Bridging Structure and Semantics via Path-level Alignment for Hierarchical Multi-Label Text Classification

Anonymous ACL submission

Abstract

Hierarchical multi-label text classification (HMTC) aims to assign documents with multiple labels organized in a predefined hierarchy, posing challenges for modeling both the hierarchical structure and the fine-grained label semantics. Existing approaches often rely on hierarchy-specific prediction architectures or hard consistency constraints, which can limit flexibility and robustness, especially for deep and imbalanced hierarchies. In this work, we propose a hierarchy-aware representation learning framework that reformulates HMTC as a path-level semantic alignment problem. We introduce PathSimNCE, a hierarchy-aware contrastive objective that aligns text with hierarchical paths using structure-based similarity, and incorporate an auxiliary description alignment objective using natural-language path descriptions generated offline by a large language model, without introducing inference-time overhead. Extensive experiments on three benchmark datasets show that our approach achieves competitive or state-of-the-art performance using a standard RoBERTa encoder and a unified classifier. Further ablation and depth-wise analyses show that hierarchy-aware and semantic supervision play complementary roles, significantly improving performance on fine-grained and rare labels.

1 Introduction

Hierarchical multi-label text classification (HMTC) aims to assign multiple labels to a document, where the labels are organized within a predefined hierarchy (Liu et al., 2023a) (Figure 1). This setting commonly arises in real-world applications, such as news topic classification (Kowsari et al., 2019), categorization of scientific publications (Chen et al., 2024a), and book genre annotation, where labels exhibit rich semantic and structural dependencies (Zhou et al., 2020). Unlike flat multi-label classification, HMTC requires models to capture not only label co-occurrence patterns and hierarchical relations, but also to effectively align textual representations with labels across different levels of semantic granularity (Kumar and Toshniwal, 2024).

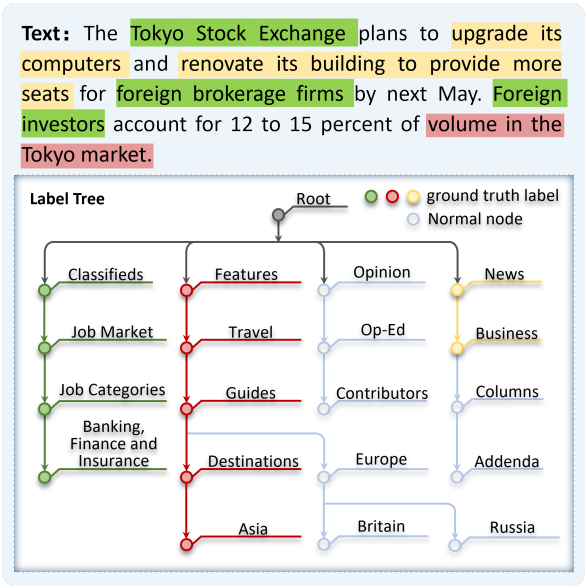

Figure 1: An example of HMTC. Labels are organized in a top-down hierarchy from coarse to fine. Labels sharing the same color correspond to nodes along the same hierarchical label path.

The hierarchical structure of HMTC poses several fundamental challenges (Zhou et al., 2020). First, real-world label hierarchies are often deep and highly imbalanced, with many fine-grained labels appearing infrequently in the training data (Zhang et al., 2024b,a). As a result, models tend to perform well on coarse-grained categories but struggle to predict rare labels at deeper levels of the hierarchy (Chalkidis et al., 2020). Second, hierarchical relationships encode rich structural information, such as parent–child and sibling relationships, which are not explicitly captured by standard label-wise objectives (Zhu et al., 2023). Finally, label semantics are often insufficiently specified by label names alone, especially for fine-grained categories,

making it difficult for models to generalize beyond superficial lexical cues (Chen et al., 2021).

Existing HMTC approaches can be broadly grouped into constraint-based and representation-based approaches (Giunchiglia and Lukasiewicz, 2020; Jiang et al., 2022). Constraint-based methods augment flat multi-label classifiers by incorporating hierarchical consistency constraints, either through structured inference or hierarchy-regularized loss functions, to enforce parent-child dependencies during prediction (Kim et al., 2024; Vaswani et al., 2025). In contrast, representation-based methods learn continuous label embeddings and explicitly encode hierarchical information, often via graph encoders, hierarchy-aware attention, or top-down decoding strategies, to capture parent–child dependencies and reduce the effective search space (Sadat and Caragea, 2022; Zhu et al., 2024). To address the long-tail issue, many studies further adopt re-weighting, sampling, or metric-learning strategies to improve performance on rare, fine-grained labels (Cao et al., 2019; Huang et al., 2021). Despite these advances, effectively aligning document representations with labels across multiple levels of semantic granularity remains challenging, particularly when labels are treated as discrete identifiers or when hierarchy modeling is tightly coupled to specific prediction architectures (Sadat and Caragea, 2022; Shen et al., 2021).

In this work, we take a different perspective and argue that hierarchical structure can be more effectively exploited during representation learning rather than at the prediction stage. We propose a hierarchy-aware representation learning framework that reformulates HMTC as a path-level semantic alignment problem. Instead of treating labels as independent nodes, we model each label as a root-to-node path in the hierarchy and align document representations with these paths in a shared embedding space. To this end, we introduce **PathSimNCE**, a hierarchy-aware contrastive learning objective that aligns text embeddings with their corresponding hierarchical paths while propagating supervision to structurally related paths through hierarchy-based similarity. This soft, structure-aware supervision encourages the model to learn representations that respect hierarchical relationships without requiring specialized decoders. In addition, we incorporate semantic description alignment by leveraging natural-language path descriptions generated offline using a large language model (LLM). These descriptions provide complementary semantic grounding during training, particularly for rare and fine-grained labels, without incurring any additional cost at inference time.

