Hybrid Early Warning System for Financial Crises in Listed Companies Using Attention Mechanism and LightGBM

Huaipeng Li, Jing Li
School of Economics and Management, Shangqiu Institute of Technology, ShangQiu 476000, China
E-mail: HuaipengLii@outlook.com

Keywords: Attention mechanism, LightGBM, financial crisis early warning, explainable machine learning, unbalanced data classification

Received: January 18, 2026

Aiming to address the accuracy and interpretability needs of early warning systems for financial crises in listed companies, this paper proposes a hybrid early warning model that integrates an attention mechanism and LightGBM. When traditional financial early warning models deal with high-dimensional and nonlinear financial data, it is often difficult to capture the differential influence of key indicators, and they are sensitive to quasi-imbalanced data. This study combines the attention mechanism and the LightGBM classifier to perform weighted processing on key financial indicators and trains the model to handle large-scale imbalanced data. The experiment uses data from ST companies and non-ST companies in the Chinese A-share market from 2015 to 2020, constructing a panel dataset of 2,560 listed companies. The model's warning accuracy on the test set is 88.67%, with an AUC value of 0.942, which is 3.2%, 4.8%, and 5.7% higher than the accuracy rates achieved by using LightGBM, XGBoost, and random forest models alone. The attention weights enhance the model's interpretability and identify key warning indicators, such as cash flow ratios and return on assets.

Povzetek: Članek predlaga hibridni model za zgodnje opozarjanje finančnih kriz, ki z mehanizmom pozornosti uteži ključne finančne kazalnike in jih klasificira z LightGBM za boljšo obravnavo visokodimenzionalnih neuravnoteženih podatkov ter večjo razložljivost.

1 Introduction

As a crucial branch of financial risk management, the early detection of financial crises in listed companies has been a longstanding concern for both academic and practical circles [1]. The financial crisis not only poses a serious threat to the continuous operation of enterprises themselves, but also may trigger a chain reaction, impacting investors' interests and financial market stability [2, 3]. Traditional early warning models mostly rely on linear statistical methods. Although they have shown certain effects in early research, their discriminant ability and generalization performance have obvious limitations when faced with high-dimensional nonlinear financial data [4]. With the rapid development of machine learning technology, the LightGBM (Light Gradient Boosting Machine) algorithm, based on gradient boosting decision trees, has been gradually introduced into the field of financial forecasting due to its high efficiency and excellent classification performance [5]. This algorithm can effectively process large-scale features and capture complex patterns, providing new possibilities for building a more accurate early warning system [6]. The core goal of this study is to construct a financial crisis early warning model for listed companies suitable for China's A-share market, and solve the problems of insufficient accuracy and weak interpretability of the existing models when dealing with high-dimensional and unbalanced financial data

There are three core problems when the existing machine learning models are applied to the financial crisis warning of listed companies in China's A-share market. First, there is a significant information asymmetry in the A-share market, and the financial data of listed companies is whitewashed, so a single model is difficult to capture the hidden crisis signals behind the data [7, 8]. Secondly, there is a certain deviation between the financial crisis label (ST) of the A-share market and the actual business situation of enterprises, and some ST enterprises are hooded due to non-operating factors [9, 10]. The existing model does not optimize this feature. Thirdly, when dealing with high-dimensional and unbalanced financial time series data of A-shares, mainstream machine learning models (such as single LightGBM and XGBoost) cannot dynamically identify key early warning features at different time nodes, resulting in insufficient timeliness and accuracy of early warning [11]. However, the existing research has not carried out the deep integration research of attention mechanism and gradient boosting tree for the above specific problems in the A-share market, which has become the core knowledge gap of this study.

Current research mostly focuses on the application of a single model or simple integration strategy, lacking systematic exploration of the deep integration of the two types of technologies [12]. Especially in the financial early warning scenario, how to effectively utilize the attention mechanism to enhance the time series modeling capabilities of LightGBM remains a topic worthy of in-depth study. This study aims to construct a collaborative early warning framework that integrates an attention mechanism and LightGBM, dynamically capturing key time series information in financial indicators through attention weights, and achieving efficient and accurate classification prediction with the help of LightGBM. This framework is not only committed to improving early warning accuracy but also focuses on the interpretability of the model, providing a clearer basis for decision-making in risk identification.

The significance of this study lies in promoting the frontier development of financial early warning methodology and providing a novel and effective technical path for risk identification under complex financial data. By introducing an attention mechanism, the model can better capture the dynamic characteristics and hidden laws underlying the formation of financial crises, thereby issuing more reliable early warning signals at an earlier stage.

2 Theoretical basis and key technical mechanism of financial crisis early warning

2.1 Theoretical framework of financial crisis early warning

The earliest research on crisis early warning primarily focused on the field of large-scale macroeconomic crises, while research on enterprise early warning at the microeconomic level began in the 1960s [13, 14]. With the widespread application and popularization of modern information technology in the financial field, financial crisis early warning has gradually become a key focus in academic circles, yielding fruitful results [15]. Financial crisis is characterized by complexity, uncertainty, and controllability, so the related theories involved are quite rich [16].

Crisis management theory primarily focuses on how enterprises respond to crises, with its core idea being to predict potential emergencies through planning, control, and other means, thereby reducing or avoiding losses as much as possible and ensuring the smooth implementation of enterprise policies [17, 18]. This theory can provide a theoretical basis for this study in the aspects of planning, control, and forecasting in financial early warning for listed companies. As a comprehensive discipline, enterprise crisis management theory encompasses knowledge and methods from various fields, including organizational structure, human resource management, and strategic and environmental analysis [19, 20].

Social and economic development typically progresses through four stages: prosperity, recession, depression, and recovery. As the economic environment continues to change, each economic entity will adjust its behavior accordingly, thus making the economic operation show a certain regularity. From a macroeconomic perspective, the economic cycle theory examines various factors that cause enterprise crises and is a theory that studies the law of cycle fluctuations in economic operations [21].

