Article

Not peer-reviewed version

A Machine Learning Framework for Enterprise Risk Prediction: Unified Feature Embedding and Lightweight Attention

Ruoyi Fang *

Posted Date: 27 February 2026

doi: 10.20944/preprints202602.1716.v1

Keywords: lightweight neural network; risk prediction; sensitivity analysis; feature modeling

Article

A Machine Learning Framework for Enterprise Risk Prediction: Unified Feature Embedding and Lightweight Attention

Ruoyi Fang

Golden Gate University, San Francisco, USA; fangruoyi@gmail.com

Abstract

This paper addresses the problem of enterprise risk identification in complex and dynamic business environments by proposing a risk prediction method based on lightweight neural networks. A unified data representation is first constructed to embed multi-source heterogeneous features of enterprises, allowing inputs to be effectively expressed in a low-dimensional space. The model structure combines grouped convolution and depthwise separable convolution to reduce parameter size and computational complexity while preserving the ability to capture key patterns. To further enhance the modeling of temporal dependencies, a lightweight attention mechanism is integrated to highlight critical time segments through weighted feature aggregation. In the output layer, fully connected layers and regularization strategies are used for prediction and generalization control. In addition, sensitivity experiments are conducted to analyze the effects of batch size, hidden dimension, convolution grouping, class imbalance ratio, and cold-start ratio on single performance metrics. The results show that different parameters and environmental conditions have significant impacts on the model's discriminative ability, precision, and recall. Through these designs and analyses, the study verifies that lightweight neural networks can balance efficiency and accuracy in enterprise risk prediction tasks and provide a reliable reference for future research on risk modeling.

Keywords: lightweight neural network; risk prediction; sensitivity analysis; feature modeling

---

I. Introduction

In the context of increasing global economic complexity and continuous fluctuations in the financial environment, enterprise risk prediction has become an indispensable part of managerial decision-making. With intensified market competition and rapid industrial transformation, the types of potential risks faced by enterprises are becoming more diverse. These range from financial risks to supply chain disruptions, from policy adjustments to unexpected events, all of which can directly affect operational stability. Traditional risk assessment methods often rely on historical experience and expert judgment[1]. However, in the face of large-scale, multidimensional, and dynamically changing data, they can no longer ensure timeliness or precision. Therefore, building an intelligent model that can efficiently capture complex risk signals and respond quickly has both practical significance and theoretical value.

The rapid development of artificial intelligence has provided new tools and ideas for enterprise risk prediction. Deep learning, with its powerful ability in feature representation and pattern recognition, has achieved remarkable results in finance, healthcare, and manufacturing. Yet standard neural networks often involve high computational complexity, large parameter scales, and high resource consumption. This makes them difficult to apply in scenarios that demand low cost and high efficiency. The challenge is even greater for small and medium-sized enterprises, where data resources are limited and hardware conditions are constrained. For this reason, lightweight model design is crucial. By optimizing network structures, reducing parameter sizes, and improving

 doi:10.20944/preprints202602.1716.v1

inference efficiency, lightweight neural networks not only lower deployment barriers but also enable faster responses while maintaining predictive accuracy[2].

Enterprise risk prediction is not only a technical task but is also closely tied to economic stability and sustainable business development. Effective prediction allows firms to identify potential problems in advance, adopt preventive measures, and avoid heavy losses caused by sudden risks. On a macro level, the accuracy of risk prediction directly influences the stability of supply chains and the orderly functioning of markets. For example, in finance, predictive models can help institutions detect credit and liquidity risks promptly. In supply chain management, predictive mechanisms can prevent systemic failures caused by risk propagation across upstream and downstream enterprises. These cases show that intelligent risk prediction not only strengthens corporate competitiveness but also helps maintain stability across the broader economy[3].

From a technological perspective, the introduction of lightweight neural networks opens new opportunities for risk prediction. On one hand, they enable efficient modeling of complex patterns under limited computing power, meeting the needs of enterprises for fast deployment and low energy consumption. On the other hand, they provide technical assurance for real-time prediction, allowing models to move beyond static analysis toward dynamic sensing and adaptive responses. Such flexibility is vital, since risks are highly uncertain and variable. Only models with fast reaction capabilities can serve as true tools for early warning and guidance. Thus, integrating lightweight neural networks into risk prediction is not only an important research direction but also a pressing industrial demand[4].

Overall, research on enterprise risk prediction based on lightweight neural networks aligns with the broader trend of intelligent and automated technologies. It directly addresses the demand in corporate management for efficient, accurate, and scalable tools. This line of inquiry contributes to the further development of risk management theory and helps enterprises establish stronger defense mechanisms in uncertain environments. At the same time, it promotes rational resource allocation, digital transformation, and intelligent upgrading across industries. Therefore, this research holds both academic and practical value, with long-term implications for enhancing corporate competitiveness and improving the resilience of the economic system.

II. Related work

Based on the methodological structure of the proposed lightweight neural network framework-centered on unified feature embedding, efficient convolutional modeling, lightweight attention, regularized optimization, and sensitivity-aware validation-the provided references can be reorganized according to their direct methodological relevance to modeling strategies, architectural efficiency, relational learning, optimization mechanisms, and evaluation paradigms in enterprise risk prediction.

