IHC-LDL: IMBALANCED HIERARCHICAL CLASSIFICATION WITH LABEL DISTRIBUTION LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Utilizing multi-label learning(MLL) to solve hierarchical classification problems is a challenging task. Recently a novel paradigm named label distribution learning (LDL) handing label polysemy well can mine the relative importance of different labels in label hierarchy. Although numerous hierarchical MLL and LDL algorithms have been proposed respectively, the combined application of the two lacks in-depth exploration. The approaches constructing the hierarchical label distribution often lack addressing feature space correlation and sample imbalance. To address this problem, we propose a LDL-based imbalanced hierarchical classification approach whose main idea is to construct and learn implicit label distribution with hierarchical information, and decompose it into three sub-tasks: label enhancement(LE), imbalanced resampling and adaptive multi-label learning. In this approach, we design a joint strategy of feature-based label distribution and hierarchical label distribution to enhance label and an adaptive weighting for multi-label learning to tackle the task with unobvious label polysemy. Experimental results on widely used open datasets show the effectiveness of our proposal.

1 INTRODUCTION

Hierarchical classification (HC), as an important branch of machine learning, has widespread application values in practical tasks, such as text classification, protein function prediction, biological sequence classification, animal species identification and object category classification. Unlike flat classification in which each category is treated independently, hierarchical classification requires the model to consider parent-child dependencies between classes. Traditional hierarchical classification methods use tree structures or Directed Acyclic Graph (DAG) to represent the inter class relationships. This complex relationship significantly increases the complexity of the task. Two key challenges in hierarchical classification field are leveraging the hierarchical dependencies between labels to improve performance while maintaining scalability across deeper or imbalanced hierarchical structures. Especially in the case of fine-grained label, because of extreme distribution imbalance of instances, it often leads to the poor effect for tail classes.

Hierarchical Multi-label Classification (HMC) is a challenging task which aims to predict multiple labels from a hierarchical structure to solve HC problem. Recently a novel paradigm named label distribution learning can effectively represent the association strength between samples and multiple labels by constructing a probability distribution in the label space, which has been validated to enhance the effect of MLL and adapt to few-shot scenarios. However, existing LDL methods face two major challenges in hierarchical classification scenarios: 1) insufficient collaborative modeling of hierarchical structure information for label distribution; 2) poor model robustness under extreme data imbalance. This paper proposes a joint learning framework which integrates feature space distribution and hierarchical label distribution, with imbalanced resampling and adaptive hierarchical relationship learning, to improve fine-grained recognition accuracy. The rest arrangement of this paper is as follows: Section 2 reviews related technologies; Section 3 elaborates on the method framework in detail; Section 4 presents experimental verification results; Section 5 summarizes the full paper.

The contributions of this paper are: (i) designing a label distribution joint construction architecture for hierarchical classification; (ii) combining imbalanced resampling methods to improve the sample imbalance phenomenon of hierarchical labels; (iii) proposing an adaptive weighting for learning

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

logical labels and label distribution as well as the hierarchical consistency constraint refinement, where the novel constructed penalty loss can effectively improve the HC learning process and handle tasks with different label polysemy; (iv) through our imbalanced hierarchical classification learning approach, the performance is improved by utilizing the given label hierarchy while considering the sample imbalance of fine-grained labels.

2 Related Works

2.1 Hierarchical Classification Methods

The most extensive research in hierarchical classification is Hierarchical Text Classification (HTC). Silla and Freitas Silla & Freitas (2011) provide the earliest comprehensive overview of hierarchical classification methods across different application domains, while pointing out key challenges such as deep hierarchical structures and data sparsity for lower-level categories which remain significant obstacles to consistent and accurate classification across multiple hierarchical levels. Zangari et al. Zangari et al. (2024) provide a recent overview of current research on hierarchical text classification through summarizing newly proposed methods and their contributions. They conduct comparative experiments on five domain-specific public datasets to evaluate the benchmark methods such as HBGL, GACaps-HTC, MATCH, HiMatch, and HiAGM against some non-hierarchical baseline methods, and conclude that HBGL outperformed the other methods in terms of F1 score and accuracy on most datasets. Fatos Torba et al. Fatos et al. (2025) propose the HLPD architecture which is built on a hybrid encoder-decoder framework. Based on the potential semantic cues between text and labels, they construct both encoders based on the T5 model, then apply a hierarchical self-attention mechanism to label embedding and performed cross-attention between labels and text, meanwhile adopt the Zero-bounded Log-sum-exp & Pairwise Ranking (ZLPR) loss function designed for multi-label classification. Ultimately, the model achieves state-of-the-art results on four datasets: RCV1-V2, AAPD, BGC, and WOS, and outperforms the HBGL model on the RCV1-V2 and WOS datasets. Yunyi Zhang et al. Zhang et al. (2025) apply Large Language Models (LLM) to hierarchical text classification. Through using the framework of leveraging taxonomy enhancement and LLM-based enhancement for category-indicative terms, refining core categories via document-category matching approaches, and utilizing label paths to guide LLMs in generating augmented text data, the trained hierarchical text classifier finally outperforms the state-of-the-art baseline method.

