Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.34374

Research on Financial Risk Prediction of Listed Companies Based on CNN-Transformer Ensemble Learning

Yimeng Wu

School of Management, Harbin Institute of Technology, Harbin, China
wuhaidong1123@163.com

Abstract. Traditional financial risk prediction models for listed companies mostly rely on annual quantitative indicators for modeling. With a long data update cycle, such models cannot effectively capture the dynamic evolution characteristics of corporate financial status within the year, making it difficult to realize timely early warning of financial risks. Quarterly financial data feature higher temporal resolution, which can accurately reflect the intra-year fluctuation pattern of corporate financial conditions and effectively improve the timeliness of financial risk identification. However, when such data are applied to risk prediction, there still exist problems including difficulty in extracting multi-dimensional feature matrices and insufficient modeling of time-series correlation features. In view of this, this paper constructs an ensemble learning model for financial risk prediction by integrating CNN and Transformer. Firstly, multi-dimensional quarterly financial risk indicators of enterprises are reconstructed into matrix-valued input features. Convolutional Neural Network (CNN) is used to mine the time-series evolution features of financial indicators, while Transformer is adopted to capture the complex correlations and global dependency features among different financial indicators. Meanwhile, an ensemble learning strategy based on stratified sampling is introduced to effectively alleviate the class imbalance caused by the scarcity of ST risk samples of listed companies and achieve a balanced trade-off between model bias and variance. The proposed model is verified with the financial risk prediction of A-share listed companies as an empirical scenario. Experimental results show that the CNN-Transformer ensemble model outperforms traditional benchmark models in prediction performance. In addition, the SHAP explainable framework is introduced to interpret the model prediction results. It is found that financial information of quarters close to the prediction point carries higher risk-discriminating weights, and indicators such as corporate debt-servicing and financing pressure as well as abnormal costs and expenses are key factors driving financial risks of listed companies.

Keywords: financial risk prediction, CNN-Transformer, ensemble learning, imbalanced samples, SHAP explainable analysis

1. Introduction

Financial risk prediction is an important research topic for capital market regulation, corporate governance optimization, and investor decision-making, with practical significance in maintaining

DOI: 10.54254/2755-2721/2026.34374

the stable operation of capital markets [1]. Once a listed company suffers financial deterioration or persistent operational difficulties, it not only directly harms the vital interests of investors and creditors and triggers investment and financing risks, but also disrupts the normal resource allocation order of the capital market and reduces market efficiency. Against the dual background of normalized capital market regulation and an increasingly complex corporate operating environment, accurately and timely identifying potential financial risks of listed companies and realizing early risk warning has become a research focus for both academia and industry [2].

With the iterative upgrading of artificial intelligence technology, machine learning and deep learning algorithms have gradually replaced traditional statistical models as mainstream tools for financial risk prediction of listed companies, owing to their strong capabilities in nonlinear fitting, feature mining, and time-series modeling. Rich research results have been accumulated worldwide. In the field of traditional machine learning, Liang et al. established a financial distress measurement system for listed companies based on machine learning methods in light of the ST system in China's capital market, confirming that ST risk prediction is a core application scenario in corporate financial distress research [3]. Raudys A and Goldstein E [4] integrated an XGBoost regressor with an LSTM network and realized effective prediction of financial risk fluctuations through multi-dimensional data transformation. On this basis, many scholars have optimized mainstream ensemble learning algorithms. Xu et al. [5] constructed a risk prediction model based on XGBoost and achieved high prediction accuracy. Wang et al. [6] introduced the LightGBM model to evaluate corporate financial risks, improving the application system of advanced machine learning algorithms in micro-enterprise financial risk prediction. To address high dimensionality, nonlinearity, and sample imbalance in financial data, researchers have further combined sampling techniques, feature optimization, and ensemble structures to improve models. Zhang [7] integrated pruning algorithms and SMOTE oversampling to optimize the random forest model, effectively improving the identification accuracy of financial distress in electronic manufacturing enterprises. Zou et al. [8] innovatively proposed an ensemble bagging-cascaded boosting framework and strengthened early warning capability by leveraging the residual fitting advantage of LightGBM. Chen et al. [9] built a hybrid ensemble model combining an oversampling data augmentation module and LightGBM, effectively solving the class imbalance problem in binary classification of financial datasets. Matsumaru M et al. [10]developed a two-stage machine learning framework integrating ensemble learning, feature selection, and resampling, suitable for modeling high-dimensional financial data and imbalanced samples. Nguyen H [11] used random forest and LightGBM ensemble algorithms to accurately predict bankruptcy risks in manufacturing and construction enterprises. With the development of deep learning, time-series modeling and attention mechanisms have been gradually introduced into financial risk research, further enhancing the ability to capture complex features. Che et al. [12] integrated financial indicators, textual announcements, and corporate correlation networks to construct a multimodal deep learning model with attention mechanisms and regularization, verifying the effectiveness of deep learning in financial distress prediction. Sun [13] combined an attention mechanism with a weighted-sampling GraphSAGE model to mine correlation features in financial networks for risk prediction. Zhang [14] proposed a GNN-GAT-LSTM hybrid model, which captured corporate correlation structure through graph neural networks to enhance financial risk identification. Korangi et al. [15] constructed a Transformer default prediction model based on multi-period panel data, breaking through the limitations of traditional static models by modeling dynamic temporal dependencies and confirming the contribution of multi-period sequential financial information to risk prediction.