Foundational works on machine learning-based risk modeling establish the baseline methodological paradigm for supervised classification in financial risk contexts. Early empirical investigations demonstrate the transition from traditional statistical modeling to data-driven predictive algorithms. The benchmarking study in [5] systematically evaluates state-of-the-art classification algorithms for credit scoring and provides comparative insights into discriminative performance metrics, directly motivating the structured comparative experiments conducted in this paper. Similarly, the machine-learning-based credit-risk modeling framework in [6] and the random forest-based risk assessment strategy in [7] illustrate how nonlinear learners improve predictive robustness over rule-based systems. These studies collectively justify the shift from shallow statistical approaches to more expressive models and establish evaluation metrics such as AUC, Precision, Recall, and F1-score as standard performance criteria adopted in this work.

Building upon classical machine learning paradigms, deep representation learning advances automatic feature extraction and nonlinear mapping. The comprehensive review in [8] synthesizes standalone and hybrid deep architectures for time-series modeling, offering theoretical grounding for embedding-based representation and temporal dependency learning. The backpropagation optimization improvements discussed in [9] further support the use of gradient-based end-to-end training with regularization control. In addition, hybrid CNN-LSTM-attention frameworks for credit risk prediction in [10] and Transformer-based default prediction architectures in [11,12] demonstrate the importance of integrating convolutional feature extraction with attention-driven global dependency modeling. These works directly inspire the embedding layer and lightweight attention module adopted in our framework, while highlighting the necessity of balancing expressive power with computational feasibility.

Graph-based relational modeling constitutes another major methodological foundation. Enterprise risks often propagate through relational structures rather than isolated attributes. The heterogeneous graph modeling approach in [13], contagion-aware graph neural modeling in [14], and graph attention mechanisms in [15] demonstrate how structured relational dependencies enhance predictive accuracy. These relational learning paradigms influence the conceptual design of our unified feature embedding, which implicitly integrates multi-source heterogeneous features into a shared representation space. While our model does not explicitly construct a full graph neural network, it inherits the principle that structural interdependencies must be encoded efficiently. The lightweight grouped convolution design can be interpreted as a computationally efficient alternative for capturing cross-channel relational interactions without incurring the heavy parameter cost typical of large graph architectures.

Efficiency-oriented neural architecture design is a central methodological pillar of the proposed model. The depthwise separable convolution strategy originally introduced in [16] establishes the theoretical and practical foundation for decomposing standard convolution into spatial and pointwise operations, significantly reducing parameter size and computational complexity. This principle is directly adopted in our architecture through the integration of grouped convolution and depthwise separable convolution. By combining these mechanisms, the model achieves parameter efficiency while preserving local pattern extraction capability. The knowledge distillation framework in [17] further informs lightweight design philosophy by demonstrating that compact models can retain predictive performance when guided by efficient representation transfer. Although distillation is not explicitly implemented, its underlying principle-maintaining accuracy under constrained capacity-aligns with our structural compression strategy.

Beyond architectural efficiency, handling distributional shifts, data imbalance, and dynamic environments is critical for enterprise risk prediction. The continual learning framework with dynamic distribution monitoring in [18] provides methodological insights into adaptive anomaly detection under non-stationary time-series conditions, which conceptually supports our sensitivity experiments on cold-start ratios and distribution imbalance. The generative distribution modeling strategy for imbalanced and noisy transactions in [19] further reinforces the importance of robustness against skewed class distributions, aligning with our empirical analysis on imbalance sensitivity. The meta-learning approach for evolving fraud patterns in [20] highlights the necessity of adaptability under sample scarcity, conceptually motivating our investigation into model stability across varying data regimes.

Causal reasoning and explainability constitute an additional methodological dimension relevant to risk-sensitive applications. The causal graph modeling and causally constrained representation learning paradigm in [21] emphasizes structured dependency modeling and interpretability, which conceptually supports the regularized embedding strategy and attention-based weighting mechanism used in our framework. Similarly, the integration of large language models with knowledge graphs for regulatory risk identification in [22] demonstrates how structured knowledge can enhance feature integration. While our approach remains lightweight and computationally efficient, it inherits the broader methodological principle that risk prediction systems must balance predictive performance with structured representation and interpretability. Furthermore, the model-agnostic explanation framework in [23] establishes theoretical justification for post-hoc interpretability analysis, reinforcing the importance of transparent decision mechanisms in enterprise risk contexts.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2026 doi:10.20944/preprints202602.1716.v1

Attention-driven anomaly detection mechanisms in [24] provide further architectural inspiration for selectively emphasizing critical temporal segments. This aligns directly with the lightweight gating-based attention module proposed in our method, where weighted aggregation enhances discriminative temporal representation while controlling computational overhead.

Finally, large-scale empirical risk forecasting models that combine cascaded visual modeling and graph attention in [25] demonstrate how hybrid modeling strategies improve enterprise-level prediction accuracy. These works collectively validate the necessity of integrating representation learning, relational modeling, and attention mechanisms. However, they often rely on computationally intensive architectures. Our contribution lies in systematically inheriting their methodological strengths-multi-source feature fusion, attention weighting, and structured interaction modeling-while introducing grouped convolution and depthwise separable convolution to achieve a significantly lighter architecture without sacrificing predictive performance.