Silla Silla & Freitas (2011), Pereira et al. Pereira et al. (2021a;b), Bang Bang et al. (2023), and other researchers categorize HTC methods into flat methods, local methods, and global methods, with specific analysis as follows:

Flat methods discard information from the hierarchical structure and transform the task into a simple classification problem. Fürnkranz et al. J et al. (2008) only assign leaf labels to text documents, ignoring non-leaf labels in the hierarchy. Johnson et al. R & T. (2015) simply treat HTC as multi-label text classification. These methods provide a simple solution for HTC, but the performance of flat methods is not ideal because they do not utilize the information of label relationships, which is crucial for the task.

Local methods use multiple classifiers, where each classifier is responsible for part of the hierarchy. Local methods can be subdivided into: Local Classifier per Node (LCN), Local Classifier per Parent Node (LCPN), and Local Classifier per Level (LCL). Nodes employ classifiers like Support Vector Machines (SVM), Convolutional Neural Networks combined with attention mechanisms to achieve better classification accuracy. Results from local classifiers at different levels are further aggregated with global classifier results to further improve performance, but they are susceptible to error propagation from top to bottom.

Global methods J et al. (2019) use a single multi-label classifier, leveraging the label hierarchy as a learning object to train and improve the performance of flat classification models. Recursive regularization Zhao et al. (2019) encourages the parameters corresponding to a label to be similar to its parent which has been shown to improve the HTC performance of SVM, logistic regression, and neural network classifiers. Yu et al. Y et al. (2021) propose a loss term that penalizes contradictions where a document is assigned a label but not its parent label, and prove that training neural networks to minimize this loss can improve performance. Giunchiglia et al.[14] propose the C-HMCNN(h)

multi-label classification method, which adds a consistency constraint layer after the multi-label output results, using label hierarchy information to produce predictions consistent with hierarchical constraints. Zhou et al. J et al. (2020), Bang et al. Bang et al. (2023), and other researchers integrate label hierarchy information with text representations through Graph Neural Networks (GNNs), Graph Attention Networks (GATs) and etc. to enable text representations contain key information of HTC.

2.2 Imbalanced Hierarchical Classification Methods

Recent research has begun to focus on imbalanced hierarchical classification. For instance, SMOTE alleviates the issue of insufficient minority class samples by generating synthetic samples through oversampling but may introduce noise. Cost-Sensitive learning and Hier-Cost methods address this issue by adjusting the classification loss weights of majority and minority classes through modified loss functions, however they struggle to handle category correlations within the hierarchical structure. Pereira et al. Pereira et al. (2021a) focus on examining three local classifier methods combined with random resampling methods to address hierarchical imbalance problems, the comparison result reveals that the integration of LCN with LI and SMOTE-TOMEK achieve the best statistically benchmark on multiple datasets. Bang et al. Bang et al. (2023) propose the Graph Attention Capsule Network (GACaps-HTC) method, which uses explicit hierarchical structure and implicit relationships of extracted labels, employing a loss term composed of focal loss and a hierarchical contradiction penalty term for optimization, alleviating the inherent label imbalance in hierarchical text classification tasks. However, these methods do not address the inherent label imbalance problem from the label hierarchy structure, resulting in relatively poor hierarchical classification performance.

2.3 Label Distribution Learning Methods

LDL Geng (2016) , which constructs a probability distribution in the label space, takes into account the relative importance of different labels and can effectively handle tasks such as multi-label and ambiguous labels. In hierarchical classification applications, Xu et al. C & X (2019) propose H-LDL method, which attempts to encode hierarchical path information and combine it with logical labels to obtain label distributions, addressing the hierarchical classification through label distribution learning, but without fully considering the impact of sample imbalance. Lin et al. Lin et al. (2023) propose a hierarchical feature selection method based on label distribution learning, HFSLDL, which employs the weighted average of label distributions obtained from Pearson correlation coefficients, label distributions obtained from the number of shared common nodes between different categories in the hierarchy, and logical labels, achieving feature selection through label distribution learning. However, the label distribution obtained from sample correlation does not fully reflect the distribution characteristics of the sample space. Xu et al. Xu et al. (2018; 2019; 2020; 2021; 2023) propose LEVI-MLP and LIVI-GCN approaches, which achieve label enhancement via latent space learning before conducting label distribution learning, and have thus yielded state-of-the-art classification performance. Nevertheless, this framework fails to specifically account for hierarchical classification. Zheng et al. Zheng et al. (2023) propose the LIBLE method, which utilizes the information bottleneck theory to decompose label-related information into label assignment information and label gap information, thereby transforming the LE problem into simultaneously learning latent representations and modeling the label gap, finally achieve performance superior to the baseline LE methods like LESC and LEVI. Jia et al. Jia et al. (2021) enhance the recovery effect of label distribution by leveraging association relationships between graph nodes. However, these methods Lu et al. (2023a;b; 2025) have not yet been specifically optimized for hierarchical classification problems.