DOI: 10.54254/2755-2721/2026.34374

Reviewing existing studies shows that machine learning and deep learning can effectively characterize the complex nonlinear relationship between financial risks and operational indicators, and multi-period time-series modeling provides a new perspective for capturing risk evolution. However, obvious room for improvement remains. Most studies still rely on low-frequency annual financial report data, ignoring quarterly intra-year financial fluctuations and failing to identify periodically accumulated potential risks. Meanwhile, for the multi-dimensional matrix features of quarterly data in the form of *time steps × indicator dimensions*, existing models can hardly meet the dual demands of local temporal feature extraction and global correlation modeling among multiple indicators. Moreover, the class imbalance caused by scarce ST risk samples is widely neglected, resulting in weak ability to identify high-risk enterprises and limited prediction accuracy and generalization. In addition, the black-box nature of deep learning models leads to a lack of interpretability of prediction results, which hinders practical applications in capital market regulation and corporate governance.

To solve the above problems, this paper constructs an ensemble learning model for financial risk prediction by fusing CNN and Transformer, using quarterly high-frequency financial data of A-share listed companies as research samples for early warning modeling. First, multi-dimensional quarterly financial indicators are reconstructed into standardized matrix inputs. CNN is used to mine local temporal evolution features of quarterly data, while the Transformer model captures global dependencies and complex correlations among financial indicators. Meanwhile, an ensemble learning strategy based on stratified sampling is introduced to specifically address class imbalance caused by insufficient ST samples, balance model bias and variance, and improve risk identification accuracy. Finally, the SHAP explainable framework is adopted to decompose the model's prediction logic, quantify the risk-discriminating weights of different quarters and indicators, and explore core driving factors of financial risk outbreaks. This study aims to provide new technical references and empirical evidence for financial risk early warning of listed companies and capital market risk prevention and control.

2. Theories and methods

The model in this paper is built on the quarterly financial time-series data of listed companies, constructing a standardized two-dimensional financial feature tensor. Given a sample set containing N listed companies, with a time-series window length T and financial risk indicator dimension D, the original financial time-series matrix of the i-th enterprise can be mathematically defined as:

$$\mathbf{X}_{\mathrm{i}} \in \mathbb{R}^{\mathrm{T}\times \mathrm{D}}, \quad i = 1,2,\ldots,N \tag{1}$$

The matrix element is strictly defined as $\mathbf{X}_{\mathrm{i}} = [\mathrm{x}_{\mathrm{i,td}}]_{\mathrm{T}\times \mathrm{D}}$, where $\mathrm{x}_{\mathrm{i,td}}$denotes the standardized value of the d-th financial indicator of the i-th enterprise in the t-th quarter. This matrix serves as a structured time-series-indicator two-dimensional input. A CNN-Transformer fusion model is constructed as the core base predictor. Convolutional operations are used to realize local nonlinear feature coupling, and the self-attention mechanism is adopted to capture global time-series dependencies, enabling deep representation learning of high-dimensional financial risk features. The complete mathematical modeling process is as follows.

A multi-layer one-dimensional convolutional neural network is applied to perform local feature mapping and nonlinear transformation on the two-dimensional financial matrix. The original input

DOI: 10.54254/2755-2721/2026.34374

feature matrix of the network is defined as:

$$\mathrm{H}_{\mathrm{i}}^{(0)}=\mathrm{X}_{\mathrm{i}} \tag{2}$$

Suppose the network stacks K convolutional layers. The k-th layer contains learnable convolution kernels $\mathrm{W}_{\mathrm{c}}^{(\mathrm{k})} \in \mathbb{R}^{\mathrm{s}\times \mathrm{c}_{\mathrm{in}}\times \mathrm{c}_{\mathrm{out}}}$, where s is the receptive field size, c_{in} is the number of input channels, and $\mathrm{c}_{\mathrm{out}}$ is the number of output channels. The full convolutional mapping formula based on weighted summation, bias correction and nonlinear activation is:

$$\mathrm{H}_{\mathrm{i}}^{(\mathrm{k})}=\sigma\left(\mathrm{H}_{\mathrm{i}}^{(\mathrm{k}-1)} * \mathrm{W}_{\mathrm{c}}^{(\mathrm{k})}+\mathrm{b}^{(\mathrm{k})}\right), \quad k=1,2,\ldots,K \tag{3}$$

where * denotes the cross-correlation operation, realizing weighted fusion of financial indicators within local windows;$\mathrm{b}^{(\mathrm{k})} \in \mathbb{R}^{\mathrm{c}_{\mathrm{out}}}$is the bias vector of the k-th convolutional layer; $\sigma(\cdot)$ is the ReLU nonlinear activation function satisfying $\sigma\left(\mathrm{x}\right)=\max\left(0,\mathrm{x}\right)$, which introduces nonlinear mapping capability to fit the complex nonlinear relationship between financial indicators and risk status. Through sliding operations with a fixed receptive field, the convolutional layers traverse all quarterly time intervals and indicator combinations to extract local fluctuation features. After K progressive convolutional mappings, the multi-channel local deep feature matrix is output:

$$\mathrm{H}_{\mathrm{i}}=\mathrm{H}_{\mathrm{i}}^{(\mathrm{K})} \in \mathbb{R}^{\mathrm{T}\times \mathrm{c}_{\mathrm{out}}} \tag{4}$$

