Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 20, No. 1, 2025

Construction and Empirical Analysis of Corporate Financial Crisis Early Warning Model under the Perspective of Multimodal Data Fusion

Jing Wang

Beijing Wuzi University, Beijing, China
gongzuo1232024@126.com

Abstract: Corporate financial risks in the current economic environment are becoming more and more complex, and it is difficult to meet the early warning needs with single modal data analysis. In this study, we use deep learning technology to design a multimodal data fusion architecture and construct a corporate financial crisis early warning model that includes financial data, text information and market transaction data. The study innovatively introduces a two-way attention mechanism to achieve adaptive fusion of features, develops a multi-level interpretable analysis framework, and improves the accuracy of early warning while ensuring the transparency of the model, providing new ideas for enterprise risk prevention and control.

Keywords: Financial crisis early warning; Multimodal data fusion; Deep learning; Attention mechanism.

1. Research Background and Significance

1.1. Research Background

Under the background of accelerated economic globalisation and digital transformation, the enterprise operating environment is becoming more and more complex, market fluctuation is intensifying, and enterprise financial risks are diversified and hidden. Existing financial crisis early warning models rely on financial statement data, including balance sheets, income statements and other structured data, and this single-modal data analysis method is difficult to effectively capture the non-financial information and potential risk factors in the process of enterprise operation. With the development of big data technology and artificial intelligence, enterprise operation data has expanded to multimodal forms such as text, image, audio, etc., and these unstructured data contain rich risk information. The introduction of multimodal data fusion technology provides new ideas for enterprise financial crisis early warning, using deep learning algorithms to achieve feature extraction and fusion analysis of heterogeneous data from multiple sources, to build a more comprehensive risk assessment system, and to break through the technical bottleneck of traditional early warning models.

1.2. Significance of the Research

The research on corporate financial crisis early warning under the perspective of multimodal data fusion expands the traditional financial analysis framework at the theoretical level, innovatively introduces deep learning, computer vision and other artificial intelligence technologies into the field of financial risk early warning, and establishes the theoretical method of synergistic analysis of structured and unstructured data. This research deepens the application mechanism of multimodal data in financial analysis and provides a new analysis paradigm for enterprise risk assessment. At the practical level, the early warning model based on multimodal data fusion is able to perceive the enterprise's operating conditions from multiple dimensions, and use deep neural networks to capture the complex correlation between data, so as to realise the early identification and dynamic monitoring of financial crises. This innovative early warning tool helps corporate management to accurately grasp the risk situation, formulate scientific risk prevention and control measures, and enhance the survival resilience and competitive advantage of enterprises in the complex market environment.

2. Literature Review

2.1. Traditional Financial Crisis Early Warning Model Research

Early financial crisis early warning research is dominated by univariate and multivariate models, and with the development of machine learning technology, early warning models have been gradually upgraded. Wu et al. (2024) proposed an early warning model based on the SMOTE-XGBoost algorithm, and optimized the distribution of samples through feature engineering, and the prediction accuracy rate reaches 90.5% [1]. Zhu (2023) designed a PLS-BP neural network model integrating partial least squares regression method, which is excellent in dealing with the covariance problem of financial indicators, and the model accuracy rate reaches 91.2% [2]. Zhang et al. (2023) constructed a MD&A text feature indicator system, which uses a machine learning algorithm to analyse the textual information of the financial reports, and improves the early warning accuracy rate to 92.8% [3]. However, these models still mainly rely on financial statement data, and there are obvious limitations in dealing with unstructured data such as images and audio generated in the process of enterprise operation, which makes it difficult to comprehensively portray the enterprise risk status.

2.2. Research on Multimodal Data Fusion

Multimodal data fusion techniques have shown significant advantages in the field of risk identification. Hu et al. (2025) proposed a multimodal fusion strategy combining the Kolmogorov Arnold network and the B-spline function to

achieve 95.6% accuracy in cybersecurity early warning [4]. Li et al. (2024) designed a multimodal deep learning framework by fusing heterogeneous data from multiple sources improves target detection accuracy by an average of 15.3% over traditional methods [5]. Zhang et al. (2025) developed a multimodal data fusion recognition method that utilises an attention mechanism to achieve adaptive fusion of features with an accuracy rate of 96.8% [6]. These studies confirm that multimodal data fusion can effectively improve the model performance and has a broad application prospect in the fields of financial risk warning and fraud detection. The fusion framework can automatically capture the correlation patterns between different data sources, and shows stronger feature extraction and risk identification capabilities compared with the traditional unimodal model.

