MHGR-Net: A Multimodal Heterogeneous Graph Transformer Approach for Compliance Risk Identification and Early Warning in Business-Finance-Law-Tax Domains

Yuan Yuan
School of Accountancy, Anhui Wenda University of Information Engineering, Hefei, China
E-mail: yyjzzlx@sina.com

Keywords: Business–Finance–Law–Tax (BFLT), intelligent risk management, graph neural network, knowledge graph, multi-source heterogeneous data

Received: February 4, 2026

To address the challenges associated with highly heterogeneous Business–Finance–Law–Tax (BFLT) data and increasingly complex risk propagation patterns, this study proposed a Multimodal and Heterogeneous Graph-based Risk Network (MHGR-Net). The framework consisted of three stages. First, a Universal Information Extraction (UIE) model integrated document layout information with semantic features. Second, a temporal heterogeneous knowledge graph was constructed to model dynamic risk associations. Third, a Heterogeneous Graph Transformer (HGT) was employed to learn risk path weights through a self-attention mechanism. The experimental dataset included more than 13,000 multimodal documents, such as contracts, financial statements, and judicial judgments, collected from 30 Chinese A-share listed companies. MHGR-Net achieved an Area Under the Curve (AUC) of 94.1% and an accuracy of 95.03%. The model required only 14 minutes for training. It outperformed mainstream baseline methods, including XGBoost, Graph Convolutional Network (GCN), and Temporal Graph Convolutional Network (T-GCN), in both predictive performance and computational efficiency. The proposed system reduced data silos and provided reliable decision support for enterprises shifting from passive compliance management to proactive and intelligent risk governance.

Povzetek: Študija predstavi MHGR-Net, ki z univerzalnim izluščanjem informacij iz večmodalnih BFLT dokumentov, časovno heterogenim grafom znanja in heterogenim grafnim Transformerjem modelira dinamične poti tveganja za proaktivnejše in natančnejše upravljanje podjetniških tveganj.

1 Introduction

Driven by the current wave of global digitalization, enterprises are undergoing a profound data-driven transformation. Massive volumes of "Business-Finance-Law-Tax" (BFLT) data—including business contracts, financial transactions, legal documents, and tax vouchers—are being generated and accumulated at an unprecedented pace [1]. These data sources are inherently heterogeneous: formats vary widely, content is fragmented, and information is often siloed, making it difficult for enterprises to develop a unified, comprehensive understanding of risk [2]. Meanwhile, the global regulatory environment is becoming increasingly stringent, compliance requirements are continually rising, and both the complexity and propagation of risks are intensifying. Even minor flaws in a single domain can cascade through intricate business processes, triggering cross-domain compliance crises.

Traditional risk management approaches are proving increasingly inadequate under these conditions. Heavy reliance on manual audits, department-specific reviews, and rule-based judgments results in limited coverage, delayed responses, and fragmented insights [3]. These methods struggle to penetrate surface-level operations and cannot effectively identify hidden risks embedded in multi-party transactions, complex equity structures, or ambiguous contract clauses. For example, a related-party transaction may appear compliant but could be used to conceal fund misappropriation or profit shifting. The true risk might only surface during a tax audit or legal dispute. By that time, the enterprise could already be at a disadvantage [4]. Therefore, breaking data silos and adopting a proactive, intelligent paradigm capable of detecting and alerting cross-domain BFLT risks has become essential for enterprise survival and sustainable development.

To address the aforementioned complex challenges, this study proposed an end-to-end intelligent system named Multimodal and Heterogeneous Graph-based Risk Network (MHGR-Net). The system was designed to enable precise identification and dynamic early warning of BFLT compliance risks. The primary objective was to eliminate data silos through technological innovation and to address three key research questions from both theoretical and empirical perspectives.

a. Multi-source integration: How can a multimodal Universal Information Extraction (UIE) model effectively integrate highly heterogeneous data sources, including contracts, financial statements, and legal documents?

b. Temporal evolution: How can a temporal heterogeneous knowledge graph be constructed to capture

dynamic risk propagation patterns across cross-domain business processes?

c. Decision robustness: Based on the heterogeneous graph transformation architecture of MHGR-Net, can the model improve the accuracy and robustness of complex risk path identification while maintaining computational efficiency?

The technical framework comprised three core components.

- First, a domain-adaptive multimodal UIE model was developed. The model incorporated industry-specific knowledge and document layout features. It transformed fragmented unstructured data, such as scanned contracts, invoices, and financial reports, into structured knowledge triples with high efficiency.
- Second, a dynamic temporal heterogeneous BFLT knowledge graph was constructed. Timestamp annotations and entity alignment algorithms were introduced. The graph represented complex enterprise relationship networks and described the temporal evolution of risk states.
- Third, the MHGR-Net deep learning model was implemented as the central analytical engine. The model drew on the architectural advantages of the Heterogeneous Graph Transformer (HGT). A self-attention mechanism was applied to learn the weights of different risk propagation paths. Compared with traditional rule-based methods and static graph models, MHGR-Net identified latent high-risk entities more accurately and predicted potential risk outbreak points more effectively. This capability supported the transition from passive compliance management to proactive and intelligent risk governance.

2 Related work

2.1 Current research on enterprise compliance risk management

Enterprise Risk Management (ERM) is an integrated framework designed to help organizations identify, assess, and respond to various risks. Recent studies have emphasized that ERM should not be treated as an isolated control mechanism. Instead, it should be deeply embedded within corporate governance structures. Gleißner and Berger [5] argued that strengthening embedded risk management was essential to improving ERM effectiveness. Tewu et al. [6] confirmed the significant positive impact of compliance practices on corporate performance in the banking sector. For start-ups, Uwamusi [7] examined strategies for optimizing legal structures under complex regulatory frameworks in order to reduce tax burdens and operational risks. Li et al. [8] conducted a systematic review of the application of Knowledge Graphs (KG) in ERM. With the rapid development of the digital economy, data-driven risk identification has become a major research focus. Li et al. [9] applied an optimized Backpropagation neural network (BPNN) to predict financial risks in listed companies. Tamascelli et al. [10] used neural networks to predict equipment failure time under specific industrial risk events. Chen [11] proposed a Deep Learning-based model for corporate financial risk prediction and intelligent early warning.

Although these studies established the importance of compliance management, existing risk prediction models remain inadequate for handling cross-domain risks in the BFLT context. Most approaches treat risk factors as independent feature variables. They fail to capture cross-domain risk propagation across multi-party transactions and complex ownership structures. In addition, many models rely primarily on single-domain financial data. As a result, they struggle to detect complex anomalies in multinational financial statements [12]. This limitation weakens their early warning capability when cross-domain chain reactions occur.