Our contributions are summarized as follows:

1. We propose a hierarchy-aware representation learning framework that injects hierarchical structure during training while remaining decoder-agnostic at inference time.
2. We introduce PathSimNCE, a path-level contrastive objective that leverages hierarchy-based similarity to provide soft, structure-aware supervision.
3. We incorporate semantic description alignment using natural-language path descriptions generated offline, enabling richer supervision without inference overhead.
4. We conduct comprehensive empirical analysis, including ablation studies and depth-wise evaluation, demonstrating the effectiveness and robustness of the proposed framework across diverse hierarchical settings.

2 Related Work

We review prior work on HMTC, focusing on constraint-based methods, representation-based hierarchical modeling, and hierarchy-aware contrastive learning.

Constraint-based approaches enforce hierarchical consistency by incorporating parent–child dependencies into prediction. These methods typically extend flat multi-label classifiers with structured inference or hierarchy-regularized losses, or adopt local strategies that train level- or node-specific classifiers and perform top-down or node-wise prediction (Kowsari et al., 2017a; Shimura et al., 2018). To reduce complexity, some variants partition the hierarchy by level, parent, or node (Stein et al., 2019; Peng et al., 2018; Krendzelak and Jakab, 2021). While these methods are conceptually simple, they are prone to error propagation, as mistakes at higher levels can cascade to downstream decisions. Moreover, hard hierarchical constraints often require task-specific inference procedures, which can be brittle in deep or noisy hierarchies. To alleviate these issues, several studies reformulate hierarchical classification as a sequential decision problem (Mao et al., 2019; Im et al., 2023). Despite their effectiveness, such approaches further entangle hierarchy modeling with specialized prediction architectures.

Representation-based methods instead encode

hierarchical structure directly into learned representations, avoiding explicit constraints at inference time. A common strategy is to learn label embeddings and propagate hierarchical information through graph-based encoders, hierarchy-aware attention, or structured message passing. For instance, Zhou et al. (2020) encode hierarchical relationships as structured constraints within a unified classifier, while Chen et al. (2021) formulate hierarchical classification as a text-to-label semantic matching problem. Other studies incorporate external knowledge or specialized geometric spaces to enhance hierarchical representation learning. Liu et al. (2023b) enrich label semantics with knowledge graphs and label attention, whereas Kumar and Toshniwal (2025) model instance-specific hierarchies in hyperbolic space, which naturally captures tree-like structures. Similarly, Zhu et al. (2023) employ hierarchy-aware tree isomorphic networks to represent complex hierarchical topologies. While these methods improve global consistency, they often rely on architecture-specific designs and remain sensitive to label imbalance.

Hierarchy-aware contrastive learning has recently emerged as an effective paradigm for HMTC, aligning texts and labels in a shared embedding space. Prior work explores how to construct hierarchy-aware positive and negative samples for contrastive objectives. For example, Zhou et al. (2025) propose hierarchy-aware negative sampling strategies, while Wang et al. (2022) inject hierarchical information into the text encoder and show improved text–label alignment. U et al. (2023) jointly model texts and labels while accounting for hierarchical structure. To address label imbalance, some studies introduce hierarchical margin constraints or rebalancing strategies within metric learning frameworks (Kim et al., 2024; Zhang et al., 2024b). Related efforts also explore alternative formulations, such as sequence generation (Jain et al., 2024), lightweight hierarchy-aware modules (Chen et al., 2024b), and combining label-wise attention with hierarchical constraints (Zhang et al., 2022). Additionally, Cai et al. (2024) demonstrate that incorporating named entity information can further enrich hierarchical representations.

In contrast to hierarchy-aware contrastive approaches that operate at the instance level or align texts with individual labels, our method performs contrastive learning at the path level, explicitly modeling each label as a root-to-node path in the hierarchy. Moreover, instead of using hard positive–negative assignments, we introduce hierarchy-aware soft targets derived from path–path similarity, allowing supervision to be smoothly propagated to structurally related paths. This formulation enables finer-grained and more flexible hierarchy modeling during representation learning, particularly for deep and imbalanced taxonomies.

3 Methodology

We propose a hierarchy-aware representation learning framework that reformulates HMTC as a path-level semantic alignment task, achieving strong performance without specialized hierarchical decoders (Figure 2). The framework encodes both texts and hierarchical paths into a shared representation space, and leverages LLM–generated path descriptions as auxiliary semantic supervision. We introduce a hierarchy-aware in-batch contrastive learning objective that aligns texts with their corresponding paths while explicitly modeling hierarchical similarity among paths. This contrastive objective is jointly optimized with a standard multi-label classification loss and a description alignment loss, allowing the model to integrate label-wise prediction, hierarchical structure, and semantic grounding in an end-to-end manner.

3.1 Task Formulation

We reformulate HMTC as a **path-level prediction problem** over a label hierarchy. Let $\mathcal{Y}$ denote a set of labels organized in a rooted hierarchical structure, where each label admits one or more root-to-node paths. In this work, we focus on hierarchies that can be represented as rooted trees or directed acyclic graphs (DAGs) with well-defined root-to-node paths.Given an input text x, the goal is to predict a set of labels $Y_x \subseteq \mathcal{Y}$, potentially spanning multiple depths of the hierarchy. Each label $y_i \in \mathcal{Y}$ is associated with one or more hierarchical paths, defined as ordered sequences $p = (y_1, y_2, \ldots, y_L)$, where y_1 is the root and y_L the target label. Under this formulation, the model identifies a set of paths $\mathcal{P}_x = \{p_1, p_2, ..., p_K\}$ for input x, such that the terminal nodes of these paths correspond to the ground-truth labels in Y_x. By operating at the path level, this formulation explicitly encodes hierarchical structure while remaining compatible with standard multi-label classification objectives.

