

Article

Graph-Regularized Low-Rank Label Correlation Learning with Label-Specific Features for Missing Labels

Tianli Li ^{1,2}, Mohammad Faizul Nasrudin ^{1,*}, Xing Peng ^{1,2,†} and Huifen Xing ^{1,2,†}

¹ Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia; p129508@siswa.ukm.edu.my (T.L.); p121329@siswa.ukm.edu.my (X.P.); p131528@siswa.ukm.edu.my (H.X.)

² School of Computer Science and Artificial Intelligence, Chaohu University, Hefei 238024, China

* Correspondence: mfn@ukm.edu.my

† These authors contributed equally to this work.

Abstract

Missing labels are common in multi-label learning and can bias both label-correlation estimation and classifier induction. Existing missing-label methods often recover incomplete supervision mainly through global label correlations. However, correlation-driven recovery alone may produce over-smoothed supervision when annotations are sparse, while label-specific discriminative evidence may be weakened. To address this problem, we propose GLCS, a graph-regularized low-rank correlation learning framework with label-specific features for multi-label learning with missing labels. GLCS first uses the observed entries as reliable supervision sources and propagates them through a learned label correlation matrix. It then jointly learns sparse label-specific predictors, low-rank label correlations, and a label graph regularizer induced by the learned correlations. In this way, global label dependencies, local label-structure consistency, and label-wise discriminative features are optimized in a unified objective. The resulting problem is solved by an alternating proximal optimization scheme with soft thresholding for sparse predictors and singular value thresholding for low-rank correlations. Experiments on twelve benchmark datasets under three missing-label ratios show that GLCS obtains strong average performance across AP, AUC, CV, HL, OE, and RL, especially under high missing rates.

Keywords: multi-label learning; missing labels; label correlation; label-specific features; low-rank learning; graph regularization

1. Introduction

Multi-label learning (MLL) aims to assign multiple relevant labels to each instance, thereby capturing semantic ambiguity and overlap widely observed in real-world data [1]. It has achieved promising performance in diverse applications, such as image recognition [2], text classification [3,4], and emotion recognition [5], where a single instance may naturally involve multiple concepts. Compared with single-label prediction, MLL provides a more expressive learning paradigm by modeling the mapping from instances to a set of correlated labels. Over the past decade, numerous algorithms have been proposed to improve MLL from the perspectives of label dependency modeling, dimensionality reduction, and discriminative feature learning [6,7].

Despite these advances, obtaining complete multi-label annotations is often prohibitively expensive in practice [8]. Consequently, the training label matrix is frequently partially observed: a portion of label entries may be missing due to annotation budget,



Academic Editor: Carson K. Leung

Received: 26 May 2026

Revised: 12 July 2026

Accepted: 24 July 2026

Published: 27 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

ambiguity, or randomness, and the unobserved entries should be treated as unknown rather than reliable negatives. To cope with such incomplete supervision, a mainstream line of research exploits latent structures in the label space and recovers informative supervision from the observed entries. A common assumption is that the unknown true label matrix has a low-rank structure, so that missing entries can be inferred via label subspace learning or matrix completion style objectives [9]. In addition to the standard missing entry settings, some works further consider harder supervision patterns where a subset of labels may be completely unobserved during training, and infer them through structured priors and distributional regularities in the label space [10,11]. Related formulations have also been extended to settings with simultaneously missing features and labels to better reflect weak supervision in practice [12]. While these approaches can leverage global label dependencies effectively, purely correlation-driven refinement may become unreliable under severe missingness: overly aggressive propagation can over-smooth the supervision, and the resulting predictors may become ambiguous when label-discriminative evidence is not sufficiently preserved.

In parallel, label-specific feature (LSF) learning signals have been increasingly recognized as a crucial ingredient for strong multi-label generalization [13]. Instead of enforcing a fully shared representation for all labels, LSF modeling aims to capture label-related discriminative factors, enabling the learner to distinguish subtle differences between labels and reduce feature interference [14]. Recent progress ranges from deep LSF representations [15] to unified frameworks that generate or refine LSF evidence for improved interpretability and robustness when labels exhibit heterogeneous semantic cues [16]. Meanwhile, emerging attempts also incorporate soft-label recovery into LSF learning to tolerate imperfect supervision [17]. However, most existing LSF methods still rely on reasonably accurate supervision; once labels are missing, biased or spurious signals may distort LSF patterns and in turn deteriorate downstream prediction.

These observations motivate a unified view of missing-label MLL: label refinement and LSF learning should be optimized together rather than treated as two separate stages. In this paper, we propose GLCS, a graph-regularized framework that jointly learns a low-rank label correlation matrix and LSF under incomplete annotations. Unlike representative missing-label MLL methods such as SGMML, LRMML, DM2L, and MSWL, which usually emphasize one specific aspect of the problem, GLCS jointly integrates correlation-based supervision propagation, low-rank label correlation learning, correlation-induced graph regularization, and sparse label-specific feature learning in one objective. Specifically, GLCS integrates: (i) correlation-based supervision propagation through a learned label correlation matrix, (ii) nuclear-norm regularization that encourages low-rank correlations under incomplete supervision, (iii) a label graph induced by the learned correlations to preserve label-structure consistency and alleviate over-smoothing, and (iv) LSF sparsity that preserves discriminative evidence while reducing feature redundancy.

Our main contributions are summarized as follows:

- We propose GLCS, a unified formulation that jointly performs correlation-based missing-label refinement, low-rank label correlation learning, graph-regularized structure preservation, and LSF learning for discriminative multi-label prediction.
- We introduce a graph regularization induced by learned correlations, complementing low-rank correlation modeling to improve robustness and alleviate over-smoothing under severe label missingness.
- Experiments on twelve benchmark datasets under three missing-label ratios show that GLCS achieves stable performance across both ranking-based and classification-based metrics.

The remainder of this paper is organized as follows. In Section 2, we review recent advances in missing-label MLL, label correlation modeling, and LSF learning. Section 3 presents the proposed GLCS framework, including the objective function and the corresponding optimization procedure. In Section 4, we describe the experimental settings, evaluation metrics, and comparative results on benchmark datasets under different missing-label rates. Additional experimental analyses further investigate statistical significance, running time, parameter sensitivity, and convergence behavior. Finally, Section 5 summarizes this work and outlines possible future directions.

2. Related Work

2.1. Missing-Label Multi-Label Learning

In real-world multi-label datasets, annotations are often incomplete, where a subset of label entries is unobserved. To learn reliable predictors under such weak supervision, existing methods mainly fall into two families: (i) two-step approaches that first complete labels and then train a classifier, and (ii) joint approaches that integrate label refinement and predictor learning into a unified optimization.

Two-step approaches typically recover a latent label matrix by exploiting structural assumptions or relational information, and then learn predictors on the recovered supervision. Representative strategies include sparse reconstruction-based completion [18], matrix completion with side information [19], and mixed-graph modeling that couples instance and label relations [20]. Margin-aware formulations further treat unobserved entries as ambiguous and enforce max-margin constraints to improve recovery reliability near decision boundaries [21,22]. To handle large-scale datasets, efficient semi-supervised frameworks and scalable probabilistic models have also been developed [23,24]. Although conceptually simple, this pipeline can suffer from error propagation: inaccuracies in completion may bias the subsequent classifier and amplify mistakes.

Joint learning approaches directly couple prediction and refinement within a single objective to alleviate error propagation. Mixed-dependency graph-based models propagate supervision over instance and label relations to learn under missing labels [25]. Semi-supervised objectives further exploit manifold structure to reduce the bias induced by severe sparsity [26], whereas safe learning provides provable guarantees that exploiting weak supervision will not degrade performance relative to a baseline [27]. Beyond correlation-based recovery, discriminant formulations explicitly sharpen label boundaries to improve separability [28], and low-rank label subspace modeling provides an effective route to capture global correlations and recover missing entries [29]. More recently, global and local structural regularization and correlation guided LSF learning have been studied to better handle imbalanced and incomplete multi-label annotations [30,31]. Overall, these studies support integrating refinement into learning. However, with severe missingness, achieving both structure consistency for stable refinement and label discriminative evidence for accurate prediction is still difficult. GLCS follows the joint-learning paradigm but further emphasizes discriminative protection via LSF sparsity, ensuring that refinement does not dilute LSF evidence when supervision is highly incomplete.

2.2. Label Correlation Modeling Under Missing Labels

Since labels are inherently dependent, exploiting label correlations is central to missing-label MLL. Probabilistic dependence modeling provides a principled view of label co-occurrence in the output space [32]. Under weak supervision, recent joint frameworks explicitly learn higher-order correlations from partially observed labels to augment supervision while inducing predictors. For example, LSML couples missing-label recovery with LSF representation learning [33], and an embedded packaging strategy further in-

tegrates higher-order dependency completion with classifier induction in a unified optimization [34]. Beyond single-view settings, correlation learning and label recovery can also be co-optimized for multi-view multi-label data with incomplete labels [35]. Moreover, factorization-based formulations can be viewed as learning a low-rank label subspace, where manifold regularization preserves instance similarity and the geometry of label relations during recovery [36]. Building on such global-structure priors, more recent studies further introduce local label relations to capture instance-dependent dependencies under weak supervision [37]. In contrast to approaches that rely on a fixed label graph or externally specified correlations, GLCS learns a low-rank correlation matrix and induces a label-graph smoothness regularizer from it, aligning structure preservation with the evolving correlation estimate and alleviating over-smoothing during refinement.

2.3. Extremely Sparse Supervision

Some studies investigate even more challenging weak supervision regimes, where only a tiny fraction of informative labels is observed. In such cases, the unobserved entries should be treated as unknown rather than reliable negatives, otherwise the learner can be heavily biased by false-negative contamination. This issue becomes particularly pronounced in single-positive annotation, where each instance is provided with only one positive label and all remaining labels are left unannotated; early work shows that debiasing the training objective under this supervision can still yield competitive multi-label predictors [38]. Subsequent studies further mitigate false negatives through semantic contrastive bootstrapping [39] or by explicitly excluding suspected false negatives during optimization [40]. Meanwhile, vision-language-based pseudo-labeling offers a complementary way to enrich supervision while keeping unobserved entries flexible [41]. Beyond the closed set assumption, open-set multi-label recognition additionally considers the emergence of unseen labels at test time [42]. Although these regimes are not identical to the random missing-entry scenario considered in this paper, they share the same core difficulty that reliable discriminative information must be preserved when supervision is extremely sparse. In this sense, GLCS is closely related to extremely sparse supervision because it refines missing labels through low-rank correlations while using sparse LSF learning to preserve label-wise discriminative evidence.

2.4. LSF Learning Under Weak Supervision

In addition to refinement based on label propagation, LSF learning extracts discriminative evidence for each label and reduces interference among labels. Early work such as LIFT demonstrates the effectiveness of LSF for multi-label classification [43]. Later studies investigate joint learning of shared and LSF [44]. Recent progress includes collaborative learning of label semantics with deep LSF representations [45], and unified frameworks that generate or refine LSF evidence to improve interpretability and robustness [46]. With incomplete annotations, combining shared and LSF representations with recovery-oriented regularization has been shown to mitigate missing supervision [47]. Related theory also provides a generalization perspective for LSF representation learning [48]. Many existing LSF methods still rely on relatively reliable supervision, or use recovered labels as fixed pseudo ground truth. GLCS carries out refinement and label-specific learning in a single procedure, allowing refined supervision and discriminative evidence to reinforce each other and reducing sensitivity to biased signals.

To summarize the above related methods and clarify the distinction of GLCS, Table 1 provides a concise comparison.

Table 1. Comparison of representative related methods.

Method	Main Idea	Difference from GLCS
SGMML [49]	Learns sparse global structures for label-specific features.	Does not jointly learn a low-rank correlation matrix or construct a correlation-induced label graph.
SLR-LSF [17]	Recovers soft labels and then learns label-specific features.	Uses a staged recovery-then-learning strategy, while GLCS jointly optimizes supervision propagation, low-rank correlation learning, graph regularization, and LSF learning.
LRMML [9]	Exploits low-rank label recovery for missing labels.	Mainly focuses on label recovery, while GLCS couples recovery with sparse label-specific predictors and graph-based structure preservation.
DM2L [28]	Improves missing-label learning through discriminant expansion and shrinkage.	Emphasizes discriminability, while GLCS explicitly propagates observed supervision through a learned correlation matrix.
MSWL [50]	Uses manifold-regularized sparse modeling for weak labels.	Relies on manifold-based structural regularization, while GLCS induces the label graph from the learned correlation matrix.
LFLI [31]	Uses correlation information for missing-label learning with label-specific features.	Closely related recent baseline; GLCS further combines $\mathcal{P}_M(Y)C$ supervision propagation, low-rank C , and a correlation-induced graph regularizer.
GLCS	Jointly learns propagated supervision, low-rank label correlations, sparse LSF predictors, and a learned label graph.	Integrates missing-label refinement, global low-rank dependency, local graph consistency, and label-specific feature sparsity in one objective.

Beyond these representative missing-label MLL methods, we further discuss the relation between GLCS and extreme multi-label classification with missing-positive learning.

2.5. Extreme Multi-Label Classification and Missing-Positive Learning

Extreme multi-label classification (XMC) is another related research direction for learning with incomplete labels. XMC usually involves very large label spaces, long-tail label distributions, and severe missing-positive annotations. In such settings, treating all unobserved labels as negatives can introduce strong false-negative bias, and prior studies have proposed propensity-based or unbiased loss functions to address this issue [51,52]. The random missing-entry setting considered in this paper is related to this line of work, but differs from XMC in scale and assumptions. In contrast, GLCS is designed for moderate-scale multi-label datasets where an explicit $l \times l$ label-correlation matrix is computationally feasible. Directly applying GLCS to XMC would be expensive because the dense correlation matrix and SVT operation scale quadratically with the number of labels. A scalable extension could approximate C using sparse label graphs, low-rank factorization, label embeddings, label-side features, or retrieval-based shortlisting.

Overall, prior studies demonstrate that (i) correlation and structure refinement is beneficial for missing-label MLL, and (ii) LSF modeling improves discriminability. However, under severe missingness, naive propagation may cause over-smoothing, and biased supervision can distort LSF patterns. The proposed GLCS addresses these issues by learning a low-rank label correlation matrix for refinement and LSF for discrimination, while further enforcing a label-graph regularization induced by learned correlations. By coupling correlation-based refinement with LSF sparsity and structure preservation, GLCS improves robustness when supervision is highly incomplete.

3. Proposed Approach

In MLL, each instance can be associated with multiple labels. Let the training data be $X \in \mathbb{R}^{n \times d}$, and the partially observed label matrix be $Y \in \mathbb{R}^{n \times l}$, where n , d , and l denote the numbers of instances, features, and labels, respectively. We consider a random missing-entry supervision scenario on the whole label matrix, where each label entry may be unobserved due to annotation randomness or budget limitation. Both positive and negative entries may be missing. An indicator mask $M \in \{0, 1\}^{n \times l}$ is introduced, where $M_{ij} = 1$ if the entry Y_{ij} is observed, and $M_{ij} = 0$ otherwise. Under a missing rate r , each label entry is observed with probability $1 - r$.