2.2 Application of knowledge graphs in financial and legal risk control

In the financial and legal domains, Knowledge Graphs have demonstrated effectiveness in fraud detection, anti-money laundering, and related-party transaction analysis. Wang et al. [13] combined Graph Convolutional Network (GCN) with knowledge graphs for stock price prediction. Shi et al. [14] constructed a knowledge fusion graph to forecast market trends. To address the integration of multi-source heterogeneous data, Yan et al. [15] proposed a practical framework for building enterprise knowledge graphs, which alleviated the data silo problem. The performance of knowledge graphs depends heavily on the quality of Information Extraction (IE). Wang and El-Gohary [16] and Zhang [17] explored Artificial Intelligence (AI)-based frameworks for tax and construction regulation IE. Shim et al. [18] introduced Large Language Models (LLMs) to process unstructured documents.

Despite these advances, State-of-the-Art (SOTA) techniques still exhibit clear limitations in semantic integration within the BFLT domain. Existing IE models often fail to simultaneously incorporate document layout features and cross-domain professional terminology. This semantic gap weakens the logical depth of the constructed knowledge graphs. Consequently, such graphs struggle to support complex compliance scenarios that require joint reasoning over legal clauses and financial data.

2.3 Graph neural networks and their application in risk analysis

Graph Neural Networks (GNNs) learn node representations by aggregating information from neighboring nodes. This mechanism aligns closely with the way risks propagate within networked systems. Wang [19] explored a financial risk identification approach based on Convolutional Neural Networks (CNNs). Mojdehi et al. [20] proposed a hybrid model that integrated topological analysis with GNNs, which improved credit risk assessment in supply chain finance.

Although GNNs possess inherent advantages in modeling non-Euclidean structured data, existing studies primarily focus on empirical metrics such as Accuracy and Area Under the Curve (AUC). They provide limited theoretical guarantees of stability. From the perspective of adaptive control, a robust early warning system should ensure finite-time convergence and resilience to uncertainty propagation. Most GNN-based models lack stability guarantees when learning path weights from dynamically evolving BFLT data. The proposed MHGR-Net addresses this limitation. It draws on principles from optimal control theory. The model incorporates temporal encoding and adaptive heterogeneous weight adjustment. This design aims to strengthen dynamic risk modeling and improve theoretical robustness.

Through a comprehensive and critical review of the above literature, Table 1 summarizes representative studies in enterprise compliance risk identification. It compares their core methodologies, application domains, advantages, and technical limitations in handling cross-domain BFLT data integration. The comparison clarifies the gaps in current SOTA research and establishes the technical positioning of this study.

Table 1: representative studies in enterprise compliance risk identification and their technical limitations

Author/Year	Core Method	Research Domain	Main Strength	Critical Limitation (SOTA Gaps)
Gleißner et al. [5]	Embedded risk framework	Corporate management	Emphasized integration of ERM and governance	Lacked concrete quantitative algorithmic implementation
Tewu et al. [6]	Empirical analysis	Banking	Verified positive correlation between compliance and performance	Limited ability to process unstructured textual risks
Uwamusi [7]	Regulatory navigation framework	Start-ups	Optimized legal and tax structures	Focused on policy guidance rather than automated identification
Li et al. [8]	Knowledge Graph review	Risk management	Systematically summarized Knowledge Graphs applications	Did not propose a dedicated BFLT architecture
Li et al. [9]	Optimized BPNN	Financial forecasting	Improved single-domain prediction accuracy	Treated risk factors as independent variables
Tamascelli [10]	Neural network failure analysis	Industrial safety	Applied AI to time-to-failure prediction	Not generalizable to complex business logic
Chen [11]	Deep learning early warning model	Financial risk	Strengthened automated financial risk prediction	Lacked cross-domain semantic association analysis
Li et al. [12]	Anomaly pattern recognition	Multinational enterprises	Detected fraud patterns in financial statements	Failed to reveal underlying legal roots of risk
Wang et al. [13]	Knowledge Graphs + GCN	Stock market	Captured relationships among market entities	Ignored temporal dynamics of risk evolution
Shi et al. [14]	GCN-LSTM	Market forecasting	Introduced preliminary temporal modeling	Limited capability in unstructured document extraction
Yan et al. [15]	Heterogeneous Knowledge Graphs construction framework	Enterprise knowledge base	Addressed heterogeneous data silos	Insufficient reasoning depth for compliance risk
Wang et al. [16]	Deep learning-based relation extraction	Construction safety	Enabled automated regulation transformation	Unable to process cross-domain knowledge such as taxation
Zhang [17]	Multimodal AI framework	Tax compliance	Improved tax data integration	Did not integrate financial–legal cross-validation
Shim et al. [18]	LLM + OmEGa	Manufacturing	Strong semantic understanding with low zero-shot cost	Lacked domain adaptation for financial and legal contexts
Wang [19]	CNN-based risk model	Financial risk	High computational efficiency	Not suitable for complex non-Euclidean graph structures
Mojdehi et al. [20]	Topological analysis + GNN	Supply chain finance	Enhanced capture of topological features	Lacked theoretical convergence guarantees
This Study	MHGR-Net	BFLT cross-domain compliance	Unified multimodality, dynamic modeling, and theoretical robustness	Filled SOTA gaps in cross-domain integration and temporal risk evolution

3 Construction of the BFLT knowledge graph and the risk identification & early warning model

3.1 System architecture and technical analysis

To enable intelligent identification and early warning of compliance risks in the BFLT domain, the system adopts a hierarchical technical architecture. This architecture consists of three layers: the Data Layer, the Model Layer, and the Application Layer. It is designed to provide end-to-end processing, from heterogeneous data ingestion to risk decision output. To visually illustrate this processing pipeline, the study presents a system workflow in the form of pseudocode and a logic path diagram (Figure 1). These depict the entire process, from multimodal data extraction, through temporal knowledge graph construction, to risk prediction using a HGT. The hierarchical design ensures modularity and scalability, allowing the system to handle complex business logic efficiently.

```
Start
Input: Extract text, tables, and image-based layout information
Output: The risk rating of each risk subject and the warning results of potential
        high-risk associations
def extract_knowledge(inputs, uie_model):
    # Perform multimodal information extraction
    entities, relations = uie_model.extract(inputs)
    return entities, relations
def construct_graph(entities, relations):
    # Build a heterogeneous temporal knowledge graph
    G = Graph()
    for e in entities:
        G.add_node(e)
    for r in relations:
        G.add_edge(r.head, r.tail, r.type, r.time)
    return G
def risk_prediction(G, htgn_model):
    # Perform risk identification using the Heterogeneous Temporal Graph Network
    embeddings = htgn_model.encode(G)
    risk_scores = htgn_model.predict(embeddings)
    return risk_scores
def active_feedback(risks, model, threshold=0.7):
    # Perform active learning for self-optimization
    low_conf = [n for n, s in risks.items() if s < threshold]
    for node in low_conf:
        new_data = retrieve_data(node)
        model.retrain(new_data)
End
```

Figure 1: Pseudocode flow of the MHGR-Net-based risk intelligent identification and early warning model.