3.2 Path and Text Representations

We encode both input texts and hierarchical paths into a shared representation space. Given an input

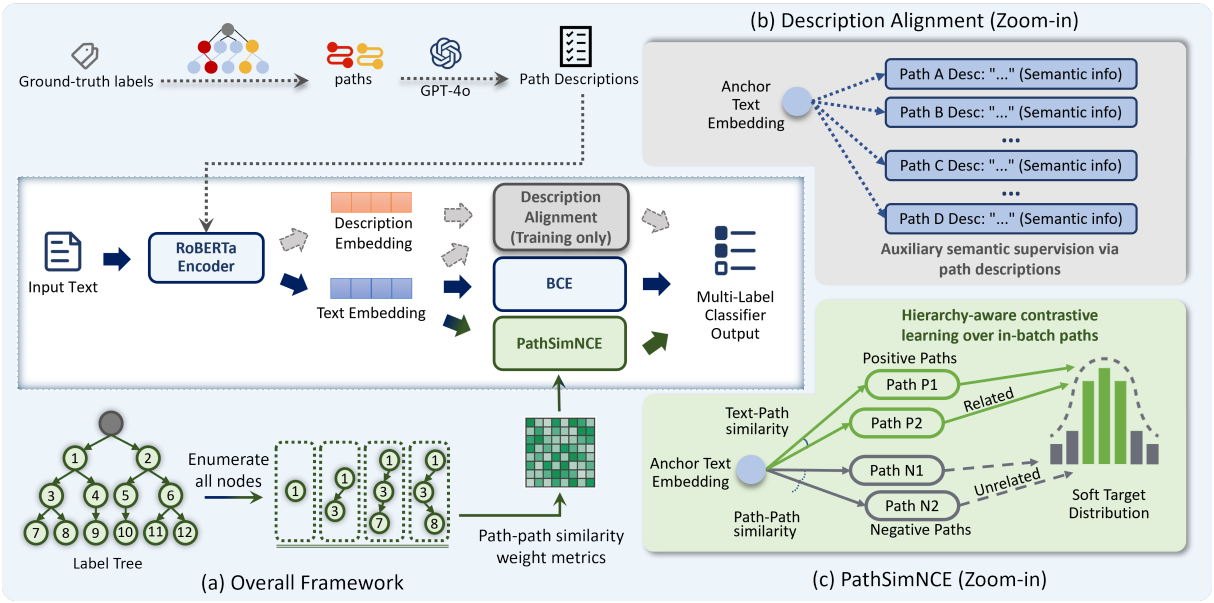

Figure 2: Overview of the proposed hierarchy-aware representation learning framework. (a) Texts are encoded using a shared RoBERTa encoder and jointly optimized with three objectives: BCE classification loss, PathSimNCE, and description alignment. Hierarchical labels are represented as root-to-node paths rather than independent nodes. (b) Description alignment aligns text embeddings with natural-language path descriptions, providing auxiliary semantic supervision during training without affecting inference. (c) PathSimNCE performs hierarchy-aware contrastive learning over in-batch paths by using both text–path and path–path similarities to construct soft targets.

text x, we obtain a contextualized text representation using RoBERTa (Liu et al., 2019):

$$\mathbf{h}_x = \text{RoBERTa}(x), \tag{1}$$

where $\mathbf{h}_x \in \mathbb{R}^d$ denotes the pooled representation. We then apply a projection head to map $\mathbf{h}_x$ into the path-alignment space:

$$\mathbf{e}_x = f_{\text{proj}}(\mathbf{h}_x). \tag{2}$$

Each hierarchical path p is represented by a learnable embedding $\mathbf{e}_p \in \mathbb{R}^d$, which serves as a structural anchor in the shared embedding space. These path embeddings are jointly optimized with the text encoder and are used to capture hierarchical relations among labels.

To incorporate semantic information beyond discrete path identifiers, we additionally generate a natural-language description d_p for each path using an LLM. These descriptions encode the hierarchical semantics of the paths and are used only during training. Each description is encoded using the same RoBERTa encoder and projection head:

$$\mathbf{e}_{d_p} = f_{\text{proj}}(\text{RoBERTa}(d_p)). \tag{3}$$

These resulting representations provide auxiliary semantic supervision, encouraging alignment between input texts and the semantic content of their corresponding hierarchical paths.

3.3 PathSimNCE: Path Similarity-Aware InfoNCE

We propose PathSimNCE, a hierarchy-aware in-batch contrastive objective that uses text–path similarities for prediction and path–path similarities to construct soft contrastive targets. Given a mini-batch of input texts $\{x_i\}_{i=1}^B$ and their associated ground-truth path sets $\{\mathcal{P}_{x_i}\}_{i=1}^B$, we construct an candidate set $\mathcal{C}$ as the union of all ground-truth paths appearing in the batch. This set serves as a shared contrastive label space for all examples in the batch. For each text-path pair, we compute the text–path similarity using scaled cosine similarity:

$$\text{sim}_{\text{text}}(x_i, p) = \frac{\mathbf{e}_{x_i}^\top \mathbf{e}_p}{\tau_{\text{h}}}, \tag{4}$$

where τ_{h} is a temperature hyperparameter. The similarities are normalized with a softmax over $\mathcal{C}$ to obtain a predicted distribution over candidate paths for each input text.