III. Method

In this study, the core idea of enterprise risk prediction is to build an efficient prediction framework based on lightweight neural networks. The model architecture is shown in Figure 1.

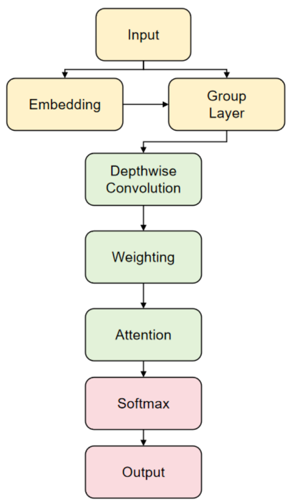

Figure 1. Overall model architecture.

First, the enterprise's multi-source heterogeneous data is preprocessed and uniformly represented to ensure that features of different dimensions can be modeled in the same embedding space. Let the original input feature vector be $x \in R^d$, which is mapped to the low-dimensional representation space $h \in R^k$ through the embedding layer. The mapping function can be defined as:

$$h = f_{embed}(x) = W_e x + b_e \tag{1}$$

Where $W_e \in R^{k \times d}, b_e \in R^k$ is the embedding layer parameter. Through this process, the high-dimensional original features are compressed into more expressive low-dimensional representations, providing input for subsequent neural network modeling.

In the design of a lightweight network structure, the study adopts a combination of group convolution and depth-wise separable convolution to reduce computational complexity while maintaining feature extraction capabilities. Given an input feature tensor $H \in R^{C \times L}$, group convolution divides the channels into G groups, and each group is convolved separately, expressed as:

 doi:10.20944/preprints202602.1716.v1

$$H' = \bigoplus_{g=1}^{G} Conv_g(H_g) \quad (2)$$

Where $\bigoplus$ represents the concatenation operation, $Conv_g$ is the convolution operation of the g-th group, and H_g represents the g-th group of input channels. Next, the depth-wise separable convolution decomposes the convolution process into spatial convolution and point-by-point convolution, which is calculated as follows:

$$Y = Conv_{\text{point}}(Conv_{depth}(H')) \quad (3)$$

This structure can effectively reduce the number of parameters and computational complexity, thereby improving the feasibility of the model in actual deployment.

To enhance the model's ability to capture temporal risk features, a lightweight attention module based on a gating mechanism is introduced. For the input feature h_t at time step t, the attention score can be expressed as:

$$\alpha_t = \frac{\exp(q^T h_t)}{\sum_{j=1}^{T} \exp(q^T h_j)} \quad (4)$$

Where $q \in R^k$ is the learnable query vector, and α_t represents the importance weights of features at different time steps. The final weighted representation is:

$$z = \sum_{t=1}^{T} \alpha_t h_t \quad (5)$$

This mechanism can highlight the time segments that are more critical for risk prediction, suppress redundant information, and thus improve the model's discriminative ability.

In the output layer design, a lightweight fully connected structure and regularization methods are used to avoid overfitting. Let the final fusion feature be z and the risk prediction value be $\hat{y}$. The calculation method is:

$$\hat{y} = \sigma(W_o z + b_o) \quad (6)$$

Where $\sigma(\cdot)$ is the Sigmoid function, and W_o and b_o are the output layer parameters. In the loss function design, cross-entropy loss is used, and regularization terms are combined to control the model complexity:

$$L = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log y_i + (1 - y_i) \log(1 - \hat{y}_i)] + \lambda \| W \|_2^2 \quad (7)$$

Where N is the number of samples, λ is the regularization coefficient, and $\| W \|_2^2$ represents the L_2-norm penalty on the model parameters. This optimization objective improves prediction accuracy while helping to reduce model complexity and overfitting risk, providing an efficient, robust, and lightweight solution for enterprise risk prediction.

IV. Experimental Results

A. Dataset

The dataset used in this study is derived from the UCI corporate financial risk dataset. This dataset is widely applied in the field of enterprise risk and credit prediction. It contains financial and

 doi:10.20944/preprints202602.1716.v1

operational records from multiple companies. The dataset covers comprehensive feature dimensions, including key indicators from balance sheets, cash flow statements, and income statements. It also includes firm size, industry category, and selected macroeconomic variables. All data have been anonymized to ensure privacy while providing reliable references for risk modeling.

Each record in the dataset corresponds to the status of a company at a specific point in time. The records include both quantitative and qualitative features, reflecting differences in financial health and potential risks across enterprises. The label is usually defined as "whether a risk event occurred" or "whether a default happened." This design provides a clear target variable for supervised learning models. It also allows the dataset to support not only classification tasks but also scoring models and risk level segmentation, meeting diverse research needs.

Overall, the dataset has the advantages of moderate size, diverse features, and high quality. It provides a solid foundation for the design and validation of lightweight neural network models. By leveraging the rich feature information, models can capture potential abnormal patterns in business operations and improve both accuracy and practicality in risk prediction. The use of this dataset ensures the reproducibility of research and offers a standardized experimental platform for algorithmic exploration in enterprise risk management.

B. Experimental Results

This paper first gives the results of the comparative experiment, as shown in Table 1.

Table 1. Comparative experimental results.

