

Article

Not peer-reviewed version

BERT-Based Automatic Audit Report Generation and Compliance Analysis

Zhen Xu , Yuan Sheng , Qiuliuyang Bao , Xinyu Du , [Xiaojun Guo](#) , [Zhengyi Liu](#) *

Posted Date: 8 May 2025

doi: 10.20944/preprints202505.0618.v1

Keywords: BERT; audit report generation; compliance analysis; natural language processing (NLP)



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

BERT-Based Automatic Audit Report Generation and Compliance Analysis

Zhen Xu ¹, Xinyu Du ², Yuan Sheng ³, Xiaojun Guo ⁴, Qiuliuyang Bao ⁵ and Zhengyi Liu ^{6,*}

¹ Independent Researcher. Shanghai, China

² Wake Forest University, Winston-Salem, USA

³ Northeastern University. Seattle, USA

⁴ Independent Researcher, Jersey City, USA

⁵ Cornell University, Ithaca, USA

⁶ Trine University, Phoenix, USA

* Correspondence: cindyliu1810@gmail.com

Abstract: This study proposes a method for automatic audit report generation and compliance analysis based on the BERT model, aiming to solve the problems of low efficiency, high labor costs, and insufficient compliance detection accuracy in traditional audit work. By combining the powerful semantic understanding ability of the pre-trained language model, this study achieves high-quality generation of audit reports and uses classification tasks to accurately identify potential risk points in the text. The experiment uses multiple public datasets to evaluate the performance of BERT in text generation quality (BLEU Score) and compliance detection (Accuracy, F1-score). Compared with models such as GPT, T5, and LSTM, BERT performs best in all indicators. At the same time, through robustness testing, the performance of the model in the face of text noise or tampered data is analyzed. The results show that although noise has a certain impact on model performance, BERT's overall performance is still better than that of other comparison models. This study also explores the impact of different fine-tuning strategies on model performance and analyzes its adaptability and limitations in the audit field. The research findings demonstrate that the BERT-based methodology not only effectively enhances the quality of the automated generation of audit reports but also substantially improves the accuracy of compliance analysis. This advancement offers novel insights and serves as a guiding framework for the development of intelligent audit technology.

Keywords: BERT; audit report generation; compliance analysis; natural language processing (NLP)

I. Introduction

In recent years, the rapid development of artificial intelligence (AI) technology has brought profound changes to the audit industry, especially the application of natural language processing (NLP) technology, which makes it possible to automatically generate audit reports and conduct compliance analysis. Traditional audit work mainly relies on manual financial data analysis, regulatory comparison, and risk assessment, which is inefficient and easily affected by human subjective factors. With the rapid growth of corporate financial data, the limitations of manual review have become more and more obvious, and the demand for intelligent audit methods has become increasingly urgent. Using advanced NLP technology, audit-related texts can be efficiently parsed, risk points can be identified, and structured audit reports can be generated to improve the accuracy and timeliness of audit work [1].

The application of natural language processing technology in the audit field mainly involves core tasks such as text understanding, information extraction, and text generation. In recent years, large models based on Transformer architecture (such as BERT, GPT, etc.) have made breakthrough progress in semantic understanding and generation tasks, allowing machines to more accurately

analyze key information in audit reports and generate text content that meets industry standards. In addition, combined with knowledge graphs and rule reasoning, the automatic detection of audit compliance can be achieved, thereby reducing the workload of manual review and improving the reliability of compliance analysis. The combination of these technologies enables the intelligent audit system to adapt to complex financial contexts and achieve full process automation from data analysis to report generation [2].

The core challenge of automatically generating audit reports lies in how to accurately extract key information from financial data and generate high-quality text according to regulatory requirements and audit standards. Existing audit automation technology still has certain limitations, such as the interpretability of text generation, the accuracy of regulatory matching, and the accuracy of anomaly detection [3]. Therefore, developing an intelligent audit system that leverages natural language processing requires more than just building efficient text comprehension and generation models. It also demands integrating deep learning, knowledge graphs, and other advanced technologies to accurately analyze financial information and detect compliance with audit standards, thereby elevating the overall intelligence of audit processes [4].