To encode hierarchical structure, we define the structural distance between two paths p and $\tilde{p}$ based on their lowest common ancestor (LCA):

$$\text{dist}(p, \tilde{p}) = |p| + |\tilde{p}| - 2\,|\text{LCA}(p, \tilde{p})|, \tag{5}$$

where $|\cdot|$ represents the path length. This distance

is converted into a path-path similarity via an exponential decay:

$$\mathrm{sim}_{\text{path}}(p, \tilde{p}) = \exp\left(-\lambda \cdot \mathrm{dist}(p, \tilde{p})\right), \quad (6)$$

where λ controls the decay rate.

For each input x_i, we then construct a soft target distribution over the candidate set $\mathcal{C}_i$. Specifically, for a candidate path $p \in \mathcal{C}_i$, we first compute an unnormalized score by aggregating its similarity to the ground-truth paths $\mathcal{P}_{x_i}$:

$$\tilde{q}_i(p) = \sum_{p' \in \mathcal{P}_{x_i}} \mathrm{sim}_{\text{path}}(p, p'), \quad (7)$$

and then normalize these scores to obtain a probability distribution:

$$q_i(p) = \frac{\tilde{q}_i(p)}{\sum_{p'' \in \mathcal{C}_i} \tilde{q}_i(p'')}. \quad (8)$$

This soft target assigns higher probability mass to paths that are structurally closer to the ground-truth paths, enabling smooth propagation of hierarchical supervision.

Finally, the PathSimInfoNCE is defined as the cross-entropy between the soft target distribution q_i and the predicted distribution induced by the text–path similarities:

$$\mathcal{L}_{\text{HCL}} = -\frac{1}{B} \sum_{i=1}^{B} \sum_{p \in \mathcal{C}} q_i(p) \log \frac{\exp(\mathrm{sim}_{\text{text}}(x_i, p))}{\sum_{p'' \in \mathcal{C}} \exp(\mathrm{sim}_{\text{text}}(x_i, p''))}. \quad (9)$$

This objective encourages each text representation to align most strongly with its ground-truth paths, while smoothly distributing probability mass to hierarchically related paths in proportion to their structural similarity.

3.4 Description Alignment

While PathSimNCE captures hierarchical structure through path-level supervision, it does not explicitly enforce alignment with the semantic meanings of the hierarchical paths. To address this limitation, we introduce a **description alignment objective** as auxiliary supervision, which aligns text representations with the natural-language descriptions associated with their ground-truth paths.

Given an input text x_i and its corresponding path set $\mathcal{P}_{x_i}$, we treat the descriptions associated with these paths as positive semantic views. The description alignment loss is defined as a multi-positive contrastive objective:

$$\mathcal{L}_{\text{DESC}} = -\frac{1}{B} \sum_{i=1}^{B} \frac{1}{|\mathcal{D}_i|} \sum_{d \in \mathcal{D}_i} \log \frac{\exp\left(\mathbf{e}_{x_i}^{\top} \mathbf{e}_d / \tau_{\text{desc}}\right)}{\sum_{d'} \exp\left(\mathbf{e}_{x_i}^{\top} \mathbf{e}_{d'} / \tau_{\text{desc}}\right)}, \quad (10)$$

where $\mathcal{D}_i$ denote the set of descriptions associated with x_i, and τ_{desc} is a temperature hyperparameter. By averaging the contrastive loss over all positive descriptions, this objective encourages each text representation to align consistently with the semantic content of its associated paths.

The description alignment loss serves as an auxiliary regularizer and is jointly optimized with the hierarchy-aware contrastive objective and the label-wise classification loss.

3.5 Joint Learning

We jointly optimize all proposed objectives end-to-end. Given an input text x_i, the model predicts label probabilities using a standard multi-label classifier, which is trained with the binary cross-entropy (BCE) loss:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} \Big[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \Big], \quad (11)$$

where $y_i \in \{0, 1\}^{|\mathcal{Y}|}$ represents the ground-truth indicator for x_i, and $\hat{y}_i$ is the predicted probability for label y.

The overall training objective jointly combines the label-wise classification loss, PathSimNCE, and the description alignment loss:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}} + \beta\, \mathcal{L}_{\text{HCL}} + \beta_{\text{desc}}\, \mathcal{L}_{\text{DESC}}, \quad (12)$$

where β and β_{desc} control the contributions of the hierarchy-aware contrastive objective and the description alignment objective, respectively.

The classification loss provides direct supervision for label prediction, PathSimNCE encodes hierarchical structure by aligning texts with relevant paths, and the description alignment loss supplies semantic grounding. All components are optimized jointly during training. Importantly, label descriptions are only used during training; at inference, the model requires only the text encoder and label classifier, introducing no additional computational overhead or auxiliary inputs.

4 Experimental settings

We evaluate our method on three HMTC benchmarks (Table 1). The **New York Times Annotated**

	# Labels	Depth	Avg # label	Avg # length
BGC	146	4	3.01	162.88
NYT	166	8	7.58	605.90
WOS	141	2	2.00	207.62

Table 1: Statistics of the datasets. For each dataset, we report the number of labels, hierarchy depth, average number of labels per document, and average document length.

Corpus (NYT) (Sandhaus) consists of news articles labeled with topics organized in an 8-level hierarchy, making it a challenging benchmark due to its deep hierarchical structure. The **Blurb Genre Collection (BGC)** (Aly et al., 2019) comprises English book blurbs annotated with multiple genres arranged in a 4-level hierarchy, and exhibits substantial label imbalance. **WOS-46985 (WOS)** (Kowsari et al., 2017b) includes scientific abstracts from the Web of Science, with labels organized in a relatively shallow 2-level hierarchy of research areas and sub-areas.