Risk management theory emphasizes the continuous management and dynamic monitoring of an enterprise's financial situation, as enterprises face unique risks and their consequences at all stages of their life cycle [22]. This theory constitutes the core foundation of this research, and financial crisis early warning is essentially a systematic control of enterprise financial risks. Improve the ability to respond to price risks at the enterprise level, and promote the improvement and construction of hardware systems at the national level. The general process of risk management includes several key stages: first, risk assessment, that is, through the analysis, classification, and judgment of risk factors, the corresponding evaluation standards are set according to the risk tolerance of enterprises, and a comprehensive assessment of specific industries is carried out to clarify the correlation among risk factors and their relative importance [23, 24]. Secondly, risk identification involves determining the risk category and then identifying the key dominant risk factors through systematic investigation and research.

2.2 Theoretical basis and applicability of attention mechanism

In deep learning, especially in natural language processing, the attention mechanism is a core technology. It simulates how humans selectively focus on important information while observing, focusing on key points, and ignoring irrelevant information. After deep learning introduces attention mechanism, the model can accurately locate key information from massive data, thereby improving performance and efficiency [25].

The mechanism employs an encoder-decoder framework, and usually the encoder will encode all the input information as a unified intermediate representation [26]. The attention mechanism is a mapping relationship, which uses a mapping matrix to obtain a vector through queries and key-value pairs. Calculate the dot product similarity between the query and the key, and normalize the processing to obtain weights through the *Softmax* function. Finally, use the weights to obtain the output result through the weighted sum of the values, as shown in formula (1).

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \quad (1)$$

Among them, Q, K, and V represent Query, Key, and Value respectively, K^T represents the transposition of K, and d_k is the dimension of Value, which is used as a scaling factor to prevent the gradient from disappearing.

2.3 Theoretical support and performance characteristics of LightGBM algorithm

Lifting algorithm is an ensemble learning method. It is easier to find a rough classification rule than an accurate classification rule given a data set. Therefore, the algorithm generates a series of weak classifiers by continuously learning from weak learners, which are then integrated to create a robust classifier [27, 28]. The boosting tree uses classification or regression trees as its weak learners and performs well. The linear aggregation of a large number of trees can well adapt to the training data and depict the complex nonlinear relationship between input and output variables. As the weak learner of the gradient boosting tree, the CART tree is a decision tree model characterized by its binary nature, where each node is precisely divided into two child nodes. The gradient boosting tree XGBoost belongs to a kind of boosting tree, which is an improvement of GBDT and essentially consists of k CART trees [29]. LightGBM is a gradient lifting framework proposed by Microsoft. It is also improved from GBDT, mainly focusing on reducing the time consumption of GBDT when processing a large amount of data. Figure 1 shows the LightGBM algorithm architecture.

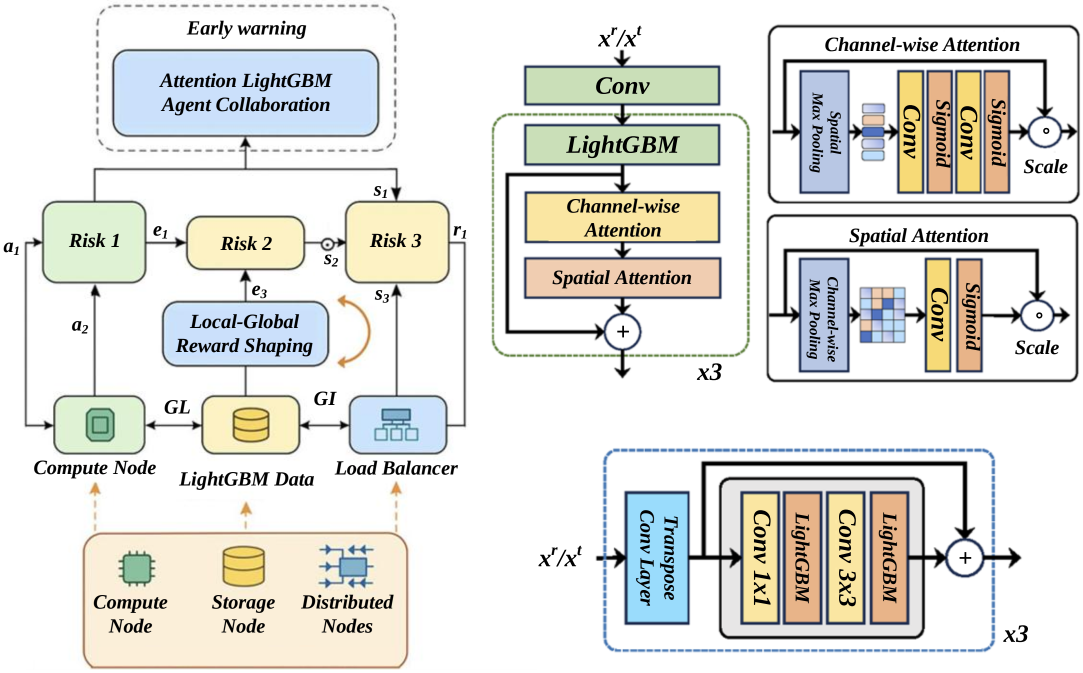

Figure 1: LightGBM algorithm architecture

The objective function construction form of XGBoostis similar to finding the optimal solution of quadratic equation of one variable. Given that the training data set T contains several sample points (x_1, y_1) to (x_n, y_n), the loss function is $l(y_i, \hat{y}_i)$, and the regularization term is $\Omega(f_k)$, the overall objective function can be expressed as Equation (2):

$$L(\varnothing)=\sum_i l(y_i,\hat{y}_i)+\sum_k \Omega(f_k) \quad (2)$$

$L(\emptyset)$ is an expression in linear space: i is the *i-th* sample; k is the *k-th* tree and $\hat{y}_i$ is the predicted value of the *i-th* sample x_i. After second-order Taylor expansion, regularization term expansion, and combining the coefficients of the first term and the quadratic term, the final objective function can be obtained as shown in equation (3). Among them, T is the number of leaf nodes, and γ is the L2 regularization coefficient.

$$L^{(t)}=\sum_{j=1}^{T}\left[G_j w_j+\frac{1}{2}\left(H_j+\lambda\right)w_j^2\right]+\gamma T \quad (3)$$

The cumulative sum of the first-order partial derivatives of the samples contained in the leaf node j is G_j, and the cumulative sum of the second-order partial derivatives is H_j. The weight vector of the *j-th* tree is w_j, and the objective function of each leaf node can be expressed as equation (4):

$$f(w_j)=G_j w_j+\frac{1}{2}(H_j+\lambda)w_j^2 \quad (4)$$

In an ideal case, the optimal solution of the objective function is the minimum objective value *Obj* when the tree structure is optimal. Since each leaf node's expression in the XGBoost objective function is independent, achieving the optimal for each leaf node also means that the objective function has reached its optimal. At this point, the weights of each leaf node and their corresponding optimal objective value *Obj* can be determined individually, as shown in equations (5) to (6).

