Journal of Computer Technology and Software
ISSN:2998-2383
Vol. 4, No. 11, 2025

A Hierarchical Encoding and Selective Attention Enhancement Framework Based on Large Language Models for Long-Document Text Classification

Zhihao Wang[1,*]
[1]Acadia University, Wolfville, Canada
*Corresponding author: Zhihao Wang; wangzhihaocom@gmail.com

Abstract: Long document text classification faces challenges such as long text spans, complex chapter structures, scattered key information, and significant interference from redundant content. Traditional methods often struggle to balance global context preservation and local key clue extraction when handling long-distance semantic dependencies and integrating multi-granularity information, thus affecting the accuracy and stability of classification. To address these issues, this paper proposes a hierarchical encoding and selective attention enhancement framework based on a large language model for long document text classification. This method first hierarchically divides the text according to the natural organization of long documents and learns segment-level semantic representations using a shared encoder. A position-aware mechanism is then introduced to preserve the internal sequential relationships and chapter structure features of the document. Subsequently, a document-level context interaction module models cross-segment dependencies and global associations between different text segments, enabling the effective transmission and integration of long-distance semantic information in a unified representation space. Furthermore, a selective attention enhancement module assigns importance to contextual representations, highlighting core content closely related to category determination and suppressing the interference of redundant information and weakly related semantics on the final decision, thus forming a more compact and discriminative document-level representation. Finally, the classification layer predicts the category based on the aggregated high-quality semantic representation. This paper proposes a method that leverages the hierarchical structure and information distribution characteristics of long documents. It organically combines hierarchical semantic modeling, cross-segment contextual interaction, and key information filtering to enhance the model's ability to understand and classify complex texts. Results show that this framework is well-suited for long document text classification scenarios, improving classification performance while maintaining structural semantic integrity, and providing an effective technical path for intelligent analysis of complex texts.

Keywords: Text semantic modeling; long-range dependency capture; segment-level information aggregation; text category discrimination

1. Introduction

With the continuous advancement of digitalization, long text data such as legal documents, financial reports, medical records, policy documents, academic papers, and industrial technical documents are experiencing rapid growth[1,2]. Compared to short texts, long documents typically contain richer thematic layers, more complex logical structures, and broader contextual dependencies, enabling them to carry more complete semantic information and decision-making basis. Long document text classification, as an important task in the field of natural language processing, aims to accurately identify the category of lengthy and structurally complex texts, providing fundamental support for applications such as information retrieval, intelligent review, risk warning, knowledge management, and decision support. Therefore, how to achieve effective semantic modeling and category determination in long-spanning contexts has become one of the key issues driving the improvement of intelligent text analysis capabilities[3].

However, long document text classification still faces significant challenges. First, the increasing document length makes it difficult for traditional text representation methods to fully preserve global semantics and key local information, easily leading to problems such as context truncation, information dilution, and semantic drift[4]. Second, long documents often have a naturally hierarchical organizational structure; different paragraphs, sentences, and words contribute differently to classification decisions, but many existing methods struggle to simultaneously address multi-level semantic aggregation and key content identification. Furthermore, long documents often contain a large amount of information that is weakly related to or even irrelevant to category determination. This redundant content can interfere with the model's capture of core semantic cues, reducing the discriminative power and stability of the classification process. In this context, relying solely on shallow representations or single-granularity modeling strategies is insufficient to meet the dual requirements of accuracy and robustness in complex long document scenarios[5].

The development of large language models has provided a new technical path for long document text understanding. With their powerful semantic modeling capabilities, contextual association capabilities, and knowledge transfer capabilities, large language models have demonstrated excellent performance in various natural language processing tasks, creating important conditions for overcoming the bottleneck of long text modeling. However, directly applying general-purpose large language models to long document classification tasks still has certain limitations. On the one hand, long document input increases computational complexity and information interaction burden, making it difficult for the model to efficiently focus on the core content that truly affects category determination[6]. On the other hand, without explicit modeling of the document's hierarchical structure, although the model possesses strong semantic understanding capabilities, it may still have shortcomings in long-range dependency integration, textual logical organization, and key information filtering. Therefore, designing more targeted, structured modeling mechanisms around the features of long documents has become an important direction for improving the task adaptability of large language models.