We implement our model in PyTorch (Paszke et al., 2019) and train on a single NVIDIA H100 80G GPU. Models are trained using the AdamW optimizer with a learning rate of $3.5e^{-5}$ for the encoder and $1e^{-4}$ for task-specific parameters. The batch size is set to 32 and training runs for up to 50 epochs with early stopping based on validation micro-F1, using a patience of 3 epochs. The temperature for contrastive objectives is set to $\tau = 0.07$. The hierarchy-aware contrastive loss weight β is set to 1.0, and the description alignment weight β_{desc} is set to 0.1. To provide semantic supervision for hierarchical paths, we generate natural-language path descriptions offline using GPT-4o (OpenAI, 2024). Details on the prompting strategy and generation settings are provided in Appendix A.1.

We evaluate model performance using micro-F1 and macro-F1, which are standard metrics for HMTC.

5 Results and Discussion

5.1 Main Results

We first compare our framework with the current state-of-the-art (SOTA) models (Table 2). Overall, our method achieves competitive or SOTA performance across all datasets while using a standard RoBERTa encoder and remaining decoder-agnostic.

On WOS, our model achieves the highest reported micro-F1 of 87.83 and a competitive macro-F1 of 81.88. Despite the shallow two-level hierarchy, our framework remains effective without relying on level-wise prediction heads or dataset-specific architectural assumptions.

On BGC, our approach establishes a new SOTA with micro-F1 of 82.93 and macro-F1 of 67.44, outperforming baselines such as HYDRA and HJCL. The improvement in macro-F1 indicates that path-level semantic alignment effectively mitigates label imbalance and improves performance on less frequent genres.

On NYT, which features a deep and fine-grained hierarchy, our method achieves a new SOTA micro-F1 of 82.16 and a macro-F1 of 71.88, closely matching the strongest existing results. Notably, this performance is obtained using a single unified classifier, demonstrating that incorporating hierarchical structure during representation learning can match the performance of more complex hierarchy-aware architectures.

5.2 Ablation Studies

We then evaluate the effectiveness and robustness of our framework by analyzing the contributions of its key components. In particular, we conduct ablation experiments along two dimensions: (1) the impact of different training objectives, and (2) the role of semantic label representations in the description alignment.

Effectiveness of Training Objectives We first examine the contribution of each loss component by comparing three variants: BCE, BCE+PathSimNCE, and BCE+DESC (Table 3). On WOS, BCE+DESC improves performance over BCE (87.01 / 79.03 vs. 86.31 / 78.58), whereas BCE+PathSimNCE reduces both micro- and macro-F1 (86.00 / 75.66). This result aligns with the limited utility of explicit hierarchical regularization in shallow hierarchies and suggests that semantic supervision provides a more robust auxiliary signal in this setting.

On BGC, BCE+DESC achieves the best overall performance, improving micro-F1 and macro-F1 to 80.83 and 65.04, compared to 80.36 / 63.97 for BCE. In contrast, BCE+PathSimNCE yields a small gain in micro-F1 (80.60) but reduces macro-F1 (61.99), indicating that hierarchy-aware contrastive regularization can trade off overall accuracy and tail-label performance on datasets with substantial label imbalance.

Model	WOS		BGC		NYT	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
HiMatch (Chen et al., 2021)	86.70	81.06	78.89	63.19	76.79	63.89
BERT+HiAGM (Wang et al., 2022)	86.04	80.19	78.62	62.98	78.64	66.76
HGCLR (Wang et al., 2022)	87.11	81.20	79.36	63.64	78.86	67.96
BERT+HTCInfoMax (Wang et al., 2022)	86.30	79.97	78.47	62.87	78.75	67.31
HYDRA Local (Karl and Scherp, 2025)	86.90	81.14	82.18	66.01	81.87	**72.43**
HJCL (U et al., 2023)	-	-	81.30	66.77	80.52	70.02
DFG (Liu et al., 2025)	87.42	82.27	81.17	66.13	-	-
Seq2Tree (Yu et al., 2022)	87.20	**82.50**	79.72	63.96	-	-
K-HTC (Liu et al., 2023b)	87.29	81.69	80.52	65.99	-	-
HILL (Zhu et al., 2024)	87.28	81.77	-	-	80.47	69.96
HALB (Zhang et al., 2024b)	87.45	82.04	-	-	79.56	69.28
HiSR (Zhou et al., 2025)	87.52	82.04	-	-	80.32	70.11
Ours	**87.83**	81.88	**82.93**	**67.44**	**82.16**	71.88

Table 2: Performance comparison to previous methods across WOS, BGC and NYT datasets. The results of HiMatch on BGC and NYT datasets come from Yu et al. (2022) and Huang et al. (2022), respectively. The results of HiAGM, HGCLR and HTCInfoMax on BGC come from Liu et al. (2025). Bold: best scores in each column.

On NYT, adding the hierarchy-aware contrastive loss yields the largest gains, improving micro-F1 from 80.26 to 81.92 and macro-F1 from 67.03 to 71.78. This suggests that explicit hierarchy-aware supervision is particularly beneficial for deep and fine-grained hierarchies with many structurally related and confusable labels. Description alignment also improves performance (81.32 / 70.10), but is less effective than PathSimNCE.

Effectiveness of Semantic Label Representations We further investigate the impact of different semantic label representations used in the description alignment objective. Specifically, we compare three variants: No Description, Temple Description, and LLM-generated Descriptions (Table 3). *No Description* disables the description alignment objective while retaining hierarchy-aware contrastive learning. *Temple Description* uses a simple template-based description constructed from the label name (e.g., "This document belongs to the category: ..."), while *LLM-generated Description* employs natural-language descriptions generated offline using GPT-4o.

