Label Distribution Imputation and Bias-Corrective Representation Learning for Incomplete and Imbalanced Label Distribution

Xiangcheng Sun[1,2], **Miaogen Ling**[3], **Han Qin**[1,2], **Guohua Lv**[1,2], **Yongbiao Gao**[1,2*]
[1] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China
[2] Shandong Provincial Key Laboratory of Computing Power Internet and Service Computing, Shandong Fundamental Research Center for Computer Science, Jinan, China
[3] School of Computer Science and Technology, Nanjing University of Information Science and Technology, Nanjing, China
10431240010@stu.qlu.edu.cn, mgling@nuist.edu.cn, {qinh, guohualv, gaoyb}@qlu.edu.cn

Abstract

Label Distribution Learning (LDL) represents each instance with a label distribution, but these distributions are often incomplete in practice due to high annotation costs and annotators' cognitive burden. Recent Incomplete Label Distribution Learning (InLDL) methods leverage global and local correlations to impute missing values, imputation errors remain systematic and non-uniform across labels. However, in the more challenging scenario of imbalanced label distributions, such non-uniform errors disproportionately harm low-degree labels, skewing learning toward dominant labels. To address the more complex challenge, we propose **L**abel **D**istribution **I**mputation and **B**ias-**C**orrective **R**epresentations (**LDIBR**) for incomplete and imbalanced label distribution learning. LDIBR performs instance-adaptive imputation conditioned on instance features and the binary observation mask. It learns a prior, a reliability-gated correction, and entry-wise fusion weights to produce a normalized imputed distribution. Additionally, LDIBR decomposes the imputed distributions into a shared low-rank core and an instance-specific residual. The core captures stable global degree patterns, while the residual isolates sample-specific deviations and imputation noise. Comprehensive experiments across multiple benchmark datasets with a 50% missing ratio validate both the effectiveness and robustness of the LDIBR approach. The code is available at https://github.com/wenhuihji/LDCBR.

1 Introduction

In supervised learning, a common classification paradigm is single-label learning (SLL) [Krizhevsky *et al.*, 2012], in which each instance is assigned a single label. When multiple labels can describe an instance, multi-label learning (MLL) [Zhang and Zhou, 2013; Tsoumakas and Katakis, 2008] treats label relevance as binary indicators. However, in real-world scenarios, labels often describe an instance to different extents rather than being simply relevant or irrelevant. To capture this ambiguity, Label Distribution Learning (LDL) [Geng, 2016] assigns a label distribution to each instance. Each element in this distribution corresponds to a description degree that quantifies the strength with which a specific label characterizes the instance. Owing to its ability to capture fine-grained label ambiguity, LDL has been successfully applied to numerous real-world tasks, including age estimation [He *et al.*, 2021; Yu *et al.*, 2023; Lee *et al.*, 2024], emotion analysis [Wang and Li, 2025; Zhao *et al.*, 2025], and facial beauty perception [Li *et al.*, 2024; Cheng *et al.*, 2024; Liu *et al.*, 2025], *etc*.

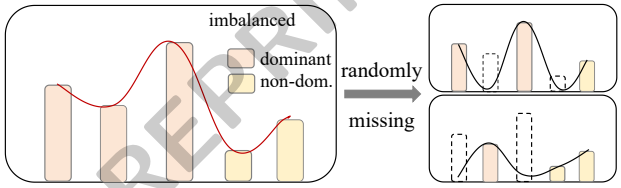


Figure 1: Illustration of the effect of label missingness under imbalanced label distributions.

LDL [Geng, 2016] is widely adopted for modeling label ambiguity. Standard LDL models are typically trained by minimizing the KL divergence between predicted and observed label distributions. WInLDL [Li and Chen, 2024] shows that more accurate modeling of label distributions improves performance. This holds when label distributions are mostly complete and description degrees exhibit low variance. However, in practice, label distributions are often incomplete. Description degrees can also vary substantially across labels, resulting in imbalanced label distributions. Under such imbalance, optimization is dominated by labels with high description degrees, while labels with low description degrees receive insufficient supervision [Gao *et al.*, 2025]. Missing entries in label distributions further exacerbate this

*Corresponding Author.

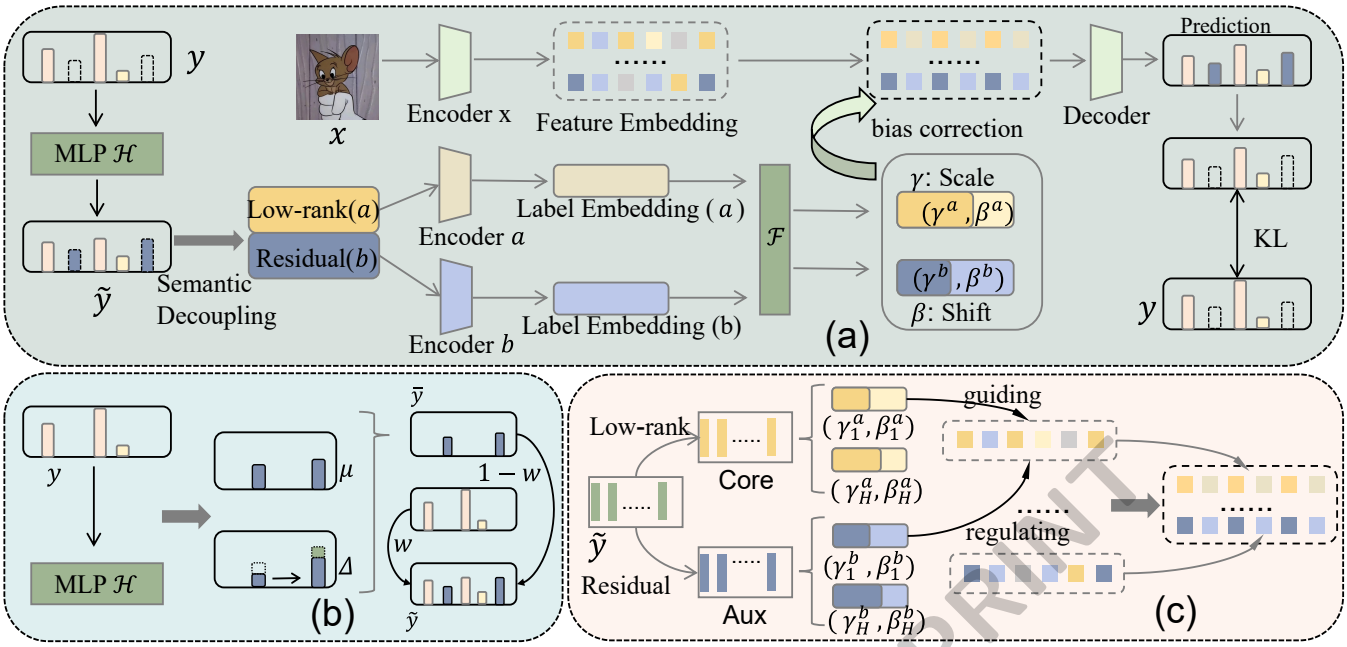


Figure 2: Overview of the proposed LDIBR framework. (a) Overall pipeline. LDIBR imputes incomplete label distributions and decomposes them into a shared low-rank core and an instance-specific residual for bias-corrective representation learning. (b) Label distribution imputation. Missing description degrees are recovered by exploiting label correlations. (c) Asymmetric feature modulation. The core and residual jointly guide representation learning to mitigate bias under imbalanced label distributions.

issue, which weakens supervision for labels with low description degrees and distorts the already skewed training signal. As illustrated in Fig. 1, the impact of missing labels becomes more pronounced under imbalanced label distributions, further skewing optimization and degrading model performance. Nevertheless, existing methods typically address label missingness or degree imbalance in isolation. There is no relevant research work that addresses the challenges in this more complex scenario.