We define the masking operator as

$$\mathcal{P}_M(A) = M \odot A, \quad (1)$$

where $\odot$ denotes the Hadamard product. Throughout this paper, unobserved entries are treated as unknown rather than reliable negatives. We learn a linear predictor

$$f(\mathbf{x}) = \mathbf{x}^\top W, \quad W \in \mathbb{R}^{d \times l}, \quad (2)$$

where the prediction score matrix for all training instances is XW .

If the complete label matrix were fully observed, a direct formulation would be to minimize the squared prediction error between XW and Y :

$$\min_W \frac{1}{2} \|XW - Y\|_F^2. \quad (3)$$

However, this objective is not suitable for missing-label learning because the complete label matrix is unavailable. Treating unobserved entries as negative labels would introduce false-negative noise and lead to biased supervision.

A straightforward way to avoid this bias is to restrict the supervision to the observed entries:

$$\min_W \frac{1}{2} \|\mathcal{P}_M(XW - Y)\|_F^2. \quad (4)$$

This masked loss avoids penalizing unobserved entries, but it uses only the available annotations and cannot infer informative supervision for missing label entries. Therefore, it does not fully exploit the dependency structure among labels.

To transfer information from observed labels to related labels, GLCS introduces a label correlation matrix $C \in \mathbb{R}^{l \times l}$. The observed label matrix is propagated through C to produce correlation-aware soft supervision:

$$\hat{Y} = \mathcal{P}_M(Y)C. \quad (5)$$

Here, $\mathcal{P}_M(Y)$ ensures that only observed entries are used as reliable sources, and C models how label information is transferred across labels. Using this propagated supervision as a surrogate target gives

$$\min_{W, C} \frac{1}{2} \|XW - \mathcal{P}_M(Y)C\|_F^2. \quad (6)$$

This term couples predictor learning and missing-label refinement: W learns from the propagated soft supervision, while C determines the label-to-label propagation pattern. The propagated term $\mathcal{P}_M(Y)C$ is used as correlation-aware soft supervision rather than as a fully recovered ground-truth label matrix. Although $\mathcal{P}_M(Y)$ uses only observed entries as propagation sources, an inaccurate C may still introduce noise or amplify initial annotation errors. To reduce this risk, GLCS regularizes C with the nuclear norm, constrains propagation through the correlation-induced graph regularizer, and stabilizes C by nonnegative projection with unit diagonal entries.

In multi-label learning, different labels usually depend on different feature subsets. If all labels share the same feature space without discrimination, irrelevant features may interfere with label prediction, especially when the observed annotations are sparse. To preserve label-wise discriminative evidence, we impose an element-wise ℓ_1 regularization on W :

$$\min_{W, C} \frac{1}{2} \|XW - \mathcal{P}_M(Y)C\|_F^2 + \lambda_1 \|W\|_1. \quad (7)$$

Since each column of W corresponds to one label, the sparsity term encourages each label to select its own relevant feature subset. This provides the label-specific feature component of GLCS and reduces feature interference among labels.

The correlation matrix should capture stable label dependencies rather than arbitrary noisy relations. Multi-label data often contain latent low-dimensional label structures, and such structures can be exploited to make supervision propagation more robust. We therefore regularize C with the nuclear norm:

$$\min_{W,C} \frac{1}{2} \|XW - \mathcal{P}_M(Y)C\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 \|C\|_*. \quad (8)$$

The nuclear norm $\|C\|_*$ promotes a low-rank correlation structure, suppresses noisy dependencies, and improves the stability of label propagation under incomplete supervision.

Low-rank correlation learning and graph regularization are complementary. The low-rank constraint on C models global latent label dependencies and suppresses noisy correlations, whereas the graph regularizer preserves local consistency among strongly related labels in the predictor space. Unlike classical structural regularization based on a fixed or externally defined graph, GLCS constructs the label graph from the learned correlation matrix, so the structure used for smoothing is adaptively coupled with the missing-label refinement process. This adaptive design is preferable when reliable external label relations are unavailable under incomplete annotations. Specifically, we first symmetrize C :

$$\bar{C} = \frac{1}{2} (C + C^\top), \quad (9)$$

define the degree matrix $\bar{D}$ by

$$\bar{D}_{ii} = \sum_{j=1}^l \bar{C}_{ij}, \quad (10)$$

and construct the graph Laplacian

$$L = \bar{D} - \bar{C}. \quad (11)$$

The corresponding graph regularizer is

$$\text{tr}(WLW^\top). \quad (12)$$

When $L = \bar{D} - \bar{C}$, this term can be equivalently written as

$$\text{tr}(WLW^\top) = \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \bar{C}_{ij} \|w_i - w_j\|_2^2, \quad (13)$$

where w_i denotes the predictor vector associated with the i -th label, i.e., the i -th column of W . This form shows that labels with stronger correlation are encouraged to have more consistent predictors. Thus, the graph regularizer preserves local label structure in the predictor space and reduces the risk of over-smoothing caused by unconstrained correlation propagation.

Combining correlation-propagated supervision, sparse label-specific prediction, low-rank correlation learning, and graph-based structure preservation, we obtain the final GLCS objective:

$$\min_{W,C} \frac{1}{2} \|XW - \mathcal{P}_M(Y)C\|_F^2 + \lambda_1 \|W\|_1 + \lambda_2 \|C\|_* + \lambda_3 \text{tr}(WLW^\top). \quad (14)$$

The first term fits the prediction scores to correlation-propagated supervision, the second term selects label-specific features through sparsity, the third term models global low-rank label correlations, and the fourth term preserves local label-structure consistency. Different from two-stage strategies that first recover labels and then train classifiers, GLCS jointly optimizes missing-label refinement and label-specific feature learning, allowing refined supervision and discriminative predictors to reinforce each other.

To stabilize correlation learning and avoid degenerate scaling, we project C after each update:

$$C \leftarrow \max(C, 0), \quad C \leftarrow C - \text{diag}(\text{diag}(C)) + I. \quad (15)$$

This projection enforces nonnegativity and fixes the diagonal entries to one, so each label remains anchored to itself while off-diagonal entries capture cross-label dependencies. Since the label graph is induced from the learned correlation matrix, its reliability depends on the quality of C . GLCS improves stability by symmetrizing C , regularizing it with the nuclear norm, and projecting it to nonnegative values with unit diagonal entries. When C is unreliable under extremely sparse observations, the induced graph may be less informative, and its influence is controlled by λ_3 .

3.1. Alternating Optimization

Directly optimizing Equation (14) is challenging because C affects both the supervision-propagation term and the Laplacian L . GLCS therefore uses an outer-loop alternating scheme. At iteration t , the Laplacian L^t is constructed from C^t . Then, W is updated with fixed (C^t, L^t) , C is updated with fixed W , and the Laplacian is reconstructed from the new C^{t+1} .

With C and L fixed, the W subproblem is

$$\min_W \frac{1}{2} \|XW - \mathcal{P}_M(Y)C\|_F^2 + \lambda_3 \text{tr}(WLW^\top) + \lambda_1 \|W\|_1. \quad (16)$$

Let the smooth part be

$$F(W) = \frac{1}{2} \|XW - \mathcal{P}_M(Y)C\|_F^2 + \lambda_3 \text{tr}(WLW^\top). \quad (17)$$

Since L is symmetric by construction, its gradient is

$$\nabla F(W) = X^\top(XW - \mathcal{P}_M(Y)C) + 2\lambda_3 WL. \quad (18)$$

A safe Lipschitz upper bound is

$$\mathcal{L}_W = \|X^\top X\|_2 + 2\lambda_3 \|L\|_2. \quad (19)$$

The proximal gradient update for W is

$$W^{k+1} = \mathcal{S}_{\lambda_1/\mathcal{L}_W} \left(W^k - \frac{1}{\mathcal{L}_W} \nabla F(W^k) \right), \quad (20)$$

where $\mathcal{S}_\tau(\cdot)$ is the element-wise soft-thresholding operator:

$$\mathcal{S}_\tau(u) = \text{sign}(u) \max(|u| - \tau, 0). \quad (21)$$

Nesterov acceleration is applied in the implementation to improve convergence speed.

With W fixed, the C -step optimizes a proximal surrogate with respect to the correlation-propagation loss and the low-rank prior. The graph term is updated indirectly by recon-

structing the Laplacian after the new C is obtained. This avoids differentiating through the Laplacian construction and keeps the update efficient. The C subproblem is

$$\min_C \frac{1}{2} \|XW - \mathcal{P}_M(Y)C\|_F^2 + \lambda_2 \|C\|_*. \quad (22)$$

Define

$$G(C) = \frac{1}{2} \|XW - \mathcal{P}_M(Y)C\|_F^2. \quad (23)$$

The gradient is

$$\nabla G(C) = \mathcal{P}_M(Y)^\top (\mathcal{P}_M(Y)C - XW). \quad (24)$$

A safe Lipschitz upper bound is

$$\mathcal{L}_C = \|\mathcal{P}_M(Y)^\top \mathcal{P}_M(Y)\|_2 \leq \|Y^\top Y\|_2. \quad (25)$$

The nuclear norm proximal update is given by singular value thresholding:

$$C^{k+1} = \text{SVT}_{\lambda_2/\mathcal{L}_C} \left(C^k - \frac{1}{\mathcal{L}_C} \nabla G(C^k) \right). \quad (26)$$

For $Z = U \text{diag}(\sigma) V^\top$, the SVT operator is

$$\text{SVT}_\tau(Z) = U \text{diag}(\max(\sigma - \tau, 0)) V^\top. \quad (27)$$

After SVT, the projection in Equation (15) is applied, and L is reconstructed by Equations (9)–(11) for the next outer iteration.

The overall optimization procedure of GLCS is summarized in Algorithm 1.

Algorithm 1 GLCS: Alternating optimization

Input: $X, Y, M, \lambda_1, \lambda_2, \lambda_3$, maximum outer iterations T , tolerance ε

Output: W, C

- 1: Initialize W^0 and C^0
 - 2: Project C^0 by Equation (15); construct L^0 by Equations (9)–(11)
 - 3: Compute F_{obj}^0 according to Equation (14)
 - 4: **for** $t = 0, 1, \dots, T - 1$ **do**
 - 5: Update W with fixed C^t and L^t
 - 6: Compute $\nabla F(W)$ by Equation (18) and set $\mathcal{L}_W$ by Equation (19)
 - 7: Perform accelerated proximal gradient steps using Equations (20) and (21)
 - 8: Update C with fixed W^{t+1}
 - 9: Compute $\nabla G(C)$ by Equation (24) and set $\mathcal{L}_C$ by Equation (25)
 - 10: Apply SVT update by Equation (26) with the operator in Equation (27)
 - 11: Project C^{t+1} by Equation (15)
 - 12: Construct L^{t+1} by Equations (9)–(11)
 - 13: Compute F_{obj}^{t+1} according to Equation (14)
 - 14: **if** $\frac{|F_{\text{obj}}^{t+1} - F_{\text{obj}}^t|}{\max(F_{\text{obj}}^t, \varepsilon)} < \varepsilon$ **then**
 - 15: **break**
 - 16: **end if**
 - 17: **end for**
-

Before analyzing the computational complexity, we briefly discuss the convergence behavior of the alternating optimization procedure. The objective in Equation (14) is not jointly convex because the graph Laplacian is reconstructed from the learned correlation matrix C ; therefore, GLCS does not guarantee global optimality. When C and L are fixed, the W -step is a standard proximal-gradient update for a smooth loss with an ℓ_1 regularizer.

When W is fixed, the C-step applies singular value thresholding to a proximal surrogate with the nuclear-norm penalty, followed by the projection in Equation (15) and Laplacian reconstruction. Under standard boundedness and block-wise surrogate-descent assumptions, any limit point, if it exists, can be interpreted as satisfying block-wise stationarity of the adopted surrogate formulation.

3.2. Complexity Analysis

Let n , d , and l denote the numbers of instances, features, and labels, respectively. Let T be the number of outer iterations, s_W be the number of proximal-gradient steps used for updating W in each outer iteration, and k be the retained rank in truncated SVD. Under dense matrix operations, each W -update step mainly involves computing XW , $\mathcal{P}_M(Y)C$, $X^\top(XW - \mathcal{P}_M(Y)C)$, and WL , with cost $\mathcal{O}(ndl + nl^2 + dl^2)$. The C-update involves dense label-space products in the correlation-propagation loss and one truncated SVT operation, with cost $\mathcal{O}(nl^2 + l^2k)$. Constructing the dense label graph and graph Laplacian costs $\mathcal{O}(l^2)$. Therefore, the total time cost over T outer iterations is $\mathcal{O}(T[s_W(ndl + nl^2 + dl^2) + nl^2 + l^2k + l^2])$. The memory cost is $\mathcal{O}(nd + nl + dl + l^2)$, where the l^2 term is mainly caused by storing the dense correlation matrix C and the graph Laplacian L . The l^2 terms indicate that the dense correlation matrix, graph Laplacian, and SVT operation become the main computational bottlenecks when the label dimension is large. Therefore, the current implementation of GLCS is intended for moderate label dimensions, while extremely large label spaces would require sparse or factorized approximations.

4. Experiments and Results Analysis

4.1. Datasets

We conduct experiments on twelve widely used multi-label benchmark datasets spanning five representative domains, including text, audio, image, music, and biology, as summarized in Table 2 (Datasets: <http://www.uco.es/kdis/mlresources/> (accessed on 23 July 2026)). These datasets vary substantially in sample size, feature dimensionality, number of labels, and label cardinality, enabling a comprehensive evaluation under diverse multi-label characteristics. We employ a five-fold cross-validation on each dataset, where four folds are used for training, and the remaining fold is used for testing. The final results are obtained by averaging over the five folds. All compared methods are evaluated on exactly the same folds to ensure a fair comparison. To simulate missing-label supervision, we use r to denote the label missing rate and randomly mask a proportion $r \in \{0.6, 0.3, 0\}$ of label entries in the complete training label matrix, treating the selected entries as unobserved regardless of whether they are positive or negative. The test label matrix remains complete for evaluation.

Table 2. Statistics of the twelve multi-label datasets.

Datasets	Instances	Features	Labels	Cardinality	Domain
Arts	5000	462	26	1.636	Text
Business	5000	438	30	1.599	Text
birds	645	260	19	1.014	Audio
Cal500	502	68	174	26.044	Music
Corel5k	5000	499	374	3.522	Image
Computers	5000	681	33	1.508	Text

Table 2. Cont.