The system's end-to-end processing logic is presented in the pseudocode of Figure 1. The workflow comprises five core stages. First, multimodal data—including text, tables, and document layout information—is ingested. Second, the UIE model is applied to perform automated knowledge extraction. Third, a heterogeneous knowledge graph with temporal dimensions is constructed. Fourth, the core MHGR-Net is used to perform risk identification and score prediction. Finally, an active feedback mechanism triggers human verification for low-confidence results and supports incremental model retraining. This closed-loop design ensures continuous optimization when processing complex BFLT data.

The model layer serves as the core of the system, hosting intelligent processing capabilities such as IE, knowledge graph construction, and risk analysis, and providing the foundation for knowledge inference, predictive analytics, and decision support. The application layer acts as the interface and feedback channel, visualizing results in a graphical and interactive manner, while supporting human-in-the-loop and closed-loop feedback mechanisms to dynamically optimize and continuously evolve the model based on new data and insights. The architecture emphasizes cross-domain semantic integration, multimodal collaborative modeling, and sustainable learning, ensuring robust adaptability in complex risk scenarios. By leveraging these three layers in a coordinated manner, the system can efficiently detect emerging risks, monitor evolving compliance patterns, and support proactive risk mitigation strategies. The structure of the BFLT risk identification and early warning system based on MHGR-Net (Multimodal + Heterogeneous Graph-based Risk Network) is illustrated in Figure 2.

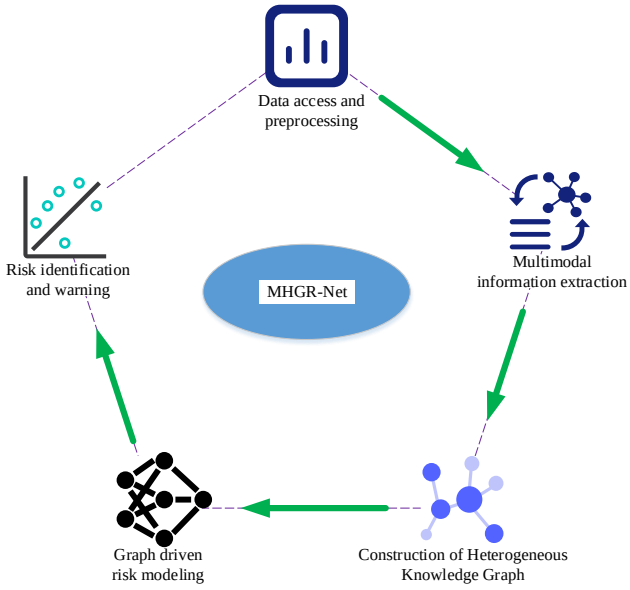

Figure 2: Illustration of the MHGR-Net-based BFLT risk identification and early warning model architecture.

Given that BFLT data involve sensitive corporate financial information and legal privacy, the system was designed in strict accordance with data ethics and security standards. During the data ingestion stage, all non-public contract details, financial statements, and tax records were de-identified, and differential privacy techniques were applied to protect sensitive attributes of specific entities. Moreover, system operations fully comply with the Data Security Law and relevant industry regulations, ensuring that data are used exclusively for research and management purposes related to compliance risk prediction. By implementing a controlled access framework, the system breaks down data silos while effectively preventing data leakage, thereby guaranteeing the compliant delivery of intelligent early-warning capabilities.

3.2 Knowledge graph construction based on multimodal IE

For the BFLT domain, a dedicated knowledge graph schema was designed, clearly defining entity types (e.g., enterprises, financial indicators, tax provisions) and relation types (e.g., "signs contract," "pays tax"). Core IE relies on a domain-adaptive multimodal UIE model [21,22]. By performing secondary pretraining on domain-specific corpora, the model significantly enhances its understanding of professional terminology and complex document structures.

When processing semi-structured and unstructured documents (e.g., PDF contracts or scanned invoices), the model uses its layout-aware capabilities to combine text, tables, and layout information. This enables accurate extraction of multimodal knowledge elements, forming triples that represent entities, their attributes, and the relationships between them.

To achieve cross-domain semantic integration, the system formalizes the input multimodal document D as a heterogeneous feature set comprising text streams, image pixels, and layout coordinates. Let D consist of N document elements:

$$D = \{d_1, d_2, \ldots, d_N\} \tag{1}$$

For each element, the system employs a deep encoder $f_{UIE}(\cdot)$ based on the LayoutLMv2 architecture. This encoder jointly models spatial layout and textual semantics via a multi-head self-attention mechanism, producing a multimodal feature vector X_i:

$$X_i = f_{UIE}(d_i), \quad i = 1, 2, \ldots, N \tag{2}$$

Based on these representations, the knowledge extraction module performs named entity recognition (NER) and relation extraction (RE) to construct a set of knowledge triples τ. To capture the temporal evolution of risks, the triples are extended into a time-aware data structure:

$$\tau = \{(h_j, r_j, t_j, \Delta T_j) \mid h_j, t_j \in E, r_j \in R\} \tag{3}$$

h_j, t_j denotes the head and tail entities, r_j represents the relation type between them, E and Z are the sets of all entity and relation types, respectively. E corresponds to the BFLT entity set, Z represents the set of relation types, and ΔT_j denotes the timestamp of the event.

After obtaining a large collection of knowledge triples, a temporal heterogeneous knowledge graph is constructed in the next step. The construction workflow of the knowledge graph is illustrated in Figure 3.

During the multi-source data fusion phase, to address inconsistencies in how the same entity is represented across different sources, the system employs a dual-path similarity evaluation model. The entity alignment function, $sim(e_a, e_b)$, is defined as:

$$sim(e_a, e_b) = \alpha \cdot sim_{attr}(e_a, e_b) + (1-\alpha) \cdot sim_{rel}(e_a, e_b) \tag{4}$$

Specifically, the attribute similarity sim_{attr} is computed using cosine similarity between entity embedding vectors, while the relation similarity sim_{rel} is calculated using the Jaccard index to evaluate the proportion of shared neighboring nodes between two entities. In this study, the weight coefficient α was set to 0.6, and the fusion decision threshold was set at 0.85. When the similarity score exceeds this threshold, entities are automatically merged. For scores in the range [0.7,0.85], the entities are routed to a manual verification pipeline for final approval, ensuring the foundational quality of the knowledge graph.

Next, timestamp annotation is performed. To capture the temporal evolution of business processes and risk events, timestamps τ are added to entity relationships, forming triples with temporal attributes (h_j, r_j, t_j, τ_j), where τ_j indicates the specific time the relation r_j occurs.

Finally, the constructed temporal heterogeneous knowledge graph is stored in a graph database to enable efficient storage and querying, supporting subsequent graph-based risk identification and early warning computations.

Through this process of multimodal IE and temporal graph construction, the system effectively integrates multi-source heterogeneous BFLT data. It builds a high-quality, dynamically evolving knowledge network that provides a solid data foundation for subsequent risk identification models.