Convolutional operations can only capture local adjacent feature correlations but cannot model long-term quarterly dependencies and global indicator relationships. To this end, the Transformer Encoder module is introduced to supplement time-series position information via positional encoding and realize global time-series dependency modeling. A learnable positional encoding matrix$\mathrm{P} \in \mathbb{R}^{\mathrm{T}\times \mathrm{c}_{\mathrm{out}}}$is defined, and time-series information is embedded through element-wise addition to obtain the initial input features of the Transformer module:

$$\mathrm{Z}_{\mathrm{i}}^{(0)}=\mathrm{H}_{\mathrm{i}}+P \tag{5}$$

The multi-head self-attention mechanism, the core of Transformer, constructs query, key and value matrices through linear projection to quantify the correlation weight between any two time-step nodes. Its complete mathematical definition is:

$$Attention\left(Q,K,V\right)=softmax\left(\frac{\mathrm{QK}^{\top}}{\sqrt{\mathrm{d}_{\mathrm{k}}}}\right)V \tag{6}$$

where the three projection matrices are obtained by linear transformation of input features:

DOI: 10.54254/2755-2721/2026.34374

$$Q = Z\mathrm{W_Q}, \quad K = Z\mathrm{W_K}, \quad V = Z\mathrm{W_V} \tag{7}$$

where $\mathrm{W_Q}, \mathrm{W_K}, \mathrm{W_V} \in \mathbb{R}^{\mathrm{c_{out}} \times \mathrm{d_k}}$ are learnable projection weight matrices, and $\mathrm{d_k}$ is the feature dimension of single-head attention. The scaling factor $\sqrt{\mathrm{d_k}}$ constrains the numerical range of high-dimensional inner products to avoid Softmax gradient vanishing. The Softmax normalization is defined as:

$$softmax\left(\mathrm{z_j}\right) = \frac{\mathrm{e}^{\mathrm{z_j}}}{\sum_{\mathrm{m=1}}^{\mathrm{T}} \mathrm{e}^{\mathrm{z_m}}} \tag{8}$$

Normalization constrains global time-series correlation weights to the interval [0,1], enabling accurate quantification of time-series dependencies. The model stacks L Transformer Encoder layers, each containing self-attention and feedforward neural network mapping. After multi-layer iterative optimization, high-order features integrating global time-series correlations are output:

$$\mathrm{Z_i} = \mathrm{Z_i^{(L)}} \in \mathbb{R}^{\mathrm{T} \times \mathrm{d_{model}}} \tag{9}$$

To eliminate redundant information in the time dimension and aggregate high-dimensional time-series features into fixed-dimensional enterprise-level representation vectors, global average pooling is adopted for dimensionality reduction:

$$\mathrm{r_i} = Pool\left(\mathrm{Z_i}\right) = \frac{1}{\mathrm{T}} \sum_{\mathrm{t=1}}^{\mathrm{T}} \mathrm{Z_{i,t}} \tag{10}$$

where $\mathrm{Z_{i,t}}$ is the high-order feature vector of the t-th time step, and $\mathrm{r_i} \in \mathbb{R}^{\mathrm{d_{model}}}$ is the aggregated global risk representation of the enterprise. The representation vector is fed into a fully connected classification layer, and the financial risk prediction probability is obtained via the Sigmoid activation function:

$$\mathrm{p_i} = \sigma\left(\mathrm{W_{fc}} \mathrm{r_i} + \mathrm{b_{fc}}\right) \tag{11}$$

where $\mathrm{W_{fc}}$ is the weight matrix of the fully connected layer and $\mathrm{b_{fc}}$ is the bias term. The Sigmoid function $\sigma(\mathrm{x}) = 1/(1 + \mathrm{e}^{-\mathrm{x}})$

maps the output to the probability interval [0,1].

Binary cross-entropy is used as the loss function for model training. For a training batch with n_b samples, the global loss function is expressed as:

DOI: 10.54254/2755-2721/2026.34374

$$\mathscr{L} = -\frac{1}{n_b}\sum_{i=1}^{n_b}\left[y_i \ln p_i + (1-y_i)\ln(1-p_i)\right] \tag{12}$$

where $y_i \in \{0,1\}$ is the true label: $y_i = 1$ indicates financial risk, and $y_i = 0$ indicates normal financial status. The model minimizes the loss function using gradient descent. The optimal parameter solution is:

$$\theta^* = \arg\min_{\theta}\mathscr{L}(\theta) \tag{13}$$

where θ denotes all learnable parameters of the CNN-Transformer model. Parameters are iteratively updated via backpropagation to complete the convergence training of the base learner.

Financial risk prediction datasets typically suffer from class distribution shift. The overall sample set can be divided into majority-class normal samples and minority-class risky samples. Let the full dataset be D, the majority subset be D_0, and the minority subset be D_1, satisfying $D = D_0 \cup D_1$ and $D_0 \cap D_1 = \emptyset$, with sample sizes satisfying $|D_0| \gg |D_1|$. Traditional random sampling distorts the prior class distribution, causing base learners to bias toward the majority class and resulting in excessively low recall for minority classes. To address this, a stratified sampling mechanism is introduced to strictly ensure that the class distribution of each subset is consistent with the overall distribution.

The prior class probabilities of the full dataset are defined as:

$$\pi_0 = \frac{|D_0|}{|D|}, \quad \pi_1 = \frac{|D_1|}{|D|} \tag{14}$$

where π_0,π_1 represent the overall proportions of normal and risky enterprises, respectively. When generating the b-th training subset D_b with total sample size n, stratified sampling is performed according to the overall prior ratio. The numbers of majority and minority samples in the subset satisfy:

$$|D_{b0}| = \lfloor n \cdot \pi_0 \rfloor, \quad |D_{b1}| = \lfloor n \cdot \pi_1 \rfloor \tag{15}$$

where D_{b0} and D_{b1} denote the majority and minority sets in the subset, ensuring $\frac{|D_{b0}|}{|D_b|} \approx \pi_0$ and $\frac{|D_{b0}|}{|D_b|} \approx \pi_0$. This mathematically avoids sampling distribution shift and guarantees that each training subset follows the true overall data distribution.

