Confidence-Aware With Prototype Alignment for Partial Multi-label Learning

Weijun Lv[1], **Yu Chen**[1], **Xiaozhao Fang**[1*], **Xuhuan Zhu**[1], **Jie Wen**[2*], **Guoxu Zhou**[1], **Sixian Chan**[3]
[1]School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[2]Shenzhen Key Laboratory of Visual Object Detection and Recognition, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China
[3]College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
{lvweijun0201, chenyu9265324}@163.com, {xzhfang168, xuhuanz_zz}@126.com
jiewen_pr@126.com, gx.zhou@gdut.edu.cn, sxchan@zjut.edu.cn

Abstract

Label prototype learning has emerged as an effective paradigm in Partial Multi-Label Learning (PML), providing a distinctive framework for modeling structured representations of label semantics while naturally filtering noise through prototype-based label confidence estimation. However, existing prototype-based methods face a critical limitation: class prototypes are the biased estimates due to noisy candidate labels, particularly when positive samples are scarce. To this end, we first propose a mutually class prototype alignment strategy bypassing noise interference by introducing two different transformation matrices, which makes the class prototypes learned by the fuzzy clustering and candidate label set mutually alignment for correcting themselves. Such alignment is also passed on to the fuzzy memberships label in turn. In addition, to eliminate noise interference in the candidate label set during the classifier learning, we use the learned permutation matrix to transform the fuzzy memberships label for learning a label reliability indicator matrix accompanied by the candidate label set. This makes the label reliability indicator matrix absolutely prevent the occurrence of numerical values located in non-label and simultaneously eliminate the introduction of incorrect label as much as possible. The resulting indicator matrix guides a robust multi-label classifier training process, jointly optimizing label confidence and classifier parameters. Extensive experiments demonstrate that our proposed model exhibits significant performance advantages over state-of-the-art PML approaches.

1 Introduction

Multi-Label Learning (MLL) is an important branch of machine learning that allows an instance to belong to multiple categories simultaneously, with widespread applications in image annotation[1, 2], text categorization[3], and medical diagnosis [4]. However, in practical application scenarios, obtaining precisely annotated data is often difficult due to high annotation costs and inherent ambiguity[5, 6, 7, 8]. Annotators frequently provide a candidate label set when uncertain, containing both ground-truth labels and noisy labels (i.e., false positive labels mistakenly included). For example, in image annotation tasks, objects with similar visual appearances may be incorrectly labeled, which will introduce interference to the classification prediction. To address this issue, Partial Multi-Label Learning (PML) [9] has emerged as a new weakly supervised learning framework. The

*Corresponding Author

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

 https://doi.org/10.52202/085713-5670

main task of PML is to learn from this uncertain supervisory information to accurately predict the ground-truth labels for unknown instances.

PML research primarily focuses on identifying ground-truth labels from the candidate label set, known as label disambiguation. Existing methods mostly adopt explicit learning strategies, directly removing noisy labels from candidate label sets through various techniques. For instance, some methods use low-rank and sparse decomposition to separate ground-truth labels from noisy ones [10, 11]; others estimate the credibility of candidate labels through label propagation or label distribution learning [12, 13, 14, 15]; while some utilize feature information to identify noisy labels [16, 5], where the practice of using feature information to impose correlation constraints or manifold constraints on labels is the most popular [17, 18, 19]. Recently, feature selection has become a popular approach for disambiguation [20, 21, 22]. Some researchers leverage label correlations or cluster assignments to assist disambiguation [9, 23, 24, 25, 26, 19]. Additionally, some methods introduce complementary classifiers that simultaneously leverage both positive and negative label information under the noise sparsity assumption [27]. Another line of research adopts a phased processing approach: first purifying noisy labels through methods like granular ball construction [25] or label propagation [12], then applying structured learning strategies such as confidence score regression or pairwise classification paradigms [28, 29] to the refined labels. These explicit learning approaches, although effective to some extent, still face limitations when handling complex noise patterns. On one hand, methods based on sparsity assumptions often fail when real-world annotation scenarios produce non-sparse noise distributions, leading to error accumulation during classifier training. On the other hand, methods based on label confidence or manifold learning struggle with samples lacking sufficient neighborhood information or when positive samples are scarce, making it difficult to fully fit the data distribution.

Recently, prototype-based learning[9, 25, 30, 19] has emerged as a promising direction that constructs class prototypes to guide label confidence estimation. As a cornerstone of prototype-based approaches, clustering learning serves as an important tool for exploratory data analysis [31, 32], naturally generating class prototypes without requiring prior labels while efficiently revealing inherent data structure [33, 34]. Building upon fuzzy clustering principles, FBD-PML [19] advances this direction by mining the correlation between sample instances and labels while learning confidence values under sample manifold assumptions. However, existing prototype-based methods face a critical limitation: prototypes derived from noisy candidate labels inevitably deviate from true semantic centers—a representation bias problem pervasive under imperfect supervision[35, 36, 37, 38]. This fundamental weakness becomes particularly pronounced when positive samples are scarce, as the limited reliable supervision further exacerbates prototype distortion. This prototype bias propagates through confidence estimation, ultimately degrading classification performance. Moreover, existing methods struggle to effectively bridge the semantic gap between unsupervised clustering-derived prototypes and weakly supervised label-based prototypes, leaving the potential complementarity between these two prototype spaces largely unexplored.

Addressing these challenges, this paper proposes CAPML (**C**onfidence-**A**ware with Prototype Matching for **P**artial **M**ulti-label **L**earning), a novel method that leverages unsupervised clustering to bypass noise interference while enhancing weakly-supervised semantic representations through effective prototype space alignment. Specifically, unlike traditional prototype-based methods that solely rely on noisy candidate labels, the proposed approach introduces a transformation mechanism that successfully bridges the gap between clustering-derived prototypes and label-based prototypes, enabling discovery of their intrinsic correspondence despite noisy supervision. Subsequently, a confidence-aware process is designed to convert fuzzy membership degrees into label reliability indicators, guiding classifier training with sparse $\ell_{2,1}$-norm constraints. Concurrently, the prototype alignment mechanism is also utilized to guide the refinement of label confidence estimation. Finally, the enhanced confidence values and learned classifiers work jointly to predict labels for unknown instances. The main contributions of this paper are summarized as follows:

- This paper provides the first investigation into prototype misalignment between prototypes derived from fuzzy clustering and prototypes computed from candidate label sets in PML tasks, introducing a transformation mechanism that successfully bridges these two prototype spaces and enables effective alignment despite noisy supervision.
- This paper designs a confidence-aware process that converts fuzzy label membership degrees into label reliability indicator values, guiding classifier training with sparse $\ell_{2,1}$-norm constraints that enhance feature selection while reducing overfitting to noisy labels.

- Extensive empirical evaluation demonstrates the proposed method's efficacy in resolving label ambiguity and prototype misalignment problems in PML, achieving superior generalization performance even with high noise rates and sparse positive samples.

This section introduces the concept of PML, its evolution, related work, and how our approach differs from existing methods. The remainder of this paper is organized as follows: Section 2 details the principles and optimization algorithm of CAPML. Section 3 and 4 provide comprehensive experimental results and analysis. Finally, Section 5 gives a conclusion and Section 6 illustrates the limitations of the method proposed.

2 Method

In PML, let $\mathcal{X} \subset \mathbb{R}^d$ denotes the d-dimensional feature space and $\mathcal{Y} = \{\boldsymbol{l}_1, \boldsymbol{l}_2, ..., \boldsymbol{l}_q\}$ denotes the label space with q class labels. The training dataset $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)|1 \leq i \leq n\}$ contains n examples, where $\boldsymbol{x}_i \in \mathcal{X}$ is the i-th instance and $\boldsymbol{y}_i \in \{0,1\}^q$ represents its corresponding label vector. Let $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{d\times n}$ denotes the feature matrix and $\mathbf{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n]^\top \in \{0,1\}^{n\times q}$ represents the candidate label matrix containing wrong annotation. Here, $y_{ij} = 1$ indicates the i-th instance is annotated with the j-th label, and $y_{ij} = 0$ indicates otherwise. Each instance correspond to a set of candidate labels with unrelated labels incorrectly labeled as 1, which is called noisy label. The goal of PML is to learn a classification function $f : \mathcal{X} \to 2^{\mathcal{Y}}$ that minimizes the effect of noisy label information and makes accurate label predictions. Figure .1 (right) illustrates the overall architecture of CAPML.