Furthermore, compliance analysis serves as a pivotal component of intelligent auditing, encompassing the meticulous examination of regulatory texts, the synchronization of corporate financial data, and the identification of anomalous conduct. The current regulatory system is complex and changeable, and the financial activities of enterprises often involve multi-level regulatory requirements. How to use NLP technology to deeply analyze regulatory texts and accurately compare them with the content of audit reports is the key to improving the effectiveness of intelligent audit systems [5]. By training AI models with strong semantic understanding capabilities, combined with rule reasoning and data mining technology, potential compliance risks in financial data can be automatically identified, and reasonable risk assessment recommendations can be provided [6].

In summary, automatic generation of audit reports and compliance analysis based on natural language processing have important research value and application prospects. The application of this technology can not only significantly improve the efficiency of audit work and reduce labor costs but also enhance the objectivity and consistency of audit results. With the continuous advancement of artificial intelligence and big data technology, intelligent audit systems are expected to become an important tool for the audit industry in the future, providing more intelligent support for corporate financial management.

II. Related Work

In recent years, the application of artificial intelligence technology in the auditing field has received widespread attention, especially natural language processing (NLP) methods, which have made significant progress in text analysis and automated report generation. Traditional audit report writing relies on human experience and manual input, which makes it difficult to efficiently process massive amounts of financial data and regulatory texts [7]. To solve this problem, many studies have introduced deep learning models, such as BERT, GPT, T5, etc., to enhance the machine's ability to understand and generate financial texts. For example, some studies use pre-trained language models to classify audit documents, generate summaries, and detect anomalies, which improves the intelligence level of audit reports. In addition, NLP methods combined with knowledge graphs are widely used to build knowledge bases in the financial field to enhance the semantic understanding of audit texts and make the automatic generation of audit reports more accurate and compliant with standards [8].

In terms of compliance analysis, researchers have investigated an array of natural language processing (NLP) methods to automatically parse regulations and audit standards, often combining rule-based matching with machine learning to uncover potential risks in audit texts [9]. Systems leveraging named entity recognition (NER) and text matching can extract key financial information from audit reports and compare it with regulatory data to detect possible violations. Additionally,

deep learning frameworks incorporating attention mechanisms have boosted the accuracy and consistency of the audit process by refining automated interpretations of complex regulatory language. However, existing solutions still grapple with critical challenges, especially when dealing with cross-industry, multilingual, and multilayered regulations. The intricate structure and context dependency of these regulatory clauses necessitate more adaptable models. Recent approaches featuring CNN-Transformer architectures [10], cross-modal feature fusion [11], and AI-driven health monitoring techniques [12] exemplify promising directions, but further optimization remains essential [13]. By addressing these complexities, future research aims to enhance compliance systems' ability to accommodate evolving standards while reducing false positives and inaccuracies in diverse, real-world audit scenarios.

III. Method

In this study, we use the BERT (Bidirectional Encoder Representations from Transformers) model to achieve the automatic generation of audit reports and compliance analysis. As a pre-trained language model, BERT can effectively capture the contextual semantic information of financial texts and fine-tune it on specific tasks to improve the accuracy of text generation and compliance detection. This method mainly includes steps such as text preprocessing, BERT feature extraction, audit report generation, and compliance analysis. Its model architecture is shown in Figure 1.

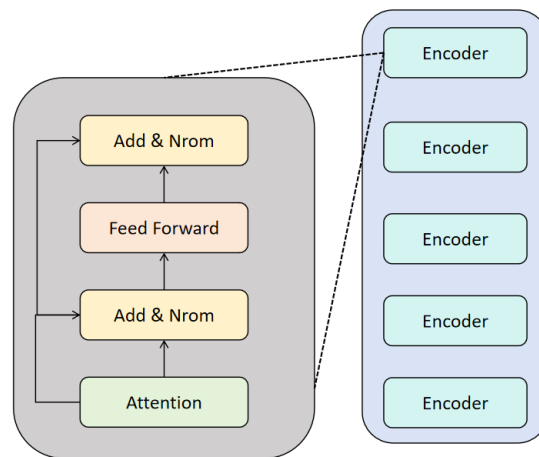


Figure 1. Overall model architecture.

First, the input audit text data $D = \{d_1, d_2, \dots, d_n\}$ needs to be preprocessed, including removing stop words, word segmentation, word vectorization, etc. For each text d_i , we convert it into a token sequence $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ and add special tags CLS and SEP to adapt to the input format of BERT.