3 Methodology

We propose an imbalanced hierarchical classification method based on label distribution learning for multi-granularity hierarchical classification application scenarios. Firstly, a label enhancement module combining feature-label distribution, hierarchical-label distribution and logic label is constructed to generate label distributions. Secondly, a mainstream hierarchical classification architecture is adopted to support multi-label classification across multiple levels, and imbalanced sample

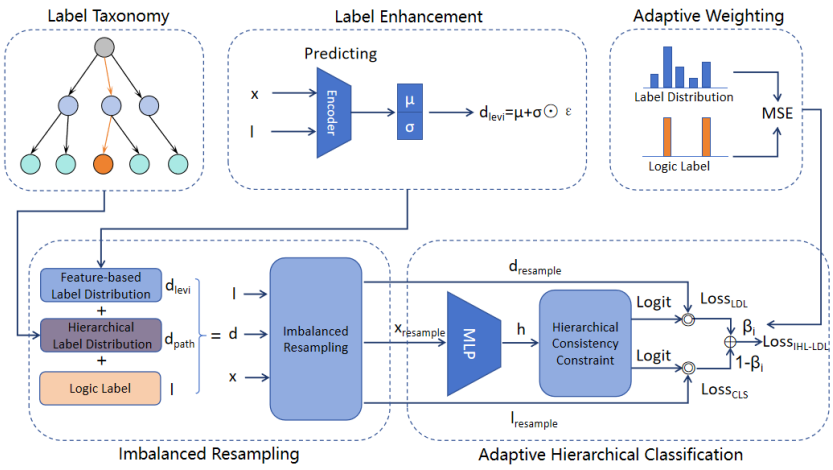


Figure 1: Architecture of the proposed method. Multi-Layer Perceptron (MLP) is used as our classification backbone network. Mean Squared Error (MSE) is adopted to calculate the disagreement between the logical labels and the label distribution.

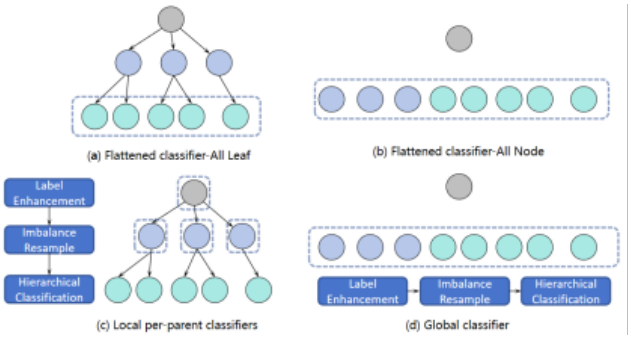


Figure 2: Architecture of the hierarchy method.

augmentation and hierarchical consistency learning techniques are used to jointly learn feature space and label space under hierarchical data imbalance. Finally, a weighting mechanism between logic label and label distribution enhances the adaptability to various scenarios. The implementation block diagram of the proposed method is shown in Figure 1.

3.1 Problem Definition and Model Structure

Given a hierarchical classification task, let the category set be C, its hierarchical structure be a tree graph T, where each node represents a category, and the edges represent parent-child relationships. For a sample x_i, its true logical label is $l_i \in [0,1]^{|C|}$, the constructed label distribution is denoted as $d_i = (d_i^1, d_i^2, \cdots, d_i^{|C|})$, satisfying $\sum_{j=1}^{|C|} d_i^j = 1$. The traditional classification objective is to approximate l_i using a learning model $f(x)$. In contrast, the classification objective based on label distribution learning aims to approximate d_i through a learning model $f(x)$, while using the category correlations inherent in the hierarchical structure.

To evaluate which model architecture best leverages label distribution learning to capture sample-label correlations in hierarchical classification tasks, this paper employs the following model architectures to conduct a comparative analysis.

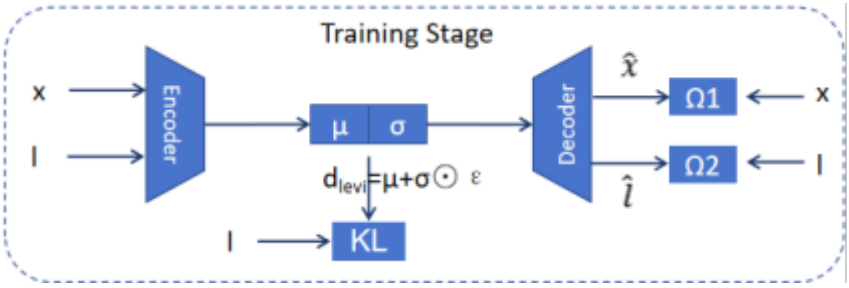


Figure 3: LEVI-based Label Enhancement Method.

3.2 Label Enhancement

It has been empirically proven that the label distribution can better represent the correlation between sample features than logical labels. To leverage this advantage, this paper constructs a joint label enhancement module that integrates the feature-label distribution, the hierarchical-label distribution, and the logic label vector. The feature-label distribution leverages correlations within the sample feature space, while the hierarchical-label distribution focuses on correlations within the label space. To ensure the output of authentic category assignments, logical labels are also combined.

After comparison, the label distribution of the sample feature space is enhanced using the LEVI-MLP method proposed by Xu et al. Xu et al. (2020), as shown in Figure 3:

The loss function is formulated as follows:

$$\begin{aligned} KL &= \frac{1}{2}\left(\mu^T\mu + \mathrm{tr}(\sigma^2 I) - c - \log|\sigma^2 I|\right) \\ \Omega_1 &= \frac{1}{2}\|x - \hat{x}\|_2^2 \\ \Omega_2 &= \frac{1}{2}\|l - \hat{l}\|_2^2 \\ \mathcal{L}_{\mathrm{LEVI}} &= \lambda_1 KL + \lambda_2\Omega_1 + \lambda_3\Omega_2 \end{aligned} \tag{1}$$

where, λ_1, λ_2, λ_3, are hyper-parameters indicating the weights of each loss term.