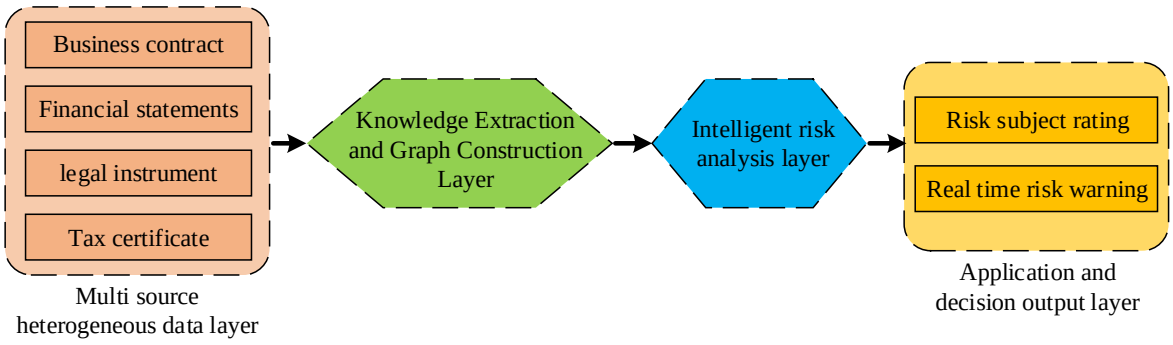

Figure 3: Illustration of knowledge graph construction steps

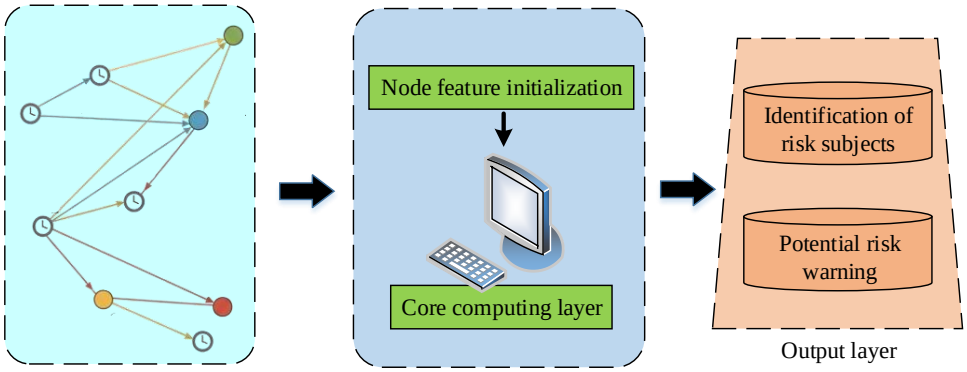

Figure 4: Schematic of the risk identification and early warning module based on heterogeneous temporal graph networks.

3.3 Risk identification and early warning module based on heterogeneous temporal graph networks

With the multi-source heterogeneous knowledge graph constructed, accurately identifying enterprise compliance risks and predicting potential risks becomes a key system challenge. The constructed knowledge graph is formally defined as a temporal heterogeneous graph G, as shown in Equation (5):

$$G=(V,E,T,\Phi) \tag{5}$$

Where V is the set of nodes, encompassing multiple entity types (e.g., enterprises, legal representatives, contracts); E is the set of typed edges, representing diverse relationships; T is the set of corresponding timestamps, indicating the occurrence time of node or edge events; and Φ is a type mapping function defining node and edge categories. Along the temporal dimension, the graph exhibits dynamic evolution, as risk states may change over time. To model this, the design incorporates a time encoding mechanism TE (·), integrating event time information into node representations to strengthen temporal dependency modeling [23].

The core model is a heterogeneous GNNs based on graph attention mechanisms, applied to risk identification, as illustrated in Figure 4.

In this network, the first step is node feature initialization. Each node $v_i \in V$ is initialized with representation $h_i^{(0)}$, which integrates both entity attributes and text-encoded information.

Subsequently, in the multi-head heterogeneous graph attention layer, attention weights are calculated for each node type $t \in T_V$ and edge type $r \in T_E$, as defined in Equation (6):

$$\alpha_{ij}^{r}=\frac{\exp\left(\text{LeakyReLU}\left(a_r^{\top}\left[W_r h_i \| W_r h_j\right]\right)\right)}{\sum_{k\in N_i^r}\exp\left(\text{LeakyReLU}\left(a_r^{\top}\left[W_r h_i \| W_r h_k\right]\right)\right)} \tag{6}$$

Where h_i, h_j denote the node representation; $\|$ indicates vector concatenation; W_r is the linear transformation matrix corresponding to edge type r; a_r is the attention vector parameter for edge type r; and N_i^r represents the neighbor set of node i under relation r.

Next, temporal encoding fusion is introduced. For each edge timestamp τ_{ij}, a positional encoding $TE(\cdot)$ is incorporated into the computation, as shown in Equation (7):

$$\tilde{h}_i=\sigma\left(\sum_{r\in R}\sum_{j\in N_i^r}\alpha_{ij}^{r}\cdot\left(W_r h_j+TE\left(\tau_{ij}\right)\right)\right) \tag{7}$$

where σ is a nonlinear activation function. Through multi-layer iterations, the model captures complex node relationships and temporal evolution patterns. In the specific implementation of Equation (7), this study employs a Sinusoidal Time Encoding (STE) mechanism to construct TE(·). This approach is inspired by the positional encoding used in the Transformer architecture. It maps continuous time intervals into 128-dimensional embedding vectors using sine and cosine functions at different frequencies. This design not only captures the temporal interval characteristics between business activities but also enables the model to adaptively learn both the periodic and abrupt patterns of risk evolution. By adding this time vector to the node feature representations, MHGR-Net can effectively distinguish between "historical stale risks" and "real-time compliance threats", thereby significantly enhancing the modeling of temporal dependencies.

Finally, the system performs risk entity identification. Framed as a node classification task, the model classifies enterprise nodes into multiple risk categories, outputting the risk level $y_i \in \{0,1,...,C-1\}$. The loss function is defined as a cross-entropy loss L_{risk} in Equation (8):

$$L_{risk}=-\sum_{i\in V_{ent}}\sum_{c=0}^{C-1}\mathbb{1}[y_i=c]\log\hat{p}_{i,c} \tag{8}$$

Where V_{ent} is the set of enterprise nodes; y_i is the true risk label of enterprise i; and $\hat{p}_{i,c}$ denotes the predicted probability that enterprise i belongs to risk level c. In the node classification task, this study categorizes

the risk level y_i into four grades according to regulatory standards and business logic: L0 (Robust, no compliance concerns), L1 (Attention, minor management gaps), L2 (Warning, significant signs of noncompliance), and L3 (Critical, major legal or tax risks triggered). Given the extreme scarcity of L2 and L3 samples in real-world scenarios (resulting in a long-tail data distribution), to prevent the model from overfitting to the high-frequency L0 class, a Weighted Cross-Entropy (WCE) mechanism was incorporated into the logistic regression. By dynamically adjusting the loss weights based on the inverse of class sample sizes, the model is forced to assign greater attention to the high-risk minority classes, thereby ensuring heightened sensitivity for detecting extreme compliance risks.