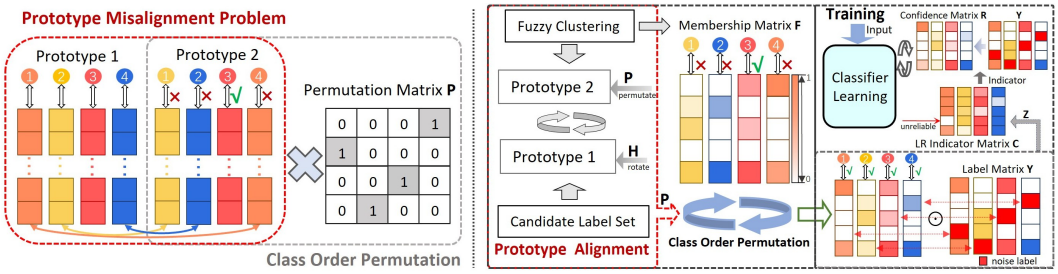


Figure 1: Overview of the CAPML framework.
Left: The prototype misalignment problem—unsupervised prototypes (Prototype 1) and weakly supervised prototypes (Prototype 2) lack optimal correspondence that require permutation matrix **P** for class order alignment.
Right: The two-stage pipeline: (1) *Prototype learning and alignment*—fuzzy clustering produces membership matrix **F** and unsupervised Prototype 1 while weakly supervised Prototype 2 is derived from candidate label set, then permutation matrix **P** establishes optimal correspondence, transforming memberships into label reliability (LR) indicator matrix **C** via element-wise product with **Y** and normalization; (2) *Confidence-aware classifier training*—indicator matrix **C** guides the learning of label confidence matrix **R**, which in turn supervises classifier training to effectively suppress noise labels (marked in red) in **Y** and obtain proper predictions.

2.1 Class Prototype Learning and Alignment

Our approach begins with unsupervised prototype learning to capture the intrinsic structure of the data, independent of potentially noisy label information. We employ an improved fuzzy clustering approach incorporating entropy regularization[39, 40] to obtain well-distributed class prototypes and reliable fuzzy label membership degrees. The unified objective function is formulated as:

$$\min_{\mathbf{F}\mathbf{1}_c=\mathbf{1}_n,\mathbf{F}\geq 0,\mathbf{M}} \sum_{i=1}^{n}\sum_{j=1}^{c} f_{ij}\|\boldsymbol{x}_i - \boldsymbol{m}_j\|_2^2 + \lambda \sum_{i=1}^{n}\sum_{j=1}^{c} f_{ij}\log f_{ij}, \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{d\times c}$ denotes the class prototype matrix, $\mathbf{F} \in \mathbb{R}^{n\times c}$ represents the fuzzy label membership matrix indicating the association strength between instances and class prototypes, and λ controls the entropy regularization. The first term quantifies weighted clustering quality, while

https://doi.org/10.52202/085713-5670

the entropy term prevents degenerate solutions and enables flexible membership distributions that can accommodate the multi-label nature by smoothly transitioning from unimodal to multimodal assignments. Through alternating optimization, the update formulas for $\mathbf{M}$ and $\mathbf{F}$ as follows:

$$\boldsymbol{m}_j = \frac{\sum_{k=1}^{n} f_{kj}\boldsymbol{x}_k}{\sum_{k=1}^{n} f_{kj}},\ f_{ij} = \frac{e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{m}_j\|_2^2}{\lambda}}}{\sum_{k=1}^{c} e^{-\frac{\|\boldsymbol{x}_i - \boldsymbol{m}_k\|_2^2}{\lambda}}}. \tag{2}$$

The membership degree f_{ij} follows a softmax-like formulation, where instances have higher membership degrees to closer prototypes. The parameter λ controls the "softness" of the assignments—larger values result in more uniform distributions, while smaller values lead to more decisive assignments. While the prototypes M derived from fuzzy clustering lack explicit semantic meaning, we can leverage them by setting $c = q$ to match the number of label classes. To incorporate label semantic information, we construct class prototypes $\mathbf{O} \in \mathbb{R}^{d\times q}$ from the candidate label matrix $\mathbf{Y}$:

$$\boldsymbol{o}_j = \frac{\sum_{k=1}^{n} y_{kj}\boldsymbol{x}_k}{\sum_{k=1}^{n} y_{kj}}. \tag{3}$$

Each prototype $\boldsymbol{o}_j$ is computed as the centroid of instances associated with the j-th label in $\mathbf{Y}$. Despite the presence of noise, $\mathbf{O}$ can be viewed as weakly-supervised prototype[41] that still contains valuable semantic information. Although $\mathbf{M}$ and $\mathbf{O}$ are derived from different principles—unsupervised clustering and weakly supervised label aggregation respectively—they fundamentally capture the same underlying class structure.

Then a critical problem emerges: *although both* $\mathbf{M}$ *and* $\mathbf{O}$ *capture class-level representation, their orderings are inherently misaligned due to the arbitrary indexing produced by clustering.* For example, as shown in Figure .1 (left), only the 3-rd prototype in $\mathbf{M}$ correctly corresponds to its counterpart in $\mathbf{O}$, while the remaining 1-st, 2-nd, 4-th unsupervised prototypes improperly correspond to positions 4-th, 1-st, 2-nd respectively. To address this issue, we introduce a permutation matrix $\mathbf{P} \in \{0,1\}^{q\times q}$ to align these orders:

$$\min_{\mathbf{P}} \|\mathbf{MP} - \mathbf{O}\|_F^2, \quad s.t.\ \mathbf{P}\mathbf{1}_q = \mathbf{1}_q,\ \mathbf{P}^\top\mathbf{1}_q = \mathbf{1}_q,\ \mathbf{P} \in \{\mathbf{0},\mathbf{1}\}^q. \tag{4}$$

This formulation seeks the optimal one-to-one mapping between unsupervised and weakly supervised prototypes by minimizing the Frobenius norm of their difference. The constraints ensure that $\mathbf{P}$ is a valid permutation matrix, with exactly one entry of 1 in each row and column.

Theorem 1: *When the fuzzy clustering successfully captures the underlying class structure, there exists an optimal permutation matrix* $\mathbf{P}^* \in \{0,1\}^{q\times q}$ *such that the alignment error satisfies:*

$$\|\mathbf{MP}^* - \mathbf{O}\|_F = O\left(\epsilon\sqrt{q} + \sqrt{\frac{q^2}{n}}\right), \tag{5}$$

where ϵ *measures the label noise level (specifically, the fraction of noisy labels in* $\mathbf{Y}$*),* n *is the number of training instances, and* q *is the number of classes.*

The proof is provided in Appendix. This theorem reveals that the alignment error is governed by two factors: (1) the label noise level ϵ in the candidate set, and (2) the finite sample effect $\sqrt{q^2/n}$, which diminishes as more training data becomes available. Importantly, the bound suggests that even with moderate noise, the permutation matrix $\mathbf{P}$ can establish meaningful correspondence when sufficient data is present.