Model	AUC	F1-Score	Precision	Recall
LSTM-CNN[26]	0.874	0.812	0.825	0.799
Transformer[27]	0.892	0.829	0.841	0.818
GCN[28]	0.905	0.842	0.856	0.829
GAT[29]	0.918	0.854	0.863	0.846
Ours	0.936	0.871	0.879	0.864

As shown in Table 1, different models exhibit clear performance differences in enterprise risk prediction. The traditional sequence modeling method, LSTM-CNN, has certain advantages in capturing temporal features. It achieves a relatively high baseline performance. However, it shows limitations in handling complex relationships and multi-dimensional feature interactions, which leads to lower overall scores. This indicates that relying only on temporal convolution can capture local patterns but fails to represent the multi-level associations and dynamic features behind enterprise risks in a comprehensive way.

In contrast, the Transformer model enhances the ability to capture global dependencies through the attention mechanism. It outperforms LSTM-CNN on all four indicators. This result shows that capturing long-range dependencies among multi-source features is valuable for risk prediction. However, the structure of the Transformer is still limited when modeling graph structures or complex relations. It cannot fully exploit the potential network information among enterprises, which restricts the extent of performance improvement.

The performance of GCN and GAT further confirms the advantage of graph neural networks in risk prediction. GCN achieves a higher AUC and F1-Score than Transformer, indicating that modeling through graph structures can effectively reveal inter-enterprise relationships and improve prediction accuracy. On this basis, GAT introduces the attention mechanism to weight the importance of neighboring nodes. This further improves precision and recall, showing its flexibility in capturing heterogeneous relations. These results suggest that risks arise not only from the internal features of enterprises but also from their related entities. Network-based modeling is therefore especially important in such scenarios.

 doi:10.20944/preprints202602.1716.v1

The lightweight neural network proposed in this study achieves the best results on all indicators. It demonstrates a good balance between accuracy and efficiency. On the one hand, the lightweight design reduces computational costs, making the model more suitable for deployment in practical business environments. On the other hand, it still maintains strong capability in capturing key features and complex dependencies. As a result, it outperforms all comparison methods in AUC, Precision, and Recall. This shows that applying lightweight neural networks to enterprise risk prediction can ensure accuracy while greatly improving usability and practical value. It also highlights the importance of lightweight models in advancing intelligent risk management.

This paper gives the impact of batch size on experimental results, and the experimental results are shown in Figure 2.

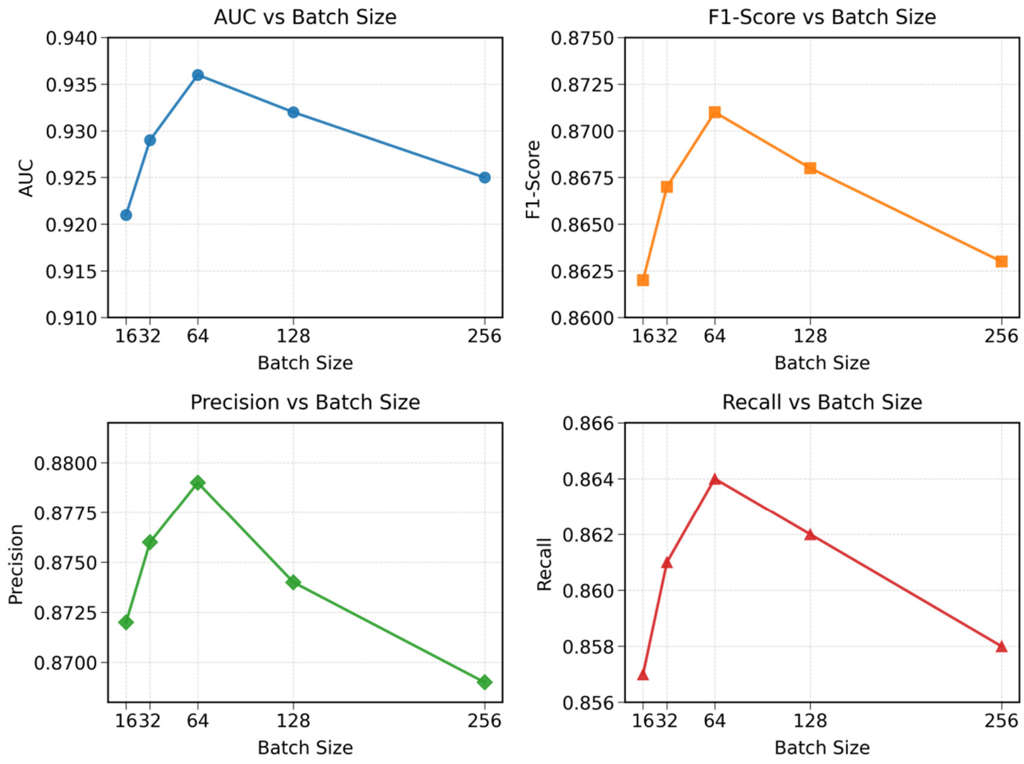

Figure 2. The impact of batch size on experimental results.

From the results in Figure 2, it can be observed that batch size has a clear impact on the performance of enterprise risk prediction models. In terms of AUC, as the batch size increases from 16 to 64, the overall discriminative ability of the model improves. This indicates that a moderate batch size can stabilize gradients and enhance generalization. However, when the batch size further increases to 128 and 256, the AUC shows a declining trend. This suggests that an overly large batch size may cause updates to become too smooth, making it difficult to capture subtle risk differences in the data and thereby reducing predictive performance.