For potential risk early warning, formulated as a link prediction task, the model predicts whether potential high-risk edges exist in the graph. The binary cross-entropy loss L_{link} is expressed in Equation (9):

$$L_{link} = -\sum_{(i,j)\in E^+} \log\sigma(s_{ij}) - \sum_{(i,j)\in E^-} \log\left(1-\sigma(s_{ij})\right) \quad (9)$$

Where E^+, E^- are sets of positive and negative sample edges, respectively; s_{ij} is the scoring function for edge (i, j), typically defined as the dot product of node embeddings or the output of an MLP; and $\sigma(\cdot)$ is the Sigmoid activation function. For the potential risk early-warning task, which is formulated as a link prediction problem, the system constructs the training set using a Negative Sampling strategy, with a positive-to-negative sample ratio of 1:5. In terms of performance evaluation, besides the conventional ROC-AUC metric, this study specifically incorporates Precision@K to measure the practical accuracy of high-risk predictions. During model deployment, a risk-trigger threshold of 0.75 is set. Once a predicted score exceeds this threshold, the system automatically identifies the most contributive risk propagation paths via an attention mechanism. This design of "explainable paths" provides compliance officers with intuitive audit-trail evidence and supports a human-in-the-loop decision-making process, enabling enterprises to take precise interventions based on the causal insights generated by the model.

The overall loss combines both tasks, as given in Equation (10):

$$L = \gamma L_{risk} + (1-\gamma) L_{link} \quad (10)$$

Where $\gamma \in [0,1]$ is the task weight coefficient.

To meet the strict compliance audit requirements for traceability and interpretability of predictions, MHGR-Net does not operate as a black-box model. The system integrates explainable AI (XAI) tools, such as GNNExplainer, which automatically identify and extract the subgraph structures and feature paths that contribute most to a specific risk score. For instance, when the system detects a potential tax-evasion risk in a cross-border related-party transaction, it highlights the relevant contract clauses, cash-flow nodes, and temporal evolution cues. This causal traceability capability provides actionable insights for management and generates "audit-trail" evidence, ensuring that the early-warning results are admissible in legal and tax practice. Consequently, it addresses the critical challenge of trustworthiness in deploying deep learning models in high-stakes compliance scenarios.

3.4 Experimental evaluation

To comprehensively validate the effectiveness of the MHGR-Net model in cross-domain BFLT compliance scenarios, this study constructed an industry-representative experimental dataset. The dataset covers 30 A-share main-board listed companies in China, spanning the years 2020 to 2024. Data were collected and integrated from three primary sources: the China Judgments Online (https://wenshu.court.gov.cn/), the WIND financial terminal (https://www.wind.com.cn/), and publicly disclosed corporate information. The dataset comprises 1,240 legal judgments, 450 annual financial reports, 3,200 commercial contracts, and over 8,500 tax vouchers. Table 2 summarizes the statistical characteristics of the dataset. After multimodal IE and semantic alignment, the resulting heterogeneous knowledge graph contains approximately 45,120 entity nodes and 128,450 relational edges. The dataset exhibits a long-tail distribution consistent with real-world business logic. High-risk warning cases (levels L2 and L3) account for roughly 12.4%, providing a realistic and stringent environment for testing the model's ability to handle class imbalance.

Table 2: Statistics of the experimental dataset

Statistical Metric	Value	Notes
Number of listed companies	30	A-share main-board enterprises
Total number of documents	13,390	Contracts, financial reports, judgments, tax vouchers
Total knowledge graph nodes	45,120	Entities including companies, executives, indicators, clauses
Total knowledge graph edges	128,450	Heterogeneous edges with temporal annotations
High-risk cases (L2–L3)	12.40%	Reflects class imbalance

Table 3: Hardware/software configurations and key training parameters.

Category	Parameter	Value
Hardware	GPU	NVIDIA A100 80GB
	CPU	Intel Xeon Gold 6226R
	Memory	256 GB
Software	OS	Windows 11 Pro 64-bit
	Deep learning framework	PyTorch 2.0
	Graph database	Neo4j 5.x

Model params	UIE pretrained model	LayoutLMv2 (finetuned)
	HTGN layers (L)	3
	Attention heads (H)	4
	Learning rate	1e-4
	Batch size	32
	Dropout	0.2
	Epochs	120

Regarding ethics and privacy protection, this study strictly adhered to the Data Security Law and relevant industry regulations. Although the data were primarily sourced from publicly available channels, all original documents ingested by the model were de-identified and deeply anonymized, removing unnecessary personal identifiers and sensitive commercial clauses. This processing approach ensures the transparency and reliability of the research findings while maximizing the security and compliance of data usage. The experiments were conducted on a high-performance GPU server. The detailed hardware/software configurations and key training parameters are summarized in Table 3.

For performance evaluation, the MHGR-Net model was compared against XGBoost [24], GCN [25], Graph Attention Network (GAT) [26], Temporal Graph Convolutional Network (T-GCN) [27], and the model proposed by Mojdehi et al. [20]. Model performance was assessed using Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), Precision, and F1-score, which collectively reflect the effectiveness of risk identification and early warning.

4 Results and discussion

4.1 Analysis of risk identification performance across different algorithms

The proposed model was compared with baseline algorithms in terms of AUC and training time, as shown in Figure 5.

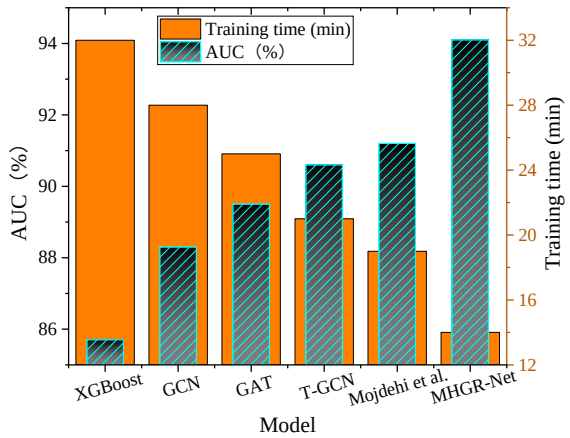

Figure 5: Comparative results of risk identification performance across different algorithms

As shown in Figure 5, the experimental results revealed notable performance differences among the evaluated models, particularly in terms of the Area Under the ROC Curve (AUC) and training time. From the perspective of the AUC metric, the proposed MHGR-Net achieved the highest accuracy in risk identification, reaching 94.1%, which was substantially superior to the benchmark models. In comparison, the classic machine learning model XGBoost achieved only 85.7%, representing the lowest performance among the tested approaches. Graph-based models demonstrated moderate improvements: GCN and GAT achieved 88.3% and 89.5%, respectively, indicating the benefits of incorporating graph structure modeling. The temporal graph model T-GCN further increased the AUC to 90.6%, while the enhanced model proposed by Mojdehi et al. [20] reached 91.2%. Overall, MHGR-Net's integration of multimodal information and heterogeneous graph structures allowed it to capture complex interdependencies across business, finance, legal, and tax data, thereby demonstrating superior accuracy, robustness, and generalization ability.