Based on this, constructing a hierarchical coding and selective attention enhancement framework for large language models targeting long document text classification tasks has clear theoretical value and practical significance. From a theoretical perspective, this direction helps to advance large language models from general semantic understanding to structured discourse modeling, strengthening their ability to uniformly characterize multi-granularity information interaction, hierarchical semantic aggregation, and key evidence extraction mechanisms[7]. From an application perspective, this direction can better adapt to the ultra-long texts, complex structured texts, and high information density texts commonly found in real-world scenarios, providing more reliable technical support for fields such as intelligent justice, financial supervision, medical information analysis, government governance, and enterprise knowledge services. Research on hierarchical coding and selective attention enhancement not only helps improve the accuracy and interpretability of long document classification but also has significant implications for promoting the in-depth application of large language models in the intelligent processing of complex texts.

Recent work on large-language-model-based retrieval and structured semantic modeling also provides useful context for long-document classification. Keyword-constrained retrieval-augmented generation demonstrates that external constraints can guide large language models toward more task-relevant evidence selection in knowledge-intensive settings[8]. Knowledge graph-augmented semantic fusion for long-text classification further indicates that structured external relations can enrich document-level representations when category cues are distributed across distant passages[9]. Frequent subgraph mining with structural prompt injection and knowledge-graph-enhanced sentiment classification both emphasize the value of explicit relational structures for adapting language models to specialized classification tasks[10,11].

Parameter-efficient and task-aware adaptation is another important direction for improving long-document classification. Task-aware regularized continual fine-tuning addresses the risk of catastrophic forgetting when large language models are adapted across changing tasks[12], while instruction-driven classification with cross-attention fusion highlights the importance of aligning task instructions, semantic features, and category decisions[13]. Retrieval and re-ranking optimization, fine-grained semantic representation learning, and hierarchical prompt learning with label-dependency modeling provide additional references for combining evidence selection, representation refinement, and structured decision constraints in complex classification scenarios[14,15,16].

2. Methodology

Long-document classification is formulated by considering a document $\mathcal{D} = \{s_1, s_2, \ldots, s_M\}$ as an ordered sequence of semantic units, where each unit s_i may correspond to a sentence, clause group, or discourse segment after length-aware partitioning, because preserving the original rhetorical organization is essential for reducing information fragmentation in extended contexts.

$$\mathcal{D} = \{s_1, s_2, \ldots, s_M\}$$

Rather than feeding the entire document into the backbone language model in a single pass, the proposed framework first constructs a hierarchical view that separates local semantic compression from global discourse aggregation, so that salient evidence can be retained while computational burden remains controllable under long-context settings.

$$\boldsymbol{h}_i = LLM_{enc}\left(s_i\right)$$

Here $\boldsymbol{h}_i \in \mathbb{R}^d$ denotes the segment-level representation generated by the shared encoder for the ith unit, and the use of parameter sharing across all segments encourages a consistent semantic space in which category-relevant clues from different positions can be compared and fused more reliably. This paper also presents the overall model architecture, as shown in Figure 1.

Beyond simple segment encoding, positional dependence across the document is injected through a hierarchical transition operator that couples semantic content with discourse order, thereby enabling the model to capture not only what is stated in each segment but also how the argument unfolds across distant textual regions.

$$\widetilde{\boldsymbol{h}}_i = \boldsymbol{h}_i + \boldsymbol{p}_i$$

Within this formulation, $\boldsymbol{p}_i$ represents the learnable structural position signal associated with segment index i, and its addition to $\boldsymbol{h}_i$ provides an explicit mechanism for

retaining long-range organizational patterns that are often decisive in legal, financial, and policy-oriented documents.

Subsequently, the sequence $\{\tilde{\boldsymbol{h}}_1, \tilde{\boldsymbol{h}}_2, \ldots, \tilde{\boldsymbol{h}}_M\}$ is passed to a document-level interaction layer, where contextual dependencies among distant segments are modeled to alleviate the semantic isolation introduced by local chunking and to reconstruct cross-segment coherence at the discourse scale.

$$\boldsymbol{z}_i = \sum_{j=1}^{M} \alpha_{ij} \boldsymbol{W}_v \tilde{\boldsymbol{h}}_j$$

Inside this aggregation process, α_{ij} measures the contribution of segment j to the contextualized representation of segment i, while $\mathbf{W}_v$ projects segment features into a value space that is more suitable for inter-segment information transfer. Unlike uniform fusion strategies that treat all contextual evidence equally, the interaction mechanism emphasizes semantically complementary regions and suppresses loosely related content, which is especially meaningful for long documents whose discriminative signals are usually sparse and unevenly distributed.