Based on stratified sampling datasets, a Bagging ensemble learning framework is constructed for multi-model fusion optimization. By decomposing bias and variance across multiple base learners and fusing outputs, the generalization stability of the model is improved. A set of B CNN-Transformer base learners $\{f_1, f_2, \ldots, f_B\}$ with homogeneous structures but heterogeneous parameters is established. The parameter set of the b-th base learner is θ_b, with independent

DOI: 10.54254/2755-2721/2026.34374

initialization and gradient updates satisfying $\theta_b \sim P(\theta)$. The predicted probability output for sample X_i by the b-th base learner is defined as:

$$p_{ib} = f_b(X_i; \theta_b), \quad b = 1,2,\ldots,B \tag{16}$$

As the number of base learners B increases, the prediction variance converges continuously, effectively alleviating overfitting and random training errors of single models. An equal-weight voting ensemble strategy is adopted to compute the arithmetic mean of prediction probabilities from all base learners, yielding the final prediction probability of the ensemble model:

$$p_i = \mathbb{E}_b\left[p_{ib}\right] = \frac{1}{B}\sum_{b=1}^{B} p_{ib} \tag{17}$$

According to the optimal classification decision rule, a risk discrimination threshold τ is set to construct the binary classification decision function:

$$\hat{y}_i = \begin{cases} 1, p_i \geq \tau \\ 0, p_i < \tau \end{cases} \tag{18}$$

This stratified Bagging strategy ensures unbiased class distribution at the sampling level and suppresses prediction variance at the ensemble level, effectively mitigating classification bias caused by imbalanced financial risk samples and achieving balanced optimization of model bias and variance.

To break the black-box nature of deep learning models and quantify the marginal contributions of each quarter and each financial indicator to risk prediction, a mathematical SHAP explainable framework is constructed based on the Shapley axiom system in game theory, enabling global attribution and mechanism analysis of model predictions. Shapley values satisfy four axioms: efficiency, symmetry, linearity, and nullity, representing the only fair allocation scheme for feature contributions. For any feature subset $S \subseteq M$, the complete formula for the Shapley value of feature m is:

$$\phi_m = \sum_{S \subseteq M \setminus \{m\}} \frac{|S|!(M-|S|-1)!}{M!}\left[f(S \cup \{m\}) - f(S)\right] \tag{19}$$

where M is the total number of feature dimensions, f(S) is the model prediction output using only feature subset S, and $\frac{|S|!(M-|S|-1)!}{M!}$ is the weight coefficient for feature combinations, used to weighted-average marginal contribution differences across all subsets to ensure unbiased and fair attribution.

An additive explainable model is built based on Shapley values, linearly decomposing the output of the complex nonlinear model into a superposition of independent contributions from each feature:

DOI: 10.54254/2755-2721/2026.34374

$$g\left(z'\right) = \phi_0 + \sum_{m=1}^{M} \phi_m z'_m \tag{20}$$

where g(z') is the explainable model output, z'_m is the standardized feature value, z'_m is the global expected baseline output (mean prediction over all samples), and ϕ_m is the SHAP marginal contribution value of the m-th feature. $\phi_m > 0$ indicates the feature increases the financial risk probability, while $\phi_m < 0$ indicates it suppresses the risk probability.

Define ϕ_{itd} as the SHAP contribution value, representing the independent marginal contribution of the d-th financial indicator in the t-th quarter for the i-th enterprise to the prediction result. To eliminate individual and temporal disturbances and quantify global feature importance, two global importance indicators are constructed based on expectations over the full dataset. Global importance in the financial indicator dimension is quantified by the overall average absolute SHAP value, reflecting the overall contribution strength of a single indicator to risk prediction:

$$I_d = \mathbb{E}_{i,t}\left[|\phi_{itd}|\right] = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} |\phi_{itd}| \tag{21}$$

Global importance in the quarterly time dimension is quantified by aggregating the mean SHAP values of all indicators in a single quarter, reflecting the risk-discriminating weight of financial information from different quarters:

$$I_t = \mathbb{E}_{i,d}\left[|\phi_{itd}|\right] = \frac{1}{ND} \sum_{i=1}^{N} \sum_{d=1}^{D} |\phi_{itd}| \tag{22}$$

The above process realizes the explainable analysis of model prediction outputs. It not only quantifies the driving effect of individual features but also mines the core influencing factors and key evolution periods of financial risks from both the indicator dimension and the time dimension.

3. Empirical analysis

3.1 Data source and experimental setup

This study takes the ST risk prediction of A-share listed companies as the experimental scenario, with data sourced from the China Stock Market & Accounting Research (CSMAR) Database. The model input consists of quarterly financial and non-financial indicators of enterprises for the four quarters of 2024, and the prediction target is whether a company will be labeled ST in 2025, where y=1 represents a risky enterprise and y=0 represents a non-risky enterprise. Taking "enterprise–quarter" as the basic unit, this study matches and merges data from balance sheets, income statements, cash flow statements, and non-financial risk early warning data to form quarterly panel data.

In terms of indicator construction, a financial risk indicator system is established from six perspectives: debt-servicing and financing pressure, profitability quality, cost and expense, cash flow, asset quality, and non-financial risk warning, including 24 secondary indicators in total. The data preprocessing procedure includes removing high-missing indicators, random forest imputation for missing values, tail shrinking of continuous variables, indicator positivization, and entropy

DOI: 10.54254/2755-2721/2026.34374

weighting. The entropy method is used to determine the weights of secondary indicators within each primary indicator and synthesize six comprehensive risk dimensions. Finally, quarterly data of each enterprise are organized into a matrix input of "4 quarters × 6 indicators". The specific indicator system and within-group weights are shown in Table 1.