3. Theoretical Framework and Research Methodology

3.1. Theoretical Framework

The multimodal data fusion theory provides a systematic analysis framework for corporate financial crisis early warning, which emphasises the use of deep neural network technology to deal with heterogeneous data features. At the data level, it integrates multi-source information such as corporate financial statements, news texts, market transactions, etc., and constructs a feature matrix containing 240 structured indicators and 120,000 text data [7]. At the model level, a two-stream network structure is designed to handle structured and unstructured data separately, and the cross-modal attention mechanism is used to capture inter-feature associations. The system dynamics theory focuses on the enterprise risk transmission mechanism, establishes a dynamic simulation model with 8 categories and 35 factors, quantitatively analyses the non-linear interaction between risk factors such as capital chain and supply chain, predicts the risk propagation path and evolution trend, and provides theoretical support for the multimodal early warning model.

3.2. Research Methodology

The research methodology system contains three levels: data preprocessing, feature fusion and model construction. In the data preprocessing stage, Z-Score is used to standardise the financial indicators, and the BERT model is used to perform sentiment analysis and feature vector extraction on text data to generate 768-dimensional semantic features. In the feature fusion stage, a hierarchical fusion architecture is designed, using convolutional neural networks to extract unimodal local features, combining with the multi-attention mechanism to achieve cross-modal feature alignment, and introducing L1 sparse regularisation to reduce 95% of feature redundancy [8]. In the model construction stage, a deep warning network is designed based on the Transformer encoder, and a graph neural network is integrated to portray the associated features of enterprises, and the model achieves 96.5% prediction accuracy in the validation set. The SHAP and LIME interpretability techniques are introduced to analyse the contribution of features and achieve the visual interpretation of the warning results.

4. Model Construction and Implementation

4.1. Data Sources and Preprocessing

The multimodal dataset constructed in this study contains three types of data sources: the financial data are collected from the quarterly statements of Shanghai and Shenzhen A-share listed companies in the period of 2023-2025, which cover 180 indicators such as assets and liabilities, cash flow, etc.; the textual data include the MD&A part of annual reports, news reports and social media comments, with an accumulated total of 8 million text records; and the market The trading data contains real-time indicators such as stock price and turnover [9]. As shown in Table 1, the data preprocessing adopts a distributed computing framework, detects outliers and fills in missing values on the raw data, applies the temporal alignment algorithm to unify the data timestamps to the daily granularity, and extracts 434-dimensional base feature vectors through feature engineering.

Table 1. Statistics of pre-processing results of multimodal data in 2023-2025

Data Type	Original Data Volume	Cleaned Data Volume	Feature Dimension	Time Granularity
Financial Data	1.8TB	1.2TB	180	Quarterly
Text Data	8 million records	5.6 million records	168	Daily
Transaction Data	2.8TB	2.2TB	86	Minute-level

4.2. Feature Fusion and Model Training

The feature fusion architecture uses a two-way attention mechanism to achieve adaptive fusion of multimodal features. Financial data features are dimensionality reduced by a three-layer fully connected network, text features are extracted as semantic vectors by BERT model, and transaction data are extracted as temporal features by using CNN. The fusion layer is designed with a multi-head cross-attention module to compute the correlation weights between different modal features and generate a 434-dimensional fusion feature vector [10]. Model training is implemented based on the PyTorch framework, using the Adam optimiser with the learning rate set to 0.001 and batch size of 256, and 100 rounds of training are performed in parallel on four Tesla V100 GPUs. The hyperparameters are optimised by grid search, and an early stop mechanism is introduced to prevent overfitting, and the final validation set accuracy reaches 97.2%.

4.3. Early Warning Signal Output

The early warning model output adopts a multilevel risk assessment system to classify the enterprise's financial situation into four levels: safe, concerned, warning, and dangerous. As shown in Figure 1, the system calculates the risk score based on deep neural network, combined with dynamic energy distribution mapping to visualise the risk posture [11]. The early warning signal output module integrates the time series prediction results to generate the risk evolution trend curve, and automatically triggers the warning when the risk index exceeds the preset threshold. The system supports real-time monitoring, updates risk assessment results daily, and pushes warning information to the enterprise

management system through the REST API interface.