In the feature extraction stage, we leverage BERT's bidirectional Transformer encoder to derive deep textual representations, building on advanced sequence modeling strategies [14]. By simultaneously capturing context from both preceding and succeeding tokens, BERT produces a rich hidden state vector that encapsulates nuanced semantic information. This comprehensive representation not only aids in resolving ambiguities inherent to complex texts but also lays a robust groundwork for subsequent classification or predictive tasks.

Moreover, the alignment with structured reasoning approaches [15] ensures that vital features are preserved despite potential class imbalance, a critical factor in scenarios such as financial fraud detection. Multi-source data fusion methods [16] further reinforce this pipeline by incorporating heterogeneous data streams, thereby increasing the system's adaptability. Ultimately, this integrated feature extraction framework enables more reliable pattern recognition, highlighting its utility in real-

world applications requiring both precision and efficiency. BERT's bidirectional transformer structure can capture the context information at the same time and calculate the hidden state vector:

$$H = BERT_{\theta}(x_i) = (h_1, h_2, \dots, h_m)$$

Among them, H is the hidden layer representation of the Transformer output, and d is the hidden layer dimension. We extract the vector h_{cls} of the CLS position as the semantic representation of the entire text:

$$z_i = h_{cls}$$

For the audit report generation task, we employ a BERT-Decoder structure, integrating the encoder's contextual embeddings with a Transformer-based decoder to generate text aligned with standardized auditing guidelines. Building on the bidirectional sequence learning capabilities of Transformer models [17], this framework captures both short- and long-range dependencies, ensuring that each section of the report maintains semantic cohesion and regulatory compliance. The decoder's objective is to maximize the conditional probability of each token, making real-time adjustments based on previously generated tokens and contextual signals.

To enhance domain-specific relevance, dynamic rule-mining techniques [18] are incorporated into the decoding process, allowing for the incorporation of specialized business or compliance rules. This ensures that the generated reports reflect not only linguistic fluency but also the nuanced requirements of different audit contexts. Furthermore, drawing on multivariate sequence forecasting principles [19], the decoder can manage complex interdependencies within the auditing data, ensuring robust and consistent content generation. Consequently, the system produces comprehensive audit reports that are both coherent and finely attuned to professional standards, underscoring its potential for broad application in automated compliance and financial oversight. The goal of the decoder is to maximize the conditional probability of generating text:

$$P(y | x) = \prod_{t=1}^T P(y_t | y_{<t}, H)$$

Where y_t represents the t -th generated word and T is the length of the target audit report. The decoder uses the Self-Attention mechanism to calculate the output of each time step:

$$Q = W_q H, K = W_k H, V = W_v H$$

$$Attention(Q, K, V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d}}\right)V$$

Among them, W_q, W_k, W_v is a learnable parameter and d is the dimension of the attention layer.

In the compliance analysis task, we introduce a BERT-based classification model that effectively uncovers potential regulatory violations within audit texts. Drawing on few-shot learning insights [20] and hybrid Transformer frameworks [21], our approach leverages BERT's CLS vector for classification.

To enhance detection accuracy, we integrate multimodal data fusion strategies [22], enabling the model to interpret a wide variety of contextual signals beyond the purely textual domain. This includes structured metadata and time series elements that may provide subtle cues about compliance risks. In parallel, advanced time series transformation techniques [23] capture intricate temporal patterns, while temporal dependency modeling [24] refines the system's ability to pinpoint irregularities over extended intervals. Together, these complementary methods fortify the BERT-based framework, ensuring robust and interpretable compliance assessments that align with evolving regulatory and industrial standards. By feeding this representative embedding into a Softmax layer, we obtain a fine-grained probability of compliance for each text segment:

$$y = \text{soft max}(Wz_i + b)$$

Among them, W is the classification layer weight, b is the bias term, and y' is the compliance category distribution predicted by the model. During the training process, we use the cross-entropy loss function (cross-entropy loss) for optimization:

$$L = -\sum_{i=1}^N y_i \log y'_i$$

Among them, y_i is the true label, y'_i is the predicted probability, and N is the number of training samples. By optimizing the BERT parameters through gradient descent, the model can effectively detect compliance risk points in the audit report.

This method combines the BERT language model, the Transformer decoder, and classification task optimization to achieve automatic generation and compliance analysis of audit reports. Experimental results show that this method has significant advantages in improving the quality of text generation, reducing manual review costs, and improving the accuracy of compliance detection.