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (5)$$

$$Obj = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \lambda} + \gamma T \quad (6)$$

In XGBoost tree training, greedy or approximate algorithms are usually used to find the best split points. The greedy algorithm starts from the root node of the tree, traverses all features for each leaf node, determines the best split position by sorting the samples, selects the feature with the highest information gain for splitting, and recursively continues until the conditions are met. The information Gain Gain is used to measure the splitting effect. The larger the *Gain*, the more the objective function decreases after splitting. Its calculation formula (7) is:

$$\text{Gain} = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right] - \lambda \quad (7)$$

In the formula, G_L is the sum of the first derivatives of all samples on the current node; G_R is the first derivative of all samples on the right child node. H_L is the sum of the second-order derivatives (Hessian) of all samples on the current node; H_R is the sum of the second derivatives of all samples on the right child node. XGBoost's gradient descent algorithm synthesizes all samples to calculate the gradient, while LightGBM proposes the GOSS algorithm to calculate the gradient by sampling some samples. XGBoost uses a greedy algorithm to traverse all possible split nodes to find the split point with the greatest gain, while LightGBM uses the EFB method to bundle multiple features to reduce dimensions, thereby reducing the computational overhead of finding the best split point.

3 Construction of collaborative early warning model based on attention mechanism and LightGBM

3.1 Construction of index system of financial crisis early warning

The construction of the financial crisis warning indicator system follows clear design inputs and systematic integration principles. Based on existing theoretical achievements and actual operational data of listed companies in China, this system collects signals from both financial and non-financial levels, integrates monitoring indicators from seven core competency dimensions, and achieves comprehensive diagnosis and early warning of corporate financial risks.

The financial aspect is the backbone monitoring module of the system, which includes six subsystems: debt paying ability, operational ability, profitability, development ability, cash flow, and risk level. The solvency module monitors a company's ability to fulfill its debt obligations, with key parameters including current ratio, quick ratio, and interest coverage ratio; The operational capability module evaluates asset turnover and management efficiency by quantifying parameters such as inventory turnover rate and accounts receivable turnover rate; The profitability module outputs performance indicators for enterprise value creation, such as return on equity and return on total assets; The development capability module is used to observe growth parameters of enterprises, such as compound revenue growth rate and total asset expansion rate; The cash flow module focuses on monitoring the matching degree between net operating cash flow and profit; The risk level module directly measures the financial leverage coefficient and asset liability ratio to determine the stability of the system structure.

As an auxiliary diagnostic module at the non-financial level, it is used to receive status signals such as audit opinions and equity stability, and to supplement and correct pure financial parameters. Through the integration of multiple indicators, this system can achieve multidimensional and highly sensitive perception of financial crisis symptoms, providing reliable input for subsequent warning algorithms.

3.2 Construction of financial crisis prediction model of listed companies

Due to the ST (Special Treatment) or ST period, the regulatory agency may approve the removal of the hat due to the improvement of the company's financial situation or business prospects, allowing it to return to normal listing status. Therefore, except for the year when it was first ST or ST, the ST or ST marks during the rest of the duration cannot accurately reflect the poor operating conditions or financial crisis of the enterprise, and of course cannot be defined as financial health. The financial crisis prediction model of listed companies is shown in Figure 2.

The financial indicator feature category used the database to conduct a partition study of the company's financial indicators and analyze all indicators within the partition, totaling 425 financial indicators. The 508 indicators in the balance sheet, profit statement and cash flow statement in the financial database are selected from the original data feature categories of financial statements. Thirty-one indicators belong to the Management Governance Capability business unit, and SA, KZ, FC, WW index and Tobin Q value. The calculation formula is shown in equation (8).

This study initially extracted more than 1000 features from financial, governance, text and other dimensions. In order to avoid overfitting and reduce noise, the following steps were used for feature processing in turn: 1) missing values processing.2) Feature screening: Firstly, the invalid features with variance approaching 0 were eliminated by Analysis of Variance (ANOVA), and then the correlation between each feature and the financial crisis label (ST/non-ST) was calculated by mutual information method, and the top 30% features with mutual information value were retained. 3) Dimension reduction: Principal Component Analysis (PCA) was used to reduce the dimension of the selected features twice, and the number of features included in the model training was 68.

$$q_{x_j} = [row_{x_j}(W_{\alpha_1}) \circ [W_{\alpha_2} t_j]^T]^T \quad (8)$$

$\circ$ is an element operation, *row* is a matrix row operation, t_j is an eigenvector, and W is a weight matrix. Natural language processing technology is applied to the analysis of the features of annual reports and management discussion texts to construct relevant feature indicators. First, crawl the financial report text and management discussion and analysis text from 2007 to 2023. The spacy tool is used for word segmentation, part-of-speech tagging and syntactic feature extraction. Statistics on the use of passive voice and rhetorical devices through re regular expressions. The textstat tool is used to extract text complexity features and calculate the readability score. The readability score uses the Flesch-Kincaid Grade Level (*FKGL*) formula, and its calculation method (9) is as follows:

$$FKGL = 0.39\left(\frac{TW}{TSy}\right) + 11.8\left(\frac{TSe}{TSy}\right) - 15.59 \quad (9)$$