Datasets	Instances	Features	Labels	Cardinality	Domain
Education	5000	550	33	1.461	Text
emotions	593	72	6	1.869	Music
Health	5000	612	32	1.662	Text
medical	978	1449	45	1.245	Text
Natural_scene	2407	294	6	1.074	Image
yeast data	2417	103	14	4.237	Biology

4.2. Evaluation Metrics

We evaluate all methods using six widely adopted multi-label metrics, including Average Precision (AP), Area Under the ROC Curve (AUC), Coverage (CV), Hamming Loss (HL), One-Error (OE), and Ranking Loss (RL). Among them, AP and AUC are higher-is-better, while CV, HL, OE, and RL are lower-is-better. The formal definitions and formulas of these metrics can be found in the standard multi-label evaluation literature [1].

4.3. Comparison Methods and Hyperparameter Settings

All experiments were conducted on a workstation equipped with a 12th Gen Intel(R) Core(TM) i7-12700H CPU at 2.30 GHz and 16.0 GB RAM, running a 64-bit Windows operating system. All algorithms were implemented and executed in MATLAB R2020a.

We compare the proposed GLCS with six representative methods for MLL with missing labels, including SGMML, SLR-LSF, LRMML, DM2L, MSWL, and LFLI. LFLI is added as a recent 2025 baseline because it addresses multi-label learning with missing labels by exploiting correlation information and label-specific features. All parameters are tuned within the ranges recommended in the original papers, as detailed below.

- (1) SGMML [49] learns LSF and constrains them with a sparse global structure. All hyperparameters are searched in $\{10^{-7}, 10^{-4}, \dots, 10^2\}$.
- (2) SLR-LSF [17] converts logical labels into soft labels by exploiting a confidence matrix and label-density information while recovering missing labels, and then learns LSF on the obtained soft labels. All trade-off parameters satisfy $\lambda_1, \lambda_2, \lambda_3 \in [10^{-10}, 10^{10}]$.
- (3) LRMML [9] recovers missing labels by constructing an auxiliary label matrix and imposing low-/high-rank constraints on the label subspace. The threshold is fixed to $\delta = 0.005$, and the hyperparameters $\lambda_R, \lambda_L, \lambda_T$ are searched in $\{10^{-10}, 10^{-9}, \dots, 10^5\}$.
- (4) DM2L [28] addresses missing-label learning by tuning a regularization weight and fixing the threshold in the optimization. The regularization parameter is searched in $\lambda_d \in \{10^{-5}, 10^{-4}, \dots, 10^5\}$ with $\delta = 0.005$; for the nonlinear variant DM2L-nl, the Gaussian kernel width is chosen from $\sigma \in \{0.5, 1, 1.5, 2\}$.
- (5) MSWL [50] learns from weakly labeled multi-label data via a manifold-regularized sparse modeling framework that exploits feature structure together with global/local label correlations. The parameters are searched in $\alpha, \beta \in \{10^{-3}, 10^{-1}, \dots, 10^3\}$ and $\gamma \in \{10^{-6}, 10^{-5}, \dots, 10^6\}$, with $\lambda = 0.1$, and the neighbor number is set to $K = q + 1$ where q is the number of labels.
- (6) LFLI [31] exploits correlation information to guide label-specific feature learning and uses instance correlations to enhance label correlations under incomplete supervision. All parameters are tuned in $[10^{-5}, 10^3]$.
- (7) GLCS learns a graph-regularized framework that jointly learns a low-rank label-correlation matrix for correlation-based supervision propagation and a sparse LSF predictor, while preserving local label structure via a label-graph smoothness regu-

larizer to alleviate over-smoothing under missing labels. All parameters are tuned in $[10^{-10}, 10^2]$.

Tables 3–8 report the results on twelve benchmark datasets under three missing-label ratios, $r \in \{0.6, 0.3, 0\}$, using AP, AUC, CV, HL, OE, and RL. All values are averaged over five-fold cross-validation and reported as mean $\pm$ standard deviation. In these tables, OM denotes out-of-memory failure; such entries are treated as unavailable when highlighting the best results and are assigned the worst rank in the statistical test and running time analysis.

To better identify where GLCS is most advantageous, we summarize its performance by missing-label rate in Table 9. Since the six evaluation metrics have different scales and optimization directions, win counts and average ranks are used as scale-independent indicators. GLCS achieves the strongest advantage under severe missing labels, with 54 wins at $r = 0.6$ and an average rank of 1.59. At $r = 0.3$, GLCS still obtains 50 wins and an average rank of 1.66. When the labels are complete ($r = 0$), the advantage decreases to 38 wins with an average rank of 2.38, because the need for correlation-based missing-label refinement becomes weaker. From the dataset perspective, GLCS performs particularly well on Health, Corel5k, Education, emotions, Computers, Business, Arts, and birds. These datasets benefit more from the joint use of correlation-based supervision propagation and sparse label-specific predictors. In contrast, the advantage of GLCS is weaker on Natural_scene and some conditions of yeast data and Cal500, suggesting that GLCS is most useful when the label space contains richer dependencies and the supervision is incomplete.

Table 3. AP $\uparrow$ results (mean $\pm$ standard deviation) on twelve multi-label datasets. Larger values indicate better performance; the best results are highlighted in bold.

Datasets	r	SGMML	SLR-LSF	LRMML	DM2L	MSWL	LFLI	GLCS
Arts	0.6	0.5153 $\pm$ 0.0129	0.5739 $\pm$ 0.0079	0.5564 $\pm$ 0.0057	0.6084 $\pm$ 0.0056	0.4806 $\pm$ 0.1792	0.5867 $\pm$ 0.0080	0.6340 $\pm$ 0.0050
	0.3	0.5728 $\pm$ 0.0116	0.5965 $\pm$ 0.0146	0.6028 $\pm$ 0.0115	0.6202 $\pm$ 0.0057	0.5391 $\pm$ 0.1307	0.6190 $\pm$ 0.0086	0.6344 $\pm$ 0.0069
	0	0.6112 $\pm$ 0.0116	0.6177 $\pm$ 0.0103	0.6287 $\pm$ 0.0102	0.6230 $\pm$ 0.0040	0.6202 $\pm$ 0.0125	0.6283 $\pm$ 0.0102	0.6371 $\pm$ 0.0072
Business	0.6	0.6184 $\pm$ 0.0175	0.8593 $\pm$ 0.0057	0.8579 $\pm$ 0.0055	0.8756 $\pm$ 0.0061	0.6178 $\pm$ 0.2275	0.8395 $\pm$ 0.0052	0.8977 $\pm$ 0.0043
	0.3	0.6931 $\pm$ 0.0103	0.8493 $\pm$ 0.0080	0.8753 $\pm$ 0.0034	0.8808 $\pm$ 0.0066	0.6391 $\pm$ 0.2139	0.8757 $\pm$ 0.0040	0.8972 $\pm$ 0.0088
	0	0.7709 $\pm$ 0.0072	0.8895 $\pm$ 0.0049	0.8886 $\pm$ 0.0043	0.8816 $\pm$ 0.0054	0.5762 $\pm$ 0.2605	0.8884 $\pm$ 0.0041	0.8976 $\pm$ 0.0066
birds	0.6	0.6934 $\pm$ 0.0088	0.7342 $\pm$ 0.0306	0.7036 $\pm$ 0.0210	0.6928 $\pm$ 0.0215	0.6913 $\pm$ 0.0367	0.7057 $\pm$ 0.0159	0.7680 $\pm$ 0.0214
	0.3	0.7361 $\pm$ 0.0051	0.6818 $\pm$ 0.0348	0.7533 $\pm$ 0.0102	0.7069 $\pm$ 0.0266	0.7325 $\pm$ 0.0177	0.7448 $\pm$ 0.0130	0.7622 $\pm$ 0.0231
	0	0.7609 $\pm$ 0.0105	0.7680 $\pm$ 0.0262	0.7656 $\pm$ 0.0089	0.7112 $\pm$ 0.0278	0.7648 $\pm$ 0.0349	0.7666 $\pm$ 0.0091	0.7681 $\pm$ 0.0227
Cal500	0.6	0.4183 $\pm$ 0.0262	0.4977 $\pm$ 0.0112	0.4949 $\pm$ 0.0079	0.4548 $\pm$ 0.0034	0.4452 $\pm$ 0.0099	0.4964 $\pm$ 0.0076	0.4980 $\pm$ 0.0084
	0.3	0.4992 $\pm$ 0.0028	0.4961 $\pm$ 0.0082	0.4993 $\pm$ 0.0072	0.4561 $\pm$ 0.0031	0.4750 $\pm$ 0.0116	0.4996 $\pm$ 0.0077	0.4963 $\pm$ 0.0135
	0	0.4956 $\pm$ 0.0064	0.4905 $\pm$ 0.0020	0.4430 $\pm$ 0.0200	0.4759 $\pm$ 0.0044	0.4908 $\pm$ 0.0071	0.5011 $\pm$ 0.0074	0.4966 $\pm$ 0.0047
Corel5k	0.6	0.2057 $\pm$ 0.0094	OM	0.1448 $\pm$ 0.0056	0.2286 $\pm$ 0.0033	0.2523 $\pm$ 0.0062	0.2357 $\pm$ 0.0076	0.2883 $\pm$ 0.0055
	0.3	0.2412 $\pm$ 0.0063	OM	0.2329 $\pm$ 0.0039	0.2598 $\pm$ 0.0035	0.2883 $\pm$ 0.0064	0.2750 $\pm$ 0.0074	0.2887 $\pm$ 0.0085
	0	0.2742 $\pm$ 0.0069	OM	0.2985 $\pm$ 0.0062	0.2743 $\pm$ 0.0029	0.3017 $\pm$ 0.0030	0.2981 $\pm$ 0.0062	0.2903 $\pm$ 0.0067
Computers	0.6	0.5616 $\pm$ 0.0062	0.6701 $\pm$ 0.0136	0.6081 $\pm$ 0.0119	0.6903 $\pm$ 0.0042	0.6099 $\pm$ 0.0938	0.6633 $\pm$ 0.0124	0.7176 $\pm$ 0.0068
	0.3	0.6242 $\pm$ 0.0069	0.6927 $\pm$ 0.0030	0.6835 $\pm$ 0.0079	0.6985 $\pm$ 0.0037	0.6841 $\pm$ 0.0060	0.6978 $\pm$ 0.0070	0.7187 $\pm$ 0.0060
	0	0.7026 $\pm$ 0.0077	0.7098 $\pm$ 0.0061	0.7120 $\pm$ 0.0069	0.7008 $\pm$ 0.0065	0.6978 $\pm$ 0.0098	0.7118 $\pm$ 0.0075	0.7186 $\pm$ 0.0089
Education	0.6	0.5180 $\pm$ 0.0148	0.5892 $\pm$ 0.0063	0.5415 $\pm$ 0.0148	0.6253 $\pm$ 0.0048	0.5774 $\pm$ 0.0072	0.5926 $\pm$ 0.0136	0.6444 $\pm$ 0.0107
	0.3	0.5647 $\pm$ 0.0131	0.5880 $\pm$ 0.0117	0.6069 $\pm$ 0.0083	0.6346 $\pm$ 0.0008	0.6182 $\pm$ 0.0060	0.6267 $\pm$ 0.0087	0.6392 $\pm$ 0.0062
	0	0.6209 $\pm$ 0.0087	0.6389 $\pm$ 0.0083	0.6395 $\pm$ 0.0092	0.6363 $\pm$ 0.0031	0.6348 $\pm$ 0.0095	0.6391 $\pm$ 0.0090	0.6401 $\pm$ 0.0120
emotions	0.6	0.7637 $\pm$ 0.0267	0.7779 $\pm$ 0.0178	0.7668 $\pm$ 0.0158	0.5484 $\pm$ 0.0092	0.7481 $\pm$ 0.0105	0.5844 $\pm$ 0.0219	0.8146 $\pm$ 0.0224
	0.3	0.7960 $\pm$ 0.0223	0.7522 $\pm$ 0.0196	0.7915 $\pm$ 0.0222	0.5617 $\pm$ 0.0230	0.7801 $\pm$ 0.0215	0.6086 $\pm$ 0.0132	0.8077 $\pm$ 0.0197
	0	0.7984 $\pm$ 0.0269	0.7353 $\pm$ 0.0247	0.8035 $\pm$ 0.0154	0.6033 $\pm$ 0.0115	0.7922 $\pm$ 0.0164	0.6043 $\pm$ 0.0098	0.8071 $\pm$ 0.0180
Health	0.6	0.6620 $\pm$ 0.0099	0.6912 $\pm$ 0.0142	0.6979 $\pm$ 0.0105	0.7775 $\pm$ 0.0057	0.7409 $\pm$ 0.0083	0.7330 $\pm$ 0.0069	0.8038 $\pm$ 0.0057
	0.3	0.7040 $\pm$ 0.0057	0.6820 $\pm$ 0.0129	0.7625 $\pm$ 0.0081	0.7838 $\pm$ 0.0041	0.6951 $\pm$ 0.1475	0.7739 $\pm$ 0.0090	0.8022 $\pm$ 0.0080
	0	0.7783 $\pm$ 0.0092	0.7220 $\pm$ 0.0102	0.7874 $\pm$ 0.0086	0.7865 $\pm$ 0.0047	0.7777 $\pm$ 0.0082	0.7870 $\pm$ 0.0085	0.8028 $\pm$ 0.0093
medical	0.6	0.8136 $\pm$ 0.0102	0.8943 $\pm$ 0.0062	0.7886 $\pm$ 0.0251	0.8728 $\pm$ 0.0103	0.3767 $\pm$ 0.2615	0.8498 $\pm$ 0.0115	0.8997 $\pm$ 0.0073
	0.3	0.8726 $\pm$ 0.0121	0.8757 $\pm$ 0.0074	0.8724 $\pm$ 0.0102	0.8781 $\pm$ 0.0071	0.3767 $\pm$ 0.2615	0.8849 $\pm$ 0.0172	0.9010 $\pm$ 0.0100
	0	0.8822 $\pm$ 0.0086	0.8221 $\pm$ 0.0088	0.8994 $\pm$ 0.0150	0.8773 $\pm$ 0.0055	0.4808 $\pm$ 0.1921	0.9014 $\pm$ 0.0137	0.8994 $\pm$ 0.0109
Natural_scene	0.6	0.7718 $\pm$ 0.0067	0.7566 $\pm$ 0.0100	0.8372 $\pm$ 0.0160	0.8221 $\pm$ 0.0094	0.7465 $\pm$ 0.0096	0.7565 $\pm$ 0.0107	0.7826 $\pm$ 0.0131
	0.3	0.7653 $\pm$ 0.0069	0.7509 $\pm$ 0.0042	0.8195 $\pm$ 0.0176	0.8215 $\pm$ 0.0112	0.7850 $\pm$ 0.0197	0.7677 $\pm$ 0.0090	0.7849 $\pm$ 0.0138
	0	0.7645 $\pm$ 0.0126	0.7119 $\pm$ 0.0087	0.8584 $\pm$ 0.0215	0.8333 $\pm$ 0.0083	0.8197 $\pm$ 0.0156	0.7798 $\pm$ 0.0090	0.7800 $\pm$ 0.0149
yeast data	0.6	0.6106 $\pm$ 0.0080	0.7446 $\pm$ 0.0082	0.7482 $\pm$ 0.0092	0.7164 $\pm$ 0.0063	0.5850 $\pm$ 0.0101	0.7496 $\pm$ 0.0092	0.7568 $\pm$ 0.0093
	0.3	0.6936 $\pm$ 0.0062	0.7481 $\pm$ 0.0069	0.7554 $\pm$ 0.0085	0.7165 $\pm$ 0.0078	0.5979 $\pm$ 0.0190	0.7549 $\pm$ 0.0099	0.7571 $\pm$ 0.0065
	0	0.7050 $\pm$ 0.0113	0.7490 $\pm$ 0.0074	0.7570 $\pm$ 0.0082	0.7522 $\pm$ 0.0071	0.6039 $\pm$ 0.0140	0.7573 $\pm$ 0.0082	0.7578 $\pm$ 0.0099

Table 4. AUC $\uparrow$ results (mean $\pm$ standard deviation) on twelve multi-label datasets. Larger values indicate better performance; the best results are highlighted in bold.