In terms of training efficiency, MHGR-Net also showed remarkable advantages. Its average training time was only 14 minutes, considerably shorter than the 32 minutes required for XGBoost and 28 minutes for GCN. This improvement indicates that the architectural and inference optimizations of MHGR-Net not only enhanced recognition performance but also significantly reduced computational costs. Consequently, the model is highly suitable for real-world deployment scenarios, particularly in contexts where rapid decision-making and limited computational resources are critical. These results highlight MHGR-Net's potential to support proactive, intelligent risk management with both high precision and practical efficiency.

As summarized in Table 4, MHGR-Net significantly outperformed all baseline models across the key evaluation metrics. A T-test analysis confirmed that the performance improvement over the second-best model (Mojdehi et al. [20]) reached a statistically significant level ($p < 0.05$), strongly demonstrating the substantive gains achieved through multimodal integration and temporal modeling. Furthermore, a sensitivity analysis of the model parameters was conducted. The selection of three graph convolutional layers balances the extraction of deep topological features while avoiding over-smoothing. The use of four attention heads was found to optimally capture heterogeneous semantics across BFLT domains in parallel. This configuration achieved a 94.1% AUC while reducing training time to 14 minutes, providing an optimal trade-off between accuracy and efficiency.

Table 4. Comparison of main experimental results

Model/ Algorithm	Accuracy (%)	AUC (%)	Precision (%)	F1-score (%)	Training Time (min)
XGBoost	80.95	85.7	78.4	79.2	32
GCN	83.89	88.3	82.1	81.5	28
GAT	85.68	89.5	83.5	84.1	30
T-GCN	88.30	90.6	84.0	86.2	26

Mojdehi et al. [20]	89.04	91.2	85.0	87.4	24
MHGR-Net	95.03	94.1	90.4	91.6	14
p-value (vs Mojdehi et al.)	< 0.05	< 0.05	< 0.05	< 0.05	-

4.2 Accuracy analysis across different algorithms

The performance of MHGR-Net was further compared with other algorithms using Accuracy, Precision, and F1-score as evaluation metrics. The results are illustrated in Figures 6–8.

As illustrated in Figures 6–8, the proposed MHGR-Net consistently outperformed all baseline models across different training epochs, demonstrating substantial performance advantages in risk identification and early warning tasks. In terms of Accuracy, MHGR-Net achieved 87.68% by the 60th epoch and gradually stabilized at 95.03% by the 120th epoch. This performance exceeded that of XGBoost (80.95%), GCN (83.89%), GAT (85.68%), T-GCN (88.30%), and the model proposed by Mojdehi et al. [20] (89.04%), highlighting its superior ability to capture complex interdependencies in BFLT data. For Precision, MHGR-Net reached 86.2% as early as the 20th epoch, substantially higher than all other models at the same stage. By the 120th epoch, Precision further improved to 90.4%, surpassing T-GCN (84%) and Mojdehi et al.'s model (85%), demonstrating its effectiveness in correctly identifying high-risk entities. The F1-score results further confirmed the balanced performance of MHGR-Net. At the 120th epoch, the model reached 91.6%, indicating stronger robustness and generalization compared with traditional GNNs and ensemble-based methods. Overall, MHGR-Net not only converged faster but also delivered higher accuracy and greater stability, providing reliable and timely early warning outcomes. These results emphasize its wide applicability and strong practical potential for analyzing complex spatiotemporal heterogeneous data, detecting hidden risks, and supporting proactive, intelligent compliance risk management in enterprise settings.

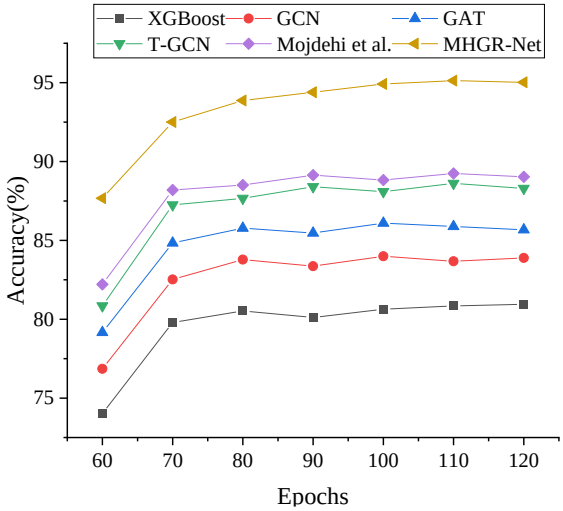

Figure 6: Accuracy comparison of risk identification and early warning across different algorithms.

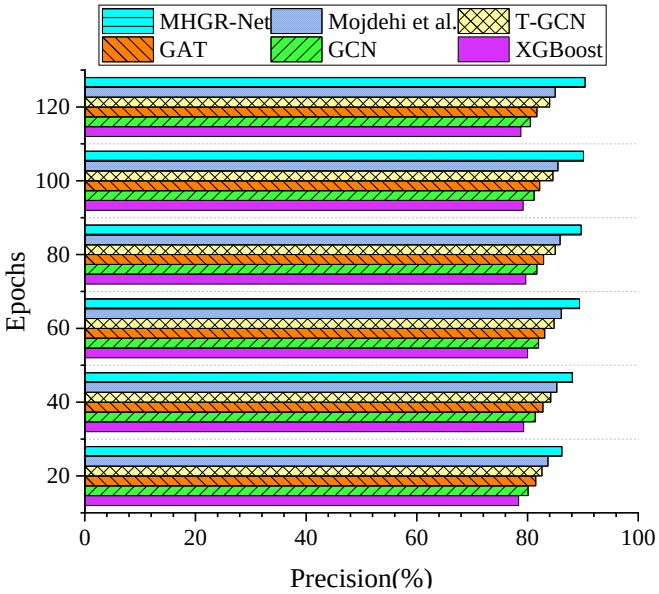

Figure 7: Precision comparison of risk identification and early warning across different algorithms.

