Zhang, Jialu Yao, Lin Liu, Jiuyong Li, Lei Li, and Xindong Wu. Partial label feature selection: An adaptive approach. *IEEE Transactions on Knowledge and Data Engineering*, 36(8):4178–4191, Aug 2024. doi: 10.1109/TKDE.2024.3365691. URL https://doi.org/10.1109/TKDE.2024.3365691.

Liang Zhao, Yanshan Xiao, Bo Liu, and Zhifeng Hao. Multi-view partial label machine. *Information Sciences*, 586:310–325, Mar 2022. doi: 10.1016/j.ins.2021.11.070. URL https://doi.org/10.1016/j.ins.2021.11.070.