Datasets	<i>r</i>	SGMML	SLR-LSF	LRMML	DM2L	MSWL	LFLI	GLCS
Arts	0.6	0.7257 $\pm$ 0.0122	0.5301 $\pm$ 0.0025	0.8374 $\pm$ 0.0083	0.8561 $\pm$ 0.0036	0.7646 $\pm$ 0.1295	0.6937 $\pm$ 0.0253	0.8575 $\pm$ 0.0030
	0.3	0.7361 $\pm$ 0.0125	0.5037 $\pm$ 0.0015	0.8536 $\pm$ 0.0074	0.8640 $\pm$ 0.0043	0.7987 $\pm$ 0.1167	0.7344 $\pm$ 0.0155	0.8562 $\pm$ 0.0029
	0	0.8461 $\pm$ 0.0064	0.5000 $\pm$ 0.0000	0.8691 $\pm$ 0.0041	0.8635 $\pm$ 0.0044	0.8690 $\pm$ 0.0039	0.7474 $\pm$ 0.0164	0.8556 $\pm$ 0.0025
Business	0.6	0.6953 $\pm$ 0.0139	0.8042 $\pm$ 0.0023	0.9410 $\pm$ 0.0032	0.9547 $\pm$ 0.0026	0.7398 $\pm$ 0.1717	0.6964 $\pm$ 0.0164	0.9503 $\pm$ 0.0029
	0.3	0.7777 $\pm$ 0.0085	0.8139 $\pm$ 0.0043	0.9509 $\pm$ 0.0018	0.9577 $\pm$ 0.0024	0.7282 $\pm$ 0.1924	0.7680 $\pm$ 0.0193	0.9499 $\pm$ 0.0025
	0	0.8566 $\pm$ 0.0049	0.8099 $\pm$ 0.0028	0.9600 $\pm$ 0.0031	0.9556 $\pm$ 0.0025	0.7257 $\pm$ 0.1963	0.8034 $\pm$ 0.0167	0.8556 $\pm$ 0.0025
birds	0.6	0.8049 $\pm$ 0.0068	0.5643 $\pm$ 0.0172	0.8546 $\pm$ 0.0127	0.8703 $\pm$ 0.0150	0.8522 $\pm$ 0.0265	0.7621 $\pm$ 0.0198	0.8729 $\pm$ 0.0153
	0.3	0.8466 $\pm$ 0.0048	0.5024 $\pm$ 0.0027	0.8913 $\pm$ 0.0095	0.8805 $\pm$ 0.0123	0.8817 $\pm$ 0.0124	0.8340 $\pm$ 0.0221	0.8686 $\pm$ 0.0133
	0	0.8815 $\pm$ 0.0151	0.6686 $\pm$ 0.0288	0.9097 $\pm$ 0.0076	0.8810 $\pm$ 0.0136	0.9029 $\pm$ 0.0168	0.8553 $\pm$ 0.0328	0.8668 $\pm$ 0.0084
Cal500	0.6	0.7255 $\pm$ 0.0257	0.5301 $\pm$ 0.0025	0.8122 $\pm$ 0.0032	0.7893 $\pm$ 0.0010	0.7731 $\pm$ 0.0032	0.5455 $\pm$ 0.0132	0.8147 $\pm$ 0.0051
	0.3	0.8125 $\pm$ 0.0044	0.5037 $\pm$ 0.0015	0.8184 $\pm$ 0.0034	0.7918 $\pm$ 0.0016	0.7939 $\pm$ 0.0073	0.5553 $\pm$ 0.0192	0.8140 $\pm$ 0.0032
	0	0.8177 $\pm$ 0.0050	0.5000 $\pm$ 0.0000	0.7514 $\pm$ 0.0201	0.7992 $\pm$ 0.0019	0.8016 $\pm$ 0.0079	0.5556 $\pm$ 0.0148	0.8147 $\pm$ 0.0019
Corel5k	0.6	0.6512 $\pm$ 0.0144	OM	0.6674 $\pm$ 0.0052	0.7312 $\pm$ 0.0024	0.7571 $\pm$ 0.0083	0.6058 $\pm$ 0.0141	0.8796 $\pm$ 0.0027
	0.3	0.6936 $\pm$ 0.0098	OM	0.7355 $\pm$ 0.0050	0.7617 $\pm$ 0.0048	0.8054 $\pm$ 0.0071	0.6390 $\pm$ 0.0075	0.8780 $\pm$ 0.0019
	0	0.7488 $\pm$ 0.0086	OM	0.8091 $\pm$ 0.0077	0.7762 $\pm$ 0.0035	0.8275 $\pm$ 0.0042	0.6571 $\pm$ 0.0078	0.8780 $\pm$ 0.0029
Computers	0.6	0.7083 $\pm$ 0.0056	0.5571 $\pm$ 0.0021	0.8418 $\pm$ 0.0085	0.9042 $\pm$ 0.0028	0.8199 $\pm$ 0.0911	0.6852 $\pm$ 0.0195	0.9056 $\pm$ 0.0027
	0.3	0.7741 $\pm$ 0.0042	0.5060 $\pm$ 0.0015	0.8866 $\pm$ 0.0042	0.9053 $\pm$ 0.0028	0.8924 $\pm$ 0.0043	0.7372 $\pm$ 0.0151	0.9064 $\pm$ 0.0024
	0	0.8932 $\pm$ 0.0035	0.6141 $\pm$ 0.0071	0.9101 $\pm$ 0.0026	0.9051 $\pm$ 0.0035	0.9032 $\pm$ 0.0027	0.7758 $\pm$ 0.0029	0.9070 $\pm$ 0.0021
Education	0.6	0.7107 $\pm$ 0.0100	0.5272 $\pm$ 0.0050	0.8548 $\pm$ 0.0033	0.8967 $\pm$ 0.0028	0.8304 $\pm$ 0.0097	0.7136 $\pm$ 0.0142	0.9161 $\pm$ 0.0020
	0.3	0.7483 $\pm$ 0.0096	0.5019 $\pm$ 0.0014	0.8895 $\pm$ 0.0038	0.9013 $\pm$ 0.0013	0.8700 $\pm$ 0.0024	0.7578 $\pm$ 0.0128	0.9165 $\pm$ 0.0029
	0	0.8956 $\pm$ 0.0049	0.5559 $\pm$ 0.0037	0.9002 $\pm$ 0.0015	0.9016 $\pm$ 0.0026	0.8996 $\pm$ 0.0028	0.7879 $\pm$ 0.0255	0.9165 $\pm$ 0.0018
emotions	0.6	0.8506 $\pm$ 0.0233	0.5205 $\pm$ 0.0074	0.8304 $\pm$ 0.0097	0.5801 $\pm$ 0.0113	0.8548 $\pm$ 0.0033	0.6124 $\pm$ 0.0326	0.8156 $\pm$ 0.0192
	0.3	0.8596 $\pm$ 0.0145	0.5002 $\pm$ 0.0010	0.8700 $\pm$ 0.0024	0.5888 $\pm$ 0.0245	0.8895 $\pm$ 0.0038	0.6379 $\pm$ 0.0306	0.8227 $\pm$ 0.0155
	0	0.8691 $\pm$ 0.0164	0.5000 $\pm$ 0.0000	0.8996 $\pm$ 0.0028	0.6277 $\pm$ 0.0157	0.9002 $\pm$ 0.0015	0.6413 $\pm$ 0.0329	0.8135 $\pm$ 0.0173
Health	0.6	0.7565 $\pm$ 0.0138	0.5022 $\pm$ 0.0012	0.8873 $\pm$ 0.0020	0.9353 $\pm$ 0.0020	0.9091 $\pm$ 0.0047	0.7356 $\pm$ 0.0141	0.9422 $\pm$ 0.0012
	0.3	0.7965 $\pm$ 0.0075	0.5103 $\pm$ 0.0046	0.9192 $\pm$ 0.0042	0.9364 $\pm$ 0.0018	0.8633 $\pm$ 0.1411	0.7760 $\pm$ 0.0164	0.9424 $\pm$ 0.0016
	0	0.9247 $\pm$ 0.0038	0.5234 $\pm$ 0.0088	0.9388 $\pm$ 0.0027	0.9356 $\pm$ 0.0008	0.9412 $\pm$ 0.0041	0.7887 $\pm$ 0.0157	0.9427 $\pm$ 0.0020
medical	0.6	0.6512 $\pm$ 0.0144	0.5005 $\pm$ 0.0542	0.9400 $\pm$ 0.0161	0.9751 $\pm$ 0.0032	0.6817 $\pm$ 0.1624	0.8532 $\pm$ 0.0410	0.9422 $\pm$ 0.0012
	0.3	0.6936 $\pm$ 0.0098	0.5020 $\pm$ 0.1035	0.9718 $\pm$ 0.0068	0.9788 $\pm$ 0.0027	0.6817 $\pm$ 0.1624	0.8930 $\pm$ 0.0327	0.9680 $\pm$ 0.0073
	0	0.7488 $\pm$ 0.0086	0.5008 $\pm$ 0.1071	0.9843 $\pm$ 0.0059	0.9791 $\pm$ 0.0032	0.7604 $\pm$ 0.1217	0.9399 $\pm$ 0.0247	0.9681 $\pm$ 0.0116
Natural_scene	0.6	0.8377 $\pm$ 0.0118	0.5106 $\pm$ 0.0024	0.9048 $\pm$ 0.0114	0.8976 $\pm$ 0.0073	0.8289 $\pm$ 0.0117	0.8017 $\pm$ 0.0101	0.7783 $\pm$ 0.0109
	0.3	0.8431 $\pm$ 0.0137	0.5009 $\pm$ 0.0015	0.8958 $\pm$ 0.0108	0.8965 $\pm$ 0.0066	0.8602 $\pm$ 0.0146	0.8085 $\pm$ 0.0099	0.7813 $\pm$ 0.0136
	0	0.8524 $\pm$ 0.0121	0.5000 $\pm$ 0.0000	0.9158 $\pm$ 0.0104	0.9035 $\pm$ 0.0045	0.8884 $\pm$ 0.0086	0.8215 $\pm$ 0.0112	0.7806 $\pm$ 0.0085
yeast data	0.6	0.6780 $\pm$ 0.0095	0.5037 $\pm$ 0.0509	0.8200 $\pm$ 0.0084	0.7960 $\pm$ 0.0029	0.4372 $\pm$ 0.0110	0.6745 $\pm$ 0.0127	0.8111 $\pm$ 0.0075
	0.3	0.7635 $\pm$ 0.0064	0.5025 $\pm$ 0.1106	0.8257 $\pm$ 0.0088	0.7964 $\pm$ 0.0029	0.4444 $\pm$ 0.0266	0.6891 $\pm$ 0.0117	0.8106 $\pm$ 0.0106
	0	0.7793 $\pm$ 0.0081	0.5009 $\pm$ 0.1048	0.8298 $\pm$ 0.0076	0.8232 $\pm$ 0.0044	0.4523 $\pm$ 0.0295	0.6971 $\pm$ 0.0113	0.8112 $\pm$ 0.0016

Table 5. CV $\downarrow$ results (mean $\pm$ standard deviation) on twelve multi-label datasets. Smaller values indicate better performance; the best results are highlighted in bold.