However, another key challenge remains: *since* $\mathbf{Y}$ *contains noise, the weakly supervised prototypes* $\mathbf{O}$ *deviate from true class centroids.* Directly aligning $\mathbf{M}$ to $\mathbf{O}$ might propagate these noise-induced shifts. To address this discrepancy and make $\mathbf{P}$ more reliable, we introduce an orthogonal rotation matrix $\mathbf{H} \in \mathbb{R}^{q\times q}$ to allow for more flexible alignment:

$$\min_{\mathbf{P},\mathbf{H}} \|\mathbf{MP} - \mathbf{OH}\|_F^2, \quad s.t.\ \mathbf{P}\mathbf{1}_q = \mathbf{1}_q,\ \mathbf{P}^\top\mathbf{1}_q = \mathbf{1}_q,\ \mathbf{P} \in \{\mathbf{0},\mathbf{1}\}^q,\ \mathbf{H}\mathbf{H}^\top = \mathbf{I}_q. \tag{6}$$

The orthogonal rotation matrix $\mathbf{H}$ introduces geometric transformation that adapts the weakly supervised prototypes $\mathbf{O}$ to better match the unsupervised structure, while preserving their relative geometry through the constraint $\mathbf{H}\mathbf{H}^\top = \mathbf{I}_q$. This mitigates the negative impact of label noise: rather than forcing $\mathbf{M}$ to directly align with the potentially biased $\mathbf{O}$, we allow $\mathbf{O}$ to rotate in its representation space, reducing the negative influence of noise-induced biases on the alignment quality.

Constructing label reliability indicator. Having obtained the optimal permutation matrix $\mathbf{P}$ and rotation matrix $\mathbf{H}$, we establish the correspondence between $\mathbf{M}$ and $\mathbf{O}$. The permutation matrix $\mathbf{P}$ not only aligns the prototype spaces but also reveals which cluster corresponds to which label class, thereby enabling transformation of the fuzzy membership degree in $\mathbf{F}$ into label-specific reliability indicator for candidate labels.

We apply the learned permutation matrix $\mathbf{P}$ to reorder the membership matrix $\mathbf{F}$, aligning each column with its corresponding label class. To mitigate interference from misaligned entries, we perform element-wise multiplication with the candidate label matrix $\mathbf{Y}$ to retain only candidate positions, followed by row-wise min-max normalization $\mathcal{N}_{\text{minmax}}(\cdot)$ to amplify confidence contrasts:

$$\mathbf{Z} = \mathcal{N}_{\text{minmax}}(\mathbf{FP} \odot \mathbf{Y}). \tag{7}$$

This operation filters out non-candidate noise while enhancing discrimination between reliable and unreliable candidates. We then construct the label reliability (LR) indicator matrix $\mathbf{C} \in \mathbb{R}^{n \times q}$ as:

$$c_{ij} = \begin{cases} 1, & \text{if } y_{ij} = 0 \\ z_{ij}, & \text{if } y_{ij} = 1 \end{cases}. \tag{8}$$

However, this initial formulation treats all instances uniformly, ignoring that instances with more candidate labels typically contain more noise. To address this, we introduce instance-adaptive weighting based on candidate label density:

$$c_{ij} = \begin{cases} \|\mathbf{y}_i\|_1, & \text{if } y_{ij} = 0 \\ (q - \|\mathbf{y}_i\|_1) * z_{ij}, & \text{if } y_{ij} = 1 \end{cases}. \tag{9}$$

This refined formulation assigns differentiated indicator values based on candidate label density, which ensures that instances with denser candidate sets—which statistically contain more false positives—receive more conservative reliability estimates.

2.2 Confidence-Aware Label Disambiguation

The LR indicator matrix $\mathbf{C}$ plays a crucial role in our label disambiguation process. A higher indicator value c_{ij} indicates a higher probability that r_{ij} is a true label rather than noise. With the label reliability (LR) indicator matrix $\mathbf{C}$ constructed, we now formulate the confidence-aware objective for joint classifier learning and label disambiguation::

$$\min_{\mathbf{W},\mathbf{R}} \|\mathbf{C} \odot (\mathbf{Y} - \mathbf{R})\|_F^2 + \|\psi(\mathbf{X})^\top \mathbf{W} - \mathbf{R}\|_F^2 + \alpha \|\mathbf{Y} - \mathbf{R}\|_1 + \beta \|\mathbf{W}\|_{2,1}, \ s.t. \ \mathbf{R} \geq 0, \tag{10}$$

where $\psi(\cdot) : \mathbb{R}^d \to \mathbb{R}^h$ is a feature mapping function that transforms input features to a kernel space for better separability, $\mathbf{W} \in \mathbb{R}^{h \times q}$ represents the classifier parameters, $\mathbf{R} \in \mathbb{R}^{n \times q}$ denotes the label confidence matrix, and α, β are regularization hyperparameters. The first term implements reliability-weighted refinement through element-wise multiplication. For non-candidates, the weight $c_{ij} = 1$ penalizes any deviation from zero in $\mathbf{R}$, constraining $r_{ij} \approx 0$. For candidates, varying weights create differentiated penalties: high-reliability positions (large c_{ij}) are tightly constrained to $\mathbf{Y}$, while low-reliability positions (small c_{ij}) receive weaker constraints, allowing the classifier consistency and ℓ_1 terms to guide their refinement. The ℓ_1 norm term further reduces the influence of noise in candidate label set $\mathbf{Y}$, while the $\ell_{2,1}$ norm term enhances the classifier's discriminative power by emphasizing features with high discriminative capacity across multiple labels.

2.3 Optimization

The proposed approach optimization is divided into two parts. After obtaining $\mathbf{O}$ and $\mathbf{F}$ through Eq(2), we need to learn to obtain an effective permutation matrix $\mathbf{P}$ to obtain LE indicator matrix $\mathbf{C}$. Then $\mathbf{C}$ is reused for further refinement of the label confidence matrix $\mathbf{R}$ to guide the learning of the classifier $\mathbf{W}$. We optimize each variable by adopting an alternating iterative way.

Update H, fix P. We can obtain the following optimization problem about variant $\mathbf{H}$.

$$\min_{\mathbf{H}} \|\mathbf{A} - \mathbf{OH}\|_F^2, \quad s.t. \ \mathbf{HH}^\top = \mathbf{I}_q, \tag{11}$$

where $\mathbf{A} = \mathbf{MP}$. Eq. (11) presents a standard orthogonal Procrustes problem [42], which we efficiently solve following the optimization approach detailed in [43].

 https://doi.org/10.52202/085713-5670

Algorithm 1: Training Process of CAPML

Input: The PML training dataset $\mathcal{D}$; parameters λ, α, β; max iterations T_0, T_1, T_2; Unseen sample $\hat{\mathbf{x}}$.
Output: the predicted label for unseen sample $\hat{\mathbf{y}}$.

```
// Stage One: Prototype Learning and Alignment
```
Initialize membership matrix $\mathbf{F} = \mathbf{Y}$ and prototypes $\mathbf{M}$ via Eq. (2). **for** $t = 1$ *to* T_0 **do**
  Update $\mathbf{M}$ and $\mathbf{F}$ via Eq. (2); `// Fuzzy clustering for unsupervised prototype`
end
Compute $\mathbf{O}$ via Eq. (3); `// Compute supervised prototype from candidate labels`
Initialize permutation matrix $\mathbf{P} = \mathbf{I}_q$ and rotation matrix $\mathbf{H} = \mathbf{I}_q$.
for $t = 1$ *to* T_1 **do**
  Update $\mathbf{H}$ by solving orthogonal Procrustes problem in Eq. (11)
  Update $\mathbf{P}$ using Hungarian algorithm in Eq. (12); `// Prototype alignment optimization`
end
Compute label reliability indicator matrix $\mathbf{C}$ via Eq. (9) ; `// Construct LR indicator matrix`

```
// Stage Two: Classifier Learning
```
Initialize $\mathbf{W} = \mathbf{0}_q$, $\mathbf{R} = \mathbf{Y}$, and auxiliary variable $\mathbf{Q} = \mathbf{0}_{n\times q}$.
for $t = 1$ *to* T_2 **do**
  Update $\mathbf{W}$ using Eq. (13)
  Update $\mathbf{R}$ using multiplicative rule in Eq. (16)
  Update $\mathbf{Q}$ using soft-thresholding via Eq. (19)
  if $\|\mathbf{W}^{(t)} - \mathbf{W}^{(t-1)}\|_F^2 + \|\mathbf{R}^{(t)} - \mathbf{R}^{(t-1)}\|_F^2 < 10^{-5}$ **then**
    break;
  end
end
return the predicted label $\hat{\mathbf{y}}$

Update P, fix H. Substitute the $\mathbf{H}$ obtained from the last iteration to Eq. (6), and let $\hat{\mathbf{M}} = \mathbf{OH}$, we can obtain the following optimization problem.

$$\min_{\mathbf{P}} \|\mathbf{MP} - \hat{\mathbf{M}}\|_F^2, \quad s.t.\ \mathbf{P}\mathbf{1}_q = \mathbf{1}_q,\ \mathbf{P}^\top \mathbf{1}_q = \mathbf{1}_q,\ \mathbf{P} \in \{\mathbf{0}, \mathbf{1}\}^{q\times q}, \tag{12}$$

Due to the binary constraints on $\mathbf{P}$, direct optimization is infeasible. We solve this assignment problem efficiently using the Hungarian algorithm via MATLAB's *matchpairs* function [44].