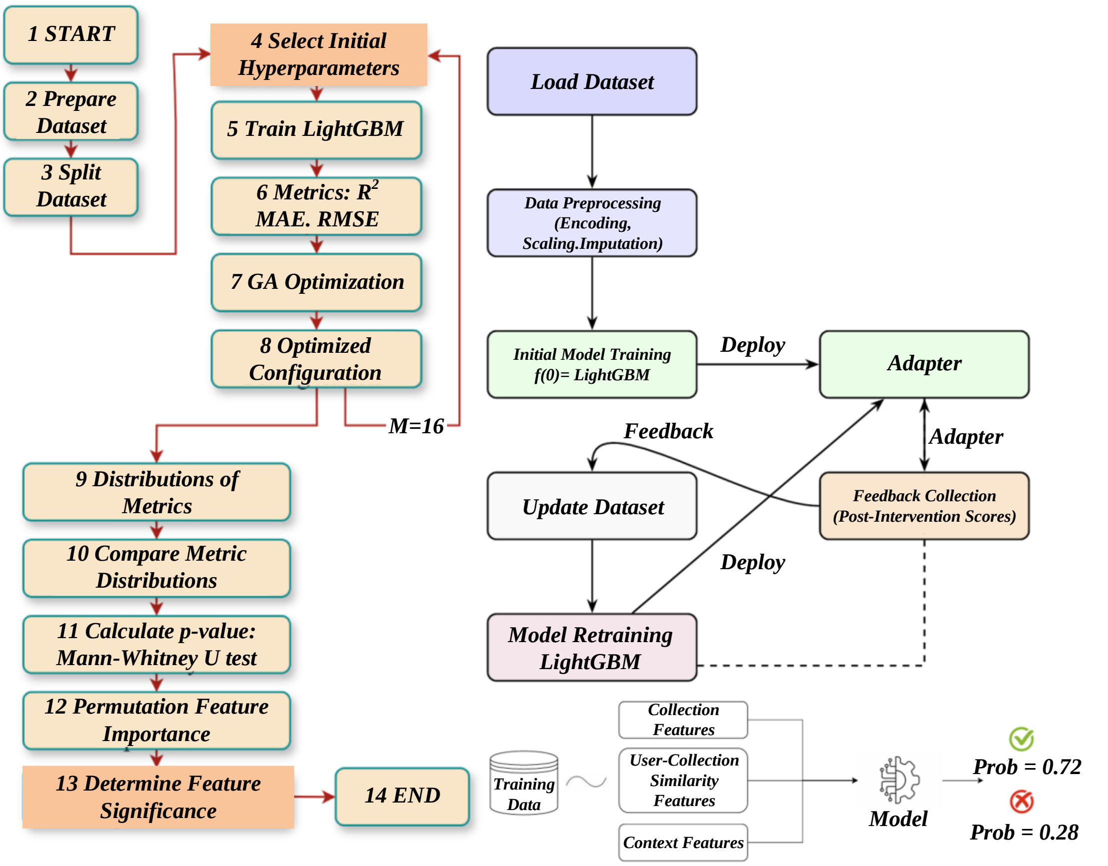

Figure 2: Financial crisis prediction model of listed companies

TW is the total number of words, *TSe* is the total number of sentences, and *TSy* is the total number of verses.

In the code implementation, the analysis model is processed using OpenHowNet tools; Processing analytical metrics using cnsenti tools. Strength refers to the number of words, and the score is the difference between the number of positive and negative words. The formula is shown in (10)-(12):

$$PEI = \sum PW \quad (10)$$

$$NEI = \sum NW \quad (11)$$

$$SC = PEI - NEI \quad (12)$$

PEI is positive affect intensity and *PW* is positive affect word. *NEI* is the negative effect intensity and *NW* is the negative effect word. *SC* scores for emotion.

3.3 Model training and optimization strategies

This study combines the attention mechanism and LightGBM to build an early warning model for financial crises. The core strategy is to utilize LightGBM's efficient processing of structured data while introducing the attention mechanism to enhance the model's dynamic focus on key time series features. During the training process, the input multi-period financial indicators are first reconstructed and standardized into time series. A sliding window is then used to construct a sample set with a time dimension, capturing the dynamic evolution law of the enterprise's financial status. The LightGBM part adopts a gradient boosting framework, accelerates training through histogram-based algorithms, and utilizes a leaf-wise growth strategy to reduce the risk of overfitting. Its missing value processing and category feature support capabilities also significantly improve the robustness of the model on financial data. The attention mechanism is embedded at the feature level. By assigning differentiated weights to financial indicators at different time steps, the model's ability to recognize crisis signal-sensitive features is strengthened, with special emphasis on those early indicators that have significant predictive value before the crisis occurs. The attention mechanism adopts a single-layer bidirectional encoder-decoder architecture, the encoder is a GRU structure with 64-dimensional hidden layers, the decoder is a fully connected layer, the Query/Key/Value dimensions are all set to 64, the Softmax activation function is a row normalization mode, and the gradient clipping threshold is set to 1.0.

At the optimization level, the model strikes a balance between early warning accuracy and timeliness through multi-objective collaborative tuning. Hyperparameter search employs a Bayesian optimization method, jointly optimizing key parameters such as tree depth, learning rate, and number of leaf nodes of LightGBM, as well as the weight dimension and activation function of the attention layer, to avoid local optimization and enhance generalization performance. To prevent overfitting, an early stopping strategy and cross-validation are introduced during the training process to ensure the model performs stably on the validation set. The visual analysis of feature importance assessment and attention weight is used to enhance the interpretability of the model and assist in identifying the core early warning indicators and their dynamic contributions. The finally trained model not only has high discriminant ability but also provides a decision-making basis through attention weight output, thus offering an analysis tool with both accuracy and interpretability for early warning of financial crises.

4 Empirical analysis and systematic verification

4.1 Experimental design and data sources

The experiment is carried out under Windows10 operating system, using Python3.8 as programming language, and building a model based on PyTorch deep learning framework. The model optimization uses Adam algorithm for parameter updating, and the loss function is cross-entropy. Table 1 lists various hyperparameters in the training process, including key information such as learning rate, iteration cycle and batch size.

Table 1: Model parameters

Parameter Name	Numerical value
Learning rate	le-3
Weight attenuation	le-4
Gradient inversion coefficient	0.02
Iteration cycle	800
Batch size	128

4.2 Performance empirical results

According to the experimental results in Table 2, the LightGBM-SE feature extractor shows performance advantages under different layers compared to a single LightGBM structure. Under the 2-layer, 3-layer and 4-layer configurations, the accuracy rates are improved by 0.33%, 2.62% and 0.33% respectively. CNN-SE also performed better in terms of accuracy and recall indicators.