Datasets	<i>r</i>	SGMML	SLR-LSF	LRMML	DM2L	MSWL	LFLI	GLCS
Arts	0.6	0.3359 ± 0.0141	0.2530 ± 0.0059	0.2351 ± 0.0098	0.2175 ± 0.0059	0.3027 ± 0.1195	0.2648 ± 0.0092	0.1698 ± 0.0056
	0.3	0.3187 ± 0.0131	0.3076 ± 0.0143	0.2207 ± 0.0098	0.2058 ± 0.0062	0.2637 ± 0.1062	0.2223 ± 0.0090	0.1686 ± 0.0054
	0	0.2000 ± 0.0064	0.1817 ± 0.0081	0.2025 ± 0.0059	0.2064 ± 0.0067	0.1944 ± 0.0044	0.2048 ± 0.0056	0.2185 ± 0.0118
Business	0.6	0.3886 ± 0.0222	0.1672 ± 0.0038	0.1112 ± 0.0054	0.0910 ± 0.0054	0.3132 ± 0.1866	0.1492 ± 0.0119	0.0639 ± 0.0024
	0.3	0.2886 ± 0.0143	0.1474 ± 0.0082	0.0971 ± 0.0053	0.0842 ± 0.0046	0.3063 ± 0.1921	0.0986 ± 0.0091	0.0653 ± 0.0047
	0	0.1869 ± 0.0083	0.0822 ± 0.0028	0.0822 ± 0.0075	0.0873 ± 0.0049	0.3173 ± 0.2023	0.0827 ± 0.0076	0.0654 ± 0.0045
birds	0.6	0.2239 ± 0.0146	0.1842 ± 0.0255	0.2090 ± 0.0164	0.1812 ± 0.0168	0.2067 ± 0.0326	0.1883 ± 0.0158	0.1401 ± 0.0180
	0.3	0.1881 ± 0.0152	0.2112 ± 0.0131	0.1662 ± 0.0166	0.1687 ± 0.0128	0.1696 ± 0.0115	0.1564 ± 0.0102	0.1471 ± 0.0202
	0	0.1555 ± 0.0218	0.1457 ± 0.0287	0.1429 ± 0.0172	0.1684 ± 0.0137	0.1475 ± 0.0284	0.1428 ± 0.0170	0.1509 ± 0.0117
Cal500	0.6	0.8821 ± 0.0274	0.7506 ± 0.0330	0.7825 ± 0.0158	0.8577 ± 0.0127	0.8543 ± 0.0161	0.7538 ± 0.0154	0.7479 ± 0.0146
	0.3	0.7652 ± 0.0165	0.7608 ± 0.0102	0.7563 ± 0.0170	0.8511 ± 0.0114	0.8072 ± 0.0190	0.7495 ± 0.0203	0.7471 ± 0.0103
	0	0.7426 ± 0.0188	0.8721 ± 0.0038	0.8978 ± 0.0425	0.8267 ± 0.0157	0.7880 ± 0.0104	0.7449 ± 0.0173	0.7480 ± 0.0153
Corel5k	0.6	0.6935 ± 0.0168	OM	0.6199 ± 0.0103	0.5754 ± 0.0041	0.5339 ± 0.0167	0.5979 ± 0.0138	0.2859 ± 0.0080
	0.3	0.6255 ± 0.0109	OM	0.5465 ± 0.0052	0.5238 ± 0.0086	0.4315 ± 0.0094	0.4998 ± 0.0133	0.2853 ± 0.0052
	0	0.5406 ± 0.0135	OM	0.4355 ± 0.0118	0.4940 ± 0.0054	0.3826 ± 0.0080	0.4408 ± 0.0119	0.2851 ± 0.0059
Computers	0.6	0.3365 ± 0.0073	0.1784 ± 0.0116	0.2144 ± 0.0057	0.1021 ± 0.0036	0.2226 ± 0.0900	0.2024 ± 0.0099	0.1038 ± 0.0053
	0.3	0.2687 ± 0.0060	0.2248 ± 0.0023	0.1629 ± 0.0041	0.1010 ± 0.0034	0.1446 ± 0.0058	0.1548 ± 0.0055	0.1019 ± 0.0045
	0	0.1300 ± 0.0015	0.1252 ± 0.0044	0.1330 ± 0.0029	0.1012 ± 0.0040	0.1312 ± 0.0032	0.1349 ± 0.0031	0.1019 ± 0.0013
Education	0.6	0.3453 ± 0.0122	0.1966 ± 0.0027	0.2227 ± 0.0119	0.1514 ± 0.0023	0.1974 ± 0.0042	0.2204 ± 0.0077	0.0984 ± 0.0014
	0.3	0.3006 ± 0.0089	0.2586 ± 0.0102	0.1807 ± 0.0028	0.1433 ± 0.0022	0.1555 ± 0.0051	0.1733 ± 0.0070	0.0980 ± 0.0029
	0	0.1416 ± 0.0061	0.1448 ± 0.0082	0.1475 ± 0.0058	0.1438 ± 0.0043	0.1406 ± 0.0024	0.1500 ± 0.0057	0.0978 ± 0.0030
emotions	0.6	0.3213 ± 0.0333	0.3176 ± 0.0101	0.3265 ± 0.0154	0.5129 ± 0.0125	0.3404 ± 0.0193	0.4867 ± 0.0195	0.2923 ± 0.0279
	0.3	0.3053 ± 0.0257	0.3364 ± 0.0257	0.3034 ± 0.0270	0.5038 ± 0.0140	0.3122 ± 0.0123	0.4735 ± 0.0188	0.2865 ± 0.0208
	0	0.2941 ± 0.0285	0.4571 ± 0.0223	0.2964 ± 0.0253	0.4769 ± 0.0119	0.3014 ± 0.0137	0.4679 ± 0.0142	0.2935 ± 0.0184
Health	0.6	0.3120 ± 0.0148	0.1644 ± 0.0042	0.1833 ± 0.0037	0.1249 ± 0.0031	0.1536 ± 0.0061	0.1837 ± 0.0067	0.0803 ± 0.0014
	0.3	0.2590 ± 0.0044	0.1762 ± 0.0045	0.1463 ± 0.0052	0.1220 ± 0.0014	0.1885 ± 0.0136	0.1362 ± 0.0050	0.0798 ± 0.0012
	0	0.1122 ± 0.0028	0.1896 ± 0.0091	0.1206 ± 0.0033	0.1231 ± 0.0013	0.1095 ± 0.0082	0.1222 ± 0.0035	0.0798 ± 0.0029
medical	0.6	0.0669 ± 0.0072	0.0465 ± 0.0022	0.0801 ± 0.0200	0.0404 ± 0.0051	0.3435 ± 0.1671	0.0606 ± 0.0136	0.0803 ± 0.0014
	0.3	0.0514 ± 0.0068	0.0458 ± 0.0018	0.0428 ± 0.0101	0.0343 ± 0.0042	0.3435 ± 0.1671	0.0395 ± 0.0097	0.0395 ± 0.0086
	0	0.0422 ± 0.0052	0.0255 ± 0.0035	0.0260 ± 0.0083	0.0338 ± 0.0051	0.2623 ± 0.1298	0.0251 ± 0.0080	0.0396 ± 0.0143
Natural_scene	0.6	0.2119 ± 0.0103	0.2124 ± 0.0051	0.0959 ± 0.0108	0.1000 ± 0.0049	0.1600 ± 0.0118	0.2183 ± 0.0111	0.1993 ± 0.0081
	0.3	0.2108 ± 0.0099	0.2188 ± 0.0053	0.1022 ± 0.0087	0.1010 ± 0.0043	0.1329 ± 0.0125	0.2096 ± 0.0120	0.1966 ± 0.0122
	0	0.2048 ± 0.0128	0.4582 ± 0.0058	0.0862 ± 0.0082	0.0950 ± 0.0040	0.1091 ± 0.0079	0.1986 ± 0.0111	0.1963 ± 0.0060
yeast data	0.6	0.6304 ± 0.0117	0.4865 ± 0.0048	0.4737 ± 0.0105	0.4898 ± 0.0030	0.2403 ± 0.0120	0.4722 ± 0.0069	0.4632 ± 0.0146
	0.3	0.5429 ± 0.0120	0.4722 ± 0.0048	0.4629 ± 0.0086	0.4894 ± 0.0030	0.2302 ± 0.0244	0.4570 ± 0.0075	0.4637 ± 0.0108
	0	0.5299 ± 0.0106	0.4693 ± 0.0056	0.4534 ± 0.0087	0.4866 ± 0.0030	0.2219 ± 0.0115	0.4535 ± 0.0089	0.4641 ± 0.0063

Table 6. HL↓ results (mean ± standard deviation) on twelve multi-label datasets. Smaller values indicate better performance; the best results are highlighted in bold.

Datasets	r	SGMML	SLR-LSF	LRMML	DM2L	MSWL	LFLI	GLCS
Arts	0.6	0.0719 ± 0.0038	0.0599 ± 0.0012	0.0597 ± 0.0018	0.0650 ± 0.0016	0.0617 ± 0.0006	0.0661 ± 0.0032	0.0527 ± 0.0015
	0.3	0.0704 ± 0.0008	0.0625 ± 0.0012	0.0553 ± 0.0013	0.0596 ± 0.0015	0.0613 ± 0.0019	0.0649 ± 0.0028	0.0527 ± 0.0006
	0	0.0705 ± 0.0022	0.0629 ± 0.0008	0.0534 ± 0.0015	0.0577 ± 0.0017	0.0604 ± 0.0006	0.0665 ± 0.0030	0.0526 ± 0.0012
Business	0.6	0.0470 ± 0.0023	0.0233 ± 0.0012	0.0274 ± 0.0016	0.0291 ± 0.0017	0.0512 ± 0.0023	0.0281 ± 0.0015	0.0237 ± 0.0004
	0.3	0.0439 ± 0.0021	0.0218 ± 0.0011	0.0254 ± 0.0011	0.0281 ± 0.0016	0.0501 ± 0.0036	0.0258 ± 0.0011	0.0240 ± 0.0010
	0	0.0381 ± 0.0018	0.0280 ± 0.0008	0.0246 ± 0.0012	0.0276 ± 0.0017	0.0497 ± 0.0036	0.0253 ± 0.0007	0.0226 ± 0.0012
birds	0.6	0.0678 ± 0.0051	0.0667 ± 0.0051	0.0537 ± 0.0044	0.0567 ± 0.0025	0.0719 ± 0.0027	0.0588 ± 0.0038	0.0474 ± 0.0048
	0.3	0.0567 ± 0.0017	0.0732 ± 0.0048	0.0533 ± 0.0031	0.0556 ± 0.0022	0.0709 ± 0.0011	0.0548 ± 0.0049	0.0482 ± 0.0052
	0	0.0619 ± 0.0074	0.0567 ± 0.0035	0.0516 ± 0.0031	0.0560 ± 0.0017	0.0652 ± 0.0080	0.0583 ± 0.0079	0.0482 ± 0.0056
Cal500	0.6	0.1988 ± 0.0108	0.1497 ± 0.0016	0.1377 ± 0.0036	0.1437 ± 0.0019	0.1491 ± 0.0016	0.1694 ± 0.0039	0.1376 ± 0.0032
	0.3	0.1962 ± 0.0066	0.1497 ± 0.0032	0.1372 ± 0.0041	0.1439 ± 0.0018	0.1488 ± 0.0036	0.1831 ± 0.0044	0.1371 ± 0.0032
	0	0.1933 ± 0.0045	0.1497 ± 0.0024	0.1429 ± 0.0030	0.1419 ± 0.0022	0.1465 ± 0.0035	0.1919 ± 0.0042	0.1377 ± 0.0027
Corel5k	0.6	0.0155 ± 0.0009	OM	0.0160 ± 0.0007	0.0112 ± 0.0001	0.0094 ± 0.0000	0.0137 ± 0.0004	0.0092 ± 0.0001
	0.3	0.0170 ± 0.0008	OM	0.0097 ± 0.0001	0.0100 ± 0.0000	0.0094 ± 0.0001	0.0165 ± 0.0009	0.0092 ± 0.0001
	0	0.0179 ± 0.0014	OM	0.0093 ± 0.0001	0.0096 ± 0.0000	0.0094 ± 0.0000	0.0172 ± 0.0005	0.0092 ± 0.0000
Computers	0.6	0.0492 ± 0.0023	0.0342 ± 0.0016	0.0455 ± 0.0006	0.0492 ± 0.0026	0.0448 ± 0.0009	0.0405 ± 0.0006	0.0326 ± 0.0009
	0.3	0.0454 ± 0.0013	0.0353 ± 0.0012	0.0346 ± 0.0007	0.0467 ± 0.0024	0.0437 ± 0.0010	0.0398 ± 0.0011	0.0328 ± 0.0010
	0	0.0463 ± 0.0020	0.0372 ± 0.0007	0.0334 ± 0.0008	0.0460 ± 0.0023	0.0432 ± 0.0015	0.0397 ± 0.0013	0.0323 ± 0.0006
Education	0.6	0.0521 ± 0.0031	0.0371 ± 0.0008	0.0439 ± 0.0026	0.0444 ± 0.0009	0.0435 ± 0.0010	0.0469 ± 0.0020	0.0368 ± 0.0014
	0.3	0.0495 ± 0.0017	0.0376 ± 0.0012	0.0384 ± 0.0011	0.0398 ± 0.0005	0.0429 ± 0.0005	0.0461 ± 0.0012	0.0370 ± 0.0009
	0	0.0513 ± 0.0016	0.0390 ± 0.0012	0.0370 ± 0.0010	0.0379 ± 0.0004	0.0416 ± 0.0007	0.0467 ± 0.0011	0.0368 ± 0.0007
emotions	0.6	0.2588 ± 0.0139	0.2991 ± 0.0116	0.2436 ± 0.0172	0.4017 ± 0.0136	0.3021 ± 0.0073	0.3533 ± 0.0224	0.1964 ± 0.0112
	0.3	0.2678 ± 0.0163	0.3114 ± 0.0099	0.2079 ± 0.0190	0.3957 ± 0.0266	0.2847 ± 0.0175	0.4181 ± 0.0267	0.1959 ± 0.0062
	0	0.2214 ± 0.0167	0.3114 ± 0.0099	0.1984 ± 0.0142	0.3279 ± 0.0128	0.2783 ± 0.0327	0.4537 ± 0.0385	0.1963 ± 0.0154
Health	0.6	0.0472 ± 0.0009	0.0423 ± 0.0017	0.0430 ± 0.0009	0.0376 ± 0.0018	0.0505 ± 0.0013	0.0414 ± 0.0013	0.0309 ± 0.0004
	0.3	0.0415 ± 0.0021	0.0411 ± 0.0015	0.0341 ± 0.0010	0.0358 ± 0.0018	0.0494 ± 0.0020	0.0360 ± 0.0011	0.0303 ± 0.0014
	0	0.0417 ± 0.0024	0.0388 ± 0.0022	0.0326 ± 0.0008	0.0356 ± 0.0018	0.0468 ± 0.0006	0.0345 ± 0.0014	0.0313 ± 0.0005
medical	0.6	0.0272 ± 0.0008	0.0113 ± 0.0003	0.0176 ± 0.0016	0.0134 ± 0.0005	0.0272 ± 0.0007	0.0160 ± 0.0011	0.0109 ± 0.0004
	0.3	0.0223 ± 0.0007	0.0133 ± 0.0002	0.0122 ± 0.0011	0.0127 ± 0.0004	0.0272 ± 0.0007	0.0125 ± 0.0011	0.0100 ± 0.0009
	0	0.0246 ± 0.0012	0.0161 ± 0.0006	0.0114 ± 0.0016	0.0127 ± 0.0006	0.0277 ± 0.0010	0.0103 ± 0.0011	0.0101 ± 0.0013
Natural_scene	0.6	0.2119 ± 0.0081	0.2424 ± 0.0032	0.1534 ± 0.0216	0.1321 ± 0.0030	0.1790 ± 0.0027	0.2105 ± 0.0052	0.1804 ± 0.0068
	0.3	0.2211 ± 0.0096	0.2468 ± 0.0028	0.1299 ± 0.0059	0.1182 ± 0.0069	0.1756 ± 0.0021	0.2096 ± 0.0082	0.1805 ± 0.0067
	0	0.2115 ± 0.0070	0.2472 ± 0.0040	0.1114 ± 0.0052	0.1158 ± 0.0030	0.1766 ± 0.0040	0.2018 ± 0.0130	0.1799 ± 0.0017
yeast data	0.6	0.3365 ± 0.0165	0.2081 ± 0.0055	0.2043 ± 0.0048	0.2283 ± 0.0024	0.1195 ± 0.0065	0.2073 ± 0.0041	0.2024 ± 0.0059
	0.3	0.3280 ± 0.0107	0.2050 ± 0.0075	0.2032 ± 0.0043	0.2291 ± 0.0021	0.1127 ± 0.0105	0.2106 ± 0.0072	0.2022 ± 0.0073
	0	0.3063 ± 0.0082	0.2100 ± 0.0077	0.1999 ± 0.0035	0.2048 ± 0.0026	0.1104 ± 0.0071	0.2187 ± 0.0081	0.2023 ± 0.0027

Table 7. OE↓ results (mean ± standard deviation) on twelve multi-label datasets. Smaller values indicate better performance; the best results are highlighted in bold.