Update W, fix R. The subproblem regarding $\mathbf{W}$ can be obtained:

$$\min_{\mathbf{W}} \|\psi(\mathbf{X})^\top \mathbf{W} - \mathbf{R}\|_F^2 + \beta\|\mathbf{W}\|_{2,1} \Leftrightarrow \min_{\mathbf{W}} \|\psi(\mathbf{X})^\top \mathbf{W} - \mathbf{R}\|_F^2 + \beta \operatorname{tr}\left(\mathbf{W}^\top \mathbf{D}\mathbf{W}\right), \tag{13}$$

where the diagonal elements of $\mathbf{D}$ are computed as $D_{ii} = 1/\sqrt{\|w_i\|_2^2} (\forall i = 1, 2, 3, ..., h)$. Taking the derivative of Eq. (13) w.r.t. $\mathbf{W}$ and setting it to 0, we can get the following equation:

$$\mathbf{W} = (\psi(\mathbf{X})\psi(\mathbf{X})^\top + \beta\mathbf{D})^{-1}\psi(\mathbf{X})\mathbf{R}, \tag{14}$$

Update R, fix W. We solve the optimization involving the non-convex $\ell_{2,1}$ norm and non-negative constraint $\mathbf{R} \geq 0$ by applying the Lagrange multiplier method, yielding the following Lagrangian function:

$$\min_{\mathbf{R}} \|\mathbf{C} \odot (\mathbf{Y} - \mathbf{R})\|_F^2 + \|\psi(\mathbf{X})^\top \mathbf{W} - \mathbf{R}\|_F^2 + \alpha\|\mathbf{Q}\|_1 + \mu/2\|\mathbf{Y} - \mathbf{R} - \mathbf{Q}\|_F^2 - \operatorname{tr}\left(\mathbf{\Theta}\mathbf{R}^\top\right), \tag{15}$$

where μ is a number large enough and $\mathbf{\Theta}$ represents the Lagrange multiplier. Taking the derivative of Eq. (15) and setting the derivative to zero. Then based on condition of Karush-Kuhn-Tucker (KKT), it can be given that: $\mathbf{\Theta} \odot \mathbf{R} = \mathbf{0}$, that is, $\theta_{ij} r_{ij} = 0$. Fix $\mathbf{Q}$, then take the derivative of Eq. (15) w.r.t. $\mathbf{R}$ and setting it to 0, we can get the following equation: We can get the update rules for $\mathbf{R}$:

$$r_{ij}^{(t+1)} = r_{ij}^{(t)} \frac{\mathbf{B}_{ij}}{\mathbf{A}_{ij} + eps}, \tag{16}$$

where $\mathbf{A}_{ij} = (2\mathbf{R} + 2\mathbf{C} \odot \mathbf{R} \odot \mathbf{C} + \mu(\mathbf{R} + \mathbf{E}))_{ij}$ and $\mathbf{B}_{ij} = (2\psi(\mathbf{X})^\top \mathbf{W}\mathbf{W} + 2\mathbf{C} \odot \mathbf{Y} \odot \mathbf{C} + \mu\mathbf{Y})_{ij}$, With $\mathbf{R}$ fixed, we can get the sub-optimization of $\mathbf{Q}$:

$$\min_{\mathbf{Q}}\ \mathcal{L}(\mathbf{Q}) = \alpha\|\mathbf{Q}\|_1 + \frac{\mu}{2}\|\mathbf{Y} - \mathbf{R} - \mathbf{Q}\|_F^2, \tag{17}$$

which is a typical LASSO regression problem [45], and we apply PGD algorithm to optimize it. The proximal operator of Eq. (17) is:

$$\text{prox}_{th(\cdot)}(\mathbf{Q}) = arg\min_{\mathbf{Q}} \|\mathbf{Q} - \mathbf{Z}\|_F^2 + \frac{\alpha}{\mu\mathbf{L}}\|\mathbf{Q}\|_1 , \tag{18}$$

where $\mathbf{Z} = \mathbf{Q}^t - \frac{1}{\mathbf{L}}\nabla\mathcal{L}(\mathbf{Q}^{(t)})$, $\mathbf{Q}^{(t)}$ represents the solution of $\mathbf{Q}$ from the t-th iteration, and $\nabla\mathcal{L}(\mathbf{Q})$ is the gradient of the objective function $\mathcal{L}(\mathbf{Q})$, $\mathbf{L}$ is the Lipschitz constant of $\nabla\mathcal{L}(\mathbf{Q})$ and t denotes the number of iteration. Eq. (18) can be iteratively updated by the soft-thresholding operator [46]:

$$q_{ij}^{(t+1)} = Soft[q_{ij}^{(t)} - \frac{1}{\mathbf{L}}\nabla\mathcal{L}(q_{ij}^t, \frac{\alpha}{\mu\mathbf{L}})], \tag{19}$$

where $Soft[b, \nu] = sign(b)max\{|b| - \nu, 0\}$. In addition, the Lipschitz constant of $\nabla\mathcal{L}(\mathbf{Q})$ is 1, so we set $\mathbf{L} = 1$. The overall pseudo code of CAPML is summarized in Algorithm 1 .

3 Experiment

3.1 Experimental Setup

Datasets To evaluate the generalization performance of our proposed CAPML approach, we conducted experiments on 10 datasets, including 6 real-world PML datasets[49] and 18 synthetic PML datasets generated from seven multi-label datasets[47, 48]. For clarity, the detailed characteristics of these datasets are shown in Table 1. Specifically, the synthetic datasets are derived from multi-label datasets by adding noise to the labels. For each instance, a portion of irrelevant labels is randomly picked as candidate labels along with the relevant ones. Taking the $birds$ dataset as an example, which originally has 1.01 ground-truth labels per instance ($avg.\#GLs$), we created three noisy variants with 3, 4, and 5 candidate labels per instance ($avg.\#CLs$) by randomly injecting approximately 1.99, 2.99, and 3.99 false positive labels per instance, respectively.

Table 1: Characteristics of experimental data sets.