Table 1. Financial risk prediction indicator system and within-group weights by entropy method

Primary Indicator	Secondary Indicators and Within-Group Weights
Debt-Servicing and Financing Pressure	Cash Ratio (0.0416); Asset-Liability Ratio (0.3480); Short-Term Loan Dependence (0.6104)
Profitability and Profit Structure	ROA (0.0182); Operating Profit Margin (0.0248); Non-Operating Profit/Loss Dependence (0.4756); Comprehensive Income Deviation (0.4815)
Cost, Expense and Impairment Abnormality	Management Expense Ratio (0.4552); Financial Expense Ratio (0.2664); Asset Impairment Loss Ratio (0.0558); Bad Debt & Provision Intensity (0.2226)
Cash Flow Authenticity and Capital Chain Pressure	Net Operating Cash Flow / Operating Income (0.0426); Net Operating Cash Flow / Current Liabilities (0.0178); Cash Collection Ratio of Sales (0.0122); Difference Ratio of Operating Cash Receipts and Payments (0.0339); Refinancing Pressure (0.8934)
Operation and Asset Quality	Accounts Receivable / Operating Income (0.3289); Other Receivables Ratio (0.3464); Inventory Ratio (0.1664); Soft Asset Ratio (0.1584)
Quality and Risk Warning	Restated Financial Report (0.2411); Independent Director Objection Frequency (0.3996); Management Turnover Rate (0.1817); Suspension Duration (0.1775)

Since the number of ST enterprises is significantly smaller than that of non-ST enterprises, the SMOTE oversampling method is adopted in the training set, combined with stratified sampling under the Bagging framework to alleviate the class imbalance problem. For sample splitting, the test set accounts for 20% of all samples, and the remaining samples are further divided into training and validation sets. A CNN-Transformer model is constructed as the base learner. The CNN module is set with 12 output channels and a kernel size of 2 to extract local combined features from the quarterly matrix. The Transformer Encoder has 1 layer, 3 attention heads, and a hidden dimension of 48 to capture temporal dependencies across quarters. The model is trained with a batch size of 32, 70 epochs, a learning rate of 0.001, the Adam optimizer, and binary cross-entropy loss. In the main experiment, an ensemble model composed of 5 CNN-Transformer sub-models is built for each random split, and the average predicted probability of sub-models is taken as the final output. Furthermore, the number of sub-models is varied to investigate the effect of ensemble size on model performance.

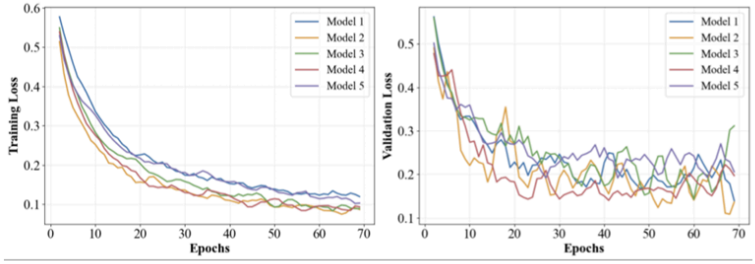
To verify the model effectiveness, Random Forest (RF), XGBoost, and LightGBM are selected as benchmark methods. Model performance is evaluated using Accuracy, Precision, Recall, F1-score, and AUC. Among them, Recall, F1-score, and AUC are mainly used to measure the ability to identify minority-class risky enterprises.

3.2 Analysis of experimental results

To examine the training stability of the CNN-Transformer ensemble model, the loss curves of each sub-model on the training and validation sets are recorded, as shown in Figure 1. It can be observed that as training epochs increase, the training loss of each sub-model shows an overall downward trend, and the validation loss decreases synchronously with consistent directions, indicating that the

DOI: 10.54254/2755-2721/2026.34374

model can stably complete parameter updating and feature learning without obvious training oscillation or divergence.



(a) Training loss curves (b) Validation loss curves

Figure 1. Training and validation loss curves of CNN-Transformer sub-models

To avoid biased evaluation results caused by a single sample split, 10 groups with different random seeds are used to repeatedly split the training and test sets. Random Forest, XGBoost, LightGBM, and the proposed CNN-Transformer ensemble model are trained under each split. All evaluation metrics are reported in the form of "mean ± standard deviation", where the mean reflects the average prediction performance and the standard deviation reflects the fluctuation across different splits. The experimental results are shown in Table 2.

Table 2. Prediction performance comparison of different models

Model	Accuracy	Precision	Recall	F1	AUC
RF	0.904±0.010	0.855±0.025	0.627±0.039	0.723±0.033	0.930±0.011
XGBoost	0.912±0.008	0.811±0.032	0.733±0.029	0.769±0.021	0.939±0.009
LightGBM	0.903±0.009	0.825±0.029	0.654±0.036	0.729±0.028	0.933±0.011
Ensemble CNN-Transformer	0.954±0.011	0.815±0.036	0.982±0.012	0.898±0.022	0.998±0.001

As shown in Table 2, the CNN-Transformer ensemble model achieves the best performance in Accuracy, Recall, F1, and AUC, reaching 0.954, 0.982, 0.898, and 0.998 respectively, demonstrating strong advantages in overall prediction, risky sample identification, and class separation. In contrast, although Random Forest and LightGBM have relatively high Precision, their Recall is significantly lower, indicating conservative predictions and a tendency to miss risky enterprises. XGBoost performs moderately well but still inferior to the proposed model. Given that financial risk early warning prioritizes reducing missed detections of risky samples, the advantages of the CNN-Transformer ensemble model in Recall, F1, and AUC better reflect its practical early warning value.