The label distribution of the sample feature space d_{levi} is derived as follows:

$$d_{\mathrm{levi}} = \mu + \sigma \odot \varepsilon \tag{2}$$

where, μ, σ represent the weights of the pre-trained MLP encoding network. The hierarchical label distribution is calculated by considering the number of common nodes in the paths between labels, as defined below:

$$\begin{aligned} b_j &= \left|\mathrm{path}(l) \cap \mathrm{path}(l_j)\right| \\ d_{\mathrm{path}} &= \frac{b_j}{\sum_{j=1}^{C} b_j} \end{aligned} \tag{3}$$

The final label distribution is obtained via the following calculation method:

$$d = (1 - \alpha_1 - \alpha_2)l + \alpha_1 d_{\mathrm{levi}} + \alpha_2 d_{\mathrm{path}} \tag{4}$$

where, α_1, α_2 are hyper-parameters indicating the weights of the label distribution of the sample feature space and hierarchical label space.

3.3 Imbalanced Resampling

To address the inherent label imbalance problem in hierarchical classification, especially the sample sparsity issue of fine-grained labels near leaf nodes, Bang Bang et al. (2023), Zhao et al. Zhao et al. (2023; 2024), and others researchers propose using focal loss and its improved versions such as

class-balanced focal loss and distribution-balanced focal loss. Pereira et al. Pereira et al. (2021a) focus on exploring random resampling methods to address hierarchical imbalance problems, and their comparative experiments conclude that the combination of LCN with LI and SMOTE-TOMEK yielded the best benchmark performance. Inspired by this finding, we also adopt the resampling techniques such as RandomOverSampler G et al. (2004), SMOTE N et al. (2002), and TOMEK I (1976), as well as their combinations, in imbalanced hierarchical classification tasks to generate new samples for leaf-node categories with few labeled instances.

3.4 Adaptive Hierarchical Classification

This subsection discusses the loss terms used to optimize IHL-LDL. Besides being implicitly embedded in feature and label representations, many scholars Dasgupta & Namboodiri (2016) have also focused on constructing loss functions to utilize hierarchical structure. Yu et al. Y et al. (2021) impose constraints on parent-child output probabilities, penalizing cases where parent label probability is greater than the maximum child label probability. Bang et al. Bang et al. (2023) improve upon this to prevent parent and child labels from outputting the same probability, but both weight the contradiction penalty loss and classification loss. Eleonora et al. Giunchiglia & Lukasiewicz (2020) add a Maximum Constraint Module(MCM) network layer to the multi-label output, imposing hierarchical constraints on each label. By taking the maximum probability among the current label and all its descendant labels, hierarchical consistency output among multiple labels is achieved, followed by binary cross-entropy classification loss calculation, resulting in better performance. Inspired by the above, we propose adding a hierarchical consistency constraint module(HCCM) after our MLP classification network as follows:

$$HCCM_{\mathrm{A}} = \max_{\mathrm{B}\in \mathrm{D_A(k)}} (h_{\mathrm{B}}) + \delta \tag{5}$$

where, h_B is the probability output for label before the network, $D_A(k)$ is the set of current and nearest k-level descendant labels for label A (not all descendants), and $\delta \in R^+$ represents the maximum allowed difference in classification probabilities between two hierarchical levels.

Label distribution learning is suitable for datasets with label ambiguity and annotator disagreement, whereas logical labeling is more appropriate for datasets with consistent labels and clear semantics. To enhance adaptability to different datasets, this paper proposes an approach for constructing a loss function that using adaptive weights between logical label classification loss and label distribution learning loss, which effectively addresses the learning effects in various application scenarios. The specific method is as follows:

Label Distribution Learning Loss ($\mathcal{L}_{LDL}$): Taking into account hierarchical label constraint relationships, the label distribution learning loss is constructed using the Kullback-Leibler(KL) divergence loss function as follows:

$$\begin{aligned} \mathcal{L}_{\mathrm{LDL}} &= \mathrm{KLDiv}\left(HCCM', d'\right) \\ &= \sum HCCM' \log\left(HCCM'/d'\right) \\ HCCM' &= HCCM_d / \sum\left(HCCM_d\right) \\ d' &= d / \sum(d) \end{aligned} \tag{6}$$

where, $HCCM', d'$ are the normalized results of the probability output $HCCM_d$ and the multi-label distribution d, respectively. Based on the output of the HCCM, is constructed using an improved binary cross-entropy loss function, which is defined as follows:

$$\begin{aligned} HCCM_d = \sum_{\mathrm{A}\in \mathrm{C}} & -d_{\mathrm{A}} \ln\left(\max_{\mathrm{B}\in \mathrm{D_A(k)}} (d_{\mathrm{B}} h_{\mathrm{B}}) + \delta\right) \\ & - (1 - d_{\mathrm{A}}) \ln\left(1 - HCCM_{\mathrm{A}}\right) \end{aligned} \tag{7}$$

where, $D_A(k)$ retains the same meaning as introduced earlier.

Table 1: Hierarchical Classification Datasets.