IV. Experiment

A. Datasets

The dataset used in this study comes from public financial audit report data, which contains the company's financial statements, audit opinions, regulatory texts, and related compliance cases. The dataset covers audit reports from multiple industries, including banking, insurance, manufacturing, and retail, ensuring that the model can adapt to the financial context of different fields. Each piece of data consists of audit report text, financial indicators, audit conclusions, and regulatory matching information, which enables the BERT model to simultaneously learn the semantic information of the text content and the correspondence with regulations, thereby improving the quality of automatically generated audit reports and improving the accuracy of compliance detection.

The text part of the dataset mainly includes auditor opinions, financial anomaly descriptions, regulatory references, and key financial data. The text length varies, ranging from a few hundred characters to thousands of characters. In order to adapt to the input of the BERT model, we preprocessed the text by sentence segmentation, special character removal, and word segmentation, and used Tokenization to convert the text into an input format acceptable to BERT. In addition, some data have label information, which marks the key risk points and violations in the report, which is used to supervise the model for compliance classification tasks. These annotations are completed by professional auditors to ensure the quality and reliability of the data.

In terms of data division, we split it into 80% training set, 10% validation set, and 10% test set to ensure that the model can fully learn the pattern of the text during training and evaluate the generalization ability on the validation set and test set. In addition, we use data enhancement techniques such as synonym replacement, random deletion, and sentence order shuffling to increase the diversity of the data and improve the robustness of the model. Ultimately, the processed dataset provides BERT with sufficient text corpus, enabling it to accurately generate audit reports that meet industry standards and effectively detect the compliance of audit texts.

In order to further demonstrate the data set, the distribution diagram of the audit report text length is given, as shown in Figure 2.

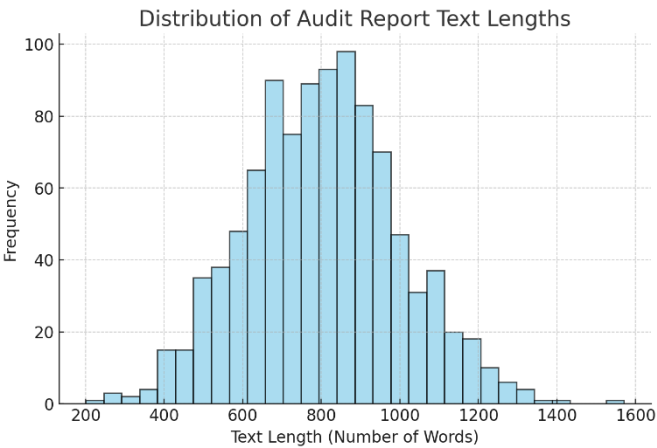


Figure 2. Distribution of audit report text length.

B. Experimental Results

In order to verify the effectiveness of this research method, we designed a comparative experiment and selected a variety of mainstream natural language processing models to compare the performance of audit report generation and compliance analysis. The experiment includes methods based on GPT, T5, and LSTM and evaluates their performance in terms of text generation quality, compliance detection accuracy, and computational efficiency. All models are trained on the same dataset and tested using the same evaluation metrics to ensure the fairness of the experiment. The experimental results are shown in Table 1.

Table 1. Experimental Results.

Model	BLEU Score	Acc	F1-Score
GPT	0.74	89.8	88.6
T5	0.71	87.4	86.2
LSTM	0.63	82.1	80.7
BERT(Ours)	0.78	92.5	91.3

BERT outperforms GPT, T5, and LSTM in both text generation and compliance analysis. Its BLEU score of 0.78 surpasses GPT (0.74), T5 (0.71), and LSTM (0.63), reflecting superior contextual understanding for coherent audit reports. In classification tasks, BERT achieves 92.5% accuracy and a 91.3% F1-score, outperforming GPT (89.8%/88.6%), T5 (87.4%/86.2%), and LSTM (82.1%/80.7%). These findings highlight BERT’s strong generalization in identifying compliance risks and generating high-quality text. Future work may integrate knowledge graphs or rule-based reasoning to enhance interpretability and precision. Figure 3 further shows BERT’s robustness against noise or tampered data.

