Transactions on Computational and Scientific Methods | Vo. 4, No. 6, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Hierarchical Semantic-Structural Encoding for Compliance Risk Detection with LLMs

Ying Qin
University of Illinois at Urbana-Champaign, Champaign, USA
yingqinuiuc@gmail.com

Abstract: This paper addresses the challenges of structural complexity, semantic density, and task diversity in financial regulatory texts. It proposes a risk identification method based on large language models that integrates structure awareness and task adaptation. The method builds basic semantic representations using a pre-trained language model. It also introduces a hierarchical semantic-structural encoding mechanism to explicitly capture logical relationships among clause numbers, substructure hierarchies, and responsible entities in regulatory texts. A dynamic task adaptation module is incorporated to construct task-aware representations and multi-task branches. This allows the model to distinguish between various risk types, such as compliance gaps and responsibility conflicts. To evaluate the performance of the proposed method, a risk identification dataset based on real financial regulatory documents is constructed. Sensitivity experiments are conducted across several dimensions, including structural integrity disturbance, sampling ratio variation, and encoding depth change. The experimental results show that the method achieves high accuracy and robustness. At the same time, it demonstrates strong task adaptability and structural awareness. This provides effective technical support for complex semantic understanding and risk factor modeling in financial regulatory texts.

Keywords: Semantic structure modeling, task perception mechanism, financial compliance analysis, risk factor identification

1. Introduction

Financial regulation plays a central role in maintaining market order, preventing systemic risks, and protecting the rights of financial consumers in modern financial systems. As financial services become more complex and financial instruments more diverse, the structure of regulated entities grows increasingly intricate[1]. At the same time, the speed and scope of information dissemination have expanded exponentially. To address these challenges, regulatory authorities in various countries have continuously issued new regulations, guidelines, and policy documents. These documents cover banking, securities, insurance, payment systems, and other domains. They are often highly technical and mandatory, and their content is directly tied to the compliant operation and strategic planning of financial institutions. Accurately understanding, identifying, and responding to the potential risks embedded in these texts has become a key issue in both regulatory practice and financial operations[2].

Against this backdrop, financial regulatory documents exhibit a high-density textual nature[3]. They include large volumes of legal terminology, industry-specific norms, and contextual reasoning requirements. Traditional rule-based or shallow analytical models are increasingly showing limitations. They struggle with adaptability, scalability, and semantic understanding. This is especially evident when dealing with frequently updated document systems and multi-domain, multi-scale expression structures. Traditional methods often fail to extract critical risk points or to respond effectively to ambiguities, contradictions, or policy conflicts. The lack of such recognition capability hinders regulatory efficiency and presents major challenges for financial institutions in adjusting their policy responses[4].

In recent years, large language models have made breakthrough progress in natural language processing. They demonstrate strong abilities in language understanding, semantic reasoning, and text generation. These models provide new solutions for handling complex textual tasks[5]. In high-demand areas such as law, healthcare, and finance, large language models show stronger generalization and structural perception compared to earlier approaches. They can process documents with complex semantic layers and tightly connected logic. In the context of financial regulation, large language models have the potential to extract key elements from unstructured texts and to conduct semantic summarization and inference. This can improve the accuracy and efficiency of risk identification and support the intelligent transformation of regulatory response mechanisms[6].

Introducing large language models into the task of identifying risks in financial regulatory documents is not only a technological innovation but also a methodological advancement. This approach shifts the focus from static rule-based systems to dynamic, context-aware modeling techniques that are capable of understanding complex regulatory language. By building language modeling mechanisms with task awareness, models are able to align semantic representations with specific regulatory objectives. This allows for the precise identification of risk-related elements in the text, including potential violations of policy, areas where regulatory guidance is lacking, and operational challenges that may arise during implementation. These insights are especially important in documents where risk expressions are subtle, conditional, or context-dependent.

Furthermore, large language models possess strong knowledge transfer capabilities due to their pretraining on vast and diverse text corpora. This enables them to be efficiently fine-tuned for application across different financial subdomains, such as banking, insurance, and securities. As a result, they can support unified modeling for cross-domain risk perception, bridging semantic gaps between regulatory texts of varying styles and scopes. This flexibility enhances the responsiveness of risk identification systems to evolving policy landscapes. It also contributes to more forward-looking regulatory practices, reduces the dependency on manual review, and fosters greater automation in compliance workflows [7].

From the perspectives of financial system stability and regulatory modernization, exploring risk identification methods based on large language models is a meaningful enhancement to the current regulatory technology system. It represents a key step toward practical and scalable intelligent regulation. This research addresses technical bottlenecks and efficiency issues in current regulatory practice[8]. It also provides theoretical foundations and methodological tools for building high-trust and high-transparency regulatory technology platforms. In the context of accelerating financial digital transformation and improved digital regulatory infrastructure, research on risk identification based on large language models holds both theoretical and practical significance.

2. Related work

2.1 Large Language Model

Large language models have taken a central role in recent advances in natural language processing[9]. Through unsupervised pretraining on large-scale corpora, these models significantly enhance understanding of language semantics, contextual logic, and cross-sentence reasoning[10]. Compared to traditional methods based on feature engineering or shallow syntactic rules, large language models offer stronger capabilities in context modeling and language generation. They show superior performance across many complex language tasks. In particular, for tasks such as text classification, information extraction, and question answering, these models enable end-to-end learning. This reduces reliance on manual rules and improves adaptability and robustness. Their deep neural structures can capture long-range dependencies in language. They also allow for more accurate modeling of phenomena like word sense disambiguation and coreference resolution. This provides strong support for semantic understanding in various domains[11,12].

In financial applications, the semantic awareness of large language models is especially valuable. Financial texts are typically characterized by a high density of domain-specific terminology, complex syntactic constructions, and embedded logical relationships. These features often make the texts difficult to process using traditional models, which rely heavily on fixed patterns or shallow representations. As a result, such models frequently encounter challenges such as information loss, semantic ambiguity, and misinterpretation when applied to financial documents [13,14]. In contrast, large language models are equipped with the ability to capture long-range dependencies and contextual meaning across multiple levels of abstraction. By encoding context globally and generating dynamic, task-relevant representations, they are able to detect subtle semantic nuances and understand the logical progression between clauses and entities. This supports a more accurate and fine-grained interpretation of complex financial content.

Moreover, these capabilities enable large language models to automatically extract structured elements from unstructured financial texts, such as named entities, regulatory clauses, and responsibility assignments. They also provide an effective means of identifying latent risk signals, policy inconsistencies, and semantic anomalies that may not be explicitly stated [15]. Beyond single-domain processing, large language models can leverage their pretraining on diverse textual sources to integrate cross-domain knowledge. With appropriate fine-tuning, they can be adapted to the unique linguistic and structural characteristics of financial regulatory language. This enhances their generalization performance when faced with domain-specific expressions, policy variations, and regulatory updates in complex financial texts.

Despite the strong performance of large language models across tasks, their controllability and interpretability remain key concerns in high-risk and high-stakes scenarios. In financial regulation, the text content often involves policy implementation, risk assessment, and legal compliance. Model outputs must be accurate and grounded in clear reasoning. Traceable semantic paths are also necessary[16,17]. Therefore, enhancing the structural transparency and reasoning explainability of large language models is a critical direction for future development. At the same time, regulatory documents usually have loose structures and high information density. Domain-specific training is needed to improve the model's sensitivity to policy language. This is essential for advancing real-world applications in regulatory settings.

2.2 Risk Identification of Financial Regulatory Documents

The task of risk identification in financial regulatory documents has long been a critical component of compliance and regulatory technology. Its goal is to automatically detect potential risk indicators, execution difficulties, logical conflicts, or policy gaps from normative texts. The complexity of this task lies not only in the density of terminology and semantic depth but also in the extensive need for cross-sentence reasoning and contextual dependency[18,19]. Early approaches often relied on manually crafted rules or keyword-matching techniques. However, such methods struggle to cover the diverse expressions in policy texts. When dealing with loosely structured or ambiguous content, they are prone to false positives and false negatives. This makes it difficult to meet real-world demands for high accuracy and high coverage. With the development of natural language processing, some studies have explored machine learning models to capture more complex semantic patterns. Yet, due to limitations in feature representation, these methods still face challenges in handling domain-specific language variations and dynamic contexts[20].

In recent years, researchers have recognized that risk identification in regulatory documents is a task with a high semantic load and strong structural dependencies. This has led to strategies that combine semantic modeling with structural analysis[21]. One line of work introduces domain dictionaries, dependency syntax, or semantic graphs to enhance language understanding and improve sensitivity to implicit risks. These methods improve the recognition of term relations and directive semantics. However, they rely heavily on predefined domain knowledge and rule design. As a result, they are less effective in adapting to rapid changes in policy language and handling multiple tasks simultaneously. Moreover, regulatory texts often include non-explicit expressions of risk, such as conditional statements, hypothetical clauses, and multi-agent constraints. These add semantic ambiguity and increase modeling complexity[22].

To address these challenges, current research is shifting toward multi-task modeling, structure-aware learning, and deep semantic reasoning. The aim is to build risk identification systems with better generalization and adaptability. By using joint task modeling, different types of risk signals can be detected and associated in a shared representation space. This improves both the coverage and consistency of identification. Some studies have also explored encoding structural cues from text, such as paragraph hierarchy, directive relationships, and responsible entities. This helps enhance the model's understanding of logical constraints and institutional context. In this setting, building a mechanism that integrates semantic representation, structural modeling, and domain adaptation has become a key approach for automating the analysis of financial regulatory documents. It also lays a solid foundation for the refinement and intelligence of regulatory technology.

3. Method

This study proposes a structure-aware large language model framework (Structure-Aware Large Language Model Framework, SA-LLMF) for risk identification in financial regulatory documents. The goal is to effectively model the complex logical structures and deep semantic dependencies in policy texts. The framework introduces two key innovations. First, it designs a hierarchical semantic-structural encoding mechanism (Hierarchical Semantic-Structural Encoding, HSSE) tailored to the regulatory text. This mechanism explicitly integrates structural cues such as clause numbering, logical conditions, and responsible entities with contextual semantics. It enhances the model's ability to represent non-linear risk paths. Second, it introduces a dynamic task adaptation module (Dynamic Task Adaptation Module, DTAM). This module uses a multi-task perception approach to distinguish and model different types of risks. It improves the model's ability to detect vague expressions, implicit logical relations, and cross-paragraph risk factors. Together, these components address the limitations of existing methods in handling structural heterogeneity and semantic uncertainty. The detailed structure of the proposed model is illustrated in Figure 1.

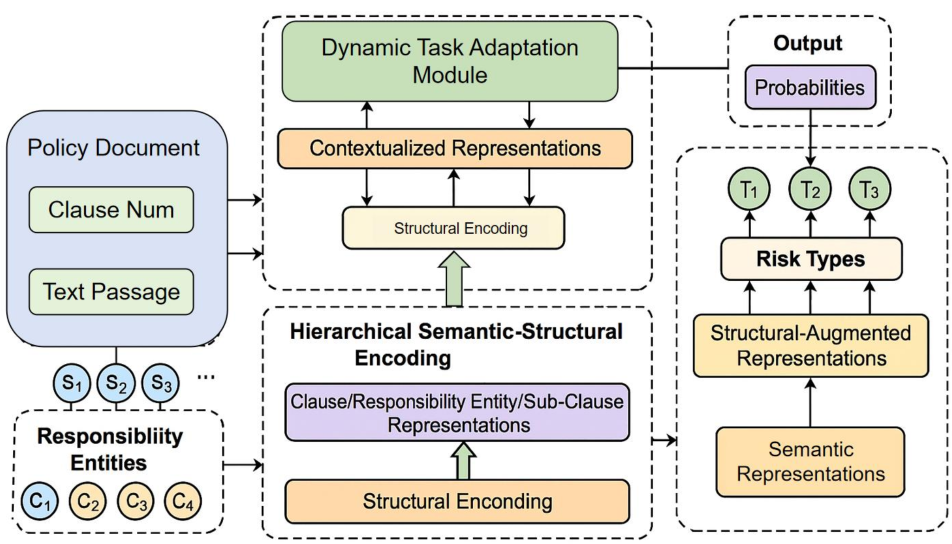

Figure 1. Overall model architecture diagram

3.1 Hierarchical Semantic-Structural Encoding

In this method, Hierarchical Semantic-Structural Encoding aims to perform multi-level fusion modeling of structural information and semantic content in financial regulatory documents. Specifically, the module is designed to encode clause-level, sub-clause-level, and entity-level structural cues — such as regulatory numbering, logical nesting, and responsibility attribution—and integrate them with the contextual semantic representations generated by the base language model. This fusion allows the model to jointly capture both the hierarchical organization and the semantic dependencies present in regulatory texts. The module architecture, which includes components for structural embedding, semantic alignment, and hierarchical aggregation, is illustrated in Figure 2.

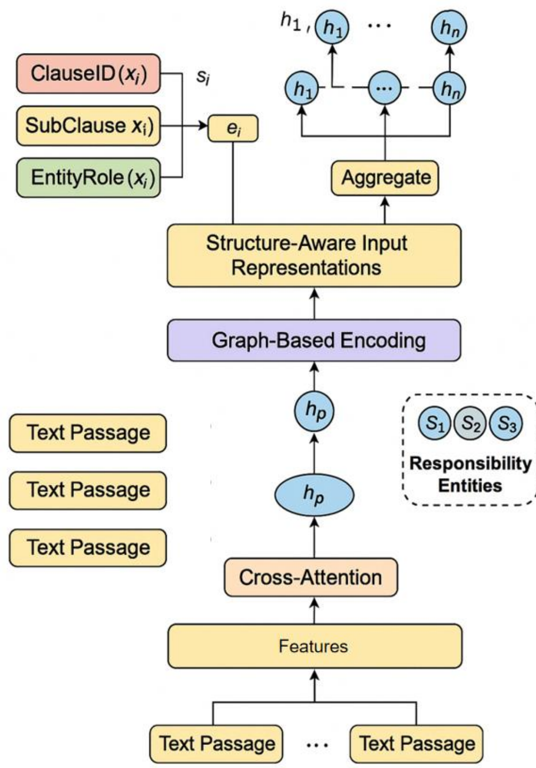

Figure 2. HSSE module architecture

Since regulatory documents usually have a clear clause numbering system, sub-clause division, and responsible entity labeling, we first introduce explicit structure embedding to construct a structure vector s_i for each text unit x_i, which includes information such as clause position, substructure identification, and entity role:

$$s_i = [ClauseID(x_i); SubClauseID(x_i); EntityRole(x_i)]$$

This structure vector will be concatenated with the original word vector e_i and fed into the encoder to form a structure-aware input representation $h_i^{(0)}$:

$$h_i^{(0)} = Concat(e_i, s_i)$$

In order to handle the dependencies between multi-level structural units in the document, this paper introduces a graph-based encoding mechanism to construct a connection graph $G = (V, E)$ between clauses, where V represents each clause or responsible entity node and E represents its structure or logical association. On this basis, the structural attention mechanism is used to update the representation of each node:

$$h_i^{(l+1)} = \sigma\Big(\sum_{j \in N(i)} a_{ij}^{(l)} W^{(l)} h_j^{(l)} \Big)$$

Where $a_{ij}^{(l)}$ is the attention weight calculated based on the structural relationship, $W^{(l)}$ is the learnable weight matrix of the lth layer, and σ is the nonlinear activation function.

When constructing cross-level representations, an aggregation function is used to aggregate the sub-clause representations to form the parent clause representation, thereby supporting global modeling across structures. Suppose a clause consists of its m sub-clauses, which can be represented as:

$$\widehat{h}_p = Aggregate(\{h_{p_1}, h_{p_2}, ..., h_{p_m}\})$$

The Aggregate function can be selected as mean pooling, weighted summation, or gated fusion mechanism to meet the modeling requirements of different structure-dependent scenarios.

Finally, to combine local semantic details with global structural context, we introduce a cross-layer attention mechanism to interactively encode the clause-level representation and the responsible entity-level representation to obtain the fused semantic structure joint representation z_i:

$$z_i = CrossAttn(h_i^{clause}, h_j^{entity}), \forall (i, j) \in P$$

P represents the set of all structural pairs, and CrossAttn represents the cross-type attention calculation, which is used to capture the deep semantic relationship between regulatory instructions and responsibility attribution. This joint representation will be used as the input of the subsequent task perception module for further risk-type modeling and output.

3.2 Dynamic Task Adaptation Module

Dynamic Task Adaptation Module (DTAM) aims to dynamically adjust the shared semantic structure representation and perform task-specific modeling based on the feature differences of different risk identification subtasks. Its module architecture is shown in Figure 3.

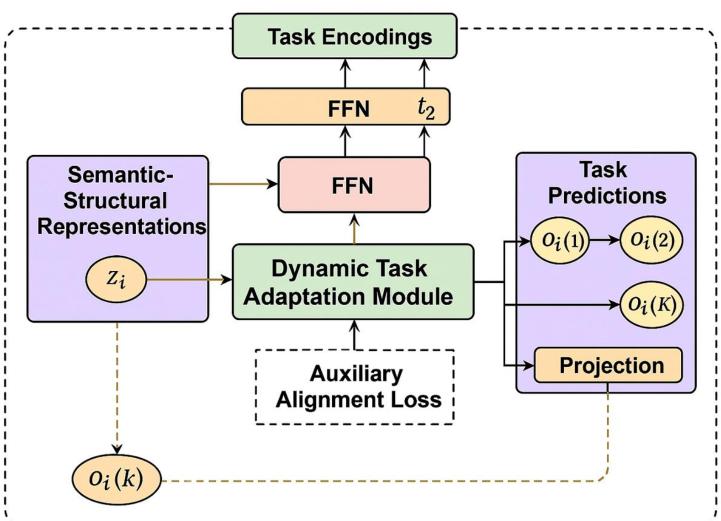

Figure 3. DTAM module architecture

In order to achieve task-sensitive representation transformation, we first introduce the task encoding vector t_k to represent the semantic features of the k-th risk subtask. For any input joint representation z_i, it is fused with the task encoding to generate a task-adaptive representation:

$$r_i^{(k)} = FFN([z_i; t_k; z_i \otimes t_k])$$

Where $[\cdot]$ represents the vector concatenation operation, $\otimes$ is the element-wise product, and $FFN()$ is the feedforward neural network structure, which is used to learn the nonlinear interactions between tasks.

To enhance the module's ability to jointly model multiple tasks, DTAM introduces a gated routing mechanism to control the degree of response of different task branches to representations through the task attention score $\beta_i^{(k)}$:

$$\beta_i^{(k)} = \frac{\exp(W_t^{(k)} z_i)}{\sum_{j=1}^{K} \exp(W_t^{(j)} z_i)}$$

Where $W_t^{(k)}$ is the trainable parameter matrix specific to each task, and K is the total number of tasks. This mechanism implements a soft selection of multi-task preferences, allowing the model to be moderately decoupled on a shared basis.

After the task-specific representation is constructed, DTAM maps the high-dimensional representation to the corresponding task space through the projection head to obtain the task prediction representation $o_i^{(k)}$ for subsequent risk judgment:

$$o_i^{(k)} = W_o^{(k)} r_i^{(k)} + b_o^{(k)}$$

$W_o^{(k)}$ and $b_o^{(k)}$ are the weight and bias parameters of task k, respectively, which are used to construct a task-specific linear output layer.

Considering the potential shared features and semantic overlap between tasks, DTAM further designs auxiliary consistency constraints to regulate the distribution of representations between tasks and minimize the KL divergence of output representations between different tasks, which is defined as follows:

$$L_{align} = \sum_{k1 \neq k2} KL(p^{(k_1)} \,||\, p^{(k_2)})$$

This regularization term encourages information interaction and alignment between multiple tasks, which helps to improve the consistency of global representation while preserving task differences, thereby enhancing the generalization ability and robustness of the risk identification framework.

4. Experimental Results

4.1 Dataset

This study uses the RegData US Financial Regulations Corpus as the primary data source. The dataset was developed by a U.S.-based regulatory technology research institution. It includes financial regulatory texts issued by multiple federal agencies, covering subfields such as banking, securities, insurance, and consumer finance. The dataset compiles rules and clauses from the 1970s to the present. The texts follow a standardized structure, making them well-suited for semantic modeling and structural analysis in risk identification tasks.

The dataset organizes regulatory content at the clause level. It includes structural metadata such as rule numbers, issuing agencies, publication dates, and regulated entities. Some documents also contain policy background notes and references to responsible parties. The dataset supports the modeling and extraction of multi-level structures, including clauses, sub-clauses, and entities. The language style is consistent and reflects the typical characteristics of financial legal texts. These include dense logic, concentrated terminology, and clear hierarchical nesting. The content accurately represents the complexity of financial regulatory language.

To meet the requirements of risk identification, the original RegData texts were preprocessed and cleaned. This process included removing formatting symbols, standardizing terminology, and splitting nested clause structures. The study selected financial subfields with frequent regulatory changes from the past decade (2012 – 2022) as the modeling subset. The original structural information was preserved for structure-aware representation. This dataset provides a reliable and realistic corpus foundation for semantic modeling and multi-task risk type classification in the proposed framework.

4.2 Experimental setup

In the experimental setup, this study uses BERT as the base language model and enhances it with structure-aware and task-adaptive components to build a complete risk identification framework. Specifically, BERT processes the preprocessed regulatory clause text. It uses a multi-layer Transformer architecture to generate contextual semantic representations. These representations are then fused with structural embedding features and serve as the input to downstream modules. The model adopts a dual-scale input strategy at the word level and clause level to balance detailed text information with structural semantic integrity.

To enhance the model's structure awareness, the experiment introduces a multi-layer structural encoding module and a graph modeling mechanism. These components model clause numbers, sub-clause relationships, and responsible entities in a structured format. The structured information is fused with BERT's semantic vectors through a cross-attention mechanism. In addition, to address the diversity of risk types, a multi-task modeling strategy is applied. Based on the structure-aware representations, a dynamic task adaptation module is connected to distinguish among different risk categories, such as compliance risks, execution blind spots, and responsibility conflicts. Each task has its linear head for prediction.

The training process uses the Adam optimizer. The initial learning rate is set to 2e-5. The batch size is 16, and the maximum number of training epochs is 20. To improve stability and generalization, dropout regularization is applied. An early stopping strategy is also used. The training data is divided by clause to ensure that each sample contains complete structural annotations. All experiments are conducted in a standard GPU environment to ensure the reproducibility and scalability of the model in real regulatory text scenarios.

4.3 Experimental Results

1) *Comparative experimental results*

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1: Comparative experimental results

Method	Accuracy	F1-Score	AUC
Transforemr[23]	82.6%	80.9%	86.3%

LSTM[24]	78.4%	75.1%	81.7%
1DCNN[25]	76.2%	72.8%	79.3%
RoBERTa[26]	84.1%	82.7%	88.1%
Ours	87.9%	85.4%	91.2%

In terms of overall performance comparison, the proposed structure-aware large language model significantly outperforms other baseline models on three core metrics: Accuracy, F1-Score, and AUC. This confirms the model's strong adaptability and generalization in the task of risk identification from financial regulatory documents. Notably, the model achieves an F1-Score of 85.4%, which reflects a good balance between precision and recall. This indicates that the model can not only accurately identify high-risk information but also effectively cover diverse types of risks, greatly improving recall performance in real-world applications.

Compared to traditional sequence models such as LSTM and 1D-CNN, the proposed method shows clear advantages in structural modeling. LSTM and CNN often fail to capture cross-sentence dependencies and hierarchical constraints among responsible entities in regulatory texts. This leads to incomplete extraction of risk information. By introducing hierarchical structure modeling at the clause, sub-clause, and entity levels, the proposed model better understands logical jumps and structural mappings in regulatory language. This results in significantly higher accuracy and AUC values.

When compared with mainstream pre-trained models such as Transformer and RoBERTaa, the advantage of the proposed model lies in its integration of a task-aware dynamic adaptation module. This enhancement improves robustness and task separation in multi-task risk classification scenarios. Although RoBERTaa performs well in semantic representation, its lack of explicit structural modeling limits its ability to capture structural cues such as nested clauses, logical hierarchies, and role mappings in complex regulatory texts.

It is worth noting that the improvement in AUC indicates a clearer boundary in distinguishing between positive and negative risk samples. This shows that the proposed model is more sensitive in identifying marginal risk instances. Such capability is critical in financial regulation, where many high-risk signals are expressed in subtle or implicit ways. Through the dual mechanism of structural enhancement and task adaptation, the model can extract high-risk signals from multi-layer structures and semantics. This significantly improves the overall practicality and decision-support capacity of the system.

2) Ablation Experiment Results

This paper further gives the results of the ablation experiment as shown in Table 2.

Table 2: Ablation Experiment Results

Method	**Accuracy**	**F1-Score**	**AUC**
Baseline	83.2%	80.1%	86.7%
+HSSE	85.3%	82.5%	88.6%
+DTAM	84.6%	81.8%	87.9%
Ours	87.9%	85.4%	91.2%

The results in the table show that when the model does not include structural modeling or task adaptation and relies only on basic BERT semantic modeling, its performance on Accuracy, F1-Score, and AUC is relatively weak. This indicates that pure language representation is not sufficient to handle the complex structural relations and risk semantics in financial regulatory texts. Although BERT has strong contextual awareness, it struggles to identify information and distinguish risks effectively when dealing with features like clause nesting and unclear responsibility references.

After introducing the Hierarchical Semantic-Structural Encoding module, the model shows significant improvement, especially with an increase of more than two percentage points in the F1 score. This demonstrates that structure awareness plays a crucial role in capturing the organizational patterns of elements in regulatory language. It enables the model to detect semantic dependencies across clauses and hierarchical levels. As a result, the extraction and boundary recognition of risk factors become more accurate. This improvement confirms that structural encoding is not just a supplement to text representation but a necessary component for modeling complex policy semantics.

When only the Dynamic Task Adaptation Module is added, the model performance also improves, mainly in the finer identification of risk types. The task-aware mechanism allows the model to dynamically adjust to the characteristics of different risk sub-tasks. This avoids overfitting in single-task settings or redundant representation. The module enhances the model's adaptability when handling multidimensional and uncertain risk expressions. The improvement in AUC further indicates stronger capability in boundary detection.

Finally, when both modules are integrated, the model achieves the best results across all three metrics. The joint modeling of structure and task provides complementary strengths. It improves the organization of semantic content and strengthens the model's ability to express task-specific objectives. These results validate the effectiveness of the proposed approach in handling multi-level structure representation and multi-task learning. It offers a viable path for addressing the high semantic density and task complexity in financial regulatory documents.

3) The impact of changes in the number of coding layers on risk identification accuracy

This paper also gives the impact of changing the number of encoding layers on the risk identification accuracy, and the experimental results are shown in Figure 4.

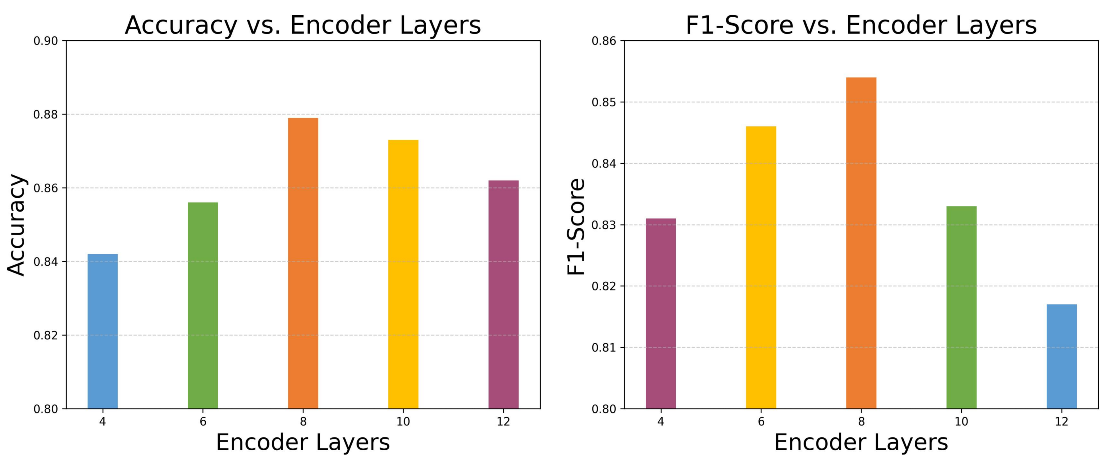

Figure 4. The impact of changes in the number of coding layers on risk identification accuracy

As shown in Figure 4, the number of encoding layers has a significant impact on the model's risk identification performance. The optimal points vary across different evaluation metrics. Specifically, when the number of encoding layers is set to 8, the model achieves the highest Accuracy. This suggests that a moderately deep structure better captures complex semantic associations in regulatory clauses. It also helps the model understand deeper relationships across clause hierarchies.

However, for the F1-Score, the optimal performance occurs at 6 layers. Beyond this point, increasing the number of layers leads to a decline in F1. This indicates that while deeper models can enhance semantic abstraction, they may also introduce redundant representations. This can reduce the model's generalization ability on boundary cases, affecting the balance between precision and recall. The effect is more noticeable in multi-task recognition scenarios, suggesting that model depth must match task complexity.

Overall, models with 4 and 12 layers perform poorly on both metrics. When the structure is too shallow, the model lacks sufficient understanding. When too deep, it may cause information dilution or overfitting. This is especially problematic for financial regulatory texts, which are loosely structured but logically dense. Too many layers may obscure directional relationships between key clauses, leading to missed or misclassified risk signals.

Therefore, these experimental results highlight the importance of structural tuning. They confirm that, after introducing structure-aware mechanisms, there is an optimal balance between encoding depth and semantic extraction. For regulatory texts organized by multi-layer logic, controlling the model depth supports both expressive power and accurate identification of structural relations.

4) Interference test of data sampling ratio on the ability to identify multiple risk types

This paper also presents an interference test on the data sampling ratio's effect on the ability to identify multiple risk types, and the experimental results are shown in Figure 5.

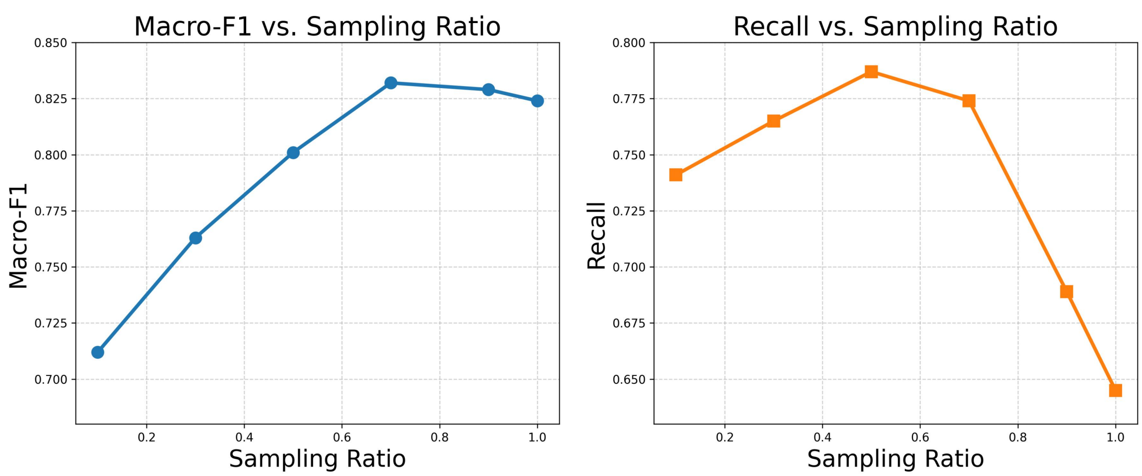

Figure 5. Interference test of data sampling ratio on the ability to identify multiple risk types

As shown in Figure 5, the data sampling ratio has a nonlinear effect on the performance of multi-risk type identification. As the sampling ratio increases, the model's performance on Macro-F1 first rises and then slightly declines. The peak occurs around a sampling ratio of 0.7. This indicates that with a moderate amount of data, the model can effectively learn the semantic distribution of various risk types while maintaining a balanced overall prediction.

However, further increasing the sampling ratio does not bring continuous performance gains. Instead, a slight performance drop is observed. This may be caused by semantic saturation due to data redundancy. The model may become biased toward dominant categories, which weakens its ability to learn from minority risk types. In financial regulatory tasks, such bias can result in missed detections of rare but high-risk events, reducing the actual effectiveness of the system.

In the Recall curve, the fluctuation is more pronounced. When the sampling ratio approaches the full dataset, the recall rate drops significantly. This suggests that although more data increases the model's coverage, it does not always improve its generalization for fine-grained risk types. Excessive data may introduce noise or redundant structures. This can make it harder for the model to capture the triggering conditions of boundary risks, especially in samples with strong coupling between responsible entities and clauses.

Overall, this experiment highlights the importance of controlling sample proportions in risk identification modeling. It confirms that training on a moderate amount of data helps balance semantic density and task separation. This prevents negative transfer caused by information overload. For regulatory texts with uneven distribution and complex structures, a well-designed sampling strategy supports stable model learning and leads to better performance on key evaluation metrics.

5) Experiment on the impact of lack of text structure integrity on model robustness

This paper also gives an experiment on the impact of the lack of text structure integrity on the robustness of the model, and the experimental results are shown in Figure 6. Specifically, the experiment is designed to simulate different degrees of structural degradation in regulatory texts, such as the removal or corruption of clause numbering, hierarchical nesting, and responsibility entity markers. By progressively reducing the amount of structural information available to the model, the experiment aims to assess how critical these structural signals are for maintaining accurate semantic interpretation and consistent risk factor identification. This setup allows for a detailed evaluation of the model's dependency on structural cues and its ability to adapt to varying levels of document completeness.

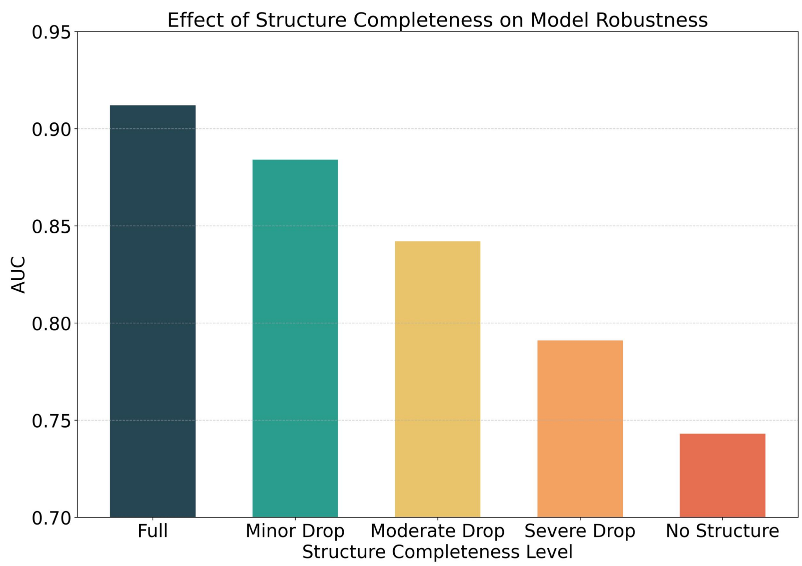

Figure 6. Experiment on the impact of lack of text structure integrity on model robustness

Figure 6 shows the trend of how structural integrity loss affects the robustness of the model. When the text structure remains intact, the model can fully leverage logical cues from clause hierarchies, responsible entities, and substructures. This leads to optimal classification performance. As structural information is gradually lost, the model's performance consistently declines. The most significant drop appears in AUC when structural information is completely removed. This confirms that structural language signals are irreplaceable for modeling financial regulatory texts.

Minor structural loss does not cause catastrophic performance degradation, but it introduces noticeable fluctuations. This suggests that while the model can tolerate some degree of structure breakage during semantic recognition, it is still affected by missing mappings between clauses and entities. These disruptions lead to unstable identification of potential risk points. The effect is especially pronounced in texts that include nested conditions or expressions of responsibility attribution.

When structural corruption reaches a moderate or severe level, the model's robustness declines sharply. Semantic representation begins to deviate from the original logical chain of the clauses. The model struggles to recover the intended policy meaning from the text. This is particularly problematic in multi-task scenarios, where the model's ability to distinguish fine-grained risk types is significantly reduced. Insufficient structural integrity also causes contextual disconnection in risk expression, which impairs precise localization and attribution of risk factors.

6) Loss function changes with epoch

At the end of this paper, a graph of the loss function changing with epoch is given, and the experimental results are shown in Figure 7.

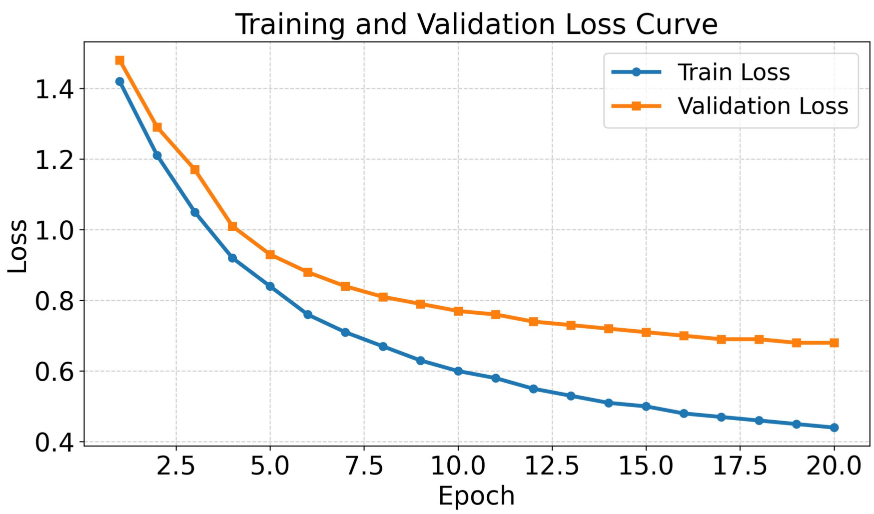

Figure 7. Loss function changes with epoch

From the training and validation loss curves, it can be observed that the model maintains a stable convergence trend throughout the training process. There is no significant oscillation or sign of overfitting. The training loss continues to decrease steadily, indicating that the model is effectively optimizing the objective function in both semantic modeling and structural alignment. This reflects good convergence behavior in the parameter space under the multi-task learning setting.

The validation loss drops sharply in the early stages and then gradually levels off. This trend suggests that the model initially captures the main semantic structures in regulatory texts, and then enters a refinement phase focused on fine-grained risk expressions and boundary type identification. The curve remains relatively stable in the later stages, showing that the model adapts well to the structural variation and task distribution in the validation set. It does not suffer from performance degradation caused by task redundancy or semantic drift.

It is also worth noting that although the validation loss is slightly higher than the training loss, the gap between them stays within a reasonable range. There is no clear divergence between the two. This indicates that the introduced structure-aware and task-adaptive mechanisms enhance representational capacity while suppressing overfitting tendencies. For complex financial regulatory texts, maintaining this balance is essential for reliably identifying cross-type and multi-level risk factors.

5. Conclusion

This paper proposes a large language model framework that integrates structure awareness and task adaptation for risk identification in financial regulatory documents. The framework incorporates a hierarchical semantic-structural encoding mechanism and a dynamic task adaptation module. It enables precise modeling of multi-level semantics, nested structures, and multiple risk types in complex regulatory texts. Given the high degree of domain specificity, logical rigor, and structural complexity of such texts, the proposed approach overcomes the limitations of traditional semantic models in structure understanding and task differentiation. It provides a path that balances accuracy, generalization, and robustness in regulatory language processing.

Experimental results show that the model outperforms baselines across multiple evaluation dimensions. This confirms the positive effect of structural information and task-aware strategies on risk identification performance. Even under challenging conditions such as incomplete structure, fluctuating sample ratios, or varying encoding depth, the method maintains high stability and expressiveness. This demonstrates its strong adaptability to real-world regulatory environments. The model's robustness also enhances its feasibility for practical deployment and provides reliable technical support for intelligent information processing in the financial regulatory domain. Furthermore, the modular design used in this study supports future method extensions. The structure-aware mechanism can be flexibly combined with graph neural networks, multi-scale encoders, or domain rule systems. The task adaptation module can also be fine-tuned or extended for specific regulatory tasks such as compliance review or responsibility labeling. This portable and composable design lays the foundation for building multi-dimensional regulatory semantic understanding systems. It also extends the model's applicability to complex scenarios such as multimodal financial text analysis, cross-document reasoning, and legal clause comparison.

6. Future work

As regulatory texts continue to grow in scale and complexity, language models for financial regulation will require further development in areas like multi-document modeling, logical and causal reasoning, and few-shot risk detection. Under conditions of severe data distribution shifts and frequent policy changes, improving model adaptability and structural generalization will be key research directions. In addition, with the advancement of regulatory technology and compliance automation, the structure-aware method proposed in this study is expected to play a greater role in policy interpretation, intelligent decision support, and financial risk control. It offers important support for building a more efficient, transparent, and intelligent financial governance system.

References

[1] Li Y, Wang S, Ding H, et al. Large language models in finance: A survey[C]//Proceedings of the fourth ACM international conference on AI in finance. 2023: 374-382.

[2] Xie Q, Han W, Chen Z, et al. Finben: A holistic financial benchmark for large language models[J]. Advances in Neural Information Processing Systems, 2024, 37: 95716-95743.

[3] Nie Y, Kong Y, Dong X, et al. A survey of large language models for financial applications: Progress, prospects and challenges[J]. arXiv preprint arXiv:2406.11903, 2024.

[4] Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. Applied soft computing, 93, 106384.

[5] Liu X Y, Wang G, Yang H, et al. Fingpt: Democratizing internet-scale data for financial large language models[J]. arXiv preprint arXiv:2307.10485, 2023.

[6] Huang A H, Wang H, Yang Y. FinBERT: A large language model for extracting information from financial text[J]. Contemporary Accounting Research, 2023, 40(2): 806-841.

[7] Zhao, Y., Du, H., Li, Q., Zhuang, F., Liu, J., & Kou, G. (2022). A Comprehensive Survey on Enterprise Financial Risk Analysis from Big Data Perspective. arXiv preprint arXiv:2211.14997.

[8] Birhane A, Kasirzadeh A, Leslie D, et al. Science in the age of large language models[J]. Nature Reviews Physics, 2023, 5(5): 277-280.

[9] Kasneci E, Seßler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education[J]. Learning and individual differences, 2023, 103: 102274.

[10] Naveed H, Khan A U, Qiu S, et al. A comprehensive overview of large language models[J]. arXiv preprint arXiv:2307.06435, 2023.

[11] Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records[J]. NPJ digital medicine, 2022, 5(1): 194.

[12] Wu S, Irsoy O, Lu S, et al. Bloomberggpt: A large language model for finance[J]. arXiv preprint arXiv:2303.17564, 2023.

[13] Peng C, Yang X, Chen A, et al. A study of generative large language model for medical research and healthcare[J]. NPJ digital medicine, 2023, 6(1): 210.

[14] Köpf A, Kilcher Y, Von Rütte D, et al. Openassistant conversations-democratizing large language model alignment[J]. Advances in Neural Information Processing Systems, 2023, 36: 47669-47681.

[15] Liu Y, Han T, Ma S, et al. Summary of chatgpt-related research and perspective towards the future of large language models[J]. Meta-radiology, 2023, 1(2): 100017.

[16] Kirchenbauer J, Geiping J, Wen Y, et al. A watermark for large language models[C]//International Conference on Machine Learning. PMLR, 2023: 17061-17084.

[17] Huang J, Gu S S, Hou L, et al. Large language models can self-improve[J]. arXiv preprint arXiv:2210.11610, 2022.

[18] Zhuang Y, Wei H. Early warning model and prevention of regional financial risk integrated into legal system[J]. PloS one, 2023, 18(6): e0286685.

[19] Dakhore M, Jenifer R D, Shamout M D, et al. The Application of Time Series Forecasting to Financial Risk Management[C]//2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2024: 1-7.

[20] Aziz L A R, Andriansyah Y. The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance[J]. Reviews of Contemporary Business Analytics, 2023, 6(1): 110-132.

[21] Khunger A. DEEP Learning for financial stress testing: A data-driven approach to risk management[J]. International Journal of Innovation Studies, 2022.

[22] Fritz-Morgenthal S, Hein B, Papenbrock J. Financial risk management and explainable, trustworthy, responsible AI[J]. Frontiers in artificial intelligence, 2022, 5: 779799.

[23] Wei Y, Xu K, Yao J, et al. Financial Risk Analysis Using Integrated Data and Transformer-Based Deep Learning[J]. Journal of Computer Science and Software Applications, 2024, 4(7): 1-8.

[24] Xu K, Cheng Y, Long S, et al. Advancing financial risk prediction through optimized LSTM model performance and comparative analysis[C]//2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE). IEEE, 2024: 1264-1270.

[25] Tan, X. W., & Kok, S. (2024). Explainable risk classification in financial reports. arXiv preprint arXiv:2405.01881.

[26] Liu Y, Ott M, Goyal N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv preprint arXiv:1907.11692, 2019.