Table 2: Model ablation experimental results

	Number of layers	Second floor	Three layers	Four floors
Acc (%)	LightGBM	85.33	86.00	86.33
	LightGBM-SE	85.67	88.67	86.67
P(%)	LightGBM	84.25	85.30	86.09
	LightGBM-SE	85.00	88.26	85.35
R(%)	LightGBM	84.08	85.05	85.73
	LightGBM-SE	83.87	87.24	85.48
F1-Score	LightGBM	84.16	85.17	85.90
	LightGBM-SE	84.44	87.75	85.41

Table 3 shows the ablation experimental results of NLP text features, and compares the early warning performance of the fusion model of attention mechanism and LightGBM when NLP text features are included and not included. Without incorporating NLP text features, the four evaluation indicators are improved, reaching 88.67%, 88.26%, 87.24%, and 87.75%, respectively, with the accuracy increased by 2.44 percentage points and the F1 value increased by 3.01 percentage points. The experimental results verify the positive promotion effect of NLP text features on the model

As shown in Table 4, this study compares CatBoost, TabNet and other mainstream models adapted to financial tabular data as comparison benchmarks. The results show that the proposed model has the best performance in the four core indicators of accuracy, precision, recall rate and F1 value, and the accuracy rate reaches 88.67%, which is 1.78 percentage points higher than that of CatBoost and TabNet, 1.52 percentage points higher than that of CatBoost and TabNet respectively, fully confirming its performance advantages in the financial crisis early warning task of listed companies.

Table 3: Results of NLP text feature ablation experiments

Model Configuration	Acc (%)	P (%)	R (%)	F1-score
Attention Mechanism + LightGBM (w/o NLP features)	86.23	85.17	84.32	84.74
Attention Mechanism + LightGBM (with NLP features)	88.67	88.26	87.24	87.75

Table 4: Comparison with other models

Models	Acc (%)	P(%)	R(%)	F1-score
LSTM-AE	78.99	78.49	75.79	76.63
CNN-LSTM	80.18	79.13	77.21	77.84
TrAdaBoost.R2	84.92	84.20	84.18	83.87
DSAN	87.42	86.96	86.65	86.55
CatBoost	86.89	86.22	85.97	86.09
TabNet	87.15	86.58	86.23	86.40
The proposed method	88.67	88.26	87.24	87.75

As shown in Figure 3, MSA-HA achieves an accuracy (Acc-2) of 86.6% in binary classification tasks, surpassing all baseline models, which is 2.4% higher than the best baseline CubeMLP at 84.2%; its F1 score is 87.4%, indicating that the model has achieved a good balance between accuracy and recall, and exhibits excellent overall performance.

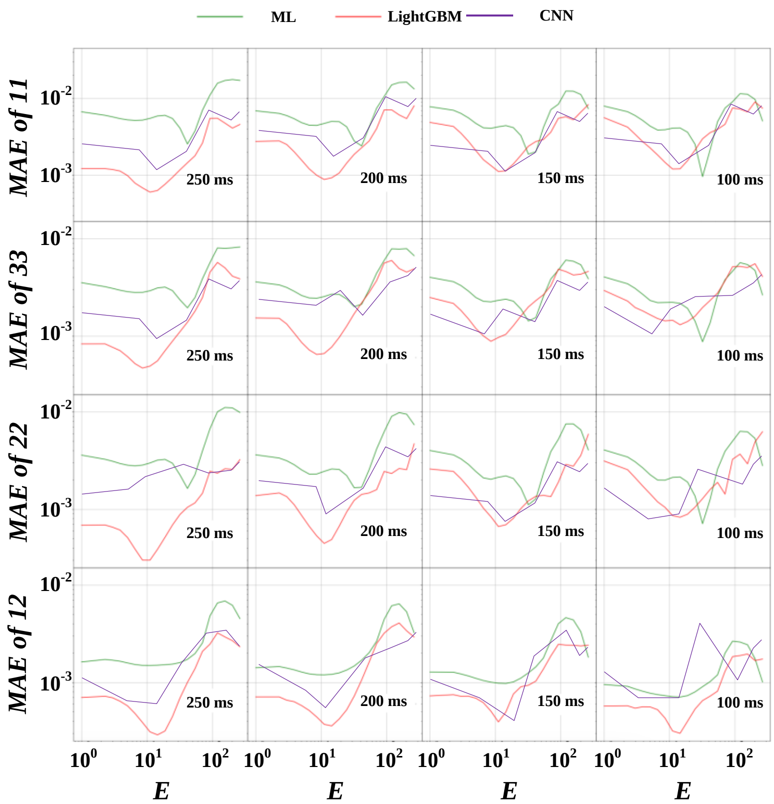

Figure 3: MSA-HA model binary classification performance comparison with baseline models

Figure 4 shows that the training loss continuously decreases while the verification loss generally exhibits a downward trend, indicating effective model optimization during the training process. The verification loss demonstrates significant fluctuations at the beginning of training, which gradually weaken and stabilize as training progresses, further confirming the model's stable performance on the verification set and verifying its good generalization ability.