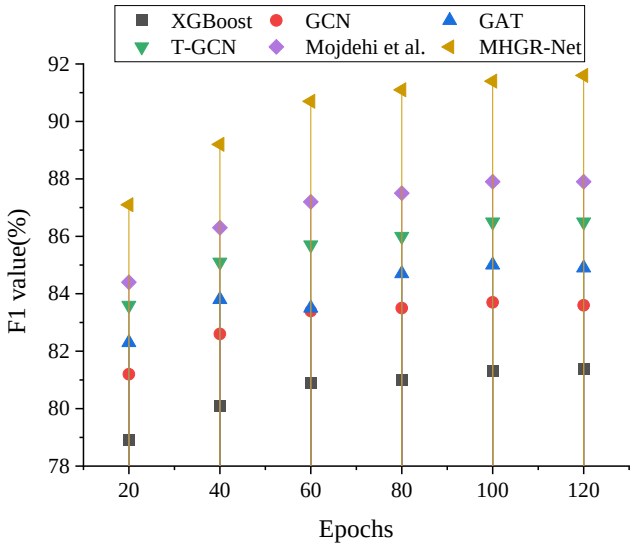

Figure 8: F1-score comparison of risk identification and early warning across different algorithms.

4.3 Ablation study

To quantitatively evaluate the contribution of each core component of MHGR-Net to risk identification performance, an ablation study was conducted, with results summarized in Table 5. Four model variants were considered: (1) w/o Multi – Multi-modal layout information removed; only plain text retained. (2) w/o Temporal – Time encoding mechanism removed, treating the knowledge graph as static. (3) w/o Hetero – The heterogeneous graph downgraded to a homogeneous graph. The results show that removing time encoding led to a 4.2% drop in AUC, highlighting the critical role of risk evolution dynamics in early warning. Removing multi-modal information caused a 5.1% decrease in Precision,

confirming the irreplaceable value of image layout features for parsing complex financial statements and contracts. These findings demonstrate that the synergistic integration of all components underpins MHGR-Net's ability to deliver precise cross-domain compliance risk alerts.

Table 5: Ablation study results

Model Variant	Accuracy (%)	AUC (%)	Precision (%)	F1-score (%)	AUC Performance Loss
MHGR-Net (Full Model)	95.03	94.1	90.4	91.6	-
w/o Multimodality	91.20	89.8	85.3	87.5	-4.3%
w/o Temporal Info	92.45	89.9	87.1	88.9	-4.2%
w/o Heterogeneity	89.50	87.5	83.2	85.0	-6.6%

4.4 Discussion

The proposed MHGR-Net achieved an AUC of 94.1% in BFLT compliance risk prediction, significantly outperforming all baseline models listed in Table 1. A detailed comparison with SOTA methods shows that, although Zhang [17] proposed a knowledge-graph-enhanced multimodal framework to improve tax data integration and compliance, MHGR-Net achieved superior recognition accuracy by incorporating a time encoding mechanism, which effectively modeled the nonlinear relationships between dynamically evolving legal provisions and fluctuating financial indicators. This cross-domain performance improvement largely results from Multimodal Representation Learning (MRL), which deeply aligns heterogeneous layout information with textual semantics. This aligns with the theoretical framework of Wang et al. [28], who demonstrated that cross-diffusion attention mechanisms facilitate the co-evolution of multimodal features, validating the importance of inter-modal interactions in understanding complex semantic scenarios.

From a deployment perspective, fintech systems must balance technological innovation with regulatory compliance. To address concerns regarding the "black-box" nature of deep learning in stringent compliance contexts, MHGR-Net integrates XAI tools. As emphasized by Anang et al. [29], ensuring the transparency of model decision logic is a core requirement for regulatory compliance. The risk propagation paths extracted by MHGR-Net provide auditable causal evidence for compliance officers and support a Human-in-the-loop decision-making framework, enabling enterprises to translate model outputs into actionable compliance defense strategies.

Nevertheless, limitations exist. Current evaluations focus primarily on 30 Chinese A-share listed companies. Future work should explore the model's generalization to SMEs under different legal systems and consider integrating LLMs to enhance semantic parsing efficiency for large-scale unstructured legal documents.

5 Conclusion

This study developed an intelligent compliance risk identification and early warning system for the BFLT domain. By integrating multimodal IE with temporal heterogeneous graph techniques, MHGR-Net successfully enabled cross-domain association analysis across business, finance, legal, and tax dimensions. Experimental results demonstrated that the model achieved over 95% accuracy in risk identification while maintaining high computational efficiency, significantly enhancing enterprises' ability to detect hidden risk paths.

Despite these achievements, the study has several limitations. The experimental dataset primarily comprises 30 Chinese A-share listed companies, which restricts the model's generalizability to resource-constrained SMEs or complex cross-border regulatory environments. Moreover, the "black-box" nature of deep learning models remains a key barrier to user trust and regulatory acceptance.

Future research will focus on three directions: (1) XAI integration – Incorporating tools such as GNNExplainer and SHAP to provide auditable logic traces for each risk alert, supporting a Human-in-the-loop decision-making framework. (2) Federated Learning exploration – Enabling joint risk modeling across multiple enterprises while preserving sensitive data privacy. (3) LLM integration – Enhancing fine-grained semantic parsing of unstructured legal texts to build a more robust global compliance monitoring system.

Acknowledgements

This paper is funded by 2024 Anhui Provincial Key Research Projects in Humanities and Social Sciences (Education Department), Research on the Construction of an Integrated Compliance Management System for "Business, Finance, Law and Taxation", (Project No.: 2024AH052528).

References

[1] A. Cahyadini, J. I. G. Hutagalung and Z. Muttaqin, "The urgency of reforming Indonesia's tax Law in the face of economic digitalization," Cogent Social Sciences, vol. 9, no. 2, p. 2285242, 2023. https://doi.org/10.1080/23311886.2023.2285242

[2] N. Xu, W. Lv and J. Wang, "The impact of digital transformation on firm performance: a perspective from enterprise risk management," Eurasian Business Review, vol. 14, no. 2, pp. 369–400, 2024. https://doi.org/10.1007/s40821-024-00264-9

[3] U. M. Tanko, "Financial attributes and corporate tax planning of listed manufacturing firms in Nigeria: moderating role of real earnings management," Journal of Financial Reporting and Accounting, vol. 23, no. 3, pp. 1024–1056, 2025. https://doi.org/10.1108/JFRA-05-2022-0198

[4] B. Solikhah, C. L. Chen, P. Y. Weng and M. A. S. Al-Faryan, "Related party transactions and tax avoidance: does government ownership play a role?" Corporate Governance: The International Journal of Business in Society, vol. 25, no. 4, pp. 763–785, 2025. https://doi.org/10.1108/CG-01-2024-0003

[5] W. Gleißner and T. Berger, "Enterprise risk management: improving embedded risk management and risk governance," Risks, vol. 12, no. 12, p. 196, 2024. https://doi.org/10.3390/risks12120196

[6] M. L. D. Tewu, I. Bernarto, S. Suwarno and P. Lisdiono, "The effect of enterprise risk management and compliance practices on the firm performance of Indonesian banking companies," Indonesian Journal of Business and Entrepreneurship, vol. 10, no. 1, pp. 52–64, 2024. https://doi.org/10.17358/ijbe.10.1.52