On WOS, semantic supervision is crucial. *No Description* achieves 86.00 in micro-F1 and 75.66 in macro-F1, while *Temple Description* yields modest improvements (86.38 / 76.20). In contrast, *LLM-generated Description* provides substantial gains, achieving 87.83 in micro-F1 and 81.88 in macro-F1, indicating that rich semantic descriptions are effective when hierarchical structure is limited.

On BGC, *No Description* attains 80.60 in micro-F1 and 61.99 in macro-F1. *Temple Description* provides only marginal improvement (80.57 / 62.10), whereas *LLM-generated Description* lead to significant improvements, boosting performance to 82.93 in micro-F1 and 67.44 in macro-F1. The pronounced increase in macro-F1 suggests that natural-language descriptions are especially beneficial for mitigating label imbalance and improving performance on infrequent genres.

On NYT, which has a deep and fine-grained hierarchy, *No Description* already yields strong performance (81.92 in micro-F1, 71.78 in macro-F1), highlighting the effectiveness of hierarchy-aware contrastive learning. *Template Description*, however, degrades performance (80.83 / 67.66), likely due to insufficient semantic or noise. *LLM-generated Description* slightly improve micro-F1 to 82.16 and maintain competitive macro-F1 (71.88), indicating that high-quality semantic descriptions can complement structural supervision without undermining hierarchical learning.

Overall, these ablations show that hierarchy-aware contrastive learning is most effective for datasets with deep and fine-grained hierarchies, while semantic description alignment provides robust and consistent gains, particularly in macro-F1. Moreover, the quality of semantic supervision is crucial: LLM-generated descriptions significantly outperform template-based ones, and combining structural and semantic supervision yields the best overall performance.

5.3 Performance across Hierarchy Depths

While overall macro-F1 provides a summary measure of performance on rare labels, it does not re-

Model	WOS		BGC		NYT	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
BCE	86.31	78.58	80.36	63.97	80.26	67.03
BCE+PathSimNCE	86.00	75.66	80.60	61.99	81.92	71.78
BCE+DESC	87.01	79.03	80.83	65.04	81.32	70.10
No Description	86.00	75.66	80.60	61.99	81.92	71.78
Template Description	86.38	76.20	80.57	62.10	80.83	67.66
LLM-generated Description	87.83	81.88	82.93	67.44	82.16	71.88

Table 3: Ablation studies. In the upper block, BCE+DESC uses LLM-generated path descriptions and does not include PathSimNCE. In the lower block, we fix the structural objective (BCE+PathSimNCE) and vary the description source (none/template/LLM) for the description alignment loss.

veal how model behavior varies across hierarchical levels. To better understand the sources of performance gains, we analyze macro-F1 as a function of hierarchy depth on NYT, focusing on the model's ability to handle increasingly fine-grained labels (Figure 3). As expected, performance decreases with increasing depth for all methods, reflecting the growing sparsity and semantic specificity of deeper labels.

However, the rate of degradation differs across training objectives. Compared to the BCE-only baseline, PathSimNCE (BCE+PathSimNCE) substantially slows the performance drop at intermediate depths (approximately depths 3–6), indicating that hierarchy-aware contrastive supervision helps discriminate among structurally related categories. Description alignment (BCE+DESC) provides more uniform gains across depths, though its improvements are generally smaller at intermediate levels. The full model, which combines structural and semantic supervision, exhibits the most robust behavior at the deepest levels: while several methods converge at depth 7, the full model achieves the highest macro-F1 at depth 8, corresponding to the rarest and most fine-grained labels.

These results complement the ablation analysis by localizing performance gains within the hierarchy and demonstrate that structural and semantic supervision not only improve overall macro-F1, but also reduce the rate at which performance degrades with increases label depth.

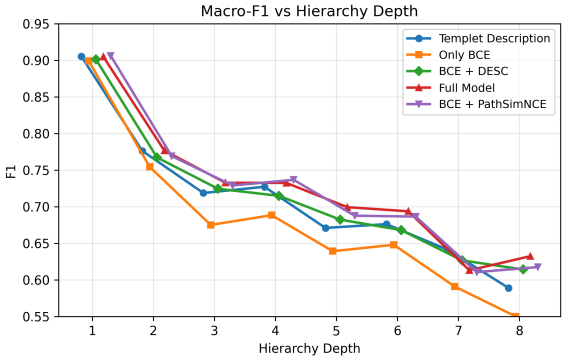

Figure 3: Macro-F1 versus hierarchy depth on NYT.

6 Conclusion

In this paper, we presented a hierarchy-aware representation learning framework for HMTC that injects hierarchical structure and semantic supervision during training while remaining decoder-agnostic at inference. By reformulating the task as a path-level semantic alignment problem, we introduced **PathSimNCE**, a hierarchy-aware contrastive objective that aligns text representations with hierarchical paths using structure-based similarity. We further incorporated **semantic description alignment** with natural-language path descriptions generated offline, which provides complementary supervision without additional inference-time cost. Extensive experiments on three benchmark datasets with varying hierarchy depths demonstrate that the proposed framework achieves competitive or SOTA performance using a standard encoder and unified classifier. Ablation studies and depth-wise analyses show that structural and semantic supervision play complementary roles. Hierarchy-aware and semantic supervision is particularly effective for deep hierarchies, while semantic descriptions provide robust gains for rare and fine-grained labels. Future work may explore adaptive or learned hierarchy similarity measures to better capture task-specific structural relationships, incorporating richer semantic supervision or external knowledge sources to enhance fine-grained label representations, and integrating the approach with specialized hierarchical architectures or dynamic label hierarchies to improve scalability and robustness.