Figure 2 visualizes the above results. Bars represent the mean values of 10 experiments, error bars represent standard deviations, and asterisks mark the optimal model for each metric. It can be seen that CNN-Transformer maintains obvious advantages in most indicators, with small standard deviations in Recall and AUC, proving that its superiority is not accidental from a single split but stable across different sample divisions.

DOI: 10.54254/2755-2721/2026.34374

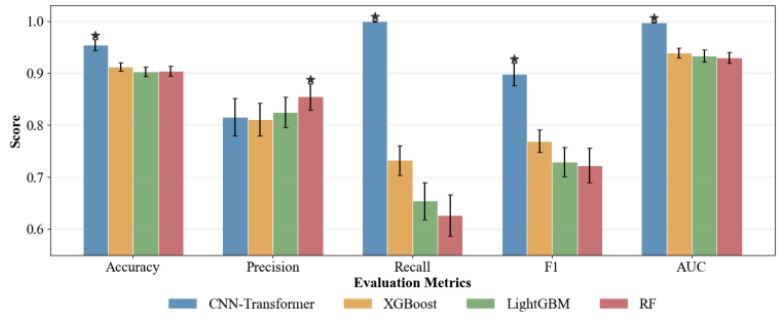


Figure 2. Prediction performance comparison under 10 random splits

Furthermore, ROC curves are plotted to compare the overall classification ability of different models, as shown in Figure 3. The ROC curve of the CNN-Transformer ensemble model is closer to the top-left corner, consistent with the high AUC value in Table 2, indicating its superior ability to distinguish ST risk samples from non-risk samples.

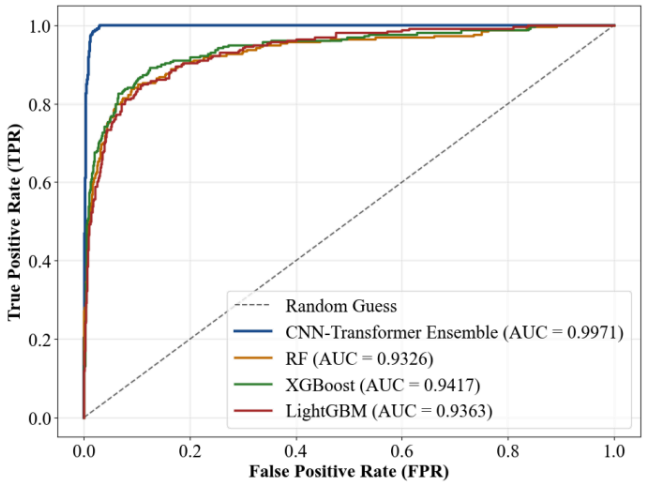


Figure 3. ROC curves of different models

After verifying the overall prediction performance, the influence of ensemble size on model performance is further investigated. Taking a single CNN-Transformer sub-model as the benchmark, the number of sub-models is set to 3, 5, and 7 respectively, and the percentage-point changes of each metric relative to the single model are calculated, as shown in Figure 4.

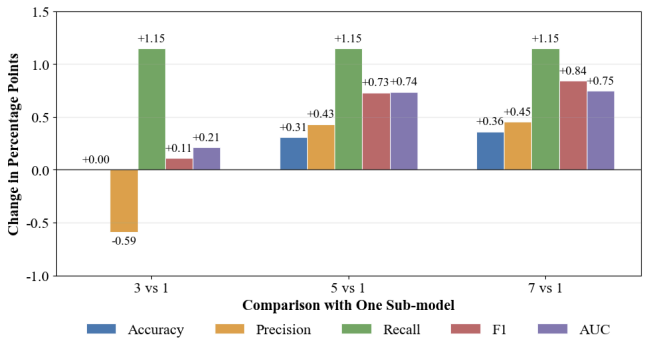


Figure 4. Performance improvements with different numbers of sub-models relative to the single model

DOI: 10.54254/2755-2721/2026.34374

It can be seen from Figure 4 that the overall performance is improved after introducing multiple sub-models. When the number of sub-models increases to 3, Recall rises by 1.15% and AUC by 0.21%. When the number reaches 5, Accuracy, Precision, F1, and AUC are further improved, with F1 increasing by 0.73% and AUC by 0.74%. When extended to 7 sub-models, the model still achieves slight improvements but with higher training costs. Considering both prediction performance and computational overhead, 5 CNN-Transformer sub-models are used to construct the ensemble model in the main experiment.

To further analyze the prediction basis of the model, the SHAP method is adopted to conduct explainable analysis on the ensemble CNN-Transformer model. Each element in the enterprise input matrix is regarded as a "quarter-indicator" combined feature, and the contributions of different variables to the prediction results are examined from three levels: combined features, quarter dimension, and indicator dimension. The SHAP results mainly reflect the model's prediction basis on the dataset and help interpret the key time points and risk dimensions emphasized by the model.

First, from the perspective of "quarter-indicator" combined features, the SHAP distribution is shown in Figure 5. The horizontal axis represents the SHAP value: a positive value indicates that the feature pushes the model to output a higher risk probability, while a negative value means the feature reduces the predicted risk probability. The color represents the magnitude of the feature value.