Datasets	r	SGMML	SLR-LSF	LRMML	DM2L	MSWL	LFLI	GLCS
Arts	0.6	0.5398 ± 0.0106	0.4904 ± 0.0137	0.5382 ± 0.0051	0.4736 ± 0.0071	0.6060 ± 0.1911	0.4998 ± 0.0074	0.4458 ± 0.0119
	0.3	0.4822 ± 0.0088	0.4668 ± 0.0140	0.4766 ± 0.0084	0.4576 ± 0.0078	0.5490 ± 0.1561	0.4636 ± 0.0085	0.4500 ± 0.0093
	0	0.4826 ± 0.0139	0.5010 ± 0.0130	0.4528 ± 0.0119	0.4526 ± 0.0038	0.4554 ± 0.0210	0.4524 ± 0.0116	0.4614 ± 0.0051
Business	0.6	0.3562 ± 0.0146	0.1628 ± 0.0123	0.1258 ± 0.0092	0.1189 ± 0.0053	0.4118 ± 0.2692	0.1292 ± 0.0071	0.1032 ± 0.0044
	0.3	0.2952 ± 0.0091	0.1528 ± 0.0153	0.1128 ± 0.0075	0.1156 ± 0.0043	0.3800 ± 0.2282	0.1136 ± 0.0067	0.1034 ± 0.0117
	0	0.2330 ± 0.0097	0.1080 ± 0.0103	0.1084 ± 0.0071	0.1124 ± 0.0036	0.4896 ± 0.3241	0.1080 ± 0.0064	0.1020 ± 0.0103
birds	0.6	0.3643 ± 0.0214	0.2946 ± 0.0329	0.3442 ± 0.0233	0.3721 ± 0.0186	0.3643 ± 0.0498	0.3798 ± 0.0357	0.2760 ± 0.0335
	0.3	0.3147 ± 0.0116	0.3705 ± 0.0585	0.2884 ± 0.0173	0.3566 ± 0.0345	0.3240 ± 0.0341	0.3132 ± 0.0200	0.2822 ± 0.0253
	0	0.2868 ± 0.0130	0.2791 ± 0.0357	0.2884 ± 0.0090	0.3512 ± 0.0393	0.2791 ± 0.0392	0.2884 ± 0.0090	0.2868 ± 0.0306
Cal500	0.6	0.1154 ± 0.0111	0.1155 ± 0.0111	0.1138 ± 0.0220	0.3144 ± 0.0155	0.2525 ± 0.0440	0.1158 ± 0.0200	0.1155 ± 0.0218
	0.3	0.1194 ± 0.0200	0.1194 ± 0.0200	0.1158 ± 0.0200	0.3194 ± 0.0133	0.2135 ± 0.0373	0.1158 ± 0.0200	0.1157 ± 0.0256
	0	0.1198 ± 0.0258	0.2214 ± 0.0268	0.2954 ± 0.0970	0.2279 ± 0.0262	0.1933 ± 0.0443	0.1158 ± 0.0200	0.1197 ± 0.0225
Corel5k	0.6	0.7376 ± 0.0149	OM	0.7818 ± 0.0080	0.7073 ± 0.0128	0.6878 ± 0.0122	0.6976 ± 0.0135	0.6588 ± 0.0150
	0.3	0.6968 ± 0.0109	OM	0.6820 ± 0.0076	0.6767 ± 0.0074	0.6622 ± 0.0197	0.6674 ± 0.0169	0.6544 ± 0.0143
	0	0.6626 ± 0.0154	OM	0.6402 ± 0.0113	0.6666 ± 0.0067	0.6500 ± 0.0115	0.6396 ± 0.0116	0.6500 ± 0.0164
Computers	0.6	0.4786 ± 0.0078	0.3416 ± 0.0114	0.4280 ± 0.0110	0.3576 ± 0.0052	0.4344 ± 0.0944	0.3874 ± 0.0179	0.3386 ± 0.0072
	0.3	0.4148 ± 0.0084	0.3460 ± 0.0122	0.3660 ± 0.0063	0.3473 ± 0.0064	0.3646 ± 0.0102	0.3574 ± 0.0110	0.3382 ± 0.0064
	0	0.3612 ± 0.0124	0.3726 ± 0.0104	0.3438 ± 0.0125	0.3441 ± 0.0080	0.3492 ± 0.0107	0.3432 ± 0.0134	0.3380 ± 0.0157
Education	0.6	0.5498 ± 0.0170	0.4564 ± 0.0123	0.5336 ± 0.0141	0.4810 ± 0.0095	0.5264 ± 0.0141	0.5002 ± 0.0152	0.6588 ± 0.0150
	0.3	0.5034 ± 0.0150	0.4720 ± 0.0920	0.4806 ± 0.0118	0.4680 ± 0.0045	0.4818 ± 0.0138	0.4710 ± 0.0130	0.4656 ± 0.0125
	0	0.4854 ± 0.0131	0.5078 ± 0.0880	0.4578 ± 0.0152	0.4637 ± 0.0084	0.4664 ± 0.0171	0.4580 ± 0.0155	0.4676 ± 0.0166
emotions	0.6	0.3471 ± 0.0493	0.3050 ± 0.0425	0.3272 ± 0.0444	0.6731 ± 0.0196	0.3643 ± 0.0133	0.6192 ± 0.0718	0.2496 ± 0.0494
	0.3	0.2781 ± 0.0515	0.3542 ± 0.0359	0.2983 ± 0.0528	0.6471 ± 0.0382	0.3067 ± 0.0387	0.5259 ± 0.0474	0.2615 ± 0.0348
	0	0.2830 ± 0.0500	0.3212 ± 0.0322	0.2681 ± 0.0282	0.5655 ± 0.0175	0.2867 ± 0.0197	0.5412 ± 0.0385	0.2663 ± 0.0321
Health	0.6	0.3440 ± 0.0188	0.3412 ± 0.0144	0.3272 ± 0.0297	0.2605 ± 0.0094	0.2968 ± 0.0120	0.3078 ± 0.0145	0.2332 ± 0.0126
	0.3	0.3008 ± 0.0121	0.3318 ± 0.0119	0.2634 ± 0.0173	0.2516 ± 0.0073	0.3506 ± 0.1552	0.2670 ± 0.0165	0.2360 ± 0.0125
	0	0.2694 ± 0.0153	0.3022 ± 0.0102	0.2502 ± 0.0174	0.2480 ± 0.0068	0.2692 ± 0.0143	0.2494 ± 0.0169	0.2352 ± 0.0151
medical	0.6	0.2097 ± 0.0262	0.1381 ± 0.0125	0.2485 ± 0.0223	0.1679 ± 0.0101	0.7323 ± 0.2930	0.1892 ± 0.0174	0.1289 ± 0.0183
	0.3	0.1606 ± 0.0122	0.1483 ± 0.0122	0.1667 ± 0.0287	0.1633 ± 0.0087	0.7323 ± 0.2930	0.1503 ± 0.0241	0.1268 ± 0.0118
	0	0.1391 ± 0.0242	0.2096 ± 0.0196	0.1431 ± 0.0300	0.1679 ± 0.0083	0.6292 ± 0.2085	0.1411 ± 0.0276	0.1329 ± 0.0221
Natural_scene	0.6	0.3500 ± 0.0101	0.3795 ± 0.0211	0.2687 ± 0.0267	0.2955 ± 0.0127	0.3954 ± 0.0116	0.3795 ± 0.0262	0.3330 ± 0.0171
	0.3	0.3650 ± 0.0177	0.3850 ± 0.0089	0.3024 ± 0.0303	0.2973 ± 0.0175	0.3389 ± 0.0295	0.3620 ± 0.0171	0.3355 ± 0.0220
	0	0.3735 ± 0.0242	0.3822 ± 0.0064	0.2323 ± 0.0429	0.2791 ± 0.0143	0.2901 ± 0.0264	0.3420 ± 0.0147	0.3405 ± 0.0270
yeast data	0.6	0.3736 ± 0.0153	0.2417 ± 0.0128	0.2412 ± 0.0209	0.2523 ± 0.0126	0.4730 ± 0.0197	0.2396 ± 0.0162	0.2214 ± 0.0185
	0.3	0.2979 ± 0.0195	0.2391 ± 0.0133	0.2338 ± 0.0097	0.2523 ± 0.0129	0.4580 ± 0.0170	0.2334 ± 0.0157	0.2234 ± 0.0098
	0	0.3178 ± 0.0198	0.2363 ± 0.0880	0.2292 ± 0.0076	0.2238 ± 0.0152	0.4580 ± 0.0251	0.2288 ± 0.0074	0.2226 ± 0.0075

Table 8. RL↓ results (mean ± standard deviation) on twelve multi-label datasets. Smaller values indicate better performance; the best results are highlighted in bold.

Datasets	<i>r</i>	SGMML	SLR-LSF	LRMML	DM2L	MSWL	LFLI	GLCS
Arts	0.6	0.2516 ± 0.0107	0.1753 ± 0.0043	0.1626 ± 0.0083	0.1458 ± 0.0040	0.2380 ± 0.1292	0.1755 ± 0.0072	0.1112 ± 0.0039
	0.3	0.2278 ± 0.0100	0.2297 ± 0.0114	0.1464 ± 0.0074	0.1380 ± 0.0046	0.2039 ± 0.1174	0.1444 ± 0.0058	0.1101 ± 0.0029
	0	0.1311 ± 0.0047	0.1523 ± 0.0044	0.1309 ± 0.0041	0.1385 ± 0.0048	0.1336 ± 0.0042	0.1325 ± 0.0039	0.1436 ± 0.0086
Business	0.6	0.2774 ± 0.0161	0.0446 ± 0.0043	0.0590 ± 0.0032	0.0466 ± 0.0029	0.2614 ± 0.1722	0.0753 ± 0.0051	0.0299 ± 0.0019
	0.3	0.1930 ± 0.0073	0.0421 ± 0.0056	0.0491 ± 0.0018	0.0436 ± 0.0026	0.2731 ± 0.1929	0.0489 ± 0.0041	0.0308 ± 0.0032
	0	0.1147 ± 0.0049	0.0396 ± 0.0013	0.0400 ± 0.0031	0.0457 ± 0.0027	0.2755 ± 0.1964	0.0404 ± 0.0033	0.0307 ± 0.0019
birds	0.6	0.1616 ± 0.0085	0.1293 ± 0.0261	0.1454 ± 0.0127	0.1297 ± 0.0150	0.1478 ± 0.0265	0.1295 ± 0.0045	0.0890 ± 0.0098
	0.3	0.1249 ± 0.0068	0.1496 ± 0.0085	0.1087 ± 0.0095	0.1195 ± 0.0123	0.1183 ± 0.0124	0.1022 ± 0.0052	0.0945 ± 0.0152
	0	0.1057 ± 0.0129	0.0910 ± 0.0157	0.0903 ± 0.0076	0.1190 ± 0.0136	0.0971 ± 0.0168	0.0902 ± 0.0074	0.0989 ± 0.0069
Cal500	0.6	0.2741 ± 0.0259	0.1812 ± 0.0083	0.1878 ± 0.0032	0.2107 ± 0.0010	0.2269 ± 0.0032	0.1827 ± 0.0038	0.1817 ± 0.0047
	0.3	0.1855 ± 0.0043	0.1828 ± 0.0028	0.1816 ± 0.0034	0.2082 ± 0.0016	0.2061 ± 0.0073	0.1805 ± 0.0042	0.1822 ± 0.0059
	0	0.1796 ± 0.0045	0.1921 ± 0.0043	0.2486 ± 0.0201	0.2008 ± 0.0019	0.1984 ± 0.0079	0.1795 ± 0.0037	0.1824 ± 0.0021
Corel5k	0.6	0.3510 ± 0.0147	OM	0.3326 ± 0.0052	0.2688 ± 0.0024	0.2434 ± 0.0084	0.2755 ± 0.0075	0.1219 ± 0.0037
	0.3	0.3078 ± 0.0098	OM	0.2645 ± 0.0050	0.2383 ± 0.0048	0.1946 ± 0.0071	0.2241 ± 0.0078	0.1216 ± 0.0030
	0	0.2527 ± 0.0087	OM	0.1909 ± 0.0077	0.2238 ± 0.0035	0.1725 ± 0.0042	0.1938 ± 0.0080	0.1219 ± 0.0028
Computers	0.6	0.2577 ± 0.0081	0.1319 ± 0.0052	0.1582 ± 0.0085	0.1021 ± 0.0036	0.1867 ± 0.0913	0.1364 ± 0.0061	0.0672 ± 0.0035
	0.3	0.1982 ± 0.0054	0.1105 ± 0.0031	0.1134 ± 0.0042	0.1010 ± 0.0034	0.1142 ± 0.0050	0.1022 ± 0.0045	0.0663 ± 0.0034
	0	0.0886 ± 0.0021	0.0901 ± 0.0036	0.0899 ± 0.0026	0.1012 ± 0.0040	0.1034 ± 0.0037	0.0914 ± 0.0025	0.0658 ± 0.0018
Education	0.6	0.2667 ± 0.0122	0.1551 ± 0.0064	0.1696 ± 0.0097	0.1034 ± 0.0029	0.1454 ± 0.0035	0.1517 ± 0.0068	0.0698 ± 0.0030
	0.3	0.2256 ± 0.0065	0.1288 ± 0.0048	0.1300 ± 0.0024	0.0988 ± 0.0013	0.1107 ± 0.0042	0.1161 ± 0.0041	0.0705 ± 0.0018
	0	0.0995 ± 0.0036	0.1012 ± 0.0082	0.1004 ± 0.0028	0.0985 ± 0.0027	0.1000 ± 0.0014	0.1022 ± 0.0026	0.0702 ± 0.0036
emotions	0.6	0.1920 ± 0.0275	0.1811 ± 0.0129	0.1913 ± 0.0146	0.4199 ± 0.0113	0.2159 ± 0.0178	0.3922 ± 0.0218	0.1478 ± 0.0144
	0.3	0.1689 ± 0.0199	0.2066 ± 0.0234	0.1675 ± 0.0234	0.4112 ± 0.0245	0.1821 ± 0.0186	0.3694 ± 0.0086	0.1554 ± 0.0113
	0	0.1592 ± 0.0251	0.1622 ± 0.0223	0.1585 ± 0.0205	0.3723 ± 0.0157	0.1706 ± 0.0143	0.3651 ± 0.0068	0.1545 ± 0.0146
Health	0.6	0.1983 ± 0.0101	0.1224 ± 0.0038	0.1127 ± 0.0020	0.0662 ± 0.0021	0.0925 ± 0.0050	0.1004 ± 0.0037	0.0393 ± 0.0013
	0.3	0.1639 ± 0.0030	0.1122 ± 0.0058	0.0808 ± 0.0042	0.0652 ± 0.0018	0.1383 ± 0.1415	0.0710 ± 0.0038	0.0397 ± 0.0015
	0	0.0575 ± 0.0023	0.1064 ± 0.0046	0.0612 ± 0.0027	0.0659 ± 0.0013	0.0604 ± 0.0044	0.0622 ± 0.0026	0.0395 ± 0.0029
medical	0.6	0.0472 ± 0.0072	0.0468 ± 0.0036	0.0600 ± 0.0161	0.0249 ± 0.0032	0.3189 ± 0.1622	0.0454 ± 0.0113	0.0224 ± 0.0047
	0.3	0.0274 ± 0.0076	0.0308 ± 0.0048	0.0282 ± 0.0068	0.0212 ± 0.0027	0.3189 ± 0.1622	0.0248 ± 0.0076	0.0219 ± 0.0044
	0	0.0158 ± 0.0060	0.0214 ± 0.0062	0.0160 ± 0.0060	0.0209 ± 0.0032	0.2398 ± 0.1216	0.0153 ± 0.0058	0.0218 ± 0.0041
Natural_scene	0.6	0.1935 ± 0.0095	0.1967 ± 0.0087	0.0952 ± 0.0114	0.1024 ± 0.0073	0.1711 ± 0.0117	0.2016 ± 0.0104	0.1822 ± 0.0148
	0.3	0.1928 ± 0.0095	0.2065 ± 0.0066	0.1042 ± 0.0108	0.1035 ± 0.0066	0.1398 ± 0.0146	0.1907 ± 0.0111	0.1771 ± 0.0151
	0	0.1901 ± 0.0123	0.2241 ± 0.0072	0.0842 ± 0.0104	0.0965 ± 0.0045	0.1116 ± 0.0086	0.1796 ± 0.0097	0.1822 ± 0.0148
yeast data	0.6	0.3322 ± 0.0108	0.1765 ± 0.0078	0.1800 ± 0.0084	0.2040 ± 0.0029	0.2665 ± 0.0059	0.1764 ± 0.0073	0.1748 ± 0.0056
	0.3	0.2456 ± 0.0065	0.1754 ± 0.0069	0.1743 ± 0.0088	0.2036 ± 0.0029	0.2544 ± 0.0258	0.1716 ± 0.0077	0.1749 ± 0.0061
	0	0.2308 ± 0.0094	0.1756 ± 0.0085	0.1702 ± 0.0076	0.1768 ± 0.0044	0.2421 ± 0.0127	0.1702 ± 0.0077	0.1744 ± 0.0101