Datsets	$\#Instances$	$\#Dim$	$\#Classes$	$avg.\#CLs$	$avg.\#GLs$	$Domain$
Mirflickr	10433	100	7	3.35	1.77	Images[1]
Music_emotion	6833	98	11	5.29	2.42	Music[1]
Msic_style	6839	98	10	6.04	1.44	Music[1]
YeastBP	6139	6139	217	5.93	5.54	Biology[1]
YeastCC	6139	6139	50	1.39	1.35	Biology[1]
YeastMF	6139	6139	39	1.04	1.01	Biology[1]
emotions	593	72	6	3, 4, 5	1.86	Music[2]
birds	645	260	19	3, 4, 5	1.01	Audio[2]
medical	978	1449	45	5, 7, 9	1.25	Text[2]
image	2000	294	5	2, 3, 4	1.23	Images[2]
yeast	2417	103	14	7, 9, 11	4.24	Biology[2]
corel5k	5000	499	374	7, 9, 11	3.52	Images[2]

[1] http://palm.seu.edu.cn/zhangml/, [2] http://mulan.sourceforge.net/datasets.html

Comparison approaches The performance of CAPML is compared with seven state-of-the-art methods, the following is a brief introduction for each comparison approach:

- fPML [16] [2019]: fPML removes noise by decomposing the candidate label matrix into two low-rank matrices and utilizing the resulting low-error approximation. [configuration: $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 10$].
- PARTICLE(PAR-MAP and PAR-VLS) [12] [2020]: A two-stage PML approach that refines candidate labels through label propagation and builds distinct predictive models [suggested configuration: $k = 10$, $\alpha = 0.9$, $thr = 10.9$].
- PML-NI [5] [2021]: Considering that fuzzy features may produce noise labels, the prediction model matrix is decomposed into truth label prediction and noise label prediction [configuration: $\lambda = 10$, $\beta = \gamma = 0.5$, $max\_iter = 500$].
- PAMB [29] [2023]: PAMB uses ECOC techniques to convert PML into a binary classification problem, avoiding the error-prone estimation of individual label confidences [configuration: $z = avg.\#CLs$, $L = 100\log_2(q)$].

https://doi.org/10.52202/085713-5670

Table 2: Comparision of CAPML with other state-of-the-art PML approaches on ***Average Precision*** (mean±std), where the best experimental performance (the larger the better) is shown in boldface.

Data Sets	avg.#CLs	CAPML	FBD-PML	LENFN	PAMB	PML-NI	PARTICLE	FPML
Mirflickr	3.35	**0.820±0.008**	0.815±0.007	0.800±0.009	0.791±0.019	0.786±0.009	0.813±0.136	0.814±0.009
Music emotion	5.29	**0.628±0.010**	0.607±0.011	0.608±0.010	0.626±0.011	0.608±0.012	0.506±0.016	0.458±0.015
Music style	6.04	0.743±0.016	0.740±0.017	**0.745±0.014**	0.741±0.007	0.739±0.015	0.657±0.012	0.566±0.090
YeastBP	5.93	**0.443±0.015**	0.406±0.021	0.423±0.017	0.356±0.022	0.404±0.022	0.168±0.016	0.328±0.012
YeastCC	1.39	**0.609±0.023**	0.584±0.011	0.603±0.019	0.556±0.024	0.454±0.025	0.348±0.016	0.458±0.032
YeastMF	1.04	**0.495±0.021**	0.431±0.017	0.484±0.023	0.405±0.014	0.418±0.016	0.228±0.012	0.326±0.009
emotions	3	**0.807±0.039**	0.783±0.008	0.783±0.035	0.804±0.017	0.777±0.028	0.747±0.035	0.663±0.020
	4	**0.787±0.027**	0.765±0.006	0.761±0.030	0.783±0.036	0.749±0.034	0.739±0.033	0.651±0.016
	5	**0.756±0.032**	0.751±0.005	0.746±0.033	0.749±0.026	0.680±0.039	0.702±0.037	0.654±0.029
birds	3	**0.627±0.058**	0.625±0.006	0.621±0.063	0.589±0.052	0.617±0.057	0.379±0.046	0.381±0.037
	4	**0.590±0.057**	0.586±0.003	0.584±0.050	0.564±0.044	0.572±0.041	0.419±0.046	0.373±0.020
	5	**0.589±0.048**	0.573±0.026	0.568±0.027	0.495±0.029	0.564±0.034	0.372±0.047	0.371±0.018
medical	5	0.876±0.027	0.864±0.024	**0.878±0.022**	0.815±0.012	0.866±0.024	0.754±0.047	0.838±0.025
	7	**0.866±0.031**	0.856±0.031	0.864±0.018	0.796±0.031	0.835±0.036	0.741±0.049	0.832±0.029
	9	**0.852±0.033**	0.842±0.020	0.851±0.028	0.771±0.011	0.798±0.031	0.715±0.022	0.817±0.019
image	2	**0.814±0.021**	0.778±0.026	0.777±0.023	0.798±0.024	0.770±0.020	0.743±0.070	0.711±0.018
	3	**0.792±0.018**	0.745±0.031	0.745±0.025	0.748±0.019	0.732±0.024	0.725±0.084	0.696±0.023
	4	**0.759±0.022**	0.691±0.011	0.671±0.027	0.711±0.026	0.653±0.011	0.668±0.091	0.670±0.022
yeast	7	0.760±0.018	0.734±0.007	0.756±0.020	**0.761±0.014**	0.746±0.017	0.754±0.013	0.732±0.016
	9	**0.755±0.021**	0.725±0.022	0.738±0.019	0.750±0.013	0.725±0.016	0.744±0.011	0.730±0.013
	11	**0.748±0.013**	0.704±0.015	0.719±0.018	0.741±0.013	0.692±0.013	0.728±0.013	0.698±0.011
corel5k	7	**0.306±0.015**	0.274±0.015	0.282±0.016	0.239±0.017	0.279±0.013	0.254±0.003	0.266±0.004
	9	**0.303±0.017**	0.267±0.015	0.273±0.015	0.230±0.008	0.273±0.013	0.234±0.004	0.264±0.001
	11	**0.299±0.017**	0.266±0.018	0.265±0.015	0.228±0.019	0.266±0.015	0.230±0.014	0.262±0.005

Table 3: Comparision of CAPML with other state-of-the-art PML approaches on ***Ranking Loss*** (mean±std), where the best experimental performance (the smaller the better) is shown in boldface.

Data Sets	avg.#CLs	CAPML	FBD-PML	LENFN	PAMB	PML-NI	PARTICLE	FPML
Mirflickr	3.35	**0.110±0.005**	0.126±0.006	0.121±0.004	0.112±0.038	0.126±0.007	0.127±0.103	0.115±0.006
Music emotion	5.29	0.236±0.007	0.249±0.012	0.245±0.012	**0.234±0.007**	0.246±0.008	0.362±0.014	0.410±0.005
Music style	6.04	**.0.135±0.010**	0.139±0.024	0.140±0.012	0.136±0.005	0.137±0.010	0.221±0.010	0.317±0.033
YeastBP	5.93	**0.203±0.009**	0.271±0.013	0.253±0.009	0.230±0.011	0.220±0.011	0.404±0.033	0.415±0.057
YeastCC	1.39	**0.167±0.015**	0.194±0.017	0.191±0.012	0.221±0.021	0.210±0.022	0.480±0.010	0.342±0.019
YeastMF	1.04	**0.218±0.017**	0.262±0.011	0.232±0.013	0.244±0.013	0.226±0.018	0.533±0.010	0.373±0.019
emotions	3	0.162±0.036	0.181±0.018	0.182±0.033	**0.160±0.022**	0.188±0.029	0.250±0.035	0.474±0.027
	4	**0.170±0.029**	0.197±0.013	0.194±0.029	0.178±0.031	0.211±0.027	0.263±0.029	0.446±0.027
	5	**0.205±0.032**	0.213±0.006	0.251±0.032	0.210±0.015	0.276±0.039	0.306±0.034	0.452±0.037
birds	3	**0.172±0.028**	0.176±0.035	0.183±0.007	0.196±0.042	0.177±0.033	0.322±0.033	0.333±0.041
	4	0.195±0.042	**0.195±0.034**	0.211±0.026	0.204±0.028	0.205±0.034	0.326±0.027	0.341±0.020
	5	**0.200±0.036**	0.207±0.038	0.223±0.012	0.229±0.025	0.219±0.036	0.359±0.036	0.328±0.018
medical	5	**0.029±0.011**	0.049±0.018	0.038±0.015	0.050±0.002	0.040±0.012	0.090±0.019	0.052±0.007
	7	**0.032±0.010**	0.047±0.016	0.045±0.016	0.062±0.019	0.052±0.013	0.111±0.022	0.058±0.009
	9	**0.038±0.013**	0.051±0.016	0.054±0.016	0.099±0.023	0.061±0.015	0.122±0.016	0.057±0.010
image	2	**0.155±0.021**	0.184±0.013	0.190±0.025	0.177±0.022	0.194±0.019	0.230±0.060	0.239±0.019
	3	**0.183±0.020**	0.206±0.011	0.214±0.028	0.217±0.015	0.230±0.024	0.261±0.070	0.254±0.018
	4	**0.213±0.022**	0.280±0.017	0.287±0.019	0.255±0.025	0.303±0.010	0.328±0.095	0.280±0.022
yeast	7	**0.173±0.013**	0.187±0.005	0.180±0.014	0.214±0.008	0.184±0.012	0.182±0.010	0.186±0.014
	9	**0.177±0.015**	0.198±0.019	0.192±0.018	0.211±0.007	0.202±0.016	0.189±0.009	0.191±0.012
	11	**0.183±0.012**	0.225±0.015	0.219±0.010	0.231±0.014	0.199±0.010	0.195±0.010	0.216±0.008
corel5k	7	0.176±0.007	**0.175±0.007**	0.223±0.012	0.310±0.034	0.215±0.008	0.317±0.008	0.255±0.013
	9	**0.184±0.007**	0.192±0.016	0.225±0.010	0.315±0.049	0.227±0.009	0.324±0.008	0.262±0.013
	11	**0.189±0.006**	0.203±0.012	0.231±0.009	0.319±0.059	0.232±0.008	0.330±0.007	0.268±0.011