In real-world scenarios, acquiring complete label distributions is often impractical due to high annotation costs. Existing incomplete label distribution learning (InLDL) methods mainly adopt two approaches: (i) mitigating missing information through suppression, smoothing, or weighted modeling (e.g., WInLDL [Li and Chen, 2024] and GLDL [Jin *et al.*, 2024]), and (ii) reconstructing distributions by exploiting label correlations via a low-rank structure or matrix factorization [Kou *et al.*, 2024; Kou *et al.*, 2025]. However, most approaches treat imputation as a standalone step and fail to address how imputation errors propagate to representation learning. In imbalanced label distribution learning (ILDL), a few dominant labels contribute most of the total description degree. As a result, optimization is driven mainly by these labels, while labels with low description degrees receive weak supervision and insufficient updates [Gao *et al.*, 2025]. Even under random missingness, imputation relies on recovered degrees whose reliability varies across entries, making explicit reliability modeling crucial. In this setting, imputation errors on entries with high description degrees may dominate optimization and bias the training signal. Imputation can also introduce reconstruction noise, which may distort optimization dynamics and amplify bias in learned representations. While recent works address imbalance through representation alignment [Zhao *et al.*, 2023] or component decoupling [Gao *et al.*, 2025], they generally apply uniform strategies and overlook the interaction between imputation errors and the degree disparity across labels.

To address these challenges, we propose Label Distribution Imputation and Bias-Corrective Representations (LDIBR), a unified framework for learning with random label missingness under imbalanced label distributions. As shown in Fig. 2, LDIBR follows a two-stage design with correlation-preserving imputation and asymmetric bias mitigation. Unlike conventional methods, it further leverages a low-rank structure for bias correction. The completed distributions are decomposed into a shared semantic core and an instance-specific residual, enabling asymmetric feature modulation that reinforces the reliable core while suppressing imputation noise. Additionally, we provide a theoretical analysis and derive a generalization error bound that links the expected risk to the imputation and representation errors, with explicit dependence on the imbalance factor and the low-rank basis rank. Extensive experiments on multiple benchmark datasets validate the effectiveness and robustness of the LDIBR.

Our main contributions are summarized as follows:

- We propose LDIBR, a unified framework for label distribution imputation and bias-corrective representation learning under random label missingness and imbalanced label distributions. To the best of our knowledge, this is the first work in this more complex scenario.
- We theoretically analyze LDIBR and derive a general-

ization error bound that explicitly characterizes the effects of imputation error, representation bias, and description degree imbalance.

- We conduct extensive experiments under varying missing ratios and imbalance levels, demonstrating the effectiveness and robustness of LDIBR.

2 Related Work

2.1 Label Distribution Learning

Label Distribution Learning (LDL) assigns each instance a label distribution, where each element denotes the degree to which a label characterizes the instance. LDL has been widely applied to tasks with inherent label ambiguity, such as age estimation [He *et al.*, 2021; Yu *et al.*, 2023; Lee *et al.*, 2024], facial beauty perception [Li *et al.*, 2024; Cheng *et al.*, 2024; Liu *et al.*, 2025], and emotion analysis [Wang and Li, 2025; Wu *et al.*, 2025; Zhao *et al.*, 2025], *etc*. Recent advances have improved LDL by exploiting richer structural information, including graph-based models that capture label correlations [Jin *et al.*, 2024] and diffusion-based methods that recover structurally coherent label distributions under noisy supervision [Hou *et al.*, 2025; Wang *et al.*, 2024]. Nevertheless, most existing LDL approaches still assume complete and balanced label distributions, which limits the applicability in practical scenarios.

2.2 Imbalanced Label Distribution Learning

Imbalanced Label Distribution Learning (ILDL) considers scenarios where description degrees vary significantly across labels, resulting in highly skewed label distributions. To mitigate performance degradation caused by label imbalance, [Zhao *et al.*, 2023] proposed Representation Distribution Alignment (RDA), which aligns feature and label distributions in a shared latent space via KL divergence. RDA alleviates the feature shift induced by imbalance at the distribution level. However, it relies on a global alignment objective and cannot adaptively adjust the contributions of labels with different description degrees during representation learning. More recently, [Gao *et al.*, 2025] introduced Decoupled Imbalanced Label Distribution Learning (DILDL), which decomposes label distributions into dominant and non-dominant components and optimizes them using separate branches. While this decoupling strategy mitigates the suppression of non-dominant labels, DILDL still relies on static weighting schemes, shared optimization hyperparameters, and fully observed label distributions. Crucially, it does not exploit distribution components to provide fine-grained, instance-adaptive regulation for representation learning.

2.3 Incomplete Label Distribution Learning

Incomplete Label Distribution Learning (InLDL) focuses on learning label distributions when only partial annotation information is available for each instance. Early InLDL methods address incompleteness by exploiting structural assumptions of label distributions, such as low-rank structures within the label distribution matrix. These structures are enforced via trace norm regularization to capture global or local correlations [Wang *et al.*, 2021; Wang and Geng, 2023]. More recent studies relax the low-rank assumption by modeling label distributions on structured manifolds or graphs and leveraging both global and local correlations to improve label distribution learning under incomplete supervision [Jin *et al.*, 2024; Wang *et al.*, 2024]. In addition, neighborhood-based inference [Zhang *et al.*, 2025], matrix completion methods [Xu *et al.*, 2025; Li *et al.*, 2025a; Chang *et al.*, 2025], and robustness-oriented learning strategies [Li and Chen, 2024] have been explored to enhance performance under incomplete supervision. While existing methods effectively mitigate annotation incompleteness through correlation mining or structural regularization, they typically treat missing entries implicitly via masking, normalization, or reconstruction objectives. This conflates truly low description degrees with missing data at the label distribution level, biasing the training signal toward labels with higher description degrees and further suppressing labels with low description degrees. Moreover, most existing approaches do not explicitly consider label missingness under imbalanced distributions, which is a more prevalent and complex problem in real-world scenarios.

3 Method

3.1 Problem Definition

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^\top \in \mathbb{R}^{N\times d}$ denotes the feature matrix of N instances, where $\mathbf{x}_i \in \mathbb{R}^d$ is the feature vector of the i-th instance. Let $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_N]^\top \in \mathbb{R}^{N\times m}$ denotes the complete label distribution matrix over m labels, where $\mathbf{d}_i = [d_{\mathbf{x}_i}^{y_1}, \ldots, d_{\mathbf{x}_i}^{y_m}]^\top$, $d_{\mathbf{x}_i}^{y_j} \in [0,1]$, and $\sum_{j=1}^{m} d_{\mathbf{x}_i}^{y_j} = 1$.