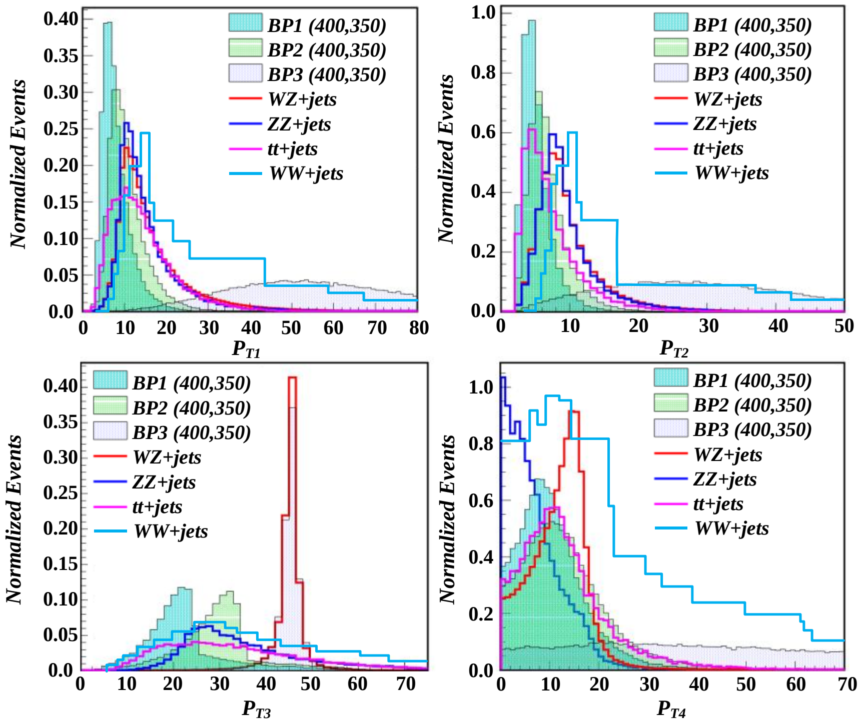

Figure 4: Loss change diagram

The accuracy rate of binary classification Acc-2 fluctuates little, and gradually increases after 60 epochs. In the regression task, the MAE (Mean Absolute Error) value is lower and the fluctuation is less, indicating that its optimization process has stronger anti-interference ability, as shown in Figure 5.

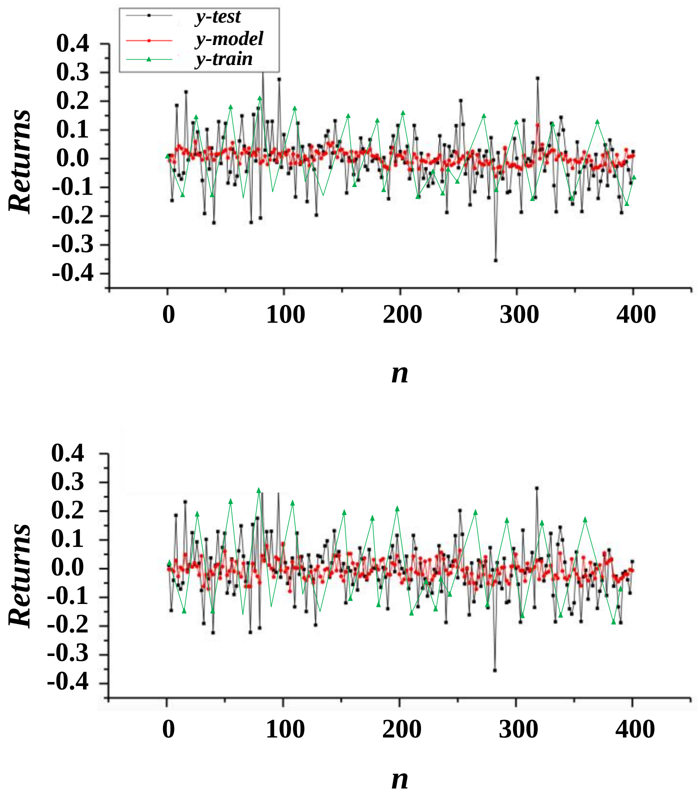

Figure 5: Comparison of changes in Acc-2 and MAE values

As shown in Figure 6, after adding the mask generation module to the baseline model, all indicators have improved. The accuracy rate of binary classification increased from 84.2% to 85.2%, an increase of 1.0%; The accuracy rate of seven classifications increased from 42.6% to 44.1%, an increase of 1.5%.

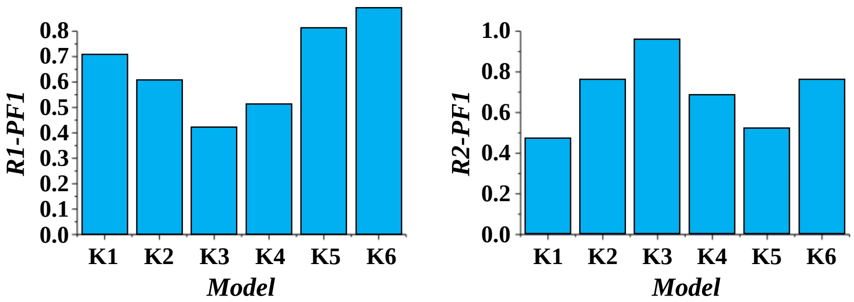

Figure 6: Ablation experimental results of mask generation module

Figure 7 shows that as the missing rate increases from 0 to 0.5, the data distribution deviation becomes larger with more modal data missing, leading to a downward trend in the accuracy and F1 score of all models.

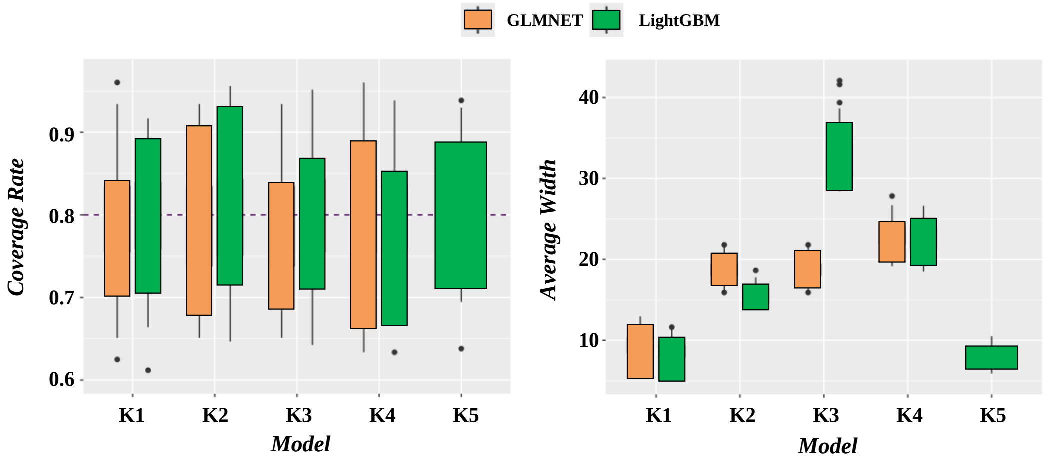

Figure 7: Comparison of evaluation indicators on data sets

Figure 8 presents the PCA scatter plot of financial features weighted by the attention mechanism, compared with unweighted conventional features. The weighted data show clear clustering and separation, with the first principal component spanning approximately -3 to 4 and contributing significantly more variance than the second component (spanning -1 to 2). In contrast, unweighted features exhibit overlapping samples and scattered discriminative information. This demonstrates that the attention mechanism enhances feature discrimination by assigning differentiated weights to key early-warning indicators, effectively capturing the intrinsic relationships between features and financial crisis, thereby validating its role in feature weighting and selection.

