

Article | Received 9 January 2026; Revised 12 March 2026; Accepted 22 May 2026; Published 27 August 2026
<https://doi.org/10.55092/pcs20260002>

An intelligent soft-voting based early warning model for loan risk

Chunhao Qiao¹, Qiqi Wang², Xinyue Peng¹, Lei Liu^{1,3,*}, Haiyun Wu¹, Shuai Wang^{1,3}, Dunwei Shi¹, Kongfen Zhu¹ and Li Shi¹

¹ School of Artificial Intelligence, Huaibei Normal University, Huaibei 235000, China

² School of Education, Huaibei Normal University, Huaibei 235000, China

³ Huaibei Key Laboratory of Digital Multimedia Intelligent Information Processing, Huaibei 235000, China

* Correspondence author; E-mail: liul@chnu.edu.cn.

Abstract: Against the backdrop of fintech driven intelligent risk control, traditional manual loan review, hindered by low efficiency and high subjectivity, struggles to meet the massive demand for credit risk assessment. This paper proposes a multi-model fusion framework based on soft-voting, integrating Feedforward Neural Network (FNN), LightGBM, and CatBoost, and combining with scorecard technology to achieve precise prediction of loan default risks and their operational implementation. Using more than 1 million loan records from Alibaba Cloud Tianchi, the study employs data cleaning, clustering binning (10 bins), standardization, and construction of 10 composite features (e.g., ratio of income to loan amount). The core features are selected using chi-square test (p), Population Stability Index (PSI), and Information Value (IV). Through exhaustive experiments, the soft-voting mechanism determines optimal weights (0.1:0.1:0.8), allowing the integrated model to achieve an AUC of 0.739, significantly improving the generalization while maintaining high accuracy. The scorecard model converts probabilities into credit scores, classifying customers into four tiers (A to D). This study transcends the limitations of single models, achieving deep integration of technical interpretability and business implementation through the dual mechanism of model fusion and scorecard. It provides financial institutions with end-to-end solutions from risk quantification to strategy execution, facilitating dynamic monitoring and management of in-lending behavioral risks.

Keywords: soft-voting; Fintech; loan risk early warning.

1. Introduction

In the wave of global financial digital transformation, the scale of the credit market continues to expand, while the limitations of traditional loan risk management have become increasingly prominent. Manual review models are inefficient and highly subjective, struggling to meet massive credit demands and exposing financial institutions to high non-performing loan risks. Against this backdrop, intelligent loan risk early-warning technology has emerged, leveraging big data analysis, artificial intelligence, and



Copyright©2026 by the authors. Published by ELSP. This work is licensed under Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

machine learning to achieve precise identification and dynamic monitoring of loan risks, becoming a critical research direction in financial risk control.

At the policy level, governments worldwide are actively promoting the adoption of intelligent risk control technologies. China's "Fintech Development Plan (2022–2025)" issued by the People's Bank of China explicitly requires financial institutions to use big data and AI to enhance risk early warning capabilities [1]. The China Banking and Insurance Regulatory Commission's "Guiding Opinions on the Digital Transformation of the Banking and Insurance Industry" further emphasizes building intelligent risk control systems to optimize full process supervision for pre-loan, in-loan, and post-loan phases [2]. Internationally, the EU's "Artificial Intelligence Act" imposes strict requirements on the interpretability of AI risk control models [3], while the United States promotes intelligent credit risk assessment through the FICO credit scoring system and tech companies such as ZestFinance [4].

From a market demand perspective, financial institutions urgently need to reduce non-performing loan ratios and improve approval efficiency to meet stricter regulatory requirements [5]. Meanwhile, small and micro enterprises and individual users expect more efficient and fair loan services, aiming to avoid biases in traditional manual reviews [6].

The current intelligent risk early warning models still have difficulty in balancing accuracy and interpretability. To address these issues, by building an interpretable ensemble model, this research aims to help banks reduce non-performing loan ratios and comply with regulatory requirements such as the "Guidelines on Stress Testing for Commercial Banks" issued by the China Banking Regulatory Commission [7].

This paper proposes an innovative method for early warning of loan risk in four dimensions: model architecture, feature engineering, risk quantification, and strategy application. The main contributions of this study are as follows.

- Traditional loan risk models heavily rely on single algorithms, failing to balance the automatic feature extraction capability of deep learning and the interpretability of machine learning models. This study innovatively constructs an FNN–LightGBM–CatBoost soft-voting ensemble framework.
- To meet regulatory requirements for algorithm transparency, a linear mapping model is proposed to convert the default probabilities output by the ensemble model into intuitive credit scores, enabling seamless integration between technical outputs and business language.
- A business logic + statistical indicators framework is developed to address the high-dimensional, sparse, and noisy characteristics of financial data, enhancing feature interpretability and predictive power.
- Breakthrough the traditional one-size-fits-all risk control model, a dynamic risk stratification–strategy matching management system is constructed to provide differentiated risk control strategies for different customer tiers.