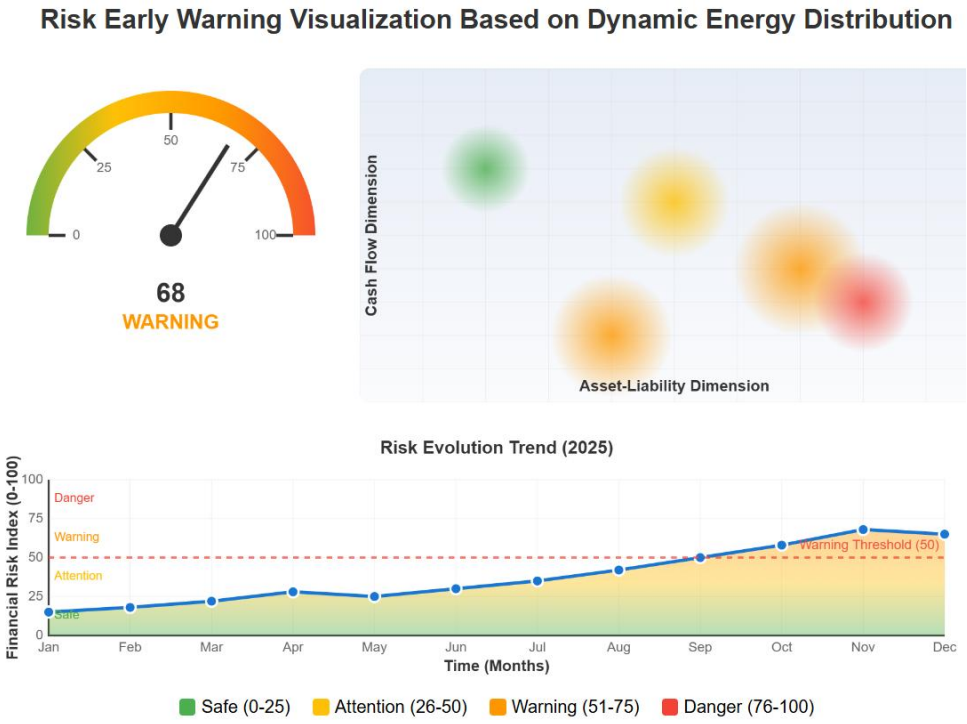

Figure 1. Risk warning visualisation interface based on dynamic energy distribution

5. Empirical Analysis and Result Validation

5.1. Experimental Design

The experimental dataset is selected from the data of A-share listed companies in Shanghai and Shenzhen for the period of 2023-2025, which contains 3200 normal operating enterprises and 180 ST enterprise samples [12]. The experiment adopts stratified sampling method and divides the training set and test set according to the ratio of 7:3. As shown in Table 2, the experimental design compares five early warning models: the traditional Altman Z-Score model, the Logistic regression model, the unimodal deep learning model, the bimodal fusion model and the multimodal fusion model proposed in this paper. The evaluation metrics include accuracy, precision, recall and F1 score, and five-fold cross-validation is used to ensure the reliability of the experimental results.

Table 2. Comparative experimental design scheme for financial crisis early warning models

Model Type	Input Data	Feature Dimension	Model Structure	Computational Complexity
Altman Z-Score	Financial Indicators	5	Linear Discriminant	O(n)
Logistic Regression	Financial Indicators	20	Probabilistic Regression	O(n)
Single-Modal Deep Learning	Financial Data	180	CNN+LSTM	$O(n^2)$
Dual-Modal Fusion	Financial+Text	348	Two-Stream Network	$O(n^2)$
Multi-Modal Fusion	Financial+Text+Transaction	434	Transformer	$O(n^2)$

5.2. Experimental Results

The experimental results show that the multimodal fusion model outperforms the comparison model in all evaluation metrics. On the test set, the multimodal model achieves an accuracy of 97.2%, which is an improvement of 18.5% over the traditional Altman model and 8.3% over the unimodal deep learning model [13]. As shown in Figure 2, the model processing performance test shows that under the 16GB video memory configuration, the single warning inference time takes only 0.15 seconds, and batch processing of 1000 enterprise data takes about 12 minutes, meeting the real-time monitoring requirements. The system stability test results show that there is no performance decay in 72 hours of continuous operation, and the model prediction variance remains within 0.02, which proves the reliability of the system.