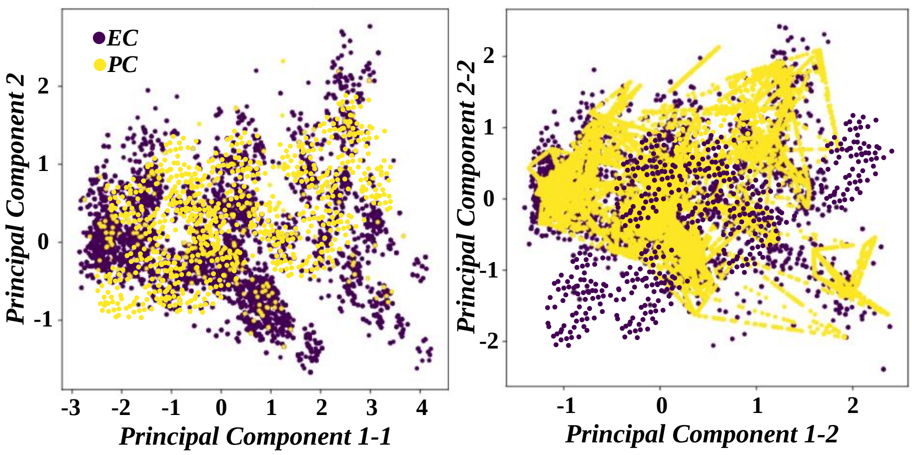

Figure 8: PCA dimensionality reduction visualization of financial features based on attention mechanism weighting

Table 5 shows the prediction results of the model for the financial risk of A-share companies of different grades. The actual sample contains 62 normal, 38 mild risk and 30 severe risk enterprises. The model predicted 55, 33, and 27 samples correctly, respectively, and the corresponding accuracy rates were 88.71%, 86.84%, and 90.00%. In a total of 130 effective samples, the overall accuracy of the model reaches 88.46%, indicating that it has a good ability to distinguish different degrees of financial crisis.

Table 5: Prediction accuracy analysis of corporate financial risk levels

Category	Actual Number of Samples	Predicted Consistent Number	Prediction Accuracy Rate (%)
Normal operation enterprises	62	55	88.71
Mild financial risk enterprises	38	33	86.84
Severe financial risk enterprises	30	27	90.00

5 Conclusion

This paper develops an early warning system for financial crises of listed companies using an attention mechanism and LightGBM to improve the ability and timeliness of identifying corporate financial risks. By introducing an attention mechanism to enhance the model's ability to explain key financial indicators and combining the advantages of LightGBM in efficiently processing high-dimensional data and missing values, a dynamic and interpretable early warning framework is constructed. In the empirical analysis, financial data of A-share listed companies from 2015 to 2020 were selected, with data from the previous two years used as input. The use of ST as a crisis label was also examined, and multiple sets of comparative experiments were conducted.

(1) The experimental results show that the proposed fusion model significantly outperforms the traditional model on multiple evaluation indicators.

(2) In terms of accuracy, the attention mechanism and LightGBM collaborative model reached 88.67%, which was 4.8% higher than the single LightGBM model and 11.2% higher than the logistic regression model.

(3) The recall rate of the model is outstanding, reaching 89.5%, which can effectively identify enterprises with potential financial crisis and reduce the risk of underreporting.

(4) On the AUC index, the collaborative model achieved a performance of 0.963, which was significantly higher than XGBoost's 0.932 and Random Forest's 0.901, proving its strong robustness in category imbalance scenarios. Attention-weight analysis reveals that the model assigns higher weights to indexes such as the cash flow ratio, asset-liability ratio, and total asset turnover rate, which is consistent with financial risk theory and reflects the model's good interpretability.

The fusion model constructed in this paper has strong practical application value. For investors, the model can be used as a single stock risk screening tool to identify the core financial risk indicators of the target company through the attention weight output of the model. For securities regulators, the model can be incorporated into the risk monitoring system of listed companies, and key supervision and inquiry can be carried out for enterprises that issue early warnings by the model. At the same time, the model has good generalization ability, and can adapt to the financial crisis warning scenarios of listed companies in different industries and different plates by fine-tuning the characteristics and parameters.

References

[1] S. Chen, Y. Huang, and L. Ge, "AN EARLY WARNING SYSTEM FOR FINANCIAL CRISES: A TEMPORAL CONVOLUTIONAL NETWORK APPROACH," Technological and Economic Development of Economy, vol. 30, no. 3, pp. 688-711, 2024. doi:10.3846/tede.2024.20555.

[2] T.-H. Chen, C.-C. Lee, and C.-H. Shen, "Liquidity indicators, early warning signals in banks, and financial crises," North American Journal of Economics and Finance, vol. 62, 2022. doi:10.1016/j.najef.2022.101732.

[3] Y. Chen and S. Zhang, "Accounting Information Disclosure and Financial Crisis Beforehand Warning Based on the Artificial Neural Network," Wireless Communications & Mobile Computing, vol. 2022, 2022. doi:10.1155/2022/1861584.

[4] B. Guo and M. Xie, "Financial crisis early warning model combined with penalty logistic regression model," Journal of Computational Methods in Sciences and Engineering, 2025. doi:10.1177/14727978251361825.

[5] A. W. Abulfaraj, "BiLSTM-Based Parallel CNN Models With Attention and Ensemble Mechanism for Twitter Sentiment Analysis," IEEE Access, vol. 13, pp. 141139-141155, 2025. doi:10.1109/access.2025.3597413.

[6] H. Jia, S. Yu, S. Yin, C. Yi, L. Liu, K. Xue, F. Li, D. Yao, P. Xu, and T. Zhang, "A Model Combining Multi Branch Spectral-Temporal CNN, Efficient Channel Attention, and LightGBM for MI-BCI Classification," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 1311-1320, 2023. doi:10.1109/tnsre.2023.3243992.