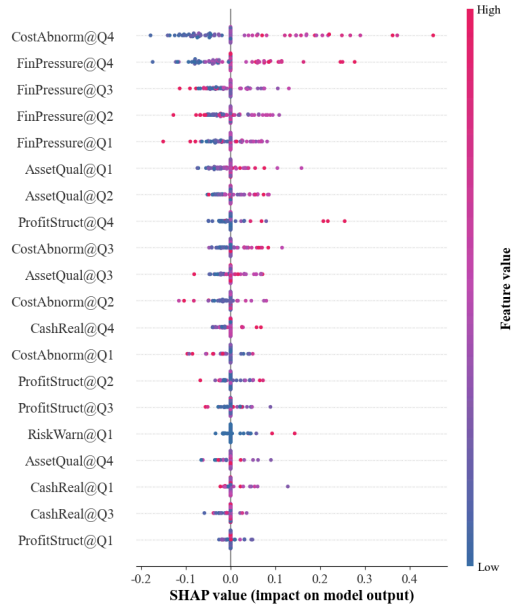


Figure 5. SHAP distribution of "quarter-indicator" combined features

It can be observed from Figure 5 that features with high contributions mainly include cost, expense and impairment abnormality in the fourth quarter, debt-servicing and financing pressure in the fourth quarter, and debt-servicing and financing pressure in the third quarter. This shows that the model does not rely on a single indicator but integrates combined information from different quarters and risk dimensions. The SHAP values are further aggregated by quarter, as shown in Figure 6.

DOI: 10.54254/2755-2721/2026.34374

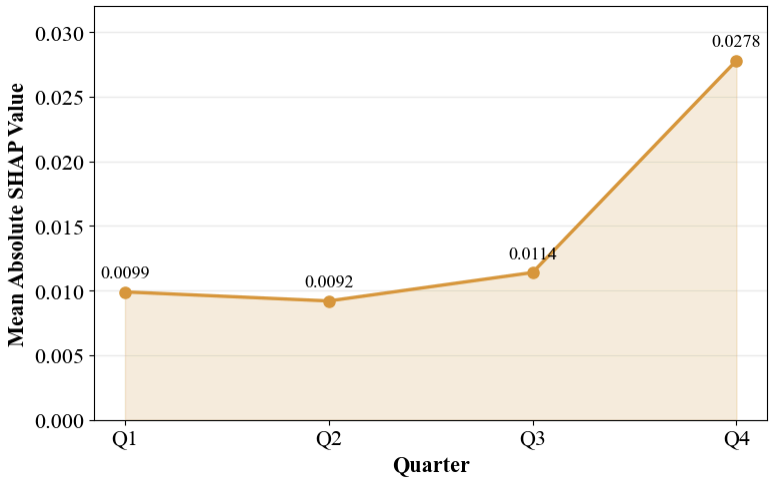


Figure 6. Average SHAP importance of different quarter

Figure 6 shows that the average absolute SHAP value of the fourth quarter is significantly higher than that of the first three quarters, indicating that the fourth-quarter information contributes the most to the model predictions in this dataset. This result is consistent with the business logic of financial risk prediction: the closer to the risk exposure time point, the more sufficient the risk signals reflected in corporate financial conditions, so the model relies more on quarterly information near the prediction period. Finally, the SHAP values are aggregated by indicator dimension, as shown in Figure 7.

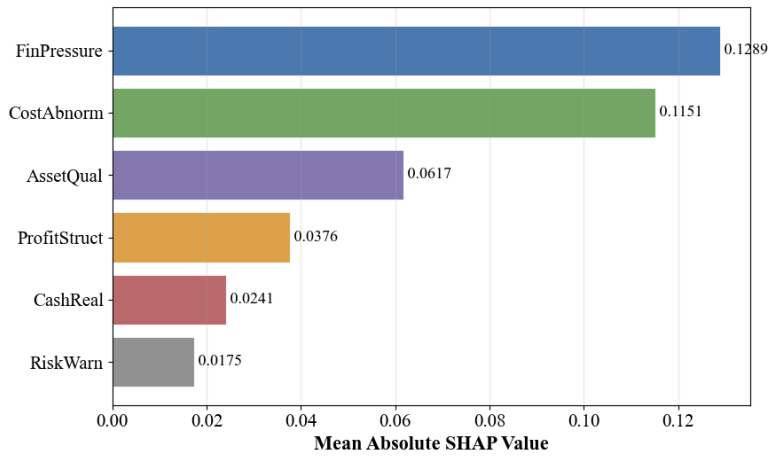


Figure 7. Average SHAP importance of the six comprehensive risk indicators

As shown in Figure 7, debt-servicing and financing pressure has the highest importance, followed by cost, expense and impairment abnormality, operation and asset quality, etc. This indicates that on the dataset of this study, the model mainly identifies ST risks based on corporate debt-servicing and financing pressure, abnormal costs and expenses, and asset quality, among which debt-servicing and financing pressure plays the strongest distinguishing role.

Overall, the SHAP analysis shows that when predicting ST risks, the model mainly focuses on the fourth-quarter information close to the prediction period, as well as key indicators such as debt-servicing and financing pressure, cost and expense, and asset quality. This demonstrates that the

DOI: 10.54254/2755-2721/2026.34374

model can not only make predictions using quarterly matrices but also provide interpretability from both temporal and indicator dimensions.