Name	Domain	Train	Test	Feature	Labels	Levels
ImCLEF	Image	10000	1006	80	97	1,8,25,63
ImCLEF07A	Image	8368	939	80	88	1,7,17,63
ImCLEF07D	Image	10000	1006	80	47	1,4,16,26
AWAphog	Animal Attribute	6405	3202	252	17	1,2,4,10
F194	Protein	7105	1420	473	202	1,7,194
DD	Protein	3020	605	473	32	1,4,27
VOC	Image	7178	5105	1000	30	1,9,20

Classification Loss ($\mathcal{L}_{CLS}$): We employ the binary cross-entropy loss function to compute the classification loss between the HCCM output and the logical labels, formulated as follows:

$$\begin{aligned} \mathcal{L}_{\mathrm{CLS}} &= \mathrm{BCELoss}\left(HCCM_l, l\right) \\ HCCM_l &= \sum_{\mathrm{A}\in\mathrm{C}} -l_{\mathrm{A}} \ln\left(\max_{\mathrm{B}\in\mathrm{D_A(k)}} (l_{\mathrm{B}} h_{\mathrm{B}}) + \delta\right) \\ &\quad - (1 - l_{\mathrm{A}}) \ln\left(1 - HCCM_{\mathrm{A}}\right) \end{aligned} \tag{8}$$

where, $D_A(k)$ has the same meaning as the term mentioned before.

To tackle the task with unobvious label polysemy, the total loss of our imbalanced hierarchical classification method is derived by an adaptive weighted sum of two loss terms:

$$\mathcal{L}_{IHL-LDL} = (1 - \beta_i)\mathcal{L}_{CLS} + \beta_i \mathcal{L}_{LDL} \tag{9}$$

where, β_i is an adaptive parameter representing the weight of the LDL loss term for instance i, calculated based on the Mean Squared Error (MSE) between the logical labels and the label distribution, which is defined as follows:

$$\beta_i = \sqrt{\frac{1}{|C|} \sum_{A\in C} (l_A - d_A)^2} \tag{10}$$

4 Experiments

4.1 Dataset Description

The effectiveness of the proposed approach is validated on several datasets used to evaluate hierarchical classification methods. These datasets come from several different real-world domains. The overall situation is shown in Table 1.

4.2 Baseline Methods

The IHL-LDL method is compared with the following baseline methods:

Flat Methods:

Flat_Leaf: A Flatten MLP multi-class classification method for leaf nodes.

Flat_Node: A Flatten MLP multi-label classification method for all non-root nodes.

Label Distribution Learning Methods:

LIBLE: A LIBLE method proposed by Zheng et al. Zheng et al. (2023), employs an information bottleneck theory to integrate label assignment information and label gap information to obtain label distribution.

HFSLDL: A hierarchical LDL method proposed by Lin et al. Lin et al. (2023), leverages neighbor enhancement, hierarchical depth, and logical labels to derive label distributions.

H-LDL: A local hierarchical LDL method proposed by Xu et al. C & X (2019), leverages common ancestor path distance, logical labels to obtain label distributions.

Resampling Method:

RF&SMT+TL: A random resample method used in this paper to appropriate Pereira et al.'s Pereira et al. (2021a) LCN with LI and SMOTE-TOMEK method which is not open-source, employs a leaf node random forest classification combined with SMOTE+TOMEK resampling strategy.

Global Method:

C-HMCNN(h): A global hierarchical classification method proposed by Eleonora et al . Giunchiglia & Lukasiewicz (2020), concatenates a consistency constraint layer after the multi-label output to utilize the label hierarchy information.

4.3 Experimental Settings

The hyper-parameters used in the label enhancement stage of this paper are as follows: epochs = 600, batch size = 16, learning rate = 0.0001, dropout=0.4, and $\lambda_1 = 0.01, \lambda_2 = 0.01, \lambda_3 = 1$, $\alpha_1 = 0.96, \alpha_2 = 0.02$. The hyper-parameters used in the classification model training stage are as follows: epochs=400, batch size=4, learning rate=0.0001, δ=0.001, k=3, dropout=0.4. After comparing imbalanced resampling methods, SMOTE+TOMEK was found a better performance when paired with the random forest classification method. In contrast, for MLP used in our approach, RandomOverSampler achieves superior performance and has thus been adopted as the final resampling strategy. The implementation of the imbalanced multi-label resampling mainly utilizes Python's imblearn and skmultilearn packages.

4.4 Evaluation Metrics

The performance of the IHL-LDL method is measured using four commonly used evaluation metrics for hierarchical classification: micro-F1, macro-F1 scores, hF1, and Tree-induced Error (TE). The micro-F1 score assigns equal weight to each example, while the macro-F1 score assigns equal weight to each label. Methods whose results are biased towards labels with a large number of examples may obtain higher micro-F1 scores while recording lower macro-F1 scores.

hF1 is the hierarchical F1 score, which is an extension of the standard classification evaluation method micro-F1, where a larger value is better. It is defined as:

$$\mathrm{hF1} = \frac{2 \cdot \mathrm{HP} \cdot \mathrm{HR}}{\mathrm{HP} + \mathrm{HR}} \tag{11}$$

where HP is the hierarchical precision and HR is the hierarchical recall, defined as:

$$\begin{aligned} \mathrm{HP} &= \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \left|\mathrm{path}\left(y_{\mathrm{i}}\right) \cap \mathrm{path}\left(\hat{y}_{\mathrm{i}}\right)\right|}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \left|\mathrm{path}\left(\hat{y}_{\mathrm{i}}\right)\right|} \\ \mathrm{HR} &= \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \left|\mathrm{path}\left(y_{\mathrm{i}}\right) \cap \mathrm{path}\left(\hat{y}_{\mathrm{i}}\right)\right|}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \left|\mathrm{path}\left(y_{\mathrm{i}}\right)\right|} \end{aligned} \tag{12}$$

where path(·) is the set of path nodes for the label.