In the InLDL setting, we introduce a binary mask $\mathbf{\Omega} \in \{0,1\}^{N\times m}$ to indicate label availability. The observed label distribution matrix $\mathbf{Y} \in \mathbb{R}^{N\times m}$ is defined as:

$$\mathbf{Y} = \mathbf{\Omega} \odot \mathbf{D}, \tag{1}$$

where $\Omega_{i,j} = 1$ if $d_{\mathbf{x}_i}^{y_j}$ is observed, and $\Omega_{i,j} = 0$ otherwise.

To characterize the label distribution imbalance, we define an imbalance factor γ based on the complete label distribution matrix $\mathbf{D}$ as:

$$\gamma = \frac{\max_{1\le j\le m} \sum_{i=1}^{N} D_{i,j}}{\min_{1\le j\le m} \sum_{i=1}^{N} D_{i,j}}, \tag{2}$$

where a larger γ indicates more severe dataset-level imbalance.

The objective is to learn a predictor $p_\theta : \mathbb{R}^d \to \Delta_m$ that infers the complete label distribution for each instance under incomplete and imbalanced supervision.

3.2 Distribution-Aware Label Completion

Inspired by [Gao *et al.*, 2025; Li and Chen, 2024], given the partially observed label distribution matrix $\mathbf{Y}$, we infer a complete label distribution for each instance by explicitly modeling instance-specific imputation dynamics.

For each instance $\mathbf{x}_i$, let $\bar{\mathbf{d}}_i = \mathbf{\Omega}_{i,:} \in \{0,1\}^m$ denote the binary observation mask, where $\bar{d}_{i,j} = 1$ if the j-th label is

observed and 0 otherwise. The corresponding missing ratio is computed as:

$$o_i = 1 - \frac{1}{m}\sum_{j=1}^{m} \bar{d}_{i,j}. \tag{3}$$

To infer missing description degrees, we employ a completion network $\mathcal{H}(\cdot)$ implemented as a multilayer perceptron. Specifically, $\mathcal{H}$ takes the concatenation of the instance feature vector and the observation mask as input and outputs four m-dimensional completion parameters:

$$(\boldsymbol{\mu}_i, \mathbf{s}_i, \boldsymbol{\Delta}_i, \mathbf{w}_i) = \mathcal{H}([\mathbf{x}_i; \bar{\mathbf{d}}_i]), \tag{4}$$

where $[\mathbf{x}_i; \bar{\mathbf{d}}_i]$ denotes concatenation. Here, $\boldsymbol{\mu}_i \in \mathbb{R}^m$ outputs the logits transformed into the prior label distribution, $\mathbf{s}_i \in \mathbb{R}^m$ encodes reliability information used to derive the variance vector $\boldsymbol{\sigma}_i^2$, $\boldsymbol{\Delta}_i \in \mathbb{R}^m$ provides a residual correction term, and $\mathbf{w}_i \in \mathbb{R}^m$ yields elementwise fusion weights for integrating observed and imputed signals.

The prior label distribution is obtained as $\mathbf{d}_i^{\text{prior}} = g(\boldsymbol{\mu}_i)$, where $g(\cdot)$ denotes the Softmax function. The corresponding variance vector $\boldsymbol{\sigma}_i^2$ is recovered from $\mathbf{s}_i$ via an exponential mapping:

$$\boldsymbol{\sigma}_i^2 = \exp(\mathbf{s}_i). \tag{5}$$

To ensure robust imputation, we adaptively control the correction strength using a reliability-aware gating mechanism:

$$\lambda_i = \text{sigmoid}\big(a - b_1 o_i - b_2 \|\boldsymbol{\sigma}_i^2\|_1\big), \tag{6}$$

where a, b_1, b_2 are learnable scalars and $\text{sigmoid}(\cdot)$ denotes the sigmoid function. The imputed label distribution signal is then formulated as:

$$\boldsymbol{\omega}_i = \mathbf{d}_i^{\text{prior}} + \lambda_i \tanh(\boldsymbol{\Delta}_i). \tag{7}$$

The refined label distribution is obtained by integrating the observed entries with the imputed signals via an adaptive weighting scheme:

$$\tilde{\mathbf{d}}_i = \mathbf{w}_i \odot \mathbf{Y}_{i,:} + (1 - \mathbf{w}_i) \odot \boldsymbol{\omega}_i. \tag{8}$$

Finally, we normalize the refined signal to obtain a valid label distribution:

$$\hat{\mathbf{d}}_i^{\text{comp}} = \frac{\max(\tilde{\mathbf{d}}_i, \epsilon)}{\mathbf{1}^\top \max(\tilde{\mathbf{d}}_i, \epsilon)}, \tag{9}$$

where $\epsilon > 0$ is a small constant for numerical stability. The resulting $\hat{\mathbf{d}}_i^{\text{comp}}$ serves as the completed label distribution for subsequent bias-corrective representation learning.

3.3 Bias-Corrective Representation Learning

Firstly, to enable bias-corrective representation learning, we decompose the completed label distribution $\hat{\mathbf{d}}_i^{\text{comp}} \in \mathbb{R}^m$ into a low-rank core and a bias-indicative residual:

$$\boldsymbol{\alpha}_i = \mathbf{B}^\top \hat{\mathbf{d}}_i^{\text{comp}}, \qquad \mathbf{d}_i^{(a)} = \mathbf{B}\boldsymbol{\alpha}_i, \qquad \mathbf{d}_i^{(b)} = \hat{\mathbf{d}}_i^{\text{comp}} - \mathbf{d}_i^{(a)}, \tag{10}$$

where $\mathbf{B} \in \mathbb{R}^{m \times r}$ is a learnable basis matrix with fixed rank r that is jointly optimized with θ, and $\boldsymbol{\alpha}_i \in \mathbb{R}^r$ denotes the corresponding latent coefficients. The core component $\mathbf{d}_i^{(a)}$ captures shared description degree patterns encoded by the low-rank basis, while the residual component $\mathbf{d}_i^{(b)}$ represents instance-specific residual variations that often manifest as distribution bias, enabling bias-corrective representation learning under incomplete and imbalanced label distributions [Kou *et al.*, 2024; Ma *et al.*, 2024; Li *et al.*, 2025b].

To ensure scale consistency, both components are normalized as:

$$\bar{\mathbf{d}}_i^{(k)} = \mathcal{N}\Big(\mathbf{d}_i^{(k)}\Big), \quad k \in \{a, b\}, \tag{11}$$

where $\mathcal{N}(\mathbf{z}) = \max(\mathbf{z}, \epsilon)/(\mathbf{1}^\top \max(\mathbf{z}, \epsilon))$ denotes the normalization operator.

The instance feature and the decoupled label distribution components are then projected into a shared latent space:

$$\mathbf{h}_i^x = \phi(\mathbf{x}_i), \qquad \mathbf{h}_i^{(a)} = \psi(\bar{\mathbf{d}}_i^{(a)}), \qquad \mathbf{h}_i^{(b)} = \psi(\bar{\mathbf{d}}_i^{(b)}), \tag{12}$$

where $\phi : \mathbb{R}^d \to \mathbb{R}^H$ and $\psi : \mathbb{R}^m \to \mathbb{R}^H$ denote the feature encoder and the label distribution encoder, respectively.