- PML-LENFN [50] [2024]: PML-LENFN improves label quality by jointly analyzing local (neighbor) and global (distant) sample relationships, paired with a hybrid classifier combining linear and nonlinear components [$\lambda_1 = 10^{-5}$, $\lambda_2 = 1$, $\lambda_3 = 10^{-5}$, $\lambda_4 = 10^{-5}$].
- FBD-PML [19] [2025]: FBD-PML performs manifold alignment by simultaneously learning the prototypes of features and labels to achieve the smooth assumption. [configuration: $\lambda_1 = 10^{-4}$, $\lambda_2 = 1$, $\lambda_3 = 10^{-3}$, $\lambda_4 = 10^{-3}$,$\lambda_5 = 10^{-2}$]

For our CAPML approach λ is set to 0.5, α and β are searched in $\{0.001,\ 0.01,\ 0.1,\ 1,\ 10,\ 100\}$. For the feature mapping function $\psi(\cdot)$, we employ the Gaussian kernel with bandwidth parameter set to the average pairwise distance between samples.

3.2 Experimental Metrics and Results

In our experiment, we evaluate the performance of CAPML and other state-of-the-art baselines using five multi-label metrics: *Hamming loss*, *Ranking loss*, *One-error*, *Coverage*, and *Average precision*.

Details of these metrics can be found in 47. And we apply ten-fold cross-validation on each PML dataset and report the mean and standard deviation for all eight comparison approaches. Due to page limitations, our experimental evaluation focuses primarily on two representative metrics: *Average Precision* and *Ranking Loss*, which are complementary metrics providing comprehensive insights into both ranking quality and classification accuracy. Results for the other three metrics are provided in the Appendix. Complete results are presented in Tables 2 and 3, respectively.

Moreover, to verify the statistical significance of CAPML's performance advantages, we summarize *win/tie/loss* counts across all evaluation metrics against each competing approach at the 0.05 significance level, which is shown in Table 4. From the experimental results and subsequent statistical analysis, we can draw several significant conclusions:

- Across all evaluation metrics, our method achieves state-of-the-art performance in 86% of cases over the entire collection of 24 datasets. From Table 2 and 3, it can be observed that in 87.5% and 83.3% of the datasets, the approach consistently outperforms in the *Average Precision* and *Ranking Loss* metrics. The statistical advantages persist across diverse data characteristics—from high-dimensional biological datasets ($YeastBP$, $YeastCC$) to low-dimensional multimedia collections (Music, Image)—suggesting the method's superior generalization performance. Even on the challenging corel5k datasets, where label spaces are particularly sparse, CA-PML maintains its statistical edge over all competitors.
- From Table 4 , CAPML demonstrates convincing statistical dominance, winning in 607 out of 720 total comparisons (84.3%) while achieving statistical ties in all remaining cases. The smallest, though still significant, improvements are observed against LENFN, which indicates our effective guidance of label reliability indicator.
- FBD-PML represents a notable baseline as it is also a prototype-based approach. Despite this similarity, CAPML consistently outperforms FBD-PML, underscoring the efficacy of mutually prototype alignment strategy and stage-wise gradual label disambiguation.

Table 4: Win/tie/loss counts of pairwise t-test (at 0.05 aignificance level) on CAPML against others

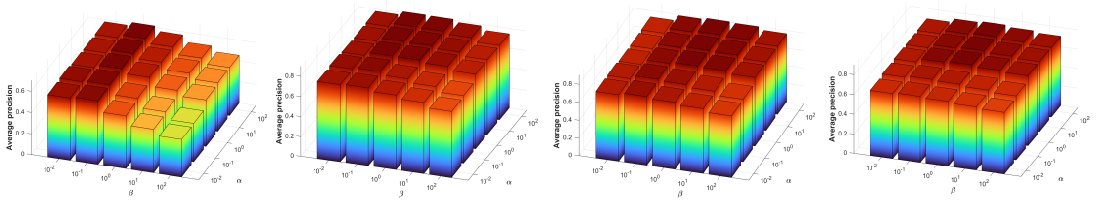
CAPML **against**	FBD-PML	LENFN	PAMB	PML-NI	PARTICLE	FPML
Hamming loss	21/3/0	16/8/0	21/3/0	20/4/0	20/4/0	18/6/0
Ranking loss	17/7/0	20/4/0	18/6/0	21/3/0	23/1/0	23/1/0
One-error	22/2/0	23/1/0	18/6/0	23/1/0	20/4/0	23/1/0
Coverage	20/4/0	17/7/0	18/6/0	19/5/0	23/1/0	22/2/0
Average precision	19/5/0	16/8/0	16/8/0	22/2/0	22/2/0	23/1/0
In Total	**99/21/0**	**93/27/0**	**91/29/0**	**105/15/0**	**108/12/0**	**109/11/0**

4 Further Analysis

Parameter Sensitivity We assess CAPML's robustness through parameter sensitivity analysis, examining how hyperparameters α and β influence the model's Average Precision performance. We vary each parameter individually while keeping the other fixed, with the results visualized as paired bar charts in Figure 2 for direct comparison. Figure 2 shows remarkable robustness of CAPML to parameter variations, with Average Precision remaining stable across a wide range of values. This stability is particularly evident for parameter α, where performance fluctuations are minimal within the reasonable range of [0.01, 100]. Similarly, when varying β, the model maintains consistent performance with only slight degradation at extreme values.

Computational Complexity The algorithm complexity is $O(T_0dnq + T_1(dq^2 + q^3) + T_2(nhq + h^2q + d^3 + nq))$. Stage one in Algorithm 1 involves SVD of the $d \times q$ matrix $O^T(MP)$ with complexity $O(dq^2)$ for the orthogonal Procrustes problem in Eq. (11), plus $O(q^3)$ for the Hungarian algorithm solving the assignment problem in Eq. (12), totaling $O(dq^2 + q^3)$ per iteration. Stage two involves $O(nhq + h^2q)$ for matrix operations in W-update including the inversion of $h \times h$ matrix, and $O(nq)$ for both R and Q updates involving element-wise operations. The dominant computational cost depends on the relative magnitudes of d, h, n, and q, but typically the $O(nhq)$ term dominates when datasets are large, making the method practically scalable.

 https://doi.org/10.52202/085713-5670



(a) *birds* (b) *image* (c) *emotions* (d) *Music style*

Figure 2: AP variations with parameters α and β on birds(avg.#CLs=2), emotions(avg.#CLs=3), image(avg.#CLs=2) and music style datasets.