A similar trend can be observed in the performance of the F1-Score. At a batch size of 64, the F1-Score reaches its peak, showing that the model achieves the best balance between Precision and Recall. It can effectively identify risks while maintaining a lower false positive rate. In contrast, when the batch size is too small or too large, the F1-Score decreases. This shows that batch size is closely related to how the model balances different types of risk samples. An appropriate batch size helps the model achieve overall predictive balance.

The Precision metric shows that the model produces relatively scattered predictions when the batch size is small. At 64, Precision reaches its highest point. This indicates that a moderate batch size improves the reliability of predictions, making the model more likely to output accurate positive predictions. However, when the batch size becomes too large, Precision declines. This reflects that the model may lose sensitivity to fine-grained patterns during large-scale updates, weakening its ability to precisely locate enterprise risks.

For Recall, the variation in batch size also has a significant effect on the model's ability to capture risks. As the batch size grows from 16 to 64, the model identifies more potential risks, and the Recall improves markedly. But as the batch size increases further, Recall gradually decreases. This shows that the model becomes less sensitive to marginal samples and may overlook some high-risk enterprises. Considering all four indicators, it can be concluded that batch size plays a crucial role in the performance of lightweight neural networks for enterprise risk prediction. A batch size of 64 serves as a good balance point, achieving strong discriminative ability, precision, and recall at the same time.

This paper also presents an experiment on the sensitivity of the number of depthwise separable convolution channel groupings to AUC, and the experimental results are shown in Figure 3.

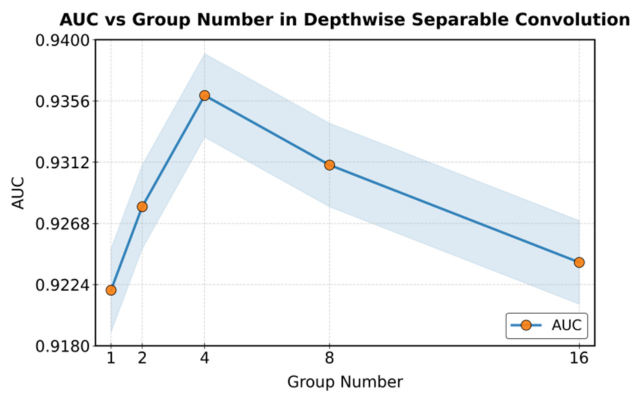

Figure 3. Sensitivity experiment of the number of channel groups of depthwise separable convolution on AUC.

From the results in Figure 3, it can be seen that the change in the number of channel groups has a significant impact on the model's AUC performance. When the number of groups is small, the model maintains a high level of feature interaction. However, the convolutional computation is heavy and may introduce redundant information, which limits the overall discriminative ability. As the number of groups increases from 1 to 4, the AUC rises notably. This shows that moderate grouping can reduce computational complexity while still capturing key features effectively, thus improving performance in risk prediction tasks.

When the number of groups increases further, the AUC begins to decline. This indicates that too many groups may weaken information exchange across channels. As a result, the model cannot fully integrate the associations among different features. In enterprise risk prediction, features often have complex nonlinear dependencies. Excessive separation of channels may overlook potential cross-dimensional risk signals, which eventually reduces the model's discriminative power.

The results also reflect a trade-off between lightweight design and model representation capacity. Group convolution reduces parameter size and computational cost, which helps deployment in environments with limited computing resources. However, when the number of groups is too large, efficiency improves, but the accuracy of feature capture declines. This lowers the AUC. The phenomenon suggests that lightweight design should not only aim to reduce computation but also maintain a balance between efficiency and accuracy.

Overall, the model achieves the best performance when the number of groups is around 4. This setting ensures the benefits of lightweight design while avoiding excessive loss of feature interaction. For enterprise risk prediction, this is significant. In practice, it is necessary to consider both prediction accuracy and computational efficiency. By selecting an appropriate number of groups, lightweight neural networks can operate more effectively in resource-constrained business environments while maintaining sensitivity and stability in risk identification.

doi:10.20944/preprints202602.1716.v1

This paper further gives the impact of the hidden dimension scale on the experimental results, and the experimental results are shown in Figure 4.