Secondly, we employ a gating mechanism to generate component-specific modulation parameters:

$$(\mathbf{g}_i^{(k)}, \boldsymbol{\beta}_i^{(k)}) = \mathcal{F}_k(\mathbf{h}_i^{(k)}), \quad k \in \{a, b\}, \tag{13}$$

where $\mathcal{F}_k(\cdot)$ is an MLP that outputs dimension-wise scaling and shifting coefficients.

Thirdly, we learn a dimension-wise responsibility mask to determine the active dimensions for modulation:

$$\mathbf{m}_i^{(a)} = \text{sigmoid}\Big(\mathcal{M}\big([\mathbf{h}_i^x; \mathbf{h}_i^{(a)}]\big)\Big), \qquad \mathbf{m}_i^{(b)} = \mathbf{1} - \mathbf{m}_i^{(a)}, \tag{14}$$

where $\mathcal{M}(\cdot)$ denotes the mask network, and bias-corrective modulation is performed via an asymmetric update rule:

$$\begin{aligned} \Delta_i^{(a)} &= \mathbf{m}_i^{(a)} \odot \big(\tanh(\mathbf{g}_i^{(a)}) \odot \mathbf{h}_i^x + \boldsymbol{\beta}_i^{(a)}\big), \\ \Delta_i^{(b)} &= \mathbf{m}_i^{(b)} \odot \big(\tanh(\mathbf{g}_i^{(b)}) \odot \mathbf{h}_i^x + \boldsymbol{\beta}_i^{(b)}\big). \end{aligned} \tag{15}$$

The final modulated feature is obtained as:

$$\tilde{\mathbf{h}}_i = \mathcal{R}\Big(\mathbf{h}_i^x + \Delta_i^{(a)} - \Delta_i^{(b)}\Big), \tag{16}$$

where $\mathcal{R}(\cdot)$ denotes a feature normalization operator.

Finally, the predicted label distribution is generated as:

$$\mathbf{h}_i^* = \mathcal{U}\Big([\mathbf{h}_i^x; \tilde{\mathbf{h}}_i]\Big), \qquad \hat{\mathbf{d}}_i^{\text{p}} = \text{softmax}(f(\mathbf{h}_i^*)), \tag{17}$$

where $\mathcal{U}(\cdot)$ denotes a fusion network and $f(\cdot)$ is the prediction head.

For analysis, we summarize the instance-wise modulation preference as:

$$\eta_i^{(a)} = \frac{1}{H}\sum_{t=1}^{H} m_{i,t}^{(a)}, \qquad \boldsymbol{\eta}_i = [\eta_i^{(a)}, 1 - \eta_i^{(a)}]. \tag{18}$$

During training, the predictor is supervised only by the observed entries of the label distribution. We minimize the masked KL divergence:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \sum_{j:\,\Omega_{i,j}=1} Y_{i,j} \log \frac{Y_{i,j}}{\hat{d}_{i,j}^{\text{p}}}. \tag{19}$$

This objective enforces consistency on observed entries. Under random missingness, the masked loss provides a consistent estimate of the full KL risk used in the theoretical analysis. The algorithm pseudocode is provided in the **Appendix**.

3.4 Theoretical Analysis

Motivated by recent Lipschitz-based generalization analyses for deep predictors [Zhang and Zhang, 2024], we analyze the generalization properties of LDIBR under incomplete and imbalanced supervision. Let P denotes the distribution over pairs $(\mathbf{x}, \mathbf{d})$, where $\mathbf{d} \in \Delta_m$ is the complete label distribution.

Definition 1. The expected risk is

$$R(\theta) = \mathbb{E}_{(\mathbf{x},\mathbf{d})\sim P}[\ell_{\mathrm{KL}}(\mathbf{d}, p_\theta(\mathbf{x}))], \tag{20}$$

and the empirical risk over N i.i.d. samples $\{(\mathbf{x}_i, \mathbf{d}_i)\}_{i=1}^N$ is

$$\hat{R}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \ell_{\mathrm{KL}}(\mathbf{d}_i, p_\theta(\mathbf{x}_i)). \tag{21}$$

The KL divergence is

$$\ell_{\mathrm{KL}}(\mathbf{d}, \mathbf{p}) = \sum_{j=1}^{m} d_j \log \frac{d_j}{p_j}, \tag{22}$$

where $\mathbf{d}, \mathbf{p} \in \Delta_m$. Since complete label distributions are not observable, LDIBR minimizes a masked KL loss on observed entries. Under random label missingness, this masked loss consistently estimates the full KL risk.

Definition 2. Let $\hat{\mathbf{d}}^{\,\mathrm{comp}}(\mathbf{x})$ and $\mathbf{h}^*(\mathbf{x})$ denote the completed label distribution and the bias-corrected representation. Define

$$\mathcal{E}_c = \mathbb{E}_{(\mathbf{x},\mathbf{d})\sim P}\left[\left\|\hat{\mathbf{d}}^{\,\mathrm{comp}}(\mathbf{x}) - \mathbf{d}\right\|_1\right], \tag{23}$$

$$\mathcal{E}_r = \mathbb{E}_{(\mathbf{x},\mathbf{d})\sim P}\left[\left\|\mathbf{h}^*(\mathbf{x}) - \mathbf{h}^\dagger(\mathbf{x},\mathbf{d})\right\|_2\right], \tag{24}$$

where $\mathbf{h}^\dagger(\mathbf{x}, \mathbf{d})$ is an oracle representation.

Assumption 1. $p_\theta(\mathbf{x}) = \mathrm{softmax}(f(\mathbf{h}^*(\mathbf{x})))$ and

$$p_\theta(\mathbf{x}) \in \Delta_m^{(\varepsilon_0)} = \{\mathbf{p} \in \Delta_m : \min_j p_j \geq \varepsilon_0\}, \tag{25}$$

where $\varepsilon_0 > 0$. Assume f is L_f-Lipschitz under ℓ_2, and softmax is L_s-Lipschitz from ℓ_2 to ℓ_1 on $\Delta_m^{(\varepsilon_0)}$ (cf. [Nguyen *et al.*, 2024]).

Theorem 1. For any $\delta \in (0, 1)$, with probability at least $1-\delta$,

$$R(\theta) \leq \hat{R}(\theta) + \frac{1}{\varepsilon_0}\mathcal{E}_c + \frac{L_s L_f}{\varepsilon_0}\mathcal{E}_r + C_1\frac{\gamma}{\sqrt{N}} + C_2\frac{r\log(2/\delta)}{N}, \tag{26}$$

where γ is the imbalance factor in Eq. (2), r is the rank of $\mathbf{B} \in \mathbb{R}^{m\times r}$, and C_1, C_2 are constants.