Ablation Study To investigate the contribution of each key component in CAPML, we conduct ablation studies by comparing our full model with two variants: (1) **CAPML-ED**, which replaces our entropy-regularized fuzzy clustering with direct Euclidean distance between instances and prototypes computed from candidate label set to derive membership degrees; (2) **CAPML-NR**, which removes the orthogonal rotation matrix $\mathbf{H}$ from the prototype alignment process. (3) **CAPML-NA**, which sets permutation matrix P to identity matrix, removing prototype alignment. (4)**CAPML-CW**, which sets label enhancement indicator matrix C to all-ones, removing confidence indicator. Figure 3 presents the comparative results across the seven of all benchmark datasets on *Average Precision* and *Ranking loss*. These results validate the effectiveness of combining entropy-regularized clustering with orthogonal transformation for prototype learning.

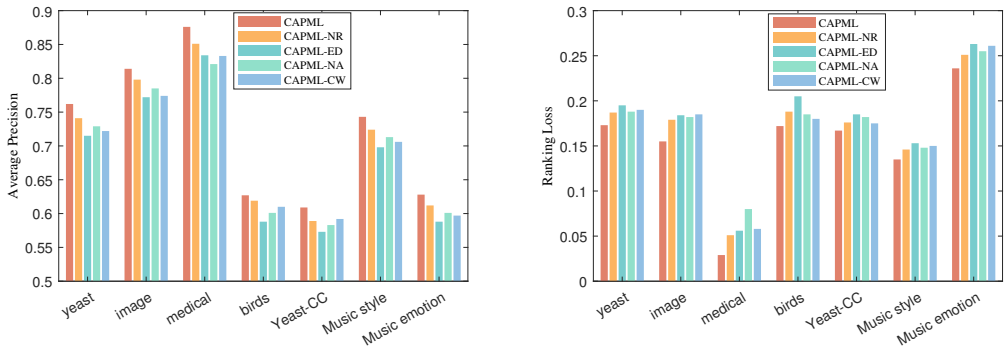


Figure 3: AP variations with parameters α and β on yeast(avg.#CLs=7), image(avg.#CLs=2), medical(avg.#CLs=5), birds(avg.#CLs=2), YeastCC, music style and music emotion datasets.

5 Conclusions

This paper introduces CAPML, a novel PML approach addressing label disambiguation through mutual prototype alignment. Unlike methods relying on noisy candidate labels alone, we align unsupervised prototypes capturing clean data structure with supervised prototypes containing semantic information. Through permutation matrices and orthogonal rotation, we transform fuzzy memberships into reliable confidence indicators operating external to classifier learning. This dual-prototype framework, combined with confidence-aware disambiguation and sparse regularization, effectively identifies true labels under challenging noise conditions. Comprehensive experiments demonstrate CAPML's significant advantages over state-of-the-art methods.

6 limitations

Despite these promising results, CAPML assumes fuzzy clustering discovers structure aligned with label semantics. Performance may degrade when high intra-class variability fragments categories into multiple clusters, when similar features cause distinct labels to merge, or when label counts considerably exceed natural cluster structures.

References

[1] Foteini Markatopoulou, Vasileios Mezaris, and Ioannis Patras. Implicit and explicit concept relations in deep neural networks for multi-label video/image annotation. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(6):1631–1644, 2019.

[2] Ying Chen, Ding Zhang, Tao Han, Xiaoliang Meng, Mianxin Gao, and Teng Wang. Label-guided cross-modal attention network for multi-label aerial image classification. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

[3] Yuyang Chai, Zhuang Li, Jiahui Liu, Lei Chen, Fei Li, Donghong Ji, and Chong Teng. Compositional generalization for multi-label text classification: A data-augmentation approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17727–17735, 2024.

[4] Jian Wang, Liang Qiao, Shichong Zhou, Jin Zhou, Jun Wang, Juncheng Li, Shihui Ying, Cai Chang, and Jun Shi. Weakly supervised lesion detection and diagnosis for breast cancers with partially annotated ultrasound images. *IEEE Transactions on Medical Imaging*, 2024.

[5] Ming-Kun Xie and Sheng-Jun Huang. Partial multi-label learning with noisy label identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3676–3687, 2021.

[6] Zhiqiang Kou, Jing Wang, Yuheng Jia, and Xin Geng. Inaccurate label distribution learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(10):10237–10249, 2024.

[7] Zhiqiang Kou, Si Qin, Hailin Wang, Mingkun Xie, Shuo Chen, Yuheng Jia, Tongliang Liu, Masashi Sugiyama, and Xin Geng. Label distribution learning with biased annotations by learning multi-label representation, 8 2025. Main Track.

[8] Zhiqiang Kou, Jing Wang, Yuheng Jia, Biao Liu, and Xin Geng. Instance-dependent inaccurate label distribution learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1):1425–1437, 2025.

[9] Ming-Kun Xie and Sheng-Jun Huang. Partial multi-label learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[10] Lijuan Sun, Songhe Feng, Tao Wang, Congyan Lang, and Yi Jin. Partial multi-label learning by low-rank and sparse decomposition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5016–5023, 2019.

[11] Zhenzhen Sun, Zexiang Chen, Jinghua Liu, Yewang Chen, and Yuanlong Yu. Partial multi-label feature selection via low-rank and sparse factorization with manifold learning. *Knowledge-Based Systems*, 296:111899, 2024.

[12] Min-Ling Zhang and Jun-Peng Fang. Partial multi-label learning via credible label elicitation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3587–3599, 2020.

[13] Ning Xu, Yun-Peng Liu, and Xin Geng. Partial multi-label learning with label distribution. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 6510–6517, 2020.

[14] Haobo Wang, Weiwei Liu, Yang Zhao, Chen Zhang, Tianlei Hu, and Gang Chen. Discriminative and correlative partial multi-label learning. In *IJCAI*, pages 3691–3697, 2019.

[15] Zhiqiang Kou, Jing Wang, Yuheng Jia, and Xin Geng. Progressive label enhancement. *Pattern Recognition*, 160:111172, 2025.

[16] Guoxian Yu, Xia Chen, Carlotta Domeniconi, Jun Wang, Zhao Li, Zili Zhang, and Xindong Wu. Feature-induced partial multi-label learning. In *2018 IEEE international conference on data mining (ICDM)*, pages 1398–1403. IEEE, 2018.

[17] Ziwei Li, Gengyu Lyu, and Songhe Feng. Partial multi-label learning via multi-subspace representation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 2612–2618, 2021.

 https://doi.org/10.52202/085713-5670

[18] Yan Hu, Xiaozhao Fang, Peipei Kang, Yonghao Chen, Yuting Fang, and Shengli Xie. Dual noise elimination and dynamic label correlation guided partial multi-label learning. *IEEE Transactions on Multimedia*, 2023.

[19] Xiaozhao Fang, Xi Hu, Yan Hu, Yonghao Chen, Shengli Xie, and Na Han. Fuzzy bifocal disambiguation for partial multi-label learning. *Neural Networks*, 185:107137, 2025.

[20] Yizhang Zou, Xuegang Hu, Peipei Li, and Yuhang Ge. Learning shared and non-redundant label-specific features for partial multi-label classification. *Information Sciences*, 656:119917, 2024.

[21] Qingqi Han, Liang Hu, and Wanfu Gao. Integrating label confidence-based feature selection for partial multi-label learning. *Pattern Recognition*, 161:111281, 2025.

[22] You Wu, Peipei Li, and Yizhang Zou. Partial multi-label feature selection with feature noise. *Pattern Recognition*, 162:111310, 2025.

[23] Peng Zhao, Shiyi Zhao, Xuyang Zhao, Huiting Liu, and Xia Ji. Partial multi-label learning based on sparse asymmetric label correlations. *Knowledge-Based Systems*, 245:108601, 2022.

[24] Lijuan Sun, Songhe Feng, Jun Liu, Gengyu Lyu, and Congyan Lang. Global-local label correlation for partial multi-label learning. *IEEE Transactions on Multimedia*, 24:581–593, 2021.