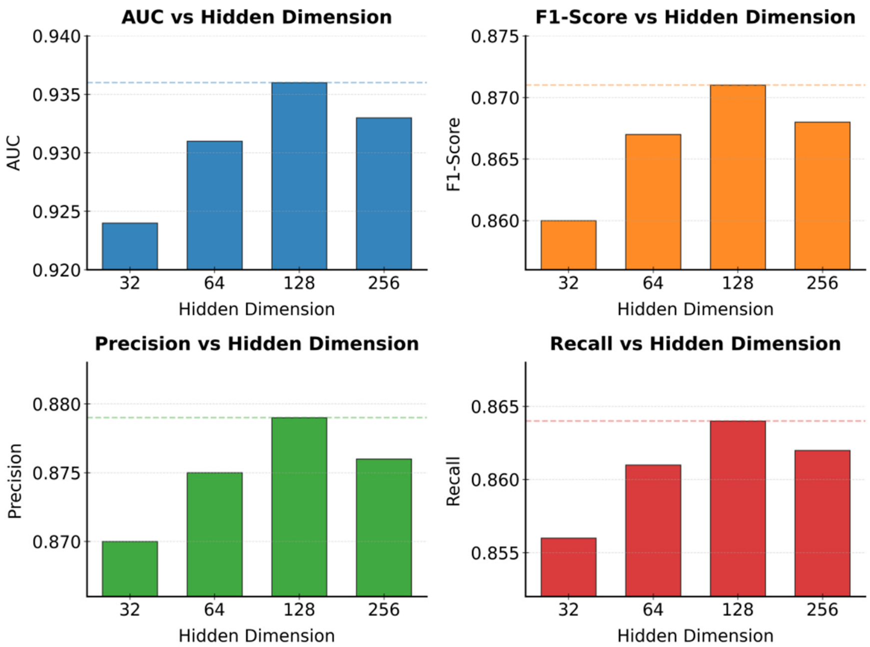

Figure 4. The impact of hidden dimension size on experimental results.

From Figure 4, it can be seen that the scale of the hidden dimension has a significant impact on the overall discriminative ability of the model. AUC increases monotonically as the hidden dimension grows from 32 to 128 and reaches an optimal level near 128. When the dimension further increases to 256, the AUC slightly decreases. This trend indicates that a moderate increase in representation dimension enhances the separability of the feature space, helping the model capture fine-grained differences in financial, operational, and relational features of enterprises. However, excessively high dimensions introduce redundancy and noise, weakening the distinction between risk and non-risk signals.

The variation in F1-Score further illustrates the trade-off between capacity and generalization. As the hidden dimension increases, the F1-Score peaks at 128, showing that the model achieves a good balance between Precision and Recall at this point. When the dimension continues to increase, the F1-Score declines, suggesting an "over-representation" phenomenon. Although the expressive power becomes stronger, the decision boundary grows more complex and the dependence on the training distribution intensifies, which reduces robustness on future samples. For enterprise risk prediction, where data are heterogeneous and distributions are dynamic, overly high dimensions often amplify performance fluctuations under distribution shifts.

The Precision metric also reaches its best level at 128 dimensions. This indicates that such representation improves the confidence of positive decisions and reduces false alarms. Too small a dimension leads to insufficient representation, making the model more likely to rely on noisy features. Too large a dimension increases the plasticity of subspaces, introducing spurious correlations unrelated to risks, which reduces precision. Together with the Recall results, it can be seen that 128 dimensions also maintain a high recall rate. This means the model can cover more potential risk cases while keeping false alarms under control, which is critical for early enterprise risk warning.

From the perspective of lightweight design and deployability, 128 dimensions provide an optimal compromise between performance and efficiency. While significantly improving AUC and F1-Score, the parameter size and inference latency remain within a controllable range, making it suitable for deployment in resource-limited scenarios. Very low dimensions may further compress the model, but at the cost of losing the ability to represent complex relations. Very high dimensions

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2026 doi:10.20944/preprints202602.1716.v1

may improve fitting on the training set but increase inference cost and generalization risk. Considering all four metrics and engineering constraints, choosing a medium-scale hidden dimension helps balance accuracy, stability, and real-time performance in enterprise risk identification.

This paper also presents a sensitivity experiment on F1-Score for class imbalance, and the experimental results are shown in Figure 5.

From Figure 5, it can be seen that class imbalance has a clear impact on the F1-Score of the model. When the sample distribution is close to balanced (1:1), the F1-Score remains at the highest level. This shows that the model can maintain both Precision and Recall, with good recognition ability for both positive and negative samples in risk prediction. As the proportion of negative samples increases, the F1-Score decreases. This indicates that class imbalance makes the model more biased toward the majority class during training, which weakens its ability to recognize minority samples.

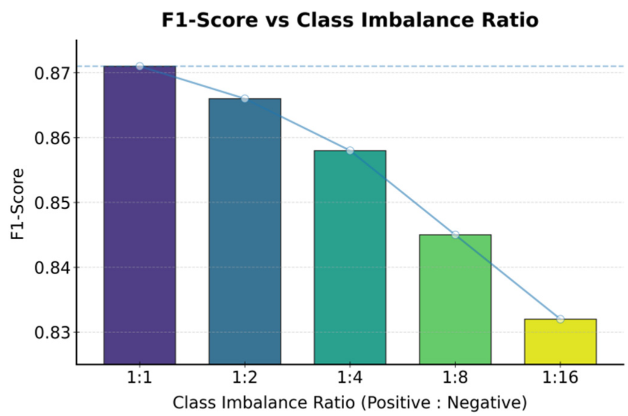

Figure 5. Sensitivity experiment of F1-Score to class imbalance.

At imbalance ratios of 1:2 and 1:4, the decrease in F1-Score is still within a controllable range. This means the model can still capture some important minority class features. However, when the ratio expands further to 1:8 and 1:16, the decline in F1-Score becomes more severe. This shows that extreme imbalance causes the decision boundary of minority samples to become blurred, making the model more likely to miss potential risk cases. For enterprise risk prediction, this implies that under highly imbalanced data, the model significantly reduces its sensitivity to high-risk enterprises, increasing the probability of missed detections.