4. Conclusions and prospects

To meet the demand for accurate financial risk early warning of listed companies, this paper proposes a financial risk prediction method based on CNN-Transformer ensemble learning, which is verified through the identification of ST risks among A-share listed companies. Departing from the traditional modeling paradigm relying on annual data, this study adopts quarterly financial data to construct a high-frequency feature system. Multi-dimensional financial indicators of four consecutive quarters are reconstructed into a "quarter × indicator" matrix input, which fully preserves the dynamic evolution information of corporate financial status within the year. The model uses CNN to extract local features and Transformer to capture long-term temporal dependencies. Combined with stratified sampling and ensemble learning, it effectively alleviates model bias caused by class imbalance in risk samples. Empirical results show that the proposed method outperforms traditional machine learning models including Random Forest, XGBoost, and LightGBM in both prediction accuracy and generalization performance. Meanwhile, SHAP-based explainable analysis quantitatively decomposes the model decision logic from temporal and indicator dimensions, verifying the rationality and interpretability of model outputs. In summary, this study organically integrates high-frequency quarterly report matrix input, CNN-Transformer temporal modeling, imbalanced sample optimization strategies, and explainable analysis, forming a complete, robust, and practically applicable financial risk prediction framework.

This research still leaves room for expansion and further improvement. In terms of model structure, graph neural networks (GNN) can be introduced to incorporate inter-enterprise relational networks such as equity connections, supply chain transactions, and industry linkages, so as to further characterize risk transmission paths and spillover effects in associated networks. For the ensemble learning framework, on the basis of the existing Bagging strategy, Boosting, Stacking and other ensemble paradigms can be compared to explore more suitable mechanisms for imbalanced financial data. In terms of information dimensions, the limitations of structured financial indicators can be broken by integrating unstructured information such as annual report texts, interim announcements, news public opinions, and management discussion and analysis (MD&A). A multimodal financial risk prediction model can thus be constructed to achieve comprehensive, dynamic, and multi-perspective identification of potential corporate risks.

References

[1] Kim, H., Cho, H. and Ryu, D. (2020) Corporate default predictions using machine learning: Literature review. Sustainability, 12(16), 6325. https: //doi.org/10.3390/su12166325

[2] El Madou, K., Marso, S., El Kharrim, M. and El Merouani, M. (2024) Evolutions in machine learning technology for financial distress prediction: A comprehensive review and comparative analysis. Expert Systems, 41(2), e13485. https: //doi.org/10.1111/exsy.13485

[3] Liang, M., Li, H. X., & Zhang, S. M. (2023). Measurement of financial distress and cross-sectional stock returns based on ST prediction. Chinese Journal of Management Science, 31(2), 138–149. https: //doi.org/10.16381/j.cnki.issn1003-207x.2020.1532

[4] Raudys, A. and Goldstein, E. (2022) Forecasting detrended volatility risk and financial price series using LSTM neural networks and XGBoost regressor. Journal of Risk and Financial Management, 15(12), 602. https: //doi.org/10.3390/jrfm15120602

[5] Xu, Y., Li, J. and Wu, A. (2024) An XGboost algorithm based model for financial risk prediction. Tehnički vjesnik, 31(6), 1898–1907. https: //doi.org/10.17559/TV-20231021001043

DOI: 10.54254/2755-2721/2026.34374

[6] Wang, D., Li, L. and Zhao, D. (2022) Corporate finance risk prediction based on LightGBM. Information Sciences, 602, 259–268. https: //doi.org/10.1016/j.ins.2022.04.058
[7] Zhang, J. (2024) Impact of an improved random forest-based financial management model on the effectiveness of corporate sustainability decisions. Systems and Soft Computing, 6, 200102. https: //doi.org/10.1016/j.sasc.2024.200102
[8] Zou, Y., Yuan, Y., Zhu, C. and Yu, C. (2025) Symmetric equilibrium bagging–cascading boosting ensemble for financial risk early warning. Symmetry, 17(10), 1779. https: //doi.org/10.3390/sym17101779
[9] Chen, W., Fang, J. and Gu, C. (2025) Class imbalance in financial distress models: A novel LightGBM-based hybrid classifier. In: Lu, P., Lu, Z. and Cheng, D. (eds.) Digital Finance: CCF China Digital Finance Conference, CDFC 2025, Shanghai, China, August 15–17, 2025, Revised Selected Papers. Communications in Computer and Information Science, vol. 2711. Singapore: Springer Nature Singapore, pp. 119–137. https: //doi.org/10.1007/978-981-95-5211-5_7
[10] Matsumaru, M. and Katagiri, H. (2025) A two-stage machine learning approach to bankruptcy prediction: Integrating full-feature modeling and optimized feature selection. Journal of Risk and Financial Management, 18(12), 662. https: //doi.org/10.3390/jrfm18120662
[11] Nguyen, H. H., Viviani, J. L. and Ben Jabeur, S. (2026) Predicting firm bankruptcy using macroeconomic and uncertainty variables: An ensemble machine learning study of the French market. The European Journal of Finance. https: //doi.org/10.1080/1351847X.2026.2661059
[12] Che, W., Wang, Z., Jiang, C. and Abedin, M. Z. (2024) Predicting financial distress using multimodal data: An attentive and regularized deep learning method. Information Processing & Management, 61(4), 103703. https: //doi.org/10.1016/j.ipm.2024.103703
[13] Sun, Y. (2025) Financial transaction network risk prediction model based on graph neural network. Procedia Computer Science, 261, 763–771. https: //doi.org/10.1016/j.procs.2025.04.403
[14] Zhang, M. (2025) GNN-GAT-LSTM: A graph neural network-based early warning model for enterprise financial risk in dynamic inter-firm networks. Informatica, 49(24), 369–384. https: //doi.org/10.31449/inf.v49i23.9868
[15] Korangi, K., Mues, C. and Bravo, C. (2023) A transformer-based model for default prediction in mid-cap corporate markets. European Journal of Operational Research, 308(1), 306–320. https: //doi.org/10.1016/j.ejor.2022.10.032