2. Related works

Loan risk monitoring is critical in finance and has attracted extensive academic attention in recent years. Ren *et al.* pointed out that in A Rural Commercial Bank, loan risk monitoring mainly relies on credit managers collecting data and visiting borrowers regularly, with risk early-warning indicators mostly qualitative. The lack of an electronic system for comprehensive quantitative and qualitative analysis leads to inaccurate risk assessment and delayed interventions, increasing non-performing loans [8]. Huang *et al.* mentioned that G Credit Union monitors inclusive finance loan risks through a combination

of on-site inspections and electronic record checks, but issues such as inadequate post-loan management and weak fund flow monitoring undermine effectiveness [9]. Zhang's study on H Bank Zhengzhou Branch found deficiencies in risk identification, assessment, and control for personal housing loans, with outdated methods unable to accurately quantify risks due to incomplete internal management [10].

These studies highlight that while some progress exists, traditional risk monitoring faces challenges: over-reliance on manual experience, lack of scientific quantification, incomplete information systems, and insufficient accuracy in adapting to dynamic market environments.

Since 1980, machine learning has advanced loan risk monitoring significantly [11,12]. Wiginton pioneered logistic regression for credit evaluation in 1980, while Makowski introduced decision trees in 1985, emphasizing their interpretability. Later, Li *et al.* compared KNN and CART algorithms for P2P default prediction, finding they outperformed SVM in specific scenarios [13]. Model fusion has become a hotspot: Lu, *et al.* used stacking to integrate logistic regression, random forest, and XGBoost, improving performance through parameter tuning [14]. Cai *et al.* constructed a heterogeneous fusion model with DeepFM, LightGBM, and CatBoost, proving superior accuracy and generalization [15].

Despite wide applications, machine learning faces challenges: linear models struggle with high-dimensional data, and tree models risk overfitting as data dimensions grow (e.g., social behavior, mobile payments) [16].

Deep learning has revolutionized risk control with automatic feature extraction. Hinton's 2006 Nature paper initiated its development [17], followed by Bequé's deep neural network for credit risk assessment in 2017 [18]. Zhou *et al.* combined unsupervised and supervised learning for credit scoring in 2018, achieving improved results [19]. CNNs excel in time-series modeling: Zhang, *et al.* enhanced fraud detection accuracy by 12% using convolutional layers to capture local anomalies in transactions [20]. Chinese scholars have also made progress: Yang *et al.* designed a stacked denoising autoencoder for high-dimensional, sparse bank data, improving model robustness [21]. Chen *et al.* compared four deep learning models, showing TabNet's superiority in performance and efficiency [22]. Li *et al.* developed an autoencoder-BP network with genetic algorithm optimization, outperforming traditional methods [23].

However, single deep learning models have limitations: CNNs lack global dependency modeling, and BP networks are sensitive to noise.

3. The proposed scheme

3.1. Data preprocessing

Missing Value Handling. Python is used to statistically analyze and visualize missing values. Attributes with a missing value ratio exceeding 50% are deleted directly. For attributes with low missing ratios or high importance, missing values are imputed using statistical measures such as the mean or median, or filled with zero. In this study, the debt-to-income ratio (DTI) and the revolving quota utilization rate (revolUtil) are filled with their respective medians.

Invalid Feature Deletion. Variables with excessive missing values, single-valued attributes (constant values across the dataset), or those with excessive categories and clear irrelevance to the target variable are removed to streamline the dataset.

Feature Binning. To enhance the stability and interpretability of the model, equal frequency binning is applied to continuous variables: total credit turnover balance (revolBal), DTI, installment,

revolUtil and credit interest rate (interestRate). This process smooths extreme values in revolBal, captures the non-linear relationship between DTI and default risk, and converts continuous variables into business interpretable interval labels to facilitate risk level classification. Binned variables are named with a bin suffix, and results are stored in both training and test datasets.

Categorical Variable Encoding. The dataset contains numerous qualitative variables alongside quantitative ones, requiring feature abstraction for downstream feature extraction and model training. Categorical, text, and date-type features are converted to numerical formats. Specifically, five variables—grade, subGrade, years of employment (employmentLength), loan disbursement month (issueDate), and the month when the borrower’s earliest credit limit was opened (earliestCreditLine)—undergo feature transformation to