Proof. Let $\mathbf{p} = p_\theta(\mathbf{x})$ and $\mathbf{p}' = \hat{\mathbf{d}}^{\,\mathrm{comp}}(\mathbf{x})$, and define $\Delta(\mathbf{x}) = \ell_{\mathrm{KL}}(\mathbf{d}, \mathbf{p}) - \ell_{\mathrm{KL}}(\mathbf{d}, \mathbf{p}')$. By Lemma 1, since $\mathbf{p}, \mathbf{p}' \in \Delta_m^{(\varepsilon_0)}$,

$$|\Delta(\mathbf{x})| \leq \frac{1}{\varepsilon_0}\|\mathbf{p} - \mathbf{p}'\|_1. \tag{27}$$

Moreover, the triangle inequality gives $\|\mathbf{p} - \mathbf{p}'\|_1 \leq \|\mathbf{p} - \mathbf{d}\|_1 + \|\mathbf{p}' - \mathbf{d}\|_1$, so taking expectation yields the completion term $\frac{1}{\varepsilon_0}\mathcal{E}_c$. Lemma 2 together with Assumption 1 further bounds the prediction deviation induced by replacing the oracle representation $\mathbf{h}^\dagger(\mathbf{x}, \mathbf{d})$ with $\mathbf{h}^*(\mathbf{x})$, which leads to the term $\frac{L_s L_f}{\varepsilon_0}\mathcal{E}_r$. The remaining generalization gap is controlled by standard concentration and complexity arguments, resulting in the $C_1\gamma/\sqrt{N}$ and $C_2 r\log(2/\delta)/N$ terms. Details are in the **Appendix**.

Dataset	Examples	Features	Labels	Factor(γ)
Flickr_LDL	11,150	168	8	55.3
Natural Scene	2,000	294	9	10.3
SCUT-FBP	1,500	300	5	17.4
Movie	7,755	1,869	5	11.0
Emotion6	1,980	260	7	10.4
RAF-ML	4,908	200	6	30.3

Table 1: Details of the datasets.

Lemma 1. *For any* $\mathbf{d} \in \Delta_m$ *and* $\mathbf{p}, \mathbf{p}' \in \Delta_m^{(\varepsilon_0)}$,

$$\left|\ell_{\mathrm{KL}}(\mathbf{d}, \mathbf{p}) - \ell_{\mathrm{KL}}(\mathbf{d}, \mathbf{p}')\right| \leq \frac{1}{\varepsilon_0}\|\mathbf{p} - \mathbf{p}'\|_1. \tag{28}$$

Proof. $\ell_{\mathrm{KL}}(\mathbf{d}, \mathbf{p}) - \ell_{\mathrm{KL}}(\mathbf{d}, \mathbf{p}') = \sum_j d_j(\log p'_j - \log p_j)$, and $|\log p'_j - \log p_j| \leq \varepsilon_0^{-1}|p'_j - p_j|$ since $p_j, p'_j \geq \varepsilon_0$. Summing and using $\sum_j d_j = 1$ completes the proof.

Lemma 2. *Let* $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{R}^H$ *and* $p_\theta(\mathbf{h}) = \mathrm{softmax}(f(\mathbf{h}))$. *Then*

$$\|p_\theta(\mathbf{h}_1) - p_\theta(\mathbf{h}_2)\|_1 \leq L_s L_f \|\mathbf{h}_1 - \mathbf{h}_2\|_2. \tag{29}$$

Proof. By Lipschitz composition, $\|p_\theta(\mathbf{h}_1) - p_\theta(\mathbf{h}_2)\|_1 \leq L_s\|f(\mathbf{h}_1) - f(\mathbf{h}_2)\|_2 \leq L_s L_f\|\mathbf{h}_1 - \mathbf{h}_2\|_2$.

4 Experiments

4.1 Experimental Settings

Datasets and Incomplete Settings. To evaluate the effectiveness of our method, we conduct experiments on six benchmark datasets: *SCUT-FBP* [Xie *et al.*, 2015], *Flickr_LDL* [Yang *et al.*, 2017], *Movie* [Geng and Hou, 2015], *Emotion6* [Peng *et al.*, 2015], *Natural Scene* [Geng, 2016], and *RAF-ML* [Li and Deng, 2019]. These datasets exhibit diverse characteristics and varying degrees of imbalanced label distributions. Their detailed statistics and imbalance factors γ (defined in Eq. (2)) are summarized in Table 1. For the incomplete setting, we adopt 10-fold cross-validation, randomly masking 50% of the training description degrees uniformly across instances.

Compared Methods. We evaluate LDIBR against seven state-of-the-art methods covering diverse challenges in label distribution learning. Specifically, WInLDL [Li and Chen, 2024] and I^2LDL [Kou *et al.*, 2024] mitigate label incompleteness via weighted optimization and consistency constraints. DILDL [Gao *et al.*, 2025] counters distribution imbalance through decoupling, whereas GLDL [Jin *et al.*, 2024], LEIC [Lu *et al.*, 2023], and UALDL [Le *et al.*, 2023] emphasize structural correlations and reliability modeling. Furthermore, LCD [Xu *et al.*, 2025] unifies label completion and learning via implicit correlation decomposition. All methods follow their original configurations.

		LDIBR	DILDL	LCD	I^2LDL	GLDL	WInLDL	LEIC	UALDL
Che.↓	*Flickr_LDL*	**0.5703±0.125**	0.7311±0.008•	0.7548±0.010•	0.6834±0.014	0.7926±0.008•	0.7660±0.053•	0.7345±0.007•	0.7367±0.008•
	Natural Scene	**0.4857±0.024**	0.6932±0.012•	0.7189±0.021•	0.5916±0.016	0.7304±0.014•	0.6933±0.015•	0.6851±0.017•	0.7183±0.017•
	SCUT-FBP	**0.4467±0.028**	0.6360±0.025•	0.6368±0.030•	0.5478±0.040	0.7058±0.022•	0.6570±0.023•	0.6249±0.023•	0.6907±0.023•
	Movie	**0.3739±0.015**	0.4946±0.006•	0.5325±0.011•	0.5582±0.041•	0.5446±0.006•	0.4398±0.009	0.4921±0.007•	0.5295±0.006•
	Emotion6	**0.5519±0.012**	0.6943±0.011•	0.7179±0.016•	0.5681±0.019	0.7332±0.014•	0.6843±0.012•	0.6919±0.010•	0.7166±0.010•
	RAF-ML	**0.3747±0.197**	0.7723±0.003•	0.6414±0.015•	0.5926±0.014•	0.7412±0.007•	0.6709±0.005•	0.5371±0.042	0.6622±0.007•
Cla.↓	*Flickr_LDL*	**2.4810±0.073**	2.7009±0.002•	2.6079±0.028	2.6884±0.002•	2.7127±0.002•	2.6363±0.008•	2.7013±0.002•	2.7029±0.002•
	Natural Scene	**2.6584±0.019**	2.7401±0.015•	2.7544±0.013•	2.7087±0.015	2.7577±0.015•	2.7397±0.015•	2.7386±0.015•	2.7527±0.016•
	SCUT-FBP	**1.9780±0.011**	2.0366±0.009•	2.0890±0.010•	2.0076±0.008	2.0582±0.007•	2.0433±0.009•	2.0388±0.011•	2.0544±0.007•
	Movie	**1.7588±0.013**	1.8263±0.006•	1.8262±0.008•	1.8671±0.018•	1.8616±0.006•	1.8002±0.007	1.8222±0.006•	1.8396±0.006•
	Emotion6	**2.4105±0.008**	2.4456±0.007•	2.4849±0.010•	2.4146±0.008	2.4592±0.009•	2.4462±0.006•	2.4502±0.007•	2.4548±0.008•
	RAF-ML	**2.0636±0.052**	2.3041±0.003•	2.3217±0.003•	2.2678±0.002•	2.3037±0.003•	2.2869±0.002•	2.2650±0.006	2.2876±0.002•
Can.↓	*Flickr_LDL*	7.2794±0.124	7.5245±0.011•	7.6105±0.022•	7.4668±0.013•	7.5907±0.011•	**7.1456±0.053**	7.7013±0.002•	7.5323±0.013•
	Natural Scene	**7.6028±0.086**	8.0184±0.061•	8.0339±0.053•	7.8342±0.060	8.0830±0.062•	8.0076±0.066•	8.0084±0.063•	8.0634±0.068•
	SCUT-FBP	**4.1593±0.046**	4.4015±0.035•	4.5062±0.036•	4.2259±0.038	4.4843±0.027•	4.4286±0.032•	4.4003±0.037•	4.4687±0.028•
	Movie	**3.5824±0.039**	3.8138±0.016•	3.8055±0.025•	3.9456±0.060•	3.9187±0.018•	3.7180±0.020	3.8007±0.017•	3.8619±0.017•
	Emotion6	**6.1057±0.034**	6.3169±0.026•	6.4180±0.040•	6.1424±0.032	6.3699±0.035•	6.3137±0.022•	6.3291±0.026•	6.3531±0.028•
	RAF-ML	**5.1700±0.192**	5.5783±0.078•	5.5220±0.014•	5.3771±0.014•	5.5440±0.016•	5.4667±0.010•	5.3406±0.036	5.4629±0.009•