[25] Wenbin Qian, Yanqiang Tu, Jintao Huang, and Weiping Ding. Partial multi-label learning via robust feature selection and relevance fusion optimization. *Knowledge-Based Systems*, 286:111365, 2024.

[26] Ke Wang, Yahu Guan, Yunyu Xie, Zhaohong Jia, Hong Ye, Zhangling Duan, and Dong Liang. Partial multi-label learning with label and classifier correlations. *Information Sciences*, 712:122101, 2025.

[27] Fuchao Yang, Yuheng Jia, Hui Liu, Yongqiang Dong, and Junhui Hou. Noisy label removal for partial multi-label learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3724–3735, 2024.

[28] Feng Li, Shengfei Shi, and Hongzhi Wang. Partial multi-label learning via specific label disambiguation. *Knowledge-Based Systems*, 250:109093, 2022.

[29] Bing-Qing Liu, Bin-Bin Jia, and Min-Ling Zhang. Towards enabling binary decomposition for partial multi-label learning. *IEEE transactions on pattern analysis and machine intelligence*, 2023.

[30] You Wu, Peipei Li, and Yizhang Zou. Partial multi-label feature selection with feature noise. *Pattern Recognition*, 162:111310, 2025.

[31] Lihua Zhou, Guowang Du, Kevin Lü, Lizheng Wang, and Jingwei Du. A survey and an empirical evaluation of multi-view clustering approaches. *ACM Computing Surveys*, 56(7):1–38, 2024.

[32] Ling Ding, Chao Li, Di Jin, and Shifei Ding. Survey of spectral clustering based on graph theory. *Pattern Recognition*, page 110366, 2024.

[33] Hossein Esfandiari, Amin Karbasi, Vahab Mirrokni, Grigoris Velegkas, and Felix Zhou. Replicable clustering. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 39277–39320. Curran Associates, Inc., 2023.

[34] Jicong Fan, Yiheng Tu, Zhao Zhang, Mingbo Zhao, and Haijun Zhang. A simple approach to automated spectral clustering. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 9907–9921. Curran Associates, Inc., 2022.

[35] Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11600–11609, 2023.

[36] Ziniu Yin, Yanglin Feng, Ming Yan, Xiaomin Song, Dezhong Peng, and Xu Wang. Roda: Robust domain alignment for cross-domain retrieval against label noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 9535–9543, 2025.

[37] Chao Su, Huiming Zheng, Dezhong Peng, and Xu Wang. Dica: Disambiguated contrastive alignment for cross-modal retrieval with partial labels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 20610–20618, 2025.

[38] Haoran Liu, Ying Ma, Ming Yan, Yingke Chen, Dezhong Peng, and Xu Wang. Dida: Disambiguated domain alignment for cross-domain retrieval with partial labels. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 3612–3620, 2024.

[39] Enrique H Ruspini. A new approach to clustering. *Information and control*, 15(1):22–32, 1969.

[40] Miyamoto Sadaaki and Mukaidono Masao. Fuzzy c-means as a regularization and maximum entropy approach. In *Proceedings of the 7th international fuzzy systems association world congress (IFSA'97)*, volume 2, pages 86–92, 1997.

[41] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[42] Peter H Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

[43] Jie Zhou, Ge Yuan, Can Gao, Xizhao Wang, Jianhua Dai, and Witold Pedrycz. Fuzzy clustering guided by spectral rotation and scaling. *IEEE Transactions on Cybernetics*, 2024.

[44] G Ayorkor Mills-Tettey, Anthony Stentz, and M Bernardine Dias. The dynamic hungarian algorithm for the assignment problem with changing costs. *Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-07-27*, 2007.

[45] Alessandro Mirone and Pierre Paleo. A conjugate subgradient algorithm with adaptive preconditioning for the least absolute shrinkage and selection operator minimization. *Computational Mathematics and Mathematical Physics*, 57:739–748, 2017.

[46] Stephen J Wright, Robert D Nowak, and Mário AT Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on signal processing*, 57(7):2479–2493, 2009.

[47] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8):1819–1837, 2013.

[48] Zhi-Hua Zhou and Min-Ling Zhang. Multi-label learning., 2017.

[49] Mark J Huiskes and Michael S Lew. The mir flickr retrieval evaluation. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, pages 39–43, 2008.

[50] Yu Chen, Yanan Wu, Na Han, Xiaozhao Fang, Bingzhi Chen, and Jie Wen. Partial multi-label learning based on near-far neighborhood label enhancement and nonlinear guidance. In *ACM Multimedia 2024*, 2024.

 https://doi.org/10.52202/085713-5670

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the paper's main contributions: (1)We make the first investigation into the prototype misalignment between prototypes derived from fuzzy clustering and prototypes computed from candidate label set in PML tasks. Our work introduces a transformation mechanism that successfully bridges these two prototype spaces, enabling effective alignment and discovery of their intrinsic correspondence relationships despite noisy supervision. (2)We design a confidence-aware process that converts fuzzy label membership degrees into label reliability indicator values, guiding our classifier training with sparse $\ell 2$,1-norm constraint that enhance feature selection while reducing overfitting to noisy labels. (3)Extensive empirical evaluation demonstrates our method's efficacy in resolving label ambiguity and prototype misalignment problem in PML, even with high noise rates and sparse positive samples, and superior generalization performance.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In the conclusion section (p.9), the paper acknowledges limitations has shown: increased computational complexity for very large datasets and potential suboptimality when label counts significantly exceed natural cluster structures.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.

- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The paper provides formal mathematical formulations with clear assumptions in Section 2 , including notation definitions and constraints. And there are some formula derivations. The optimization process is thoroughly described in Section 2 with detailed equations and algorithmic steps.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Section 3 provides comprehensive experimental details including dataset characteristics and url (Table 1), parameter settings, comparison methods, evaluation metrics, and the implementation of 10-fold cross-validation, which should be sufffcient to reproduce the results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

https://doi.org/10.52202/085713-5670

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: While the paper uses publicly available datasets (from mulan.sourceforge.net and palm.seu.edu.cn), there is no explicit mention of releasing the implementation code for the proposed CAPML method.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 4.1 speciffes implementation details including 10-fold cross-validation for all datasets, hyperparameter search ranges for α and β and comparison with baseline methods using their recommended conffgurations.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper reports mean and standard deviation for all experimental results (Tables 2 and 3) and we summarize *win/tie/loss* counts across all evaluation metrics against each competing approach at the 0.05 significance level, which is shown in Table 4.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: Due to space limitation, the paper does not provide speciffc details about computing resources such as CPU/GPU speciffcations, memory requirements, or execution times for the experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

https://doi.org/10.52202/085713-5670

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: The research focuses on algorithmic improvements for partial multi-label learning without raising ethical concerns. It uses standard public datasets and comparison methods, and does not involve sensitive data, human subjects, or applications that could cause harm.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: While our paper thoroughly addresses technical contributions and future research directions, we do not explicitly discuss potential societal impacts (either positive or negative) of the proposed method, as our work focuses primarily on algorithmic advances in the machine learning domain rather than specific applications with direct societal implications.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [No]

Justification: This paper introduces a partial multi-label learning algorithm using standard benchmark datasets, presenting no high-risk models or significant ethical concerns requiring safeguards.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper properly cites the sources of datasets used (Table 1 with footnotes to source URLs), and references original papers for comparison methods.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: Our paper introduces a new algorithm (CAPML) but does not release new datasets, code, or models requiring documentation beyond what is presented in the paper itself.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

https://doi.org/10.52202/085713-5670

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [No]

Justification: This research does not involve crowdsourcing or human subjects. It focuses on algorithmic development and evaluation using existing benchmark datasets.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The research does not involve human subjects, so IRB approval was not required for this study.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The paper does not utilize large language models in its methodology. Our CAPML method is based entirely on mathematical formulations and traditional machine learning approaches.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.