Tree-induced Error (TE) is used to measure the average shortest path distance between the true labels and predicted labels on the tree hierarchical structure, where a smaller value is better. It is defined as:

$$\mathrm{TE} = \frac{1}{n} \sum_{i=1}^{n} \gamma\left(y_i, \hat{y}_i\right) \tag{13}$$

where $\gamma(y_i, \hat{y}_i)$ is the shortest path distance between two nodes on the tree.

4.5 Results and Analysis

Table 2 presents the experimental results of the four metrics achieved by the proposed method and other baseline methods on the test datasets.

Among the 7 datasets examined in this paper, the IHL-LDL approach achieves the best results according to their MicroF1, hF1, and TE scores in 3, 2 and 2 datasets, respectively, and the top-2

Table 2: Experimental results of Hierarchical Classification Approachs. Bold and * denote the best and second best performance among all the approaches respectively.

Dateset	Approaches	micro-F1↑	macro-F1↑	hF1↑	TE↓
ImCLEF	Flat_Leaf	0.8767	0.6779*	0.9006	0.5964
	Flat_Node	0.8758	0.6674	0.9003	0.5984
	LIBLE	0.7893	0.5006	0.8673	0.9801
	HFSLDL	0.7425	0.3379	0.7913	1.2525
	H-LDL	0.7734	0.4570	0.8148	1.1113
	RF&SMT+TL	0.8320	0.5662	0.8622	0.8270
	C-HMCNN(h)	0.8787*	0.6736	**0.9039**	**0.5765**
	IHL-LDL(ours)	**0.8797**	**0.6780**	0.9033*	0.5805*
ImCLEF 07A	Flat_Leaf	0.8839*	0.6619	0.9250	0.5282
	Flat_Node	**0.8871**	0.6607	0.9267*	0.5165*
	LIBLE	0.8191	0.3493	0.8648	0.8111
	HFSLDL	0.7476	0.3066	0.8401	1.1214
	H-LDL	0.7902	0.4338	0.8664	0.9393
	RF&SMT+TL	0.8498	0.5733	0.9062	0.6603
	C-HMCNN(h)	0.8829	0.6592	0.9252	0.5272
	IHL-LDL(ours)	0.8839*	**0.6623**	**0.9292**	**0.4984**
ImCLEF 07D	Flat_Leaf	0.8121*	**0.7495**	0.8793*	0.8807*
	Flat_Node	0.8072	0.6727	0.8751	0.9165
	LIBLE	**0.8190**	0.3205	**0.8996**	**0.7071**
	HFSLDL	0.6561	0.3524	0.7675	1.7197
	H-LDL	0.6740	0.3942	0.7842	1.6004
	RF&SMT+TL	0.7803	0.6126	0.8569	1.0577
	C-HMCNN(h)	0.8042	0.6666	0.8737	0.9264
	IHL-LDL(ours)	0.8062	0.7441*	0.8754	0.9165
AWAphog	Flat_Leaf	0.7714	0.7703	0.8632	1.0943
	Flat_Node	0.7667	0.7666	0.8642	1.0862
	LIBLE	0.5915	0.5951	0.7984	1.6127
	HFSLDL	0.2164	0.1336	0.5751	3.3991
	H-LDL	0.3442	0.3249	0.6224	3.0206
	RF&SMT+TL	0.7751*	0.7733*	**0.8686**	**1.0512**
	C-HMCNN(h)	0.7661	0.7661	0.8626	1.0993
	IHL-LDL(ours)	**0.7758**	**0.7751**	0.8675*	1.0600*
F194	Flat_Leaf	0.4521	0.2459	0.7570	1.4578
	Flat_Node	0.4641	0.2549	0.7617	1.4296
	LIBLE	0.1796	0.0133	0.6585	2.0493
	HFSLDL	0.4458	0.2389	0.7585	1.4493
	H-LDL	**0.4986**	**0.3329**	**0.7791**	**1.3254**
	RF&SMT+TL	0.3092	0.3094*	0.6570	2.0578
	C-HMCNN(h)	0.4655*	0.2748	0.7655*	1.4070*
	IHL-LDL(ours)	0.4593	0.2511	0.7570	1.4578
DD	Flat_Leaf	0.7984	0.7688	0.9102	0.5388
	Flat_Node	**0.8595**	**0.8407**	**0.9366**	**0.3802**
	LIBLE	0.7703	0.6529	0.9058	0.5653
	HFSLDL	0.7504	0.6742	0.8920	0.6479
	H-LDL	0.7736	0.7578	0.8970	0.6182
	RF&SMT+TL	0.8165	0.8195	0.9212	0.4727
	C-HMCNN(h)	0.8446	0.8242*	0.9300	0.4198
	IHL-LDL(ours)	0.8463*	0.8218	0.9322*	0.4066*
VOC	Flat_Leaf	0.4889	**0.3523**	0.6949*	1.8607
	Flat_Node	0.4927*	0.3441	0.6955*	1.8349*
	LIBLE	0.4482	0.2608	0.6695	1.9530
	HFSLDL	0.4449	0.3303	0.6804	2.0098
	H-LDL	0.4431	0.3365	0.6804	2.0229
	RF&SMT+TL	0.2746	0.2601	0.5913	2.7830
	C-HMCNN(h)	0.4868	0.3462	0.6937	1.8652
	IHL-LDL(ours)	**0.4932**	0.3498*	**0.6993**	**1.8343**

Table 3: Comparison of Hierarchical Classification Ablation Results on all Datasets. Bold denotes the best performance among all the approaches.