This cross-segment interaction layer can also be viewed as a structured semantic fusion process. Knowledge-graph semantic fusion and fine-grained domain representation learning support the idea that document-level features should not be aggregated only by surface proximity, but should also reflect latent semantic relations among evidence-bearing segments[9,15].

$$\alpha_{ij} = \frac{exp\left(\left(\boldsymbol{W}_q \tilde{\boldsymbol{h}}_i\right)^{\top} \left(\boldsymbol{W}_k \tilde{\boldsymbol{h}}_j\right) / \sqrt{d}\right)}{\sum_{t=1}^{M} exp\left(\left(\boldsymbol{W}_q \tilde{\boldsymbol{h}}_i\right)^{\top} \left(\boldsymbol{W}_k \tilde{\boldsymbol{h}}_t\right) / \sqrt{d}\right)}$$

Through the query and key transformations $\boldsymbol{W}_q$ and $\boldsymbol{W}_k$, the model dynamically estimates which segments should interact more strongly, so the resulting discourse representation becomes sensitive to both semantic affinity and structural relevance instead of relying on fixed-range dependency assumptions.

Meanwhile, selective attention enhancement is introduced to identify evidence-bearing segments before final decision aggregation, because long documents often contain descriptive, procedural, or background passages whose excessive participation may obscure category boundaries. A relevance estimator, therefore, assigns an adaptive importance score to each contextualized segment, allowing the framework to preserve classification-critical information while reducing interference from redundant text.

$$\beta_i = \frac{exp\left(\boldsymbol{u}^{\top} tanh\left(\boldsymbol{W}_s \boldsymbol{z}_i + \boldsymbol{b}_s\right)\right)}{\sum_{t=1}^{M} exp\left(\boldsymbol{u}^{\top} tanh\left(\boldsymbol{W}_s \boldsymbol{z}_t + \boldsymbol{b}_s\right)\right)}$$

Under this design, the trainable vector $\boldsymbol{u}$ and projection parameters $\boldsymbol{W}_s, \boldsymbol{b}_s$ define a content-sensitive scoring function whose objective is not merely compression, but the extraction of semantically concentrated evidence aligned with the document label. From a modeling perspective, such selective weighting increases interpretive consistency, since the final document representation is constructed from a principled distribution over informative segments rather than from an indiscriminate average of all hidden states.

$$\boldsymbol{g} = \sum_{i=1}^{M} \beta_i \boldsymbol{z}_i$$

By letting $\boldsymbol{g} \in \mathbb{R}^d$ summarize the document under attention weights β_i, the framework forms a compact yet semantically focused representation that is more robust to length variation and topic digression than conventional pooling operations.

The selective attention pathway also draws on ideas from constrained retrieval and adaptive re-ranking. Keyword-constrained retrieval and dual retrieval with adaptive re-ranking both suggest that candidate evidence should be filtered and prioritized before final generation or classification decisions are made[8,14].

Finally, category prediction is produced from the document representation through a discriminative classifier, where optimization is guided by the objective of aligning hierarchical semantics, cross-segment dependencies, and evidence selectivity within a unified end-to-end learning process.

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log \widehat{y}_c$$

In this objective, $\widehat{y}_c$ denotes the predicted probability for class c and y_c is the corresponding one-hot supervision signal, so minimizing $\mathcal{L}$ encourages the model to assign higher confidence to the correct category while jointly refining the upstream hierarchical encoder and the selective attention pathway. Overall, the proposed architecture strengthens long-document classification by connecting local semantic encoding, global discourse interaction, and relevance-aware evidence extraction in a coherent manner, which is particularly valuable for tasks requiring both contextual completeness and precise identification of decisive textual cues.

For task adaptation, regularized continual fine-tuning, instruction-driven cross-attention fusion, and hierarchical prompt learning with label dependencies provide methodological references for keeping the classifier aligned with task definitions while preserving robust representations across categories[12,13,16].