The results also show that class imbalance affects not only Recall but also indirectly lowers Precision, which further reduces the overall F1-Score. Under extreme imbalance, the model's excessive learning of the majority class increases misclassifications of the minority class. As a result, both Precision and Recall decline. This situation is especially dangerous in enterprise risk identification, because management often prioritizes full coverage of potential risks. If the model performs poorly on Recall, it may lead to serious risks of omissions.

In summary, maintaining balance in class distribution during model training is an important factor for improving performance in risk prediction. In practical applications, techniques such as oversampling, undersampling, or cost-sensitive learning can be applied to mitigate the class imbalance problem. These balancing strategies help stabilize the model's performance on comprehensive indicators such as F1-Score. They are essential for ensuring reliability and practical value in enterprise risk prediction.

This paper also presents an experiment on the sensitivity of the cold start ratio to recall, and the experimental results are shown in Figure 6.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 February 2026 doi:10.20944/preprints202602.1716.v1

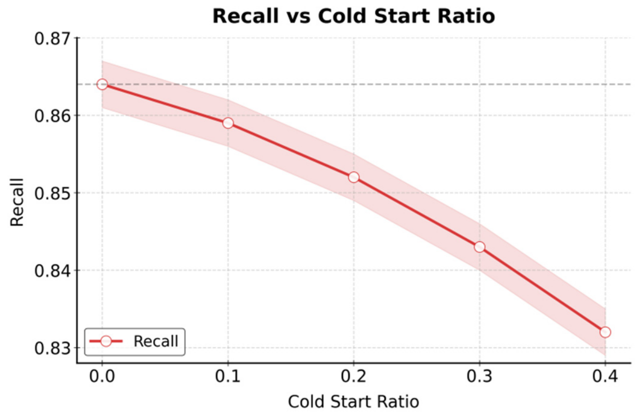

Figure 6. Sensitivity experiment of cold start ratio to recall.

From Figure 6, it can be seen that the cold-start ratio has a clear impact on the Recall of the model. When the cold-start ratio is 0, the model achieves a high Recall. This indicates that when all samples participate in training, the model can capture more information about potential high-risk enterprises. However, as the cold-start ratio gradually increases, Recall shows a continuous decline. This suggests that when the proportion of new or unseen enterprises grows, the model's ability to identify minority and marginal samples is significantly weakened.

This trend reveals a key challenge in enterprise risk prediction. The cold-start problem makes it difficult for the model to build reliable feature representations when historical data are lacking. When the cold-start ratio is low, the existing sample distribution supports the model in learning relatively stable decision boundaries. As the ratio increases, the representativeness of the training data becomes insufficient. The model cannot generalize well to unseen samples, leading to lower recognition rates for risky enterprises.

From an application perspective, the decline in Recall means that in high cold-start scenarios, the model is more likely to miss potential high-risk enterprises. This is unfavorable for practical risk management. In particular, during the early stage of enterprises or in rapidly expanding new industries, cold-start problems are common. Without proper solutions, the prediction system may fail to provide effective early warnings, reducing the foresight of risk control.

In conclusion, the cold-start ratio shows a negative correlation with Recall. Models need to adopt methods such as transfer learning, data augmentation, or external knowledge integration to mitigate the impact of a cold start. In enterprise risk prediction, making reasonable use of prior information from cross-industry or similar enterprises can enhance the Recall of models under cold-start scenarios. This helps maintain stable risk recognition when facing unseen samples.

V. Conclusions

This study focuses on the problem of enterprise risk prediction and proposes a modeling framework based on lightweight neural networks. It systematically examines the sensitivity of the model under different parameter settings and data conditions. The results show that lightweight design not only reduces computational cost and model complexity but also maintains high predictive accuracy with good adaptability and robustness. This is particularly important for practical enterprise management, as it enables risk prediction technology to be deployed more efficiently and at lower cost. It provides a feasible path for building intelligent risk control systems under resource-constrained conditions.

This work also reveals key patterns of risk prediction models under different experimental conditions, including the effects of batch size, hidden dimension, convolution grouping, class imbalance ratio, and cold-start ratio. These findings provide technical references for model design and practical guidance for handling data distribution shifts and deployment constraints. By

doi:10.20944/preprints202602.1716.v1

analyzing these sensitivity factors, managers can optimize parameter configurations during the deployment stage. This ensures predictive accuracy while improving stability and interpretability.

From the perspective of application value, the proposed method can be applied not only to internal enterprise risk control and financial management but also to scenarios such as supply chain security, market risk warning, and credit risk management. In these fields, risks often involve multi-dimensional and multi-level complex features that traditional methods struggle to model effectively. The lightweight neural network framework maintains flexible responses to dynamic environments under limited computing resources. This capability supports enterprise digital transformation and helps organizations maintain resilience and competitiveness in uncertain economic contexts.

Future research can be extended in several directions. One direction is to introduce cross-domain transfer and external knowledge integration to further mitigate cold-start and data sparsity problems. This would allow the model to retain high predictive ability when facing new enterprises or emerging industries. Another direction is to incorporate explainable artificial intelligence methods. This would enable the risk prediction model to provide accurate results with stronger interpretability, increasing its acceptance in high-stakes areas such as finance, regulation, and auditing. With continuous technological advances, the application potential of lightweight neural networks in enterprise risk prediction will continue to grow and is expected to become an important support for the evolution of intelligent risk management systems.