Dateset	Approaches	micro-F1↑	macro-F1↑	hF1↑	TE↓
ImCLEF	without LDL	0.8668	0.6699	0.8963	0.6223
	without ROS	**0.8797**	**0.6782**	**0.9062**	**0.5626**
	IHL-LDL	**0.8797**	0.6780	0.9033	0.5805
ImCLEF 07A	without LDL	0.8722	**0.6824**	0.9187	0.5730
	without ROS	**0.8850**	0.6699	0.9264	0.5186
	IHL-LDL	0.8839	0.6623	**0.9292**	**0.4984**
ImCLEF 07D	without LDL	0.8042	**0.7863**	0.8724	0.9443
	without ROS	**0.8092**	0.6574	**0.8771**	**0.8986**
	IHL-LDL	0.8062	0.7441	0.8754	0.9165
AWAphog	without LDL	0.7586	0.7589	0.8606	1.1156
	without ROS	0.7730	0.7728	**0.8687**	**1.0506**
	IHL-LDL	**0.7758**	**0.7751**	0.8675	1.0600
F194	without LDL	**0.4697**	**0.2714**	**0.7636**	**1.4183**
	without ROS	0.4486	0.2505	0.7566	1.4606
	IHL-LDL	0.4593	0.2511	0.7570	1.4578
DD	without LDL	**0.8628**	**0.8549**	**0.9372**	**0.3769**
	without ROS	0.8446	0.8185	0.9317	0.4099
	IHL-LDL	0.8463	0.8218	0.9322	0.4066
VOC	without LDL	0.4878	**0.3514**	0.6944	1.8725
	without ROS	0.4907	0.3333	0.6964	1.8441
	IHL-LDL	**0.4932**	0.3498	**0.6993**	**1.8343**

results of all terms in 5 datasets. Compared to the second-best approach, the proposed approach achieves gains of 0.10%, 0.38%, and 0.0181 in terms of MicroF1, hF1, and TE scores, respectively, and surpasses the micro-F1 score of 0.8687 on the ImCLEF dataset from Pereira et al. Pereira et al. (2021a). These results indicate that the proposed approach can improve the classification performance of fine-grained labels in hierarchical classification. It is worth noting that the LIBLE and H-LDL approachs achieved the best results in MicroF1, hF1, and TE scores respectively on the ImCLEF07D and F194 datasets, which demonstrating the effect of label distribution learning on improving hierarchical classification task.

4.6 Ablation Study

The results of the ablation study conducted on all datasets of the label enhancement and resampling mechanisms are shown in Table 3. The independent use of label enhancement or resampling mechanisms can improve micro-F1, hF1 and reduce TE scores on some datasets. In particular, the resampling mechanism can enhance macro-F1 score. The combination of the two strategies can further improve performance on most datasets (3 out of 7 datasets). These results indicate that the adoption of label enhancement and resampling mechanisms respectively yielded representation capabilities and enhanced generalization, which enabling the proposed approach to achieve better results.

5 Conclusion

This paper proposes the IHL-LDL approach, a novel global HC method consisting of three modules: label enhancement, imbalanced resampling, and adaptive multi-label learning with hierarchical consistency. Label enhancement obtains a joint distribution representation of samples and the label space. Imbalanced resampling achieves sample synthesis for sparse sample labels, improving sample balance for fine-grained labels. Adaptive weighting and hierarchical consistency learning further integrate the constraint information contained within the hierarchical structure from a global perspective. Experiments are conducted on multiple HC datasets to evaluate the effectiveness of the proposed method. The results show that compared to previous state-of-the-art methods, the pro-

posed approach achieves the best results according to MicroF1, hF1, and TE scores in 3, 2, and 2 out of the 7 datasets, respectively, with gains of 0.10%, 0.38%, and 0.0181. Furthermore, the ablation study confirms the effectiveness of each module employed in the IHL-LDL approach.

REFERENCES

Jinhyun Bang, Jonghun Park, and Jonghyuk Park. Gacaps-htc: graph attention capsule network for hierarchical text classification. *Applied Intelligence (2023)*, 53:20577–20594, 2023. (Applied Intelligence) https:////doi.org/10.1007/s10489-023-04585-6.

Xu C and Geng X. Hierarchical classification based on label distribution learning. *Proc AAAI Conf Artif Intell*, 33:5533–5540, 2019.

Riddhiman Dasgupta and Anoop M. Namboodiri. Leveraging multiple tasks to regularize fine-grained classification. *ICPR2016*, pp. 3476–3481, 2016.

Torba Fatos, Gravier Christophe, Laclau Charlotte, Kammoun1 Abderrhammen, and Subercaze Julien. Decoding the hierarchy: A hybrid approach to hierarchical multi-label text classification. *Springer Nature Switzerland AG 2025*, 2025.

Batista G, Prati R, and Monard M. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor Newsl*, 6(1):20–29, 2004.

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

Eleonora Giunchiglia and Thomas Lukasiewicz. coherent hierarchical multi label classification networks. *34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada*, 2020.

Tomek I. An experiment with the edited nearest-neighbor rule. *IEEE Trans Syst Man Cybern*, 6(6): 448–452, 1976.