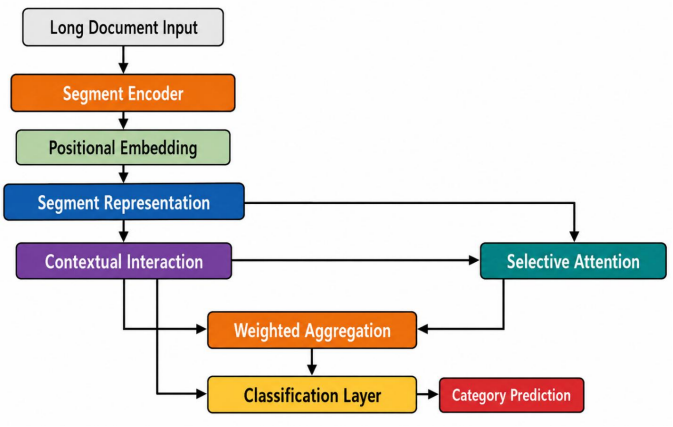

Figure 1. Overall architecture of the proposed hierarchical encoding and selective attention enhanced framework for long document classification, which progressively transforms segmented document inputs into category predictions through positional modeling, contextual interaction, and relevance-aware aggregation.

3. Experimental Results and Analysis

3.1 Dataset Introduction

This paper selects the SemEval 2019 Hyperpartisan News Detection open-source dataset as the data foundation for the long document text classification task. This dataset is designed for binary classification tasks at the news article level, aiming to determine whether a news article contains a clear bias based on its overall content. Therefore, its task format is highly compatible with the long document classification scenario. The data resources include manually annotated article-level corpora and a large-scale weakly annotated corpus. The article-level data contains 1,273 news texts, while the large-scale corpus comprises approximately 754,000 articles. The text units are all complete news articles rather than isolated sentences or short fragments, which allows them to effectively support the learning of hierarchical encoding, cross-segment semantic modeling, and selective attention mechanisms for long-distance contextual dependencies and key information filtering. Furthermore, this dataset has been released through public tasks and open data platforms, possessing clear open access attributes, high reproducibility, and a broad research base, making it suitable as a validation dataset for large language model methods for long document text classification tasks.

3.2 Experimental Results and Analysis

To compare with the hierarchical encoding and selective attention enhancement framework for large language models in long document text classification proposed in this paper, representative recent works on long document classification, hierarchical modeling, long context representation, and efficient attention mechanisms were selected as control methods. The experimental results are shown in Table 1. These methods cover techniques such as hierarchical attention networks, local convolutional aggregation, hierarchical neural modeling, interactive hierarchical Transformers, long document Transformer comparison frameworks, state-space long document modeling, and length-aware Transformers, thus comprehensively reflecting the typical research trajectory in the current field of long document text classification. Finally, the large language model used in this paper is Bert.

Table 1. Experimental results compared with other models

Method	Acc	Prec	Rec	F1
Yang et al.[17]	88.14	87.52	86.93	87.22
Liu et al.[18]	88.67	88.01	87.45	87.73
Khandve et al.[19]	89.13	88.56	88.02	88.28
Wu et al.[20]	90.04	89.48	88.91	89.19
Dai et al.[21]	90.37	89.92	89.36	89.64
Lu et al.[22]	90.88	90.15	89.84	89.99
Han et al.[23]	91.26	90.74	90.31	90.52
Ours	92.14	91.63	91.08	91.35

Overall, the proposed hierarchical encoding and selective attention enhancement framework demonstrates more stable and superior discriminative ability in long document text classification tasks. This method effectively enhances the focus and integration of key semantic content while preserving the hierarchical structure information of the document, thereby improving the accuracy and consistency of category recognition. Since long documents often contain a large amount of redundant information and significant contextual dependencies, traditional methods are prone to an imbalance between global semantic aggregation and local key information extraction. Our proposed method, by introducing hierarchical representation learning and selective attention mechanisms, better mitigates the discriminative bias caused by information interference, making the final classification results more reliable. This further demonstrates the strong applicability of this framework in complex long-text understanding scenarios.

To verify the actual role of each core component in the long document text classification process, ablation analysis was further conducted around three key aspects: hierarchical coding, document-level contextual interaction, and selective attention enhancement. The relevant results are shown in Table 2. This design can more clearly reveal the impact of different structural units on document semantic organization, long-distance dependency modeling, and key information filtering capabilities, thus demonstrating the necessity of the overall framework in maintaining hierarchical structural information and improving discriminative focusing capabilities.

Table 2. Ablation study results of the proposed framework

Setting	Acc	Prec	Rec	F1
w/o Hierarchical Encoding	90.92	90.33	89.74	90.03
w/o Contextual Interaction	91.18	90.61	90.08	90.34

w/o Selective Attention	91.47	90.95	90.42	90.68
Ours	92.14	91.63	91.08	91.35

The framework proposed in this paper exhibits good integrity and coordination, with each component playing an irreplaceable role in the long document text classification process. Hierarchical encoding can more fully preserve the hierarchical structure and multi-granular semantic information within the document, while the contextual interaction mechanism helps strengthen the association modeling between different text fragments, and selective attention further enhances the model's ability to identify and focus on key information. Through the combined effect of these modules, the constructed method can more effectively alleviate problems such as redundant content interference, large semantic spans, and scattered distribution of important clues in long texts, thereby making classification more accurate, stable, and possessing stronger overall adaptability. This also demonstrates that the proposed method has good methodological effectiveness and structural rationality in long document understanding tasks.