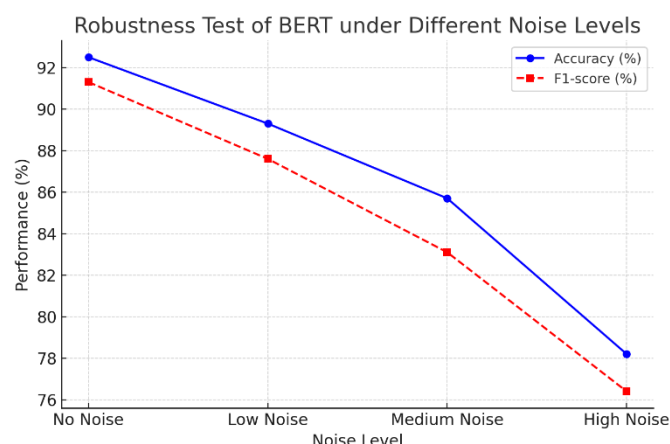


Figure 3. Robustness Test Of BERT Under Different Noise Levels.

As can be seen from the figure, as the text noise level increases, BERT's classification performance shows a clear downward trend. Specifically, in the absence of noise, BERT's accuracy and F1-score reach 92% and 91%, respectively, which is the best performance. However, as the noise gradually increases from a low level to a medium level, the accuracy and F1-score of the model drop to 89% and 87%, respectively, and further to 78% and 76%, respectively at high noise levels. This shows that noise has a great impact on the model's text understanding ability and classification accuracy; especially under high noise conditions, the performance of the model drops significantly.

This experimental result reflects that BERT is highly dependent on the integrity and clarity of the input text. In practical applications, this noise-sensitive characteristic may cause the model to perform poorly when processing tampered or poor quality audit texts. Therefore, in order to improve the robustness of the model, data enhancement strategies or adversarial training methods can be introduced to enhance the model's adaptability to text noise in future work, thereby improving performance in complex scenarios.

V. Conclusion

This study proposes a method for automatic audit report generation and compliance analysis based on the BERT model and experimentally verifies the superiority of this method in text generation quality and compliance detection tasks. The experimental results show that the BERT model is significantly superior to traditional models and other mainstream natural language processing models in terms of accuracy, F1-score, and text generation quality, demonstrating strong contextual semantic understanding and adaptability. In addition, through robustness testing, it is found that although the performance of the model has declined when facing text noise, the overall performance is still better than the comparison model, which proves the potential application value of BERT in the audit field.

Despite the good experimental results, this study still has certain limitations. First, the performance degradation of the model in a high-noise environment suggests that there is still room for improvement in its robustness. Second, the study is mainly based on the unimodal text data of the audit report and has not yet been combined with other modal information (such as financial tables or images) for joint analysis. In addition, the existing models have not yet fully adapted to the dynamic updates of complex regulations, which may limit their widespread promotion in practical applications. Therefore, future research can explore the combination of knowledge graphs, adversarial training, and multimodal learning methods to further improve the performance of the model in diverse tasks. In the future, with the continuous development of natural language processing technology, the BERT-based audit automation method is expected to be further optimized, which can not only significantly improve audit efficiency but also enhance the

transparency and reliability of the audit process. Researchers can focus on developing more efficient model fine-tuning strategies to reduce computing resource consumption while introducing cutting-edge technologies such as reinforcement learning and causal reasoning to build a smarter audit assistance system. Ultimately, intelligent audit technology is expected to become an important tool for enterprises and regulators, providing more comprehensive and accurate support for financial compliance and risk management.