Table 9. Aggregated comparison of GLCS by missing-label rate. A win means that GLCS outperforms the best competing method for a given dataset, metric, and missing-label rate.

Missing Rate	Wins	Ties	Losses	Average GLCS Rank
<i>r</i> = 0.6	54	0	18	1.59
<i>r</i> = 0.3	50	0	22	1.66
<i>r</i> = 0	38	0	34	2.38

As expected, performance generally becomes more challenging when *r* increases because fewer label entries are available during training. Under the high missing-label ratio *r* = 0.6, GLCS remains among the strongest methods on most datasets, indicating that combining correlation propagation with sparse label-specific prediction is helpful when supervision is limited. When more labels are observed at *r* = 0.3 and *r* = 0, most methods improve, while GLCS maintains stable performance across different supervision levels.

Across the ranking-based metrics, GLCS obtains strong AP and AUC on many text datasets such as Arts, Business, Computers, Education, Health, and medical. It also performs competitively on image, audio, music, and biology datasets, although some baselines may obtain the best value on individual datasets or metrics. For the lower-is-better metrics, including CV, HL, OE, and RL, GLCS often reduces label-ranking errors and label-wise misclassification rates. These results suggest that the proposed model is not only effective for ranking relevant labels near the top, but also stable for binary label prediction under incomplete supervision.

Although GLCS achieves strong overall performance, it is not dominant in all cases. On Corel5k, GLCS remains competitive, indicating that low-rank correlation learning and label-specific sparsity are useful in larger label spaces. However, its advantage is weaker on Natural_scene, which contains only six labels, and on some conditions of yeast data and Cal500, where simpler structures or alternative ranking behavior can be sufficient. Moreover, GLCS is not always the best method on AUC, where LRMML or DM2L can be competitive. These observations suggest that GLCS is most useful when label dependencies are richer and supervision is incomplete, while its relative advantage may decrease on small-label datasets or metrics less sensitive to correlation-based refinement.

4.4. Statistical Significance Analysis

To rigorously assess whether the observed performance differences among the compared algorithms are statistically significant, we conduct the non-parametric Friedman test separately for each evaluation metric. The test is performed over $N = 36$ experimental conditions, corresponding to twelve datasets under three missing-label ratios. For AP and AUC, larger values are assigned better ranks, whereas for CV, HL, OE, and RL, smaller values are assigned better ranks. Out-of-memory results are assigned the worst rank in the corresponding condition. Table 10 reports the Friedman statistics and Kendall's W for the six evaluation metrics. With seven algorithms, the critical value of the Friedman chi-square test at $\alpha = 0.05$ is 12.5916. As shown in the table, the Friedman statistics for AP, AUC, CV, HL, OE, and RL are all larger than this critical value. Therefore, the null hypothesis that all compared algorithms perform equivalently is rejected at the 95% confidence level for all six metrics. To quantify the effect size, Kendall's W is computed as $W = \chi_F^2 / [N(k - 1)]$, where $N = 36$ and $k = 7$. The resulting W values indicate that the performance differences are strong for AUC and HL, moderate for AP, and smaller but still meaningful for CV, OE, and RL.

Table 10. Friedman test statistics, critical value, and Kendall's W effect size at $\alpha = 0.05$.

Evaluation Metrics	χ_F^2	Critical Value	Kendall's W
AP	86.6250	12.5916	0.4010
AUC	128.9048		0.5968
CV	46.7262		0.2163
HL	104.6429		0.4845
OE	46.2262		0.2140
RL	56.2083		0.2602

After the Friedman test, we further apply the Nemenyi post hoc test for pairwise multiple comparisons at $\alpha = 0.05$. This test controls the family-wise comparison among all algorithms based on their average ranks. With seven algorithms and 36 experimental conditions, the critical distance is $CD = 1.5010$. Figure 1 presents the critical-difference diagrams for AP, AUC, CV, HL, OE, and RL. Algorithms connected by a horizontal line are not significantly different at the 0.05 significance level. The diagrams show that GLCS obtains the best or among the best average ranks on most metrics, especially AP, CV, HL, OE, and RL. For AUC, LRMML obtains the best average rank, while GLCS remains competitive and close to the top-ranked methods.

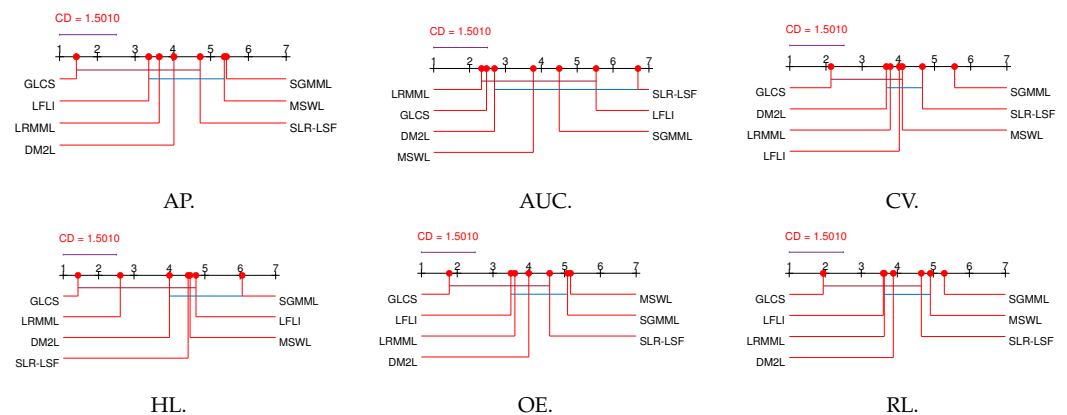


Figure 1. Nemenyi test results for different evaluation metrics. Red lines indicate the average-rank positions of individual algorithms, while blue horizontal lines connect algorithms with no statistically significant difference.

4.5. Running Time Analysis

Table 11 reports the average training time per fold at $r = 0.3$. LFLI is the fastest method on average, while MSWL and LRMML also have relatively low running time. GLCS requires 283.406 s per fold on average, which is close to DM2L and lower than SGMMML and SLR-LSF. The largest running time of GLCS occurs on Corel5k, indicating that the dense label-correlation matrix and SVT update become the main computational bottlenecks when the label dimension is large. SLR-LSF encounters an out-of-memory failure on Corel5k, which is reported as OM.

Table 11. Running time comparison at $r = 0.3$. The reported value is the mean training time per fold in seconds over five-fold cross-validation.

Dataset	SGMMML	SLR-LSF	LRMML	DM2L	MSWL	LFLI	GLCS
Arts	208.255	2221.622	186.056	328.139	34.627	76.366	197.714
Business	680.137	3681.930	247.908	377.160	40.857	108.570	203.940
birds	2.997	15.017	38.763	5.638	3.487	1.003	15.224
Cal500	2.090	2018.035	118.460	17.416	274.680	1.062	126.335
Corel5k	731.820	OM	115.448	1548.231	657.551	127.832	1956.689
Computers	716.159	4998.232	114.762	296.456	48.276	101.754	302.927
Education	6138.891	6138.891	10.621	278.931	41.895	80.490	235.780
emotions	1.972	3.464	1.761	2.188	5.248	0.283	2.891
Health	228.540	7388.981	205.117	289.804	43.747	91.657	246.128
medical	45.638	477.283	165.570	21.501	49.092	5.673	43.288
Natural_scene	17.949	18.452	8.870	8.857	5.306	3.305	18.351
yeast data	53.853	109.108	12.214	40.897	12.915	9.665	51.610
Average	735.692	2461.001	102.129	267.935	101.473	50.638	283.406

4.6. Parameter Sensitivity Analysis

We examine the sensitivity of GLCS to three trade-off parameters, λ_1 , λ_2 , and λ_3 , which weigh the ℓ_1 sparsity on the LSF predictor W , the nuclear-norm regularization on the learned correlation matrix C , and the label-graph smoothness term $\text{tr}(WLW^T)$, respectively. The analysis is conducted on the Birds dataset with $r = 0.3$. Each parameter is varied in $\{10^{-5}, 10^{-4}, \dots, 10^0\}$. To facilitate interpretation, we vary two parameters at a time while fixing the remaining one to 10^{-5} and report AP and HL in Figure 2.

Figure 2 shows that GLCS maintains competitive AP and HL over a wide range of settings, suggesting limited sensitivity to moderate changes of the regularization weights. The influence of λ_1 is comparatively mild, whereas overly large λ_1 may reduce performance due to excessive sparsification of W . The effects of λ_2 and λ_3 are more apparent. Larger λ_2 tends to improve performance by promoting a compact low-rank structure in

C , which benefits correlation-based supervision propagation. A moderate λ_3 is helpful by encouraging smoothness over the label graph, while excessively large λ_3 can lead to over-smoothing and slightly degrade AP or HL. Overall, the results indicate that GLCS is robust when λ_2 and λ_3 are selected in a moderate range and λ_1 remains non-dominant.

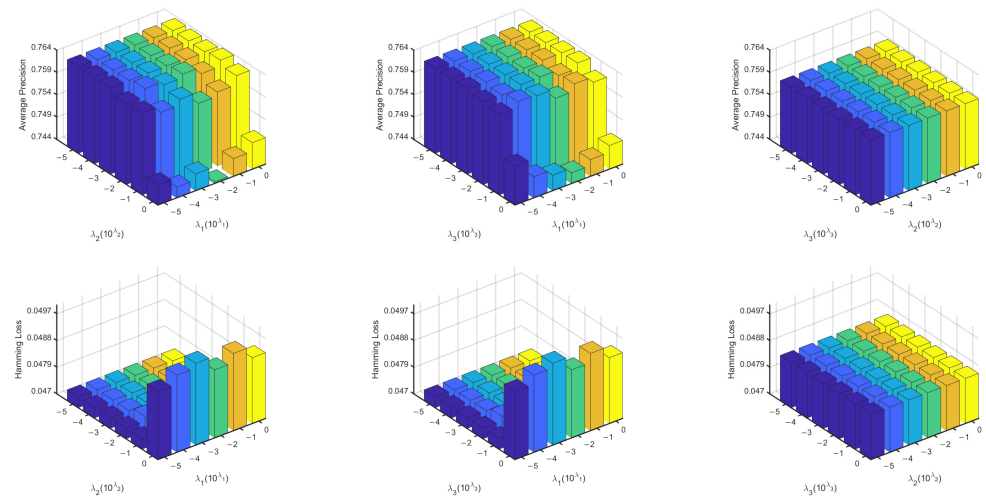


Figure 2. Parameter sensitivity analysis on the birds dataset for AP and HL.

4.7. Convergence Analysis

We analyze the convergence behavior of GLCS by monitoring the objective function value across iterations on two representative datasets, medical and birds, under $r = 0.3$. The hyperparameters are fixed to the values used in the main experiments.

As shown in Figure 3, the objective value decreases sharply in the early iterations and then changes only marginally after about 20 iterations on both datasets, indicating that the optimization procedure converges efficiently. Similar convergence trends are observed on the other benchmark datasets, suggesting that GLCS can reliably solve the proposed objective in practice.

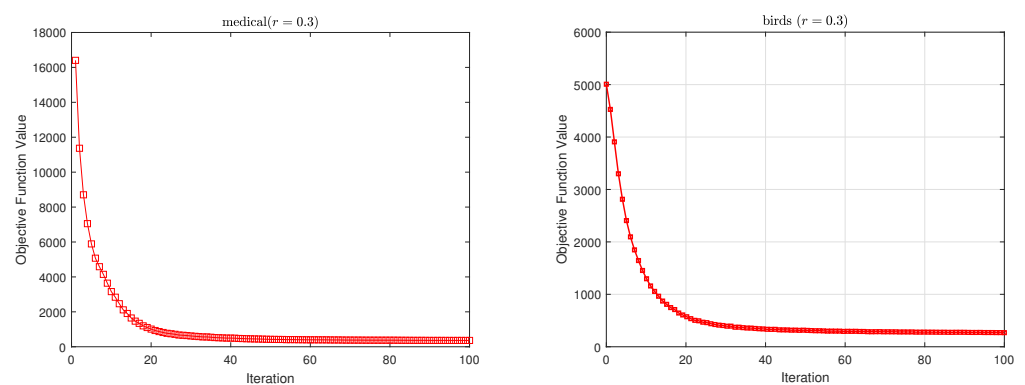


Figure 3. Convergence analysis on medical and birds under $r = 0.3$.