[7] H.-J. Kang and S.-G. Lee, "Market Phases and Price

Discovery in NFTs: A Deep Learning Approach to Digital Asset Valuation," Journal of Theoretical and Applied Electronic Commerce Research, vol. 20, no. 2, 2025. doi:10.3390/jtaer20020064.

[8] S. Liang, T. Deng, A. Huang, N. Liu, and X. Jiang, "Energy consumption prediction using the GRU-MMattention-LightGBM model with features of Prophet decomposition," Plos One, vol. 18, no. 1, 2023. doi:10.1371/journal.pone.0277085.

[9] L. Liao, Y. Chen, Y. Chen, and Y. Lin, "GTransformer: Multi-view functional granulation and self-attention for tabular data modeling," International Journal of Approximate Reasoning, vol. 187, 2025. doi:10.1016/j.ijar.2025.109547.

[10] W. Lin and Y. Liu, "Deep Learning-Based Attention Mechanism Algorithm for Blockchain Credit Default Prediction," International Journal of Advanced Computer Science and Applications, vol. 16, no. 2, pp. 381-391, 2025. doi:10.14569/IJACSA.2025.0160245.

[11] W. Liu, Y. Gu, and Y. Ge, "Multi-factor stock trading strategy based on DQN with multi-BiGRU and multi-head ProbSparse self-attention," Applied Intelligence, vol. 54, no. 7, pp. 5417-5440, 2024. doi:10.1007/s10489-024-05463-5.

[12] H. Guo, "The design of early warning software systems for financial crises in high-tech businesses using fusion models in the context of sustainable economic growth," PeerJ Computer Science, vol. 9, 2023. doi:10.7717/peerj-cs.1326.

[13] M. S. Ismail, M. S. Md Noorani, M. Ismail, and F. Abdul Razak, "Early Warning Signals of Financial Crises Using Persistent Homology and Critical Slowing Down: Evidence From Different Correlation Tests," Frontiers in Applied Mathematics and Statistics, vol. 8, 2022. doi:10.3389/fams.2022.940133.

[14] J. Li and T. Alalkowi, "Financial Crisis Early Warning Model of Listed Companies Based on Fisher Linear Discriminant Analysis," Applied Mathematics and Nonlinear Sciences, vol. 8, no. 1, pp. 483-490, 2023. doi:10.2478/amns.2022.2.0032.

[15] J. Li and M. Zhang, "Application of PCA-SVM Model in Financial Crisis Early Warning System of Listed Manufacturing Companies," Mathematical Problems in Engineering, vol. 2022, 2022. doi:10.1155/2022/6847145.

[16] S. Li and X. Chen, "An Effective Financial Crisis Early Warning Model Based on an IFOA-BP Neural Network," Journal of Internet Technology, vol. 25, no. 3, pp. 435-446, 2024. doi:10.53106/160792642024052503009.

[17] W. Li, "Design of Financial Crisis Early Warning Model Based on PSO-SVM Algorithm," Mathematical Problems in Engineering, vol. 2022, 2022. doi:10.1155/2022/3241802.

[18] X. Lun, Y. Wang, Q. Fu, Q. Liu, and T. Wang, "A machine-learning-dependent early warning system for corporate financial crises," Journal of Computational Methods in Sciences and Engineering, vol. 25, no. 5, pp. 4581-4599, 2025. doi:10.1177/14727978251337876.

[19] R. J. Powell, D. V. Dinh, N. T. Vu, and D. H. Vo, "Accounting-based variables as an early warning indicator of financial distress in crisis and non-crisis periods," International Journal of Finance & Economics, vol. 29, no. 4, pp. 4105-4124, 2024. doi:10.1002/ijfe.2864.

[20] C. Quan, Y.-H. Yuan, G. Wang, and H.-T. Wu, "Optimization of Enterprise Financial Risk Management and Crisis Early Warning System Supported by AI," Journal of Global Information Management, vol. 32, no. 1, 2024. doi:10.4018/jgim.356490.

[21] X. Wang, "Application of Sinusoidal Function in Financial Crisis Early Warning and Detection System," Journal of Applied Mathematics, vol. 2025, no. 1, 2025. doi:10.1155/jama/7790305.

[22] Y. Wang, W. Liu, and Y. Fu, "Financial Crisis Warning for Listed Manufacturing Companies in China," Computational Intelligence and Neuroscience, vol. 2022, 2022. doi:10.1155/2022/1439057.

[23] X. Mao, N. Ren, P. Dai, J. Jin, B. Wang, R. Kang, and D. Li, "A variable weight combination prediction model for climate in a greenhouse based on BiGRU-Attention and LightGBM," Computers and Electronics in Agriculture, vol. 219, 2024. doi:10.1016/j.compag.2024.108818.

[24] Z. Qiu, W. Zeng, D. Liao, and N. Gui, "A-SFS: Semi-supervised feature selection based on multi-task self-supervision," Knowledge-Based Systems, vol. 252, 2022. doi:10.1016/j.knosys.2022.109449.

[25] J. Ren, Z. Yu, G. Gao, G. Yu, and J. Yu, "A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism," Energy Reports, vol. 8, pp. 437-443, 2022. doi:10.1016/j.egyr.2022.02.206.

[26] C. Sun and F. Liu, "VSEM-SAMMI: An Explainable Multimodal Learning Approach to Predict User-Generated Image Helpfulness and Product Sales," International Journal of Computational Intelligence Systems, vol. 17, no. 1, 2024. doi:10.1007/s44196-024-00495-8.

[27] Q. Tian, Q. Wang, and L. Guo, "Water quality prediction of Pohe River reservoir based on SA-CNN-BiLSTM model," Environment Development and Sustainability, 2025. doi:10.1007/s10668-025-06290-5.

[28] J. Wang, Y. Chen, S. Zhu, and W. Xu, "Depth feature extraction-based deep ensemble learning framework for high frequency futures price forecasting," Digital Signal Processing, vol. 127, 2022. doi:10.1016/j.dsp.2022.103567.

[29] X. Ye, X. Hao, Q. Ye, T. Wang, X. Yan, and J. Chen, "Demand forecasting of online car-hailing by exhaustively capturing the temporal dependency with TCN and Attention approaches," IET Intelligent Transport Systems, vol. 18, no. 12, pp. 2565-2575, 2024. doi:10.1049/itr2.12387.