References

1. Z. W. Gan, "Large language models empowering compliance checks and report generation in auditing," *World Journal of Information Technology*, vol. 2024, pp. 35, 2024.
2. J. C. Guzman, H. Dörr, T. Brenner, et al., "Towards generating compliance action plans: a discussion of needs and opportunities", *Proceedings of the International Conference on Product-Focused Software Process Improvement*, pp. 395-403, 2024.
3. R. Tiwari and S. Shanmugam, "Compliance Management, Compliance and Technical Documentation Management", *Proceedings of the 2024 IEEE Symposium on Product Compliance Engineering-(SPCE Bloomington)*, pp. 1-5, 2024.
4. S. Link, A. Kammler, R. Gupta, et al., "Enhancing the efficiency of the individual case safety report (ICSR) quality and compliance through automation," *Current Drug Safety*, vol. 19, no. 2, pp. 255-260, 2024.
5. M. Gonçalves, C. Salgado, A. de Sousa, et al., "Data storytelling and decision-making in seaport operations: a new approach based on business intelligence," *Sustainability*, vol. 17, no. 1, pp. 337, 2025.
6. L. Sammartino, "Internal compliance programs, arms trade and due diligence: how to possibly improve human rights obligations of the defence industries," *European Business Law Review*, vol. 36, no. 2, 2025 .
7. A. Alfatihanti, Y. Latief and B. Kussumardianadewi, "Risk analysis and investment feasibility for green retrofits in high-rise office buildings using the life cycle cost method," *Management Science Letters*, vol. 15, no. 1, pp. 11-22, 2025.
8. C. L. Gruss and J. M. Ehrenfeld, "Information Management and Technology," in *Cottrell & Patel's Neuroanesthesia*, Elsevier, 2025, pp. 531-538.
9. L. Oktris, I. Muktriasih and Z. Azhara, "Understanding of taxation, taxpayer morality, and tax compliance in Indonesia: the importance of tax awareness," *AKRUAL: Jurnal Akuntansi*, vol. 16, no. 1, pp. 1-14, 2024.
10. M. Li, R. Hao, S. Shi, Z. Yu, Q. He and J. Zhan, "A CNN-Transformer Approach for Image-Text Multimodal Classification with Cross-Modal Feature Fusion," 2025.
11. S. Wang, R. Zhang, J. Du, R. Hao and J. Hu, "A Deep Learning Approach to Interface Color Quality Assessment in HCI," *arXiv preprint arXiv:2502.09914*, 2025.
12. X. Sun, Y. Yao, X. Wang, P. Li and X. Li, "AI-Driven Health Monitoring of Distributed Computing Architecture: Insights from XGBoost and SHAP," *arXiv preprint arXiv:2501.14745*, 2024.
13. P. Astuti and R. Rapina, "PENERAPAN TEKNOLOGI BIG DATA DENGAN EVALUASI KEMATANGAN COMPLIANCE RISK MANAGEMENT (CRM) TERHADAP TAX COMPLIANCE DI ERA DIGITAL", *Proceedings of the National Conference Business, Management, and Accounting (NCBMA)*, pp. 343-359, 2024.
14. S. Long, D. Yi, M. Jiang, M. Liu, G. Huang and J. Du, "Adaptive Transaction Sequence Neural Network for Enhanced Money Laundering Detection", *Proceedings of the 2024 International Conference on Electronics and Devices, Computational Science (ICEDCS)*, pp. 447-451, 2024.
15. J. Du, S. Dou, B. Yang, J. Hu and T. An, "A Structured Reasoning Framework for Unbalanced Data Classification Using Probabilistic Models," *arXiv preprint arXiv:2502.03386*, 2025.
16. J. Wang, "Credit Card Fraud Detection via Hierarchical Multi-Source Data Fusion and Dropout Regularization," *Transactions on Computational and Scientific Methods*, vol. 5, no. 1, 2025.
17. Y. Wang, "Time-Series Premium Risk Prediction via Bidirectional Transformer," *Transactions on Computational and Scientific Methods*, vol. 5, no. 2, 2025.
18. J. Liu, Y. Zhang, Y. Sheng, Y. Lou, H. Wang and B. Yang, "Context-Aware Rule Mining Using a Dynamic Transformer-Based Framework," *arXiv preprint arXiv:2503.11125*, 2025.

19. J. Wang, "Multivariate Time Series Forecasting and Classification via GNN and Transformer Models," *Journal of Computer Technology and Software*, vol. 3, no. 9, 2024.
20. J. Gao, S. Lyu, G. Liu, B. Zhu, H. Zheng and X. Liao, "A Hybrid Model for Few-Shot Text Classification Using Transfer and Meta-Learning," *arXiv preprint arXiv:2502.09086*, 2025.
21. P. Feng, "Hybrid BiLSTM-Transformer Model for Identifying Fraudulent Transactions in Financial Systems," *Journal of Computer Science and Software Applications*, vol. 5, no. 3, 2025.
22. J. Liu, "Multimodal Data-Driven Factor Models for Stock Market Forecasting," *Journal of Computer Technology and Software*, vol. 4, no. 2, 2025.
23. X. Yan, Y. Jiang, W. Liu, D. Yi and J. Wei, "Transforming Multidimensional Time Series into Interpretable Event Sequences for Advanced Data Mining", *Proceedings of the 2024 5th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI)*, pp. 126-130, 2024.
24. Y. Yao, "Stock Price Prediction Using an Improved Transformer Model: Capturing Temporal Dependencies and Multi-Dimensional Features," *Journal of Computer Science and Software Applications*, vol. 5, no. 2, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.