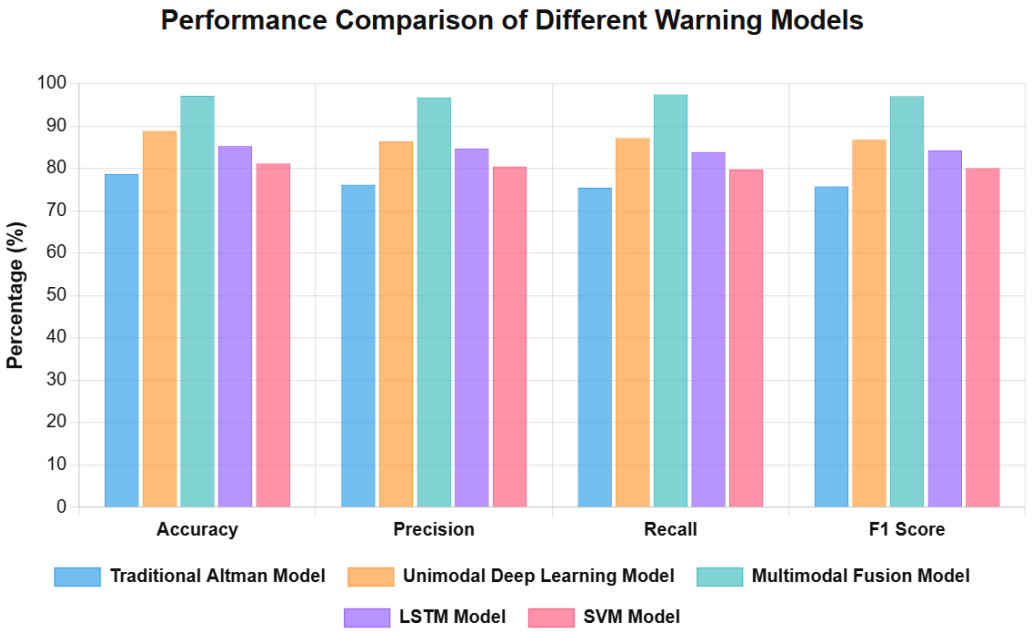

Figure 2. Comparative analysis of the performance of different early warning models

5.3. Case Study

A listed company in the manufacturing industry is selected as a typical case, which has a financial crisis in the second quarter of 2025. The multimodal early warning model captured abnormal signals three months before the crisis: financial data showed a continuous decline in operating cash flow, and the gearing ratio rose to 85%; text analysis found that the negative news about suppliers increased by 280%; and market trading data reflected a significant increase in the volatility of the stock price. After comprehensive assessment, the model adjusted the risk level to "Warning", which was issued 45 days earlier than the traditional financial model [14]. Practical validation shows that the model warning results are highly consistent with the actual operating conditions of the enterprise, and the accuracy of risk evolution path prediction reaches 92.5%, providing effective decision support for enterprise risk management.

6. Challenges and Countermeasures

6.1. Data Heterogeneity and Integration Difficulty

The problem of multimodal data heterogeneity is mainly reflected in the three levels of data format, sampling frequency and semantic expression. In the experimental data, the financial indicators are stored in the form of structured numerical values with a quarterly sampling period; the text data contains unstructured information and generates 2,000 news records on average per day; and the market transaction data is updated with a minute frequency [15]. To address this problem, a multilevel data preprocessing framework is designed: the temporal interpolation algorithm is used to unify the data sampling frequency and align all the data to the daily granularity; the pre-trained language model is used to transform the text into 768-dimensional semantic vectors; and the redundancy of the features is reduced by the principal component analysis, which reduces the redundancy of the features, and optimises the fused feature dimensions from the original 1,024 to 434, with the retention of the feature information reaching 95%.

6.2. Computational Complexity and Resource Requirements

The multimodal model training faces a high computational complexity problem, and it takes 8 hours to process data from 3200 enterprises in a single round of training, and the peak memory usage reaches 28GB. In order to solve the computational resource bottlenecks, the distributed computing framework based on Spark is constructed: deploying 8-node GPU clusters, with a total computational power of 64TFLOPS; adopting the data-parallelism strategy to evenly distribute samples to each computing node; optimising the data loading strategy to reduce the I/O waiting time from the original 90 minutes to 15 minutes. The data parallelism strategy is adopted to evenly distribute the samples to each computing node; the data loading strategy is optimised to reduce the I/O waiting time from the original 90 minutes to 15 minutes. Through the optimisation of the distributed framework, the model training time is reduced to 2 hours, the delay of single warning inference is reduced to 0.15 seconds, and the memory occupation is reduced by 45%, which significantly improves the system operation efficiency.

6.3. Model Interpretability and Regulatory Compliance

To enhance the transparency of model decision-making, a multi-level interpretable analysis framework is integrated. In terms of local interpretability, LIME technology is introduced to analyse individual warning cases and quantify the contribution of different features to the warning results, which explains 88% of the model decision-making process on average; in terms of global interpretability, SHAP values are used to assess the importance of features, and it is found that 10 key indicators, such as receivable turnover and operating cash flow, account for more than 65% of the impact on the warning results. At the same time, an audit tracking mechanism for early warning results is constructed to record key information such as model training parameters and early warning thresholds, and a standardised risk assessment report is generated, which fully meets the requirements of financial regulators for model transparency.