Table 2: Predictive performance of each method (mean ± std) on Chebyshev, Clark, and Canberra distances. ↑ / ↓ indicate that larger/smaller values are better. The best result is highlighted in **bold**, and the second-best result is underlined. A solid circle (•) indicates that LDIBR significantly outperforms the corresponding method.

Evaluation Metrics. We comprehensively evaluate the predictive performance using six standard metrics: Chebyshev, Clark, Canberra, and KL divergence as distance-based measures (↓), alongside Cosine and Intersection as similarity-based measures (↑). All results are reported as mean ± standard deviation over 10-fold cross-validation. Statistical significance is assessed via paired t-tests at the 0.05 level, where a solid circle (•) denotes that LDIBR significantly outperforms the compared baseline.

4.2 Results

As shown in Tables 2 and 3, we report quantitative comparisons under the 50% incomplete setting, where lower values indicate better performance for distance-based metrics and higher values indicate better performance for similarity-based metrics. Across distance-based metrics, LDIBR consistently outperforms competing methods on most datasets. For instance, on *Natural Scene*, LDIBR reduces the Chebyshev distance to 0.4857, corresponding to relative reductions of 17.9% over I^2LDL and 29.9% over DILDL, and further lowers the Clark distance to 2.6584. On *SCUT-FBP*, LDIBR achieves the best performance across all distance-based measures, reducing the KL divergence from 0.8815 for I^2LDL to 0.6255, a 29.0% relative reduction. Similar gains are observed on *Movie* and *Emotion6*, with improvements ranging from 15% to 30% over incomplete and imbalance-aware baselines.

Regarding similarity-based metrics, LDIBR attains the highest Cosine similarity on five datasets and improves the Intersection score on the highly imbalanced *RAF-ML* to 0.5212, yielding a 14.8% relative gain. Compared with correlation-based completion methods and imbalance-aware baselines, LDIBR consistently achieves higher similarity scores, suggesting that addressing missingness and imbalance in a unified manner is more effective. Notably, larger standard deviations are observed on *Flickr-LDL* and *RAF-ML*, with values of 0.125 and 0.197 for the Chebyshev distance, while all other datasets exhibit standard deviations below 0.03. This variability is likely due to severe label imbalance, which increases sensitivity to split-dependent missing patterns and leads to heterogeneous completion difficulty across folds. Nevertheless, by explicitly recovering unobserved description degrees, LDIBR mitigates such fluctuations and maintains strong mean performance even under pronounced distribution bias.

4.3 Ablation Study

We examine the contribution of each component in LDIBR by ablating the Distribution-Aware Label Completion module and the Bias-Corrective Representation Learning module. Tables 4 and 5 report results on *Natural Scene* and *SCUT-FBP*. The full model achieves the best performance across all metrics, and removing either module consistently degrades performance. Removing Distribution-Aware Label Completion causes the largest drop. On *Natural Scene*, KL increases to 1.421 and Cosine similarity decreases to 0.599, with Che., Cla., Can., and Int. also worsening. A similar trend is observed on *SCUT-FBP*, where KL increases to 1.034 and Cosine similarity decreases to 0.512. Removing Bias-Corrective Representation Learning leads to milder yet consistent degradation. On *Natural Scene*, KL increases to 1.012 and Cosine similarity decreases to 0.721, and on *SCUT-FBP*, KL increases to 0.781 with Cosine similarity decreasing to 0.603. When both modules are removed, performance deteriorates most severely on both datasets, confirming that the two components are complementary.

4.4 Further Analyses

We further investigate model robustness under increasing label incompleteness by gradually raising the missing ratio