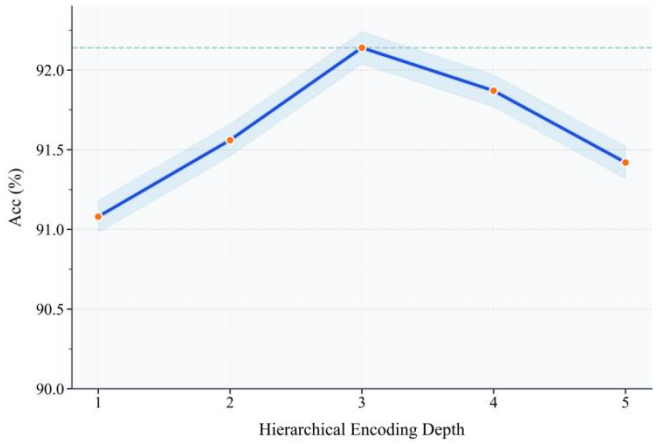

Figure 2. The impact of layered coding depth on Acc

Figure 2 demonstrates that the proposed algorithm maintains good classification performance under the hierarchical encoding depth setting, indicating its strong adaptability to modeling the hierarchical semantics within long documents. The hierarchical encoding mechanism enables the model to organize textual information of different granularities more hierarchically and retain more crucial semantic cues in long-distance contexts, thereby enhancing overall discriminative ability. Simultaneously, the selective attention and contextual interaction modules work together in the document representation learning process, allowing for a more comprehensive aggregation of important content, further demonstrating the effectiveness and rationality of the proposed method in complex long text classification tasks.

4. Conclusion

This paper addresses common challenges in long document text classification tasks, such as large context spans, complex text hierarchies, scattered distribution of key information, and severe interference from redundant content. It constructs a hierarchical encoding and selective attention enhancement framework for large language models. Starting from the structural features of long documents, this method organically combines local semantic modeling, cross-segment contextual interaction, and key information filtering. While maintaining the overall semantic continuity of the document, it further enhances the extraction of core discriminative cues. Compared to modeling methods relying solely on single-layer representations or uniform aggregation strategies, this framework places greater emphasis on the organizational relationships and contribution differences among multi-granular information within long texts, thus providing a more reasonable modeling path for the semantic understanding of complex texts. The results demonstrate that collaborative design around the structural attributes and semantic selection mechanisms of long documents can effectively improve representation quality and category discrimination capabilities in text classification tasks, and also provide targeted ideas for adapting large language models to long contextual understanding scenarios.

From a methodological perspective, the significance of this work lies not only in improving the performance of long document classification but also in further expanding the paradigm of intelligent long text processing. The hierarchical encoding mechanism emphasizes the fundamental role of text organization structure in semantic modeling, enabling the model to progressively compress and integrate information at different levels. The selective attention mechanism highlights the core value of key evidence in the final decision-making process, allowing the model to more effectively suppress irrelevant content interference and focus on retaining high-value semantic segments when faced with lengthy texts. The synergistic effect of these two mechanisms enhances the model's depth of understanding of complex texts and the stability of its decision-making, making the framework highly applicable in scenarios such as news content analysis, financial report recognition, medical document classification, educational resource organization, enterprise knowledge management, policy text processing, and academic literature organization. As industry data continues to grow in scale and text formats become increasingly complex, long document classification methods with hierarchical awareness and focus capabilities will play an increasingly important role in improving information processing efficiency, optimizing knowledge discovery quality, and supporting intelligent decision-making.

Further improvements can be made in the breadth, depth, and generalization capabilities of long document semantic modeling. On the one hand, by further integrating more refined text structure information, paragraph relationship information, and task prior knowledge, the model's ability to characterize complex logical chains, implicit topic transfer, and cross-paragraph dependency patterns can be enhanced, thereby improving its adaptability in longer texts and noisier scenarios. On the other hand, the integration of this framework

with stronger long-context modeling strategies, lightweight inference mechanisms, and multi-task learning paradigms can be further explored to balance semantic expressiveness, computational efficiency, and cross-domain transferability. Simultaneously, for real-world applications, this method can be continuously extended to areas with higher interpretability, stronger robustness, and broader coverage, enabling it not only to serve standard text classification tasks but also to provide more valuable technical support for scenarios such as complex document understanding, intelligent content moderation, risk identification, and knowledge services.