5. Conclusions

This paper studies multi-label learning with incomplete supervision and proposes GLCS, a graph-regularized low-rank correlation learning framework with label-specific features. GLCS uses observed label entries as reliable sources and propagates them through a learned low-rank correlation matrix. It further learns sparse label-specific predictors and constructs a label graph from the learned correlations to preserve local label-structure consistency. In this way, missing-label refinement and discriminative feature learning are optimized in the same framework.

Experiments on twelve benchmark datasets show that GLCS achieves stable performance under different missing-label ratios and obtains strong results on both ranking-based and classification-based metrics. The statistical tests further indicate that the observed differences are significant for most evaluation criteria. These results suggest that combining low-rank label correlation, graph regularization, and label-specific sparsity is effective for learning from incomplete multi-label annotations.

One limitation of GLCS is that it relies on iterative updates involving matrix regularization and singular value thresholding, which may increase computational cost when the number of labels is very large. Future work will explore more scalable optimization strategies, approximate low-rank updates, and extensions to more complex missing mechanisms with additional annotation noise. Another limitation is that the experiments mainly use random missing-entry simulation. Although this protocol provides a controlled way to compare different methods, it does not fully represent realistic missingness mechanisms such as missing positives, long-tail missingness, and single-positive supervision. We will evaluate GLCS under these more realistic annotation mechanisms in future work.

Author Contributions: Conceptualization, T.L. and M.F.N.; methodology, T.L.; software, T.L., X.P. and H.X.; validation, T.L., X.P. and H.X.; formal analysis, T.L.; investigation, T.L.; resources, M.F.N.; data curation, T.L.; writing—original draft preparation, T.L.; writing—review and editing, M.F.N., X.P. and H.X.; visualization, T.L.; supervision, M.F.N.; project administration, M.F.N.; funding acquisition, M.F.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the FRGS/1/2022/ICT02/UKM/02/5 research grant, the Natural Science Research Project of Anhui Educational Committee, Key Project, grant number 2024AH051339, and the Horizontal Scientific Research Project of Chaohu University, Anhui Province, China, grant number hxkt2604009.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in the MLL Resources repository at <https://www.uco.es/kdis/mlresources/> (accessed on 23 July 2026).

Acknowledgments: The authors would like to thank the providers of the publicly available multi-label learning datasets used in this study.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Zhang, M.-L.; Zhou, Z.-H. A review on multi-label learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2013**, *26*, 1819–1837. <https://doi.org/10.1109/TKDE.2013.39>.
2. Xiao, J.; Li, S.; Lin, T.; Zhu, J.; Yuan, X.; Feng, D.D.; Sheng, B. Multi-label chest x-ray image classification with single positive labels. *IEEE Trans. Med Imaging* **2024**, *43*, 4404–4418. <https://doi.org/10.1109/TMI.2024.3421644>.
3. Taha, A.Y.; Tiun, S.; Rahman, A.H.A.; Ayob, M.; Abdulameer, A.S. Unified graph-based missing label propagation method for multilabel text classification. *Symmetry* **2022**, *14*, 286. <https://doi.org/10.3390/sym14020286>.
4. Li, C.; Ab Aziz, M.J.; Yaakub, M.R. Multi-Label Text Classification Via Document Enhancement Furthermore, Label Correlations Learning. *Malays. J. Comput. Sci.* **2025**, *38*, 131–151. <https://doi.org/10.22452/mjcs.vol38no2.1>.
5. Zhang, Y.; Chen, M.; Shen, J.; Wang, C. Tailor versatile multi-modal learning for multi-label emotion recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 22 February–1 March 2022; Volume 36, pp. 9100–9108. <https://doi.org/10.1609/aaai.v36i8.20895>.
6. Li, R.; Du, J.; Ding, J.; Jia, L.; Chen, Y.; Shang, Z. Semi-supervised multi-label dimensionality reduction learning by instance and label correlations. *Mathematics* **2023**, *11*, 782. <https://doi.org/10.3390/math11030782>.

7. Hang, J.-Y.; Zhang, M.-L. Dual perspective of label-specific feature learning for multi-label classification. *ACM Trans. Knowl. Discov. Data* **2024**, *19*, 21. <https://doi.org/10.1145/3705006>.
8. Li, T.; Nasrudin, M.F.; Zhao, D.; Chen, F.; Peng, X.; Sarim, H.M. Uncovering hidden patterns: Low-rank label correlations for multi-label weak-label learning. *Int. J. Mach. Learn. Cybern.* **2025**, *16*, 1371–1387. <https://doi.org/10.1007/s13042-024-02341-x>.
9. Kumar, S.; Rastogi, R. Low rank label subspace transformation for multi-label learning with missing labels. *Inf. Sci.* **2022**, *596*, 53–72. <https://doi.org/10.1016/j.ins.2022.03.015>.
10. Huang, J.; Xu, L.; Qian, K.; Wang, J.; Yamanishi, K. Multi-label learning with missing and completely unobserved labels. *Data Min. Knowl. Discov.* **2021**, *35*, 1061–1086. <https://doi.org/10.1007/s10618-021-00743-x>.
11. Huang, M.; Zhao, Y.; Wang, Y.; Wahab, F.; Sun, Y.; Chen, C. Multi-graph multi-label learning with novel and missing labels. *Knowl.-Based Syst.* **2023**, *276*, 110753. <https://doi.org/10.1016/j.knosys.2023.110753>.
12. Hao, X.; Huang, J.; Qin, F.; Zheng, X. Multi-label learning with missing features and labels and its application to text categorization. *Intell. Syst. Appl.* **2022**, *14*, 200086. <https://doi.org/10.1016/j.iswa.2022.200086>.
13. Xu, S.; Yang, X.; Yu, H.; Yu, D.J.; Yang, J.; Tsang, E.C. Multi-label learning with label-specific feature reduction. *Knowl.-Based Syst.* **2016**, *104*, 52–61.
14. Cheng, Z.; Zeng, Z. Joint label-specific features and label correlation for multi-label learning with missing label. *Appl. Intell.* **2020**, *50*, 4029–4049. <https://doi.org/10.1007/s10489-020-01715-2>.
15. Li, J.; Zhang, C.; Zhou, J.T.; Fu, H.; Xia, S.; Hu, Q. Deep-LIFT: Deep label-specific feature learning for image annotation. *IEEE Trans. Cybern.* **2021**, *52*, 7732–7741. <https://doi.org/10.1109/TCYB.2021.3049630>.
16. Yu, Z.-B.; Zhang, M.-L. Multi-label classification with label-specific feature generation: A wrapped approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5199–5210. <https://doi.org/10.1109/TPAMI.2021.3070215>.
17. Jiang, J.; Ge, W.; Wang, Y.; Cheng, Y.; Xu, Y. Soft-label recover based label-specific features learning. *Sci. Rep.* **2024**, *14*, 23099. <https://doi.org/10.1038/s41598-024-72765-6>.
18. Lin, Z.; Ding, G.; Hu, M.; Wang, J.; Ye, X. Image tag completion via image-specific and tag-specific linear sparse reconstructions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, Oregon, 23–28 June 2013; pp. 1618–1625.
19. Xu, M.; Jin, R.; Zhou, Z. Speedup Matrix Completion with Side Information: Application to Multi-Label Learning. In Proceedings of the Advances in Neural Information Processing Systems 26 (NeurIPS 2013), Lake Tahoe, NV, USA, 5–8 December 2013; pp. 2301–2309.
20. Wu, B.; Lyu, S.; Ghanem, B. Ml-mg: Multi-label learning with missing labels using a mixed graph. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4157–4165.
21. Liu, Y.; Wen, K.; Gao, Q.; Gao, X.; Nie, F. SVM based multi-label learning with missing labels for image annotation. *Pattern Recognit.* **2018**, *78*, 307–317. <https://doi.org/10.1016/j.patcog.2018.01.022>.
22. Cheng, Y.; Qian, K.; Wang, Y.; Zhao, D. Missing multi-label learning with non-equilibrium based on classification margin. *Appl. Soft Comput.* **2020**, *86*, 105924. <https://doi.org/10.1016/j.asoc.2019.105924>.
23. Akbarnejad, A.H.; Baghshah, M.S. An efficient semi-supervised multi-label classifier capable of handling missing labels. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 229–242. <https://doi.org/10.1109/TKDE.2018.2833850>.
24. Jain, V.; Modhe, N.; Rai, P. Scalable generative models for multi-label learning with missing labels. In Proceedings of the International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017; pp. 1636–1644.
25. Wu, B.; Jia, F.; Liu, W.; Ghanem, B.; Lyu, S. Multi-label learning with missing labels using mixed dependency graphs. *Int. J. Comput. Vis.* **2018**, *126*, 875–896. <https://doi.org/10.1007/s11263-018-1085-3>.
26. Ma, J.; Chow, T.W. Robust non-negative sparse graph for semi-supervised multi-label learning with missing labels. *Inf. Sci.* **2018**, *422*, 336–351. <https://doi.org/10.1016/j.ins.2017.08.061>.
27. Wei, T.; Guo, L.-Z.; Li, Y.-F.; Gao, W. Learning safe multi-label prediction for weakly labeled data. *Mach. Learn.* **2018**, *107*, 703–725. <https://doi.org/10.1007/s10994-017-5675-z>.
28. Ma, Z.; Chen, S. Expand globally, shrink locally: Discriminant multi-label learning with missing labels. *Pattern Recognit.* **2021**, *111*, 107675. <https://doi.org/10.1016/j.patcog.2020.107675>.
29. Zhu, Y.; Kwok, J.T.; Zhou, Z.-H. Multi-label learning with global and local label correlation. *IEEE Trans. Knowl. Data Eng.* **2017**, *30*, 1081–1094. <https://doi.org/10.1109/TKDE.2017.2785795>.
30. Su, X.; Xu, Y. Imbalanced and missing multi-label data learning with global and local structure. *Inf. Sci.* **2024**, *677*, 120910. <https://doi.org/10.1016/j.ins.2024.120910>.
31. Cheng, Z.; Tan, Z. Multi-label learning for label-specific features using correlation information with missing label. *Expert Syst. Appl.* **2025**, *269*, 126491.
32. Liu, W. Copula multi-label learning. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 6334–6343.
33. Huang, J.; Qin, F.; Zheng, X.; Cheng, Z.; Yuan, Z.; Zhang, W.; Huang, Q. Improving multi-label classification with missing labels by learning label-specific features. *Inf. Sci.* **2019**, *492*, 124–146. <https://doi.org/10.1016/j.ins.2019.04.021>.

34. Zhao, D.; Tan, Y.; Sun, D.; Gao, Q.; Lu, Y.; Zhu, D. Multi-label learning of missing labels using label-specific features: an embedded packaging method. *Appl. Intell.* **2024**, *54*, 791–814. <https://doi.org/10.1007/s10489-023-05203-1>.
35. He, Z.-F.; Zhang, C.-H.; Liu, B.; Li, B. Label recovery and label correlation co-learning for multi-view multi-label classification with incomplete labels. *Appl. Intell.* **2023**, *53*, 9444–9462. <https://doi.org/10.1007/s10489-022-03945-y>.
36. Feng, L.; Huang, J.; Shu, S.; An, B. Regularized matrix factorization for multilabel learning with missing labels. *IEEE Trans. Cybern.* **2020**, *52*, 3710–3721. <https://doi.org/10.1109/TCYB.2020.3016897>.
37. Jiang, Q.; Li, P.; Zhang, Y.; Hu, X. Global and adaptive local label correlation for multi-label learning with missing labels. In Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN), Gold Coast, Australia, 18–23 June 2023; pp. 1–8. <https://doi.org/10.1109/IJCNN54540.2023.10191231>.
38. Cole, E.; Mac Aodha, O.; Lorieul, T.; Perona, P.; Morris, D.; Jojic, N. Multi-label learning from single positive labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 933–942.
39. Chen, C.; Zhao, Y.; Li, J. Semantic contrastive bootstrapping for single-positive multi-label recognition. *Int. J. Comput. Vis.* **2023**, *131*, 3289–3306. <https://doi.org/10.1007/s11263-023-01849-z>.
40. Song, G.; Kim, N.-r.; Lee, J.-S.; Lee, J.-H. Ignore: Information gap-based false negative loss rejection for single positive multi-label learning. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2024; pp. 472–488. https://doi.org/10.1007/978-3-031-72754-2_27.
41. Xing, X.; Xiong, Z.; Stylianou, A.; Sastry, S.; Gong, L.; Jacobs, N. Vision-language pseudo-labels for single-positive multi-label learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 7799–7808.
42. Wang, Y.B.; Hang, J.Y.; Zhang, M.L. Multi-label open set recognition. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 5739–5756.
43. Zhang, M.-L.; Wu, L. Lift: Multi-label learning with label-specific features. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 107–120. <https://doi.org/10.1109/TPAMI.2014.2339815>.
44. Tan, Y.; Sun, D.; Shi, Y.; Gao, L.; Gao, Q.; Lu, Y. Bi-directional mapping for multi-label learning of label-specific features. *Appl. Intell.* **2022**, *52*, 8147–8166. <https://doi.org/10.1007/s10489-021-02868-4>.
45. Hang, J.-Y.; Zhang, M.-L. Collaborative learning of label semantics and deep label-specific features for multi-label classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 9860–9871. <https://doi.org/10.1109/TPAMI.2021.3136592>.
46. Zhang, J.; Liu, K.; Yang, X.; Ju, H.; Xu, S. Multi-label learning with relief-based label-specific feature selection. *Appl. Intell.* **2023**, *53*, 18517–18530. <https://doi.org/10.1007/s10489-022-04350-1>.
47. Li, R.; Ouyang, Z.; Shang, Z.; Jia, L.; Li, X. Learning common and label-specific features for multi-label classification with missing labels. *IEEE Access* **2024**, *12*, 81170–81195. <https://doi.org/10.1109/ACCESS.2024.3411095>.
48. Zhang, Y.-F.; Zhang, M.-L. Generalization analysis for label-specific representation learning. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 104904–104933. <https://doi.org/10.52202/079017-3331>.
49. Kumar, S.; Ahmadi, N.; Rastogi, R. Multi-label learning with missing labels using sparse global structure for label-specific features. *Appl. Intell.* **2023**, *53*, 18155–18170. <https://doi.org/10.1007/s10489-022-04439-7>.
50. Zhang, J.; Li, S.; Jiang, M.; Tan, K.C. Learning from weakly labeled data based on manifold regularized sparse model. *IEEE Trans. Cybern.* **2020**, *52*, 3841–3854. <https://doi.org/10.1109/TCYB.2020.3015269>.
51. Jain, H.; Prabhu, Y.; Varma, M. Extreme Multi-label Loss Functions for Recommendation, Tagging, Ranking & Other Missing Label Applications. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 935–944. <https://doi.org/10.1145/2939672.2939756>.
52. Qaraei, M.; Schultheis, E.; Gupta, P.; Babbar, R. Convex Surrogates for Unbiased Loss Functions in Extreme Classification with Missing Labels. In Proceedings of the The Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 3711–3720. <https://doi.org/10.1145/3442381.3450139>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.