		LDIBR	DILDL	LCD	I^2LDL	GLDL	WInLDL	LEIC	UALDL
KL↓	*Flickr_LDL*	1.7792±0.989	1.6691±0.033	1.6449±0.113	**1.4310±0.029**	1.8407±0.017•	1.6165±0.112	1.6821±0.031	1.7050±0.034
	Natural Scene	**0.7766±0.049**	1.6722±0.041•	1.5589±0.012•	1.3382±0.072	1.9203±0.041•	1.7248±0.045•	1.7477±0.055•	1.8634±0.053•
	SCUT-FBP	**0.6255±0.060**	1.2462±0.073•	1.5126±0.036•	0.8815±0.088	1.5162±0.061•	1.3176±0.066•	1.2692±0.098•	1.4609±0.056•
	Movie	**0.6331±0.031**	0.8821±0.016•	0.8767±0.016•	0.9529±0.134•	1.1345±0.022•	0.7756±0.022	0.8957±0.018•	0.9815±0.020•
	Emotion6	**1.1434±0.038**	1.5650±0.032•	1.6526±0.050•	1.1758±0.053	1.7500±0.046•	1.5607±0.032•	1.6105±0.039•	1.6834±0.028•
	RAF-ML	1.1891±0.061	1.5845±0.009•	1.6873±0.013•	1.1168±0.025	1.6088±0.034•	1.3815±0.028•	**1.0740±0.084**	1.3943±0.036•
Cos.↑	*Flickr_LDL*	**0.6614±0.132**	0.4952±0.013•	0.2777±0.010•	0.5604±0.010	0.4394±0.008•	0.5412±0.035•	0.4905±0.012•	0.4836±0.013•
	Natural Scene	**0.8457±0.022**	0.4543±0.018•	0.3136±0.021•	0.5888±0.027	0.3770±0.010•	0.4381±0.017•	0.4434±0.021•	0.3925±0.015•
	SCUT-FBP	0.6752±0.028	0.5881±0.033•	0.4355±0.033•	**0.7201±0.035**	0.4839±0.027•	0.5562±0.026•	0.5793±0.040•	0.5048±0.023•
	Movie	**0.7741±0.015**	0.6864±0.007•	0.6555±0.013•	0.6585±0.061•	0.6059±0.008•	0.7223±0.009	0.6825±0.007•	0.6484±0.008•
	Emotion6	**0.6393±0.015**	0.4824±0.013•	0.4374±0.025•	0.5843±0.022	0.4179±0.014•	0.4831±0.013•	0.4673±0.014•	0.4417±0.011•
	RAF-ML	**0.7127±0.167**	0.4815±0.004•	0.4299±0.016•	0.6462±0.011•	0.4821±0.012•	0.5620±0.013•	0.6650±0.030	0.5575±0.015•
Int.↑	*Flickr_LDL*	**0.4205±0.127**	0.2573±0.007•	0.2389±0.010•	0.3074±0.014•	0.1954±0.007•	0.3253±0.053	0.2540±0.007•	0.2516±0.008•
	Natural Scene	**0.4732±0.022**	0.2385±0.007•	0.2231±0.017•	0.3655±0.016	0.1954±0.007•	0.2388±0.008•	0.2458±0.012•	0.2130±0.012•
	SCUT-FBP	**0.6459±0.028**	0.3486±0.025•	0.3571±0.028•	0.5460±0.040	0.2784±0.023•	0.3282±0.023•	0.3601±0.023•	0.2938±0.024•
	Movie	**0.6119±0.015**	0.4843±0.006•	0.4509±0.012•	0.4193±0.041•	0.4294±0.007•	0.5399±0.009	0.4872±0.007•	0.4480±0.006•
	Emotion6	**0.4066±0.011**	0.2578±0.008•	0.2522±0.018•	0.3284±0.019	0.2213±0.009•	0.2694±0.010•	0.2615±0.007•	0.2368±0.007•
	RAF-ML	**0.5212±0.199**	0.2189±0.003•	0.3532±0.015•	0.4021±0.014•	0.2486±0.008•	0.3198±0.006•	0.4542±0.042	0.3285±0.008•

Table 3: Predictive performance of each method (mean ± std) on KL divergence, Cosine similarity, and Intersection similarity. ↑ / ↓ indicate that larger/smaller values are better. The best result is highlighted in **bold**, and the second-best result is underlined. A solid circle (•) indicates that LDIBR significantly outperforms the corresponding method.

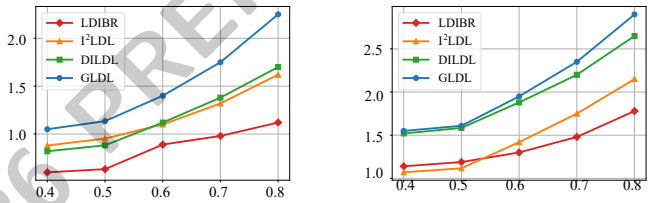
Method	Che.↓	Cla.↓	Can.↓	KL↓	Cos.↑	Int.↑
LDIBR (full)	**0.486**	**2.658**	**7.603**	**0.777**	**0.846**	**0.473**
w/o DALC	0.622	2.709	7.834	1.421	0.599	0.366
w/o BCRL	0.548	2.684	7.721	1.012	0.721	0.421
w/o Both	0.681	2.731	7.912	1.587	0.563	0.331

Table 4: Ablation results on *Natural Scene*.

Method	Che.↓	Cla.↓	Can.↓	KL↓	Cos.↑	Int.↑
LDIBR (full)	**0.447**	**1.978**	**4.159**	**0.626**	**0.675**	**0.646**
w/o DALC	0.583	2.024	4.323	1.034	0.512	0.546
w/o BCRL	0.511	2.006	4.245	0.781	0.603	0.589
w/o Both	0.641	2.047	4.384	1.211	0.468	0.498

Table 5: Ablation results on *SCUT-FBP*.

from 0.4 to 0.8 and evaluating performance using KL divergence. Fig. 3 shows that KL divergence consistently increases for all methods as the label missing ratio grows, indicating that higher levels of label missingness make accurate label distribution recovery more challenging. On the *Movie* dataset, LDIBR consistently achieves lower KL divergence across all missing ratios and exhibits a smooth degradation trend as label missingness increases, demonstrating stable behavior under progressively reduced supervision. On *RAF-ML*, although LDIBR does not achieve the lowest KL divergence at lower missing ratios, its performance degrades more slowly as the missing ratio rises and it exhibits the smallest overall increase in KL divergence under severe label missingness. Compared with baseline methods, LDIBR maintains more stable relative performance as the label missing ratio increases, particularly in high missing-ratio regimes. Experimental results demonstrate that LDIBR maintains robust performance even under increasing levels of label incompleteness, particularly in challenging scenarios where concurrent



(a) *Movie* (b) *RAF-ML*

Figure 3: KL divergence under different label missing ratios across datasets. As the missing ratio increases, all methods deteriorate, while LDIBR shows the smallest drop.

label missingness and distribution skew jointly compromise model training.

5 Conclusion

In this paper, we propose LDIBR, a unified framework for more complex label distribution learning with missingness under imbalanced label distributions. We highlight the limitations of conventional methods that treat missing description degrees as low values, thereby introducing systematic bias in label distribution recovery and representation learning. LDIBR explicitly models missing entries as unobserved description degrees, applies distribution-aware label completion, and incorporates bias-corrective representation learning to mitigate imbalance-induced bias. Additionally, we provide a theoretical generalization analysis that characterizes how completion error, representation error, and description degree imbalance jointly affect the risk. Extensive experiments on benchmark datasets demonstrate that LDIBR achieves more accurate and stable predictions, particularly under high missing ratios and severe label imbalance. Future work will focus on enhancing model robustness and extending LDIBR to hierarchical and continuous label scenarios.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under No. 62406155, the project No. ZR2024QF115 supported by Shandong Provincial Natural Science Foundation, the Innovation Capability Enhancement Project for Technology-based Small and Medium-sized Enterprises of Shandong Province under Grant No. 2025TSGCCZZB0077, the National Natural Science Foundation of China under No. 62202235 and No. 52303025, the project No. ZR2025MS750 supported by Shandong Provincial Natural Science Foundation, the Jinan Science and Technology Plan (Post-subsidy) Project under No.202430039.

References

[Chang *et al.*, 2025] Gengshuo Chang, Wei Zhang, and Lehan Zhang. Matrix completion with incomplete side information via orthogonal complement projection. In *Proceedings of the International Conference on Machine Learning*, 2025.

[Cheng *et al.*, 2024] Zebang Cheng, Zhi-Qi Cheng, Jun-Yan He, Kai Wang, Yuxiang Lin, Zheng Lian, Xiaojiang Peng, and Alexander Hauptmann. Emotion-llama: Multimodal emotion recognition and reasoning with instruction tuning. In *Proceedings of the Advances in Neural Information Processing Systems*, 2024.

[Gao *et al.*, 2025] Yongbiao Gao, Xiangcheng Sun, Miaogen Ling, Cao Tan, Yi Zhai, and Guohua Lv. Decoupled imbalanced label distribution learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2025.