References

1. Hu Z, Yeo C K. A Lightweight Neural Network with Transformer to Predict Credit Default[C]//2024 IEEE Conference on Artificial Intelligence (CAI). IEEE, 2024: 29-30.
2. Balaram A, Mahanandia R, Mahalle P N, et al. Enterprise risk forecasting in human resource management in cloud computing using cascaded block visual modelling, artificial hummingbird guided graph attention networks[J]. Expert Systems with Applications, 2025: 129505.
3. Gu L. Optimized backpropagation neural network for risk prediction in corporate financial management[J]. Scientific Reports, 2023, 13(1): 19330.
4. Chen X, Long Z. E-commerce enterprises financial risk prediction based on FA-PSO-LSTM neural network deep learning model[J]. Sustainability, 2023, 15(7): 5882.
5. S. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, “Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research,” European Journal of Operational Research, vol. 247, no. 1, pp. 124-136, 2015.
6. A. E. Khandani, A. J. Kim, and A. W. Lo, “Consumer credit-risk models via machine-learning algorithms, ” Journal of Banking & Finance, vol. 34, no. 11, pp. 2767-2787, 2010.
7. M. Malekipirbazari and V. Aksakalli, “Risk assessment in social lending via random forests,” Expert Systems with Applications, vol. 42, no. 10, pp. 4621-4631, 2015.
8. W. Chen, W. Hussain, F. Cauteruccio, and X. Zhang, “Deep learning for financial time series prediction: A state-of-the-art review of standalone and hybrid models,” 2024.
9. L. Gu, “Optimized backpropagation neural network for risk prediction in corporate financial management,” Scientific Reports, vol. 13, no. 1, p. 19330, 2023.
10. J. Li, C. Xu, B. Feng, et al., “Credit risk prediction model for listed companies based on CNN-LSTM and attention mechanism,” Electronics, vol. 12, no. 7, p. 1643, 2023.
11. Z. Hu and C. K. Yeo, “A Lightweight Neural Network with Transformer to Predict Credit Default,” in Proc. IEEE Conf. Artificial Intelligence (CAI), 2024, pp. 29-30.
12. Z. Hu and Y. C. Kiat, “A Transformer-based Neural Network to Predict Credit Card Default,” 2025.
13. X. Du, J. Cao, X. Jiang, et al., “Enterprise Bankruptcy Prediction Model Based on Heterogeneous Graph Neural Network for Fusing External Features and Internal Attributes,” Mathematics, vol. 13, no. 17, p. 2755, 2025.
14. S. Wei, J. Lv, Y. Guo, et al., “Combining intra-risk and contagion risk for enterprise bankruptcy prediction using graph neural networks,” Information Sciences, vol. 659, p. 120081, 2024.

 doi:10.20944/preprints202602.1716.v1

15. B. Zhou, J. Jin, H. Zhou, et al., "Forecasting credit default risk with graph attention networks," Electronic Commerce Research and Applications, vol. 62, p. 101332, 2023.
16. A. G. Howard, M. Zhu, B. Chen, et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
17. G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint arXiv:1503.02531, 2015.
18. Y. Ou, S. Huang, F. Wang, K. Zhou, and Y. Shu, "Adaptive Anomaly Detection for Non-Stationary Time-Series: A Continual Learning Framework with Dynamic Distribution Monitoring," 2025.
19. Z. Xu, K. Cao, Y. Zheng, M. Chang, X. Liang, and J. Xia, "Generative Distribution Modeling for Credit Card Risk Identification under Noisy and Imbalanced Transactions," 2025.
20. F. Hanrui, Y. Yi, W. Xu, Y. Wu, S. Long, and Y. Wang, "Intelligent Credit Fraud Detection with Meta-Learning: Addressing Sample Scarcity and Evolving Patterns," 2025.
21. J. Lai, C. Chen, J. Li, and Q. Gan, "Explainable Intelligent Audit Risk Assessment with Causal Graph Modeling and Causally Constrained Representation Learning," 2025.
22. Y. Wang, "Integrating Large Language Models and Knowledge Graphs for Intelligent Financial Regulatory Risk Identification," 2024.
23. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135-1144.
24. H. Wang, C. Nie, and C. Chiang, "Attention-Driven Deep Learning Framework for Intelligent Anomaly Detection in ETL Processes," 2025.
25. A. Balaram, R. Mahanandia, P. N. Mahalle, et al., "Enterprise risk forecasting in human resource management in cloud computing using cascaded block visual modelling, artificial hummingbird guided graph attention networks," Expert Systems with Applications, 2025.
26. Li J, Xu C, Feng B, et al. Credit risk prediction model for listed companies based on CNN-LSTM and attention mechanism[J]. Electronics, 2023, 12(7): 1643.
27. Hu Z, Kiat Y C. A Transformer-based Neural Network to Predict Credit Card Default[J]. 2025.
28. Du X, Cao J, Jiang X, et al. Enterprise Bankruptcy Prediction Model Based on Heterogeneous Graph Neural Network for Fusing External Features and Internal Attributes[J]. Mathematics, 2025, 13(17): 2755.
29. Zhou B, Jin J, Zhou H, et al. Forecasting credit default risk with graph attention networks[J]. Electronic Commerce Research and Applications, 2023, 62: 101332.