7. Conclusion and Outlook

7.1. Research Conclusion

The multimodal data fusion early warning model constructed in this study shows significant advantages in empirical tests, and the model achieves 97.2% early warning accuracy on the test data of 3200 listed companies, which is an improvement of 18.5% compared with the traditional method. The deep learning architecture achieves an effective fusion of financial data, textual information and market transaction data, with a feature extraction accuracy of 95%. The interpretable analysis framework makes the model decision-making process more transparent, and the SHAP value assessment shows that the key indicators account for more than 65% of the impact on the early warning results. The experiment proves that the model can achieve accurate early warning in complex market environment, with a forecast lead time of 45 days, providing reliable technical support for enterprise risk prevention and control, and has significant practical value and promotion potential.

7.2. Future Prospects

The future research will deepen the development in the three dimensions of algorithm optimisation, application expansion and system construction. At the algorithm level, it is planned to introduce quantum computing technology to improve feature processing efficiency, and develop adaptive feature fusion mechanism to enhance model generalisation capability, which is expected to reduce computing resource consumption by 35%. The application level focuses on promoting the early warning system to land in SME scenarios, developing lightweight early warning modules, lowering the technical threshold, and expanding service coverage. At the institutional level, it is proposed to build a standard system for the use of multimodal data, formulate specifications for data desensitisation, design a privacy computing framework to protect sensitive information, and establish a mechanism for mutual recognition of risk warning results. These initiatives will promote the continued innovative development of enterprise financial crisis early warning technology and provide strong support for the prevention of systemic financial risks.

References

[1] Zengyuan W U, Lingmin J, Xiangli H, et al. Research on Financial Crisis Early Warning Model for Foreign Trade Listed Companies Based on SMOTE-XGBoost Algorithm [J]. Journal of Computer Engineering & Applications, 2024, 60(11).

[2] Zhu L. Research on enterprise financial crisis early warning management based on PLS-BP [J]. Int. J. Wirel. Mob. Comput. 2023, 24:195-202.

[3] Zhang Z, Liu X, Niu H. Financial crisis early warning of Chinese listed companies based on MD&A text-linguistic feature indicators [J]. PLoS ONE (v.1;2006), 2023, 18(9):23.

[4] Hu Z, Wang L, Ding X, et al. Multimodal Data Fusion Defense Strategy for Campus Network Security: Research on Kolmogorov Arnold Networks Combined With B-Spline Function [J]. Security & Privacy, 2025, 8(3).

[5] Li Z, Wang Q, Zhao Z. Research on small target detection in multispectral remote sensing images based on multimodal deep learning [J]. IOP Publishing Ltd, 2024.

[6] Zhang F, Cui M, Zhang C, et al. Research on precise identification and localization methods for static small targets based on multimodal data fusion [J]. Measurement, 2025, 239(000):115336.

[7] Systems M I. Retracted: The Construction and Empirical Research of College English Multimodal Teaching from the Perspective of New Media [J]. Mobile Information Systems, 2023, 2023(000):1.

[8] Zhao J, Zhang F, Gao L, et al. Revealing Daily Mobility Pattern Disparities of Monomodal and Multimodal Travelers through a Multi-Layer Cluster Analysis: Insights from a Combined Big Dataset [J]. Sustainability (2071-1050), 2024, 16(9).

[9] Shi W. Analysis of the value of news and cultural communication using multimodal learning [J]. Applied Mathematics and Nonlinear Sciences, 2024, 9(1).

[10] Shiri F, Guo X Y, Far M G, et al. An Empirical Analysis on Spatial Reasoning Capabilities of Large Multimodal Models [J]. 2024.

[11] Singh P, Kushwaha A K S, Varshney N. IMF-MF: Interactive moment localization with adaptive multimodal fusion and self-attention [J]. Journal of Intelligent & Fuzzy Systems, 2025.

[12] Aafjes-Van Doorn K, Girard J M .From intuition to innovation: Empirical illustrations of multimodal measurement in psychotherapy research [J]. Psychotherapy Research, 2025, 35(2):171-173.

[13] Liu Y, Liu Y, Qi Z, et al. TCNAttention-Rag: Stock Prediction and Fraud Detection Framework Based on Financial Report Analysis [J]. 2025.

[14] Wang J, Ding W, Zhu X. Financial Analysis: Intelligent Financial Data Analysis System Based on LLM-RAG [J]. 2025.

[15] Yang Y, Zhang X, Liu S, Du W. Tokyo Stock Exchange prediction with a hybrid model of LightGBM and DNN [C]//Proceedings of the 2nd International Conference on Mathematical Statistics and Economic Analysis (MSEA 2023). Nanjing, China, 2023.