Limitations

While the proposed framework demonstrates strong empirical performance, it has several limitations. First, our hierarchy-aware contrastive learning relies on the availability of a well-defined label hierarchy with meaningful structural relationships. In scenarios where the hierarchy is noisy, incomplete, or weakly correlated with label semantics, the benefits of path-level supervision may be reduced.

Second, although semantic description alignment improves performance on fine-grained labels, its effectiveness depends on the quality of the generated path descriptions. Inaccurate or overly generic descriptions may introduce noisy supervision and limit potential gains. Finally, our approach injects hierarchical and semantic information during training and does not explicitly enforce hierarchical consistency at inference time, which may still lead to occasional inconsistencies in predicted label sets.

References

Rami Aly, Steffen Remus, and Chris Biemann. 2019. Hierarchical multi-label classification of text with capsule networks. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 323–330, Florence, Italy. Association for Computational Linguistics.

Fuhan Cai, Duo Liu, Zhongqiang Zhang, Ge Liu, Xiaozhe Yang, and Xiangzhong Fang. 2024. NER-guided comprehensive hierarchy-aware prompt tuning for hierarchical text classification. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 12117–12126, Torino, Italia. ELRA and ICCL.

Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. 2019. Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Ilias Chalkidis, Manos Fergadiotis, Sotiris Kotitsas, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. An empirical study on large-scale multi-label text classification including few and zero-shot labels. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7503–7515, Online. Association for Computational Linguistics.

Haibin Chen, Qianli Ma, Zhenxi Lin, and Jiangyue Yan. 2021. Hierarchy-aware label semantics matching network for hierarchical text classification. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4370–4379, Online. Association for Computational Linguistics.

Huiyao Chen, Yu Zhao, Zulong Chen, Mengjia Wang, Liangyue Li, Meishan Zhang, and Min Zhang. 2024a. Retrieval-style in-context learning for few-shot hierarchical text classification. *Transactions of the Association for Computational Linguistics*, 12:1214–1231.

Zhijian Chen, Zhonghua Li, Jianxin Yang, and Ye Qi. 2024b. Hilight: A hierarchy-aware light global model with hierarchical local contrastive learning. *Preprint*, arXiv:2408.05786.

Eleonora Giunchiglia and Thomas Lukasiewicz. 2020. Coherent hierarchical multi-label classification networks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9662–9673. Curran Associates, Inc.

Wei Huang, Chen Liu, Bo Xiao, Yihua Zhao, Zhaoming Pan, Zhimin Zhang, Xinyun Yang, and Guiquan Liu. 2022. Exploring label hierarchy in a generative way for hierarchical text classification. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1116–1127, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Yi Huang, Buse Giledereli, Abdullatif Köksal, Arzucan Özgür, and Elif Ozkirimli. 2021. Balancing methods for multi-label text classification with long-tailed class distribution. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8153–8161, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

SangHun Im, GiBaeg Kim, Heung-Seon Oh, Seongung Jo, and Dong Hwan Kim. 2023. Hierarchical text classification as sub-hierarchy sequence generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(11):12933–12941.

Vidit Jain, Mukund Rungta, Yuchen Zhuang, Yue Yu, Zeyu Wang, Mu Gao, Jeffrey Skolnick, and Chao Zhang. 2024. HiGen: Hierarchy-aware sequence generation for hierarchical text classification. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1354–1368, St. Julian's, Malta. Association for Computational Linguistics.

Ting Jiang, Deqing Wang, Leilei Sun, Zhongzhi Chen, Fuzhen Zhuang, and Qinghong Yang. 2022. Exploiting global and local hierarchies for hierarchical text classification. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4030–4039, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Fabian Karl and Ansgar Scherp. 2025. HYDRA: A multi-head encoder-only architecture for hierarchical text classification. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 9303–9314, Suzhou, China. Association for Computational Linguistics.

Gibaeg Kim, SangHun Im, and Heung-Seon Oh. 2024. Hierarchy-aware biased bound margin loss function for hierarchical text classification. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 7672–7682, Bangkok, Thailand. Association for Computational Linguistics.

Kamran Kowsari, Donald E. Brown, Mojtaba Heidarysafa, Kiana Jafari Meimandi, Matthew S. Gerber, and Laura E. Barnes. 2017a. Hdltex: Hierarchical deep learning for text classification. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 364–371.

Kamran Kowsari, Donald E Brown, Mojtaba Heidarysafa, Kiana Jafari Meimandi, Matthew S Gerber, and Laura E Barnes. 2017b. Hdltex: Hierarchical deep learning for text classification. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)*, pages 364–371. IEEE.

Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. 2019. Text classification algorithms: A survey. *Information*, 10(4).

Milan Krendzelak and Frantisek Jakab. 2021. Hierarchical text classification using CNNs with local approaches. *Computing and Informatics*, 39(5):907–924.

Ashish Kumar and Durga Toshniwal. 2024. Modeling text-label alignment for hierarchical text classification. In *Machine Learning and Knowledge Discovery in Databases. Research Track*, pages 163–179, Cham. Springer Nature Switzerland.

Ashish Kumar and Durga Toshniwal. 2025. HyILR: Hyperbolic instance-specific local relationships for hierarchical text classification. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 872–883, Vienna, Austria. Association for Computational Linguistics.

Renyuan Liu, Xuejie Zhang, Jin Wang, and Xiaobing Zhou. 2025. Disentangled feature graph for hierarchical text classification. *Information Processing & Management*, 62(3):104065.