Fürnkranz J, Hüllermeier E, and Loza Menc´ıa E et al. Multilabel classification via calibrated label ranking. *Mach Learn(2008)*, 73(2):133–153, 2008. (Mach Learn) https:////doi.org/10.1007/s10994-008-5064-8.

Wehrmann J, R Cerri, and Barros R C. Hierarchical multi-label classification networks. *International Conference on Machine Learning*, 35, 2019.

Zhou J, Ma C, and Long D et al. Hierarchy-aware global model for hierarchical text classification. *ACL(2020)*, pp. 1106–1117, 2020. (ACL(2020)) https:////doi.org/10.18653/v1/2020.acl-main.104.

Xiuyi Jia, Tao Wen, Weiping Ding, Huaxiong Li, and Weiwei Li. Semi-supervised label distribution learning via projection graph embedding. *Information Sciences*, 581:840–855, 2021.

Yaojin Lin, Haoyang Liu, Hong Zhao, Qinghua Hu, Xingquan Zhu, and Xindong Wu. Hierarchical feature selection based on label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):5964–5976, 2023.

Yunan Lu, Liang He, Fan Min, Weiwei Li, and Xiuyi Jia. A generative label enhancement model with gaussian mixture and partial ranking. *AAAI2023*, pp. 8975–8983, 2023a.

Yunan Lu, Weiwei Li, and Xiuyi Jia. Label enhancement via joint implicit representation clustering. *IJCAI2023*, pp. 4019–4027, 2023b.

Yunan Lu, Bowen Xue, Xiuyi Jia, and Lei Yang. Entropy-calibrated label distribution learning. *NIPS2025*, 2025.

Chawla N, Bowyer K, Hall L, and Kegelmeyer P. Smote: synthetic minority over-sampling technique. *J Artif Intell Res*, 16:321–357, 2002.

R. M. Pereira, M. G. Costa Yandre, and N. Silla Jr. Carlos. Handling imbalance in hierarchical classfication problems using local classifiers approaches. *Data Mining and Knowledge Discovery*, 2021a. (Data Mining and Knowledge Discovery) https:////doi.org/10.1007/s10618-021-00762-8.

R. M. Pereira, M.G. Costa Yandre, and N. Silla Jr. Carlos. Toward hierarchical classification of imbalanced data using random resampling algorithms. *Information Sciences*, 578:344–363, 2021b.

Johnson R and Zhang T. Effective use of word order for text categorization with convolutional neural networks. *Proc Conf North American Chapter Assoc Comput Linguist: Human lang Technol. ACL(2015)*, pp. 103–112, 2015. (ACL(2015)) https:////doi.org/10.3115/v1/N15-1011.

C.N. Silla and A.A. Freitas. A survey of hierarchical classification across different application domains. *Data Min. Knowl. Discov.*, 22:31–72, 2011.

Ning Xu, An Tao, and Xin Geng. Label enhancement for label distribution learning. *International Joint Conference on Artificial Intelligence*, pp. 1632–1643, 2018.

Ning Xu, Jiaqi Lv, and Xin Geng. Partial label learning via label enhancement. *AAAI Conference on Artificial Intelligence*, pp. 5557–5564, 2019.

Ning Xu, Jun Shu, Yun-Peng Liu, and Xin Geng. Variational label enhancement. *International Conference on Machine Learning*, 119:10597–10606, 2020.

Ning Xu, Yun-Peng Liu, and Xin Geng. Label enhancement for label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 33(4):1632–1643, 2021.

Ning Xu, Jun Shu, Renyi Zheng, Xin Geng, Deyu Meng, and Min-Ling Zhang. Variational label enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):6537–6551, 2023.

Yu Y, Sun Z, and Sun C et al. Hierarchical multilabel text classification via multitask learning. *Proc IEEE int Conf Tools artif Intell*, pp. 1138–1143, 2021. (Proc IEEE int Conf Tools artif Intell) https:////doi.org/10.1109/ICTAI52525.2021.00180.

A. Zangari, M. Marcuzzo, M. Rizzo, L. Giudice, A. Albarelli, and A. Gasparetto. Hierarchical text classification and its foundations: A review of current research. *Electronics*, 13(7):1199, 2024. (Electronics) https:////doi.org/10.3390/electronics13071199.

Yunyi Zhang, Ruozhen Yang, Xueqiang Xu, Rui Li, Jinfeng Xiao, Jiaming Shen, and Jiawei Han. Teleclass: Taxonomy enrichment and llm enhanced hierarchical text classification with minimal supervision. *Proceedings of the ACM Web Conference 2025 (WWW'25)*, pp. 11, 2025.

Hong Zhao, Qinghua Hu, Pengfei Zhu, Yu Wang, and Ping Wang. A recursive regularization based feature selection framework for hierarchical classification. *IEEE Trans. on Knowledge and Data Engineering*, 2019.

Xingyu Zhao, Yuexuan An, Ning Xu, Jing Wang, and Xin Geng. Imbalanced label distribution learning. *AAAI2023*, 2023.

Xingyu Zhao, Yuexuan An, Ning Xu, and Xin Geng. Variational continuous label distribution learning for multi-label text classification. *TKDE2024*, 2024.

Qinghai Zheng, Jihua Zhu, and Haoyu Tang. Label information bottleneck for label enhancement. *CVPR 2023*, pp. 7497–7506, 2023.