References

[1] D. Premasiri, T. Ranasinghe, and R. Mitkov, "Can model fusing help transformers in long document classification? an empirical study," in Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing, 2023, pp. 871-878.

[2] T. Czinczoll, C. Hönes, M. Schall, et al., "NextLevelBERT: masked language modeling with higher-level representations for long documents," in Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2024, pp. 4656-4666.

[3] R. A. A. Principe, N. Chiarini, and M. Viviani, "An LCF-IDF document representation model applied to long document classification," in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, 2024, pp. 1129-1135.

[4] S. S. Roy, X. Wang, R. Mercer, et al., "Graph-Tree Fusion Model with Bidirectional Information Propagation for Long Document Classification," in Findings of the Association for Computational Linguistics: 2024 Conference on Empirical Methods in Natural Language Processing, 2024, pp. 4460-4470.

[5] T. Liu, Y. Hu, J. Gao, et al., "Hierarchical Multi-modal Transformer for Cross-modal Long Document Classification," IEEE Transactions on Multimedia, 2025.

[6] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, et al., "Text classification algorithms: A survey," Information, vol. 10, no. 4, p. 150, 2019.

[7] P. Kokate, M. Sarnaik, M. Khopade, et al., "Efficient Zero-Shot Long Document Classification by Reducing Context Through Sentence Ranking," arXiv preprint arXiv:2508.17490, 2025.

[8] W. Ma and W. Zhong, "Keyword-Constrained Retrieval-Augmented Generation for Large Language Models," 2025.

[9] Z. Zhu, "Knowledge Graph-Augmented Semantic Fusion for Large Language Model-Based Long-Text Classification," 2024.

[10] Z. Xiao, C. Moretti, and Q. Li, "Frequent Subgraph Mining and Structural Prompt Injection for Large Language Model Fine-Tuning," 2025.

[11] Z. Gu and L. Nie, "Knowledge Graph-Enhanced Large Language Models for Opinion Target Sentiment Classification," 2025.

[12] J. Hu, Y. Lyu, and J. Jiang, "Task-Aware Regularized Continual Fine-Tuning for Mitigating Catastrophic Forgetting in Large Language Models," 2025.

[13] Y. Jiang and Y. L. Peng, "Instruction-Driven Language Model for Text Classification via Cross-Attention Fusion and Semantic Alignment," 2024.

[14] L. Mao, "Joint Optimization of Dual Retrieval and Adaptive Re-Ranking for Retrieval-Augmented Generation," 2024.

[15] L. Ma, E. Lin, and Z. Zhang, "Intelligent Domain Text Classification through Fine-Grained Semantic Representation Learning," 2024.

[16] R. Long, M. Zhang, and A. Hartwell, "Multi-Label Text Classification with Large Language Models via Hierarchical Prompt Learning and Label Dependency Modeling," 2024.

[17] Z. Yang, D. Yang, C. Dyer, et al., "Hierarchical attention networks for document classification," in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 1480-1489.

[18] L. Liu, K. Liu, Z. Cong, et al., "Long length document classification by local convolutional feature aggregation," Algorithms, vol. 11, no. 8, p. 109, 2018.

[19] S. I. Khandve, V. K. Wagh, A. D. Wani, et al., "Hierarchical neural network approaches for long document classification," in Proceedings of the 14th International Conference on Machine Learning and Computing, 2022, pp. 115-119.

[20] C. Wu, F. Wu, T. Qi, et al., "Hi-Transformer: Hierarchical interactive transformer for efficient and effective long document modeling," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021, pp. 848-853.

[21] X. Dai, I. Chalkidis, S. Darkner, et al., "Revisiting transformer-based models for long document classification," in Findings of the Association for Computational Linguistics: 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 7212-7230.

[22] P. Lu, S. Wang, M. Rezagholizadeh, et al., "Efficient classification of long documents via state-space models," in Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 2023, pp. 6559-6565.

[23] G. Han, J. Tsao, and X. Huang, "Length-aware multi-kernel transformer for long document classification," in Proceedings of the 13th Joint Conference on Lexical and Computational Semantics, 2024, pp. 278-290.