[Geng and Hou, 2015] Xin Geng and Peng Hou. Pre-release prediction of crowd opinion on movies by label distribution learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2015.

[Geng, 2016] Xin Geng. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1734–1748, 2016.

[He *et al.*, 2021] Huan He, Shifan Zhao, Yuanzhe Xi, and Joyce Ho. Age: Enhancing the convergence on gans using alternating extra-gradient with gradient extrapolation. In *Proceedings of the NeurIPS Workshop on Deep Generative Models and Downstream Applications*, 2021.

[Hou *et al.*, 2025] Senyu Hou, Gaoxia Jiang, Jia Zhang, Shangrong Yang, Husheng Guo, Yaqing Guo, and Wenjian Wang. Directional label diffusion model for learning from noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

[Jin *et al.*, 2024] Yufei Jin, Richard Gao, Yi He, and Xingquan Zhu. Gldl: Graph label distribution learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[Kou *et al.*, 2024] Zhiqiang Kou, Haoyuan Xuan, Jing Wang, Yuheng Jia, and Xin Geng. Towards better performance in incomplete ldl: Addressing data imbalance. *arXiv preprint arXiv:2410.13579*, 2024.

[Kou *et al.*, 2025] Zhiqiang Kou, Si Qin, Hailin Wang, Jing Wang, Mingkun Xie, Shuo Chen, Yuheng Jia, Tongliang Liu, Masashi Sugiyama, and Xin Geng. Label distribution learning with biased annotations assisted by multi-label learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2025.

[Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[Le *et al.*, 2023] Nhan Le, Khoa Nguyen, Quoc Tran, et al. Uncertainty-aware label distribution learning for facial expression recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023.

[Lee *et al.*, 2024] Taehan Lee, WooSang Shin, Jong-Hyeon Lee, Sangmoon Lee, Han-Gyeol Yeom, and Jong Pil Yun. Resolving the non-uniformity in the feature space of age estimation: A deep learning model based on feature clusters of panoramic images. *Computerized Medical Imaging and Graphics*, 112:102329, 2024.

[Li and Chen, 2024] Xiang Li and Songcan Chen. No regularization is needed: efficient and effective incomplete label distribution learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2024.

[Li and Deng, 2019] Shan Li and Weihong Deng. Blended emotion in-the-wild: Multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning. *International Journal of Computer Vision*, 127(6):884–906, 2019.

[Li *et al.*, 2024] Weiwei Li, Wei Qian, Lei Chen, and Xiuyi Jia. Sample diversity selection strategy based on label distribution morphology for active label distribution learning. *Pattern Recognition*, 150:110322, 2024.

[Li *et al.*, 2025a] Jun Li, Xiaofeng Zhu, Haibo Wang, et al. Multi-label ranking loss minimization for matrix completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[Li *et al.*, 2025b] Kaixin Li, Renrui Zhang, Xiangfei Hu, Bo Zhang, Han Xiao, Haocheng Feng, Errui Ding, Jingdong Wang, Ziyu Ma, and Xiaodan Liang. Motionpro: A precise motion controller for image-to-video generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

[Liu *et al.*, 2025] Shu Liu, Enquan Huang, Ziyu Zhou, Yan Xu, Xiaoyan Kui, Tao Lei, and Hongying Meng. Lightweight facial attractiveness prediction using dual label distribution. *IEEE Transactions on Cognitive and Developmental Systems*, 17, 2025.

[Lu *et al.*, 2023] Yunan Lu, Weiwei Li, and Xiuyi Jia. Label enhancement via joint implicit representation clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2023.

[Ma *et al.*, 2024] Xu Ma, Xiyang Dai, Jianwei Yang, Bin Xiao, Yinpeng Chen, Yun Fu, and Lu Yuan. Efficient mod-

ulation for vision networks. In *Proceedings of the International Conference on Learning Representations*, 2024.

[Nguyen *et al.*, 2024] Huy Nguyen, Pedram Akbarian, TrungTin Nguyen, and Nhat Ho. A general theory for softmax gating multinomial logistic mixture of experts. In *Proceedings of the International Conference on Machine Learning*, 2024.

[Peng *et al.*, 2015] Kuan-Chuan Peng, Tsuhan Chen, Amir Sadovnik, and Andrew C. Gallagher. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[Tsoumakas and Katakis, 2008] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications*, pages 64–74, 2008.

[Wang and Geng, 2023] Jing Wang and Xin Geng. Label distribution learning by exploiting label distribution manifold. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2):839–852, 2023.

[Wang and Li, 2025] Xianzhong Wang and Jian Li. Sbpm model for analyzing students' learning behavior based on fine grained emotion analysis and emotion assessment. *Informatica*, 49(7):187–200, 2025.

[Wang *et al.*, 2021] Ke Wang, Ning Xu, Miaogen Ling, and Xin Geng. Fast label enhancement for label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(2):1502–1514, 2021.

[Wang *et al.*, 2024] Xin Wang, Hong Chen, Si'ao Tang, Zihao Wu, and Wenwu Zhu. Disentangled representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9677–9696, 2024.

[Wu *et al.*, 2025] Sheng Wu, Dongxiao He, Xiaobao Wang, Longbiao Wang, and Jianwu Dang. Enriching multimodal sentiment analysis through textual emotional descriptions of visual-audio content. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[Xie *et al.*, 2015] Duorui Xie, Lingyu Liang, Lianwen Jin, Jie Xu, and Mengru Li. Scut-fbp: A benchmark dataset for facial beauty perception. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2015.

[Xu *et al.*, 2025] Suping Xu, Lin Shang, Furao Shen, Xibei Yang, and Witold Pedrycz. Incomplete label distribution learning via label correlation decomposition. *Information Fusion*, 113:102600, 2025.

[Yang *et al.*, 2017] Jufeng Yang, Ming Sun, and Xiaoxiao Sun. Learning visual sentiment distributions via augmented conditional probability neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017.

[Yu *et al.*, 2023] Zhen Yu, Ruiye Chen, Peng Gui, Lie Ju, Xianwen Shang, Zhuoting Zhu, Mingguang He, and Zongyuan Ge. Retinal age estimation with temporal fundus images enhanced progressive label distribution learning. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2023.

[Zhang and Zhang, 2024] Yi-Fan Zhang and Min-Ling Zhang. Generalization analysis for label-specific representation learning. *Advances in Neural Information Processing Systems*, 37:104904–104933, 2024.

[Zhang and Zhou, 2013] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2013.

[Zhang *et al.*, 2025] Bo Zhang, Meng Chen, Jun Song, et al. Normalize then propagate: Efficient homophilous regularization for few-shot semi-supervised node classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[Zhao *et al.*, 2023] Xingyu Zhao, Yuexuan An, Ning Xu, Jing Wang, and Xin Geng. Imbalanced label distribution learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[Zhao *et al.*, 2025] Qingqing Zhao, Yuhan Xia, Yunfei Long, Ge Xu, and Jia Wang. Leveraging sensory knowledge into text-to-text transfer transformer for enhanced emotion analysis. *Information Processing & Management*, 62(1):103876, 2025.