Rundong Liu, Wenhan Liang, Weijun Luo, Yuxiang Song, He Zhang, Ruohua Xu, Yunfeng Li, and Ming Liu. 2023a. Recent advances in hierarchical multi-label text classification: A survey. *Preprint*, arXiv:2307.16265.

Ye Liu, Kai Zhang, Zhenya Huang, Kehang Wang, Yanghai Zhang, Qi Liu, and Enhong Chen. 2023b. Enhancing hierarchical text classification through knowledge graph integration. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5797–5810, Toronto, Canada. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *Preprint*, arXiv:1907.11692.

Yuning Mao, Jingjing Tian, Jiawei Han, and Xiang Ren. 2019. Hierarchical text classification with reinforced label assignment. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 445–455, Hong Kong, China. Association for Computational Linguistics.

OpenAI. 2024. Gpt-4o model. `https://platform.openai.com/docs/models/gpt-4o`.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, and 2 others. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Hao Peng, Jianxin Li, Yu He, Yaopeng Liu, Mengjiao Bao, Lihong Wang, Yangqiu Song, and Qiang Yang. 2018. Large-scale hierarchical text classification with recursively regularized deep graph-cnn. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 1063–1072, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Mobashir Sadat and Cornelia Caragea. 2022. Hierarchical multi-label classification of scientific documents. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8923–8937, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Evan Sandhaus. The new york times annotated corpus. (ACQUIRED).

Jiaming Shen, Wenda Qiu, Yu Meng, Jingbo Shang, Xiang Ren, and Jiawei Han. 2021. TaxoClass: Hierarchical multi-label text classification using only class names. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4239–4249, Online. Association for Computational Linguistics.

Kazuya Shimura, Jiyi Li, and Fumiyo Fukumoto. 2018. HFT-CNN: Learning hierarchical category structure for multi-label short text categorization. In *Proceedings of the 2018 Conference on Empirical Methods*

in Natural Language Processing, pages 811–816, Brussels, Belgium. Association for Computational Linguistics.

Roger Alan Stein, Patricia A. Jaques, and João Francisco Valiati. 2019. An analysis of hierarchical text classification using word embeddings. *Information Sciences*, 471:216–232.

Simon Chi Lok U, Jie He, Víctor Gutiérrez-Basulto, and Jeff Z. Pan. 2023. Instances and labels: Hierarchy-aware joint supervised contrastive learning for hierarchical multi-label text classification. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8858–8875, Singapore. Association for Computational Linguistics.

Ashwin Vaswani, Yashas Samaga, Gaurav Aggarwal, Praneeth Netrapalli, and Narayan Hegde. 2025. All mistakes are not equal: Comprehensive hierarchy aware multilabel predictions (champ). In *Pattern Recognition*, pages 264–282, Cham. Springer Nature Switzerland.

Zihan Wang, Peiyi Wang, Lianzhe Huang, Xin Sun, and Houfeng Wang. 2022. Incorporating hierarchy into text encoder: a contrastive learning approach for hierarchical text classification. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7109–7119, Dublin, Ireland. Association for Computational Linguistics.

Chao Yu, Yi Shen, and Yue Mao. 2022. Constrained sequence-to-tree generation for hierarchical text classification. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, page 1865–1869, New York, NY, USA. Association for Computing Machinery.

Chaoqun Zhang, Linlin Dai, Chengxing Liu, and Longhao Zhang. 2024a. Hgbl: A fine granular hierarchical multi-label text classification model. *Neural Processing Letters*, 57(1):1.

Jun Zhang, Yubin Li, Fanfan Shen, Chenxi Xia, Hai Tan, and Yanxiang He. 2024b. Hierarchy-aware and label balanced model for hierarchical text classification. *Knowledge-Based Systems*, 300:112153.

Xinyi Zhang, Jiahao Xu, Charlie Soh, and Lihui Chen. 2022. La-hcn: Label-based attention for hierarchical multi-label text classification neural network. *Expert Systems with Applications*, 187:115922.

Jie Zhou, Chunping Ma, Dingkun Long, Guangwei Xu, Ning Ding, Haoyu Zhang, Pengjun Xie, and Gongshen Liu. 2020. Hierarchy-aware global model for hierarchical text classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1106–1117, Online. Association for Computational Linguistics.

Juncheng Zhou, Lijuan Zhang, Yachen He, Rongli Fan, Lei Zhang, and Jian Wan. 2025. A novel negative sample generation method for contrastive learning in hierarchical text classification. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 5645–5655, Abu Dhabi, UAE. Association for Computational Linguistics.

He Zhu, Junran Wu, Ruomei Liu, Yue Hou, Ze Yuan, Shangzhe Li, Yicheng Pan, and Ke Xu. 2024. HILL: Hierarchy-aware information lossless contrastive learning for hierarchical text classification. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4731–4745, Mexico City, Mexico. Association for Computational Linguistics.

He Zhu, Chong Zhang, Junjie Huang, Junran Wu, and Ke Xu. 2023. HiTIN: Hierarchy-aware tree isomorphism network for hierarchical text classification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7809–7821, Toronto, Canada. Association for Computational Linguistics.

A Example Appendix

We used the following prompt to generate the path descriptions using GPT-4o.

A.1 Prompt

```
Abstract:
"{abstract}"

Hierarchical Label Path:
{label_path_str}

Task:
Generate a detailed description that:
- Defines each label in the hierarchy and their
    relationships.
- Explains how the abstract content corresponds
    to each label.
- Avoids introducing external or unrelated
    information.
- Uses professional and academic language.
- Is approximately 200 words in length.
```