[7] J. A. Uwamusi, "Navigating complex regulatory frameworks to optimize legal structures while minimizing tax liabilities and operational risks for startups," International Journal of Research Publication and Reviews, vol. 6, no. 2, pp. 845–861, 2025. https://doi.org/10.55248/gengpi.6.0225.0736

[8] P. Li, Q. Zhao, Y. Liu, C. Zhong, J. Wang and Z. Lyu, "Survey and prospect for applying knowledge graph in enterprise risk management," Computers, Materials & Continua, vol. 78, no. 3, pp. 3825–3865, 2024. https://doi.org/10.32604/cmc.2024.046851

[9] X. Li, J. Wang and C. Yang, "Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy," Neural Computing and Applications, vol. 35, no. 3, pp. 2045–2058, 2023. https://doi.org/10.1007/s00521-022-07377-0

[10] N. Tamascelli, G. E. Scarponi, M. T. Amin, Z. Sajid, N. Paltrinieri and F. Khan, "A neural network approach to predict the time-to-failure of atmospheric tanks exposed to external fire," Reliability Engineering & System Safety, vol. 245, p. 109974, 2024. https://doi.org/10.1016/j.ress.2024.109974

[11] W. Chen, "Enterprise financial risk prediction and intelligent early warning model based on deep learning," Discover Artificial Intelligence, vol. 5, no. 1, p. 227, 2025. https://doi.org/10.1007/s44163-025-00497-1

[12] Y. Li, Y. Zhou and Y. Wang, "Deep learning-based anomaly pattern recognition and risk early warning in multinational enterprise financial statements," Journal of Sustainability, Policy, and Practice, vol. 1, no. 3, pp. 40–54, 2025.

[13] T. Wang, J. Guo, Y. Shan, Y. Zhang, B. Peng and Z. Wu, "A knowledge graph–GCN–community detection integrated model for large-scale stock price prediction," Applied Soft Computing, vol. 145, p. 110595, 2023. https://doi.org/10.1016/j.asoc.2023.110595

[14] Y. Shi, Y. Wang, Q. Qu and Z. Chen, "Integrated GCN-LSTM stock prices movement prediction based on knowledge-incorporated graphs construction," International Journal of Machine Learning and Cybernetics, vol. 15, no. 1, pp. 161–176, 2024. https://doi.org/10.1007/s13042-023-01817-6

[15] C. Yan, X. Fang, X. Huang, C. Guo and J. Wu, "A solution and practice for combining multi-source heterogeneous data to construct enterprise knowledge graph," Frontiers in Big Data, vol. 6, p. 1278153, 2023. https://doi.org/10.3389/fdata.2023.1278153

[16] X. Wang and N. El-Gohary, "Deep learning-based relation extraction and knowledge graph-based representation of construction safety requirements," Automation in Construction, vol. 147, p. 104696, 2023. https://doi.org/10.1016/j.autcon.2022.104696

[17] T. Zhang, "A knowledge graph-enhanced multimodal AI framework for intelligent tax data integration and compliance enhancement," Frontiers in Business and Finance, vol. 2, no. 2, pp. 247–261, 2025. https://doi.org/10.71465/fbf384

[18] M. Shim, H. Choi, H. Koo, K. Um, K. H. Lee and S. Lee, "OmEGa (Ω): Ontology-based information extraction framework for constructing task-centric knowledge graph from manufacturing documents," Advanced Engineering Informatics, vol. 64, p. 103001, 2025. https://doi.org/10.1016/j.aei.2024.103001

[19] Z. Wang, "Application of CNN-based financial risk identification and management convolutional neural networks in financial risk," Systems and Soft Computing, vol. 7, p. 200215, 2025. https://doi.org/10.1016/j.sasc.2025.200215

[20] K. F. Mojdehi, B. Amiri and A. Haddadi, "A novel hybrid model for credit risk assessment of supply chain finance based on topological data analysis and graph neural network," IEEE Access, vol. 13, pp. 13101–13127, 2025. https://doi.org/10.1109/ACCESS.2025.3528373

[21] S. Chen, S. Liu and J. Liu, "Type-specific modality alignment for multi-modal information extraction," IEEE Signal Processing Letters, vol. 31, pp. 1525–1529, 2024. https://doi.org/10.1109/LSP.2024.3396705

[22] X. Li, J. Jin, Y. Zhou, Y. Zhang, P. Zhang, Y. Zhu and Z. Dou, "From matching to generation: A survey on generative information retrieval," ACM Transactions on Information Systems, vol. 43, no. 3, p. 83, 2025. https://doi.org/10.1145/3722552

[23] G. Cai, X. Zheng, J. Guo and W. Gao, "Real-time identification of borehole rescue environment situation in underground disaster areas based on multi-source heterogeneous data fusion," Safety Science, vol. 181, pp. 106690, 2025. https://doi.org/10.1016/j.ssci.2024.106690

[24] S. Yao, Q. Wu, Q. Kang, Y. W. Chen and Y. Lu, "An interpretable XGBoost-based approach for Arctic navigation risk assessment," Risk Analysis, vol. 44, no. 2, pp. 459–476, 2024. https://doi.org/10.1111/risa.14175

[25] X. Li, X. Li, J. Jia, L. Li, J. Yuan and Y. Gao, "A high accuracy and adaptive anomaly detection model with dual-domain graph convolutional network for insider threat detection," IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1638–1652, 2023. https://doi.org/10.1109/TIFS.2023.3245413

[26] X. Liu, J. Lu, X. Chen, Y. H. C. Fong, X. Ma and F. Zhang, "Attention based spatio-temporal graph convolutional network with focal loss for crash risk evaluation on urban road traffic network based on multi-source risks," Accident Analysis & Prevention, vol. 192, p. 107262, 2023. https://doi.org/10.1016/j.aap.2023.107262

[27] R. Reka, R. Karthick, R. S. Ram and G. Singh, "Multi head self-attention gated graph convolutional network based multi-attack intrusion detection in MANET," Computers and Security, vol. 136, pp. 103526, 2024. https://doi.org/10.1016/j.cose.2023.103526

[28] X. Wang, X. Wang, B. Jiang, J. Tang and B. Luo, "Mutualformer: Multi-modal representation learning via cross-diffusion attention," International Journal of Computer Vision, vol. 132, no. 9, pp. 3867–3888, 2024. https://doi.org/10.1007/s11263-024-02067-x

[29] A. N. Anang, O. E. Ajewumi, T. Sonubi, K. C. Nwafor, J. B. Arogundade and I. J. Akinbi, "Explainable AI in financial technologies: Balancing innovation with regulatory compliance," International Journal of Science and Research Archive, vol. 13, no. 1, pp. 1793–1806, 2024. https://doi.org/10.30574/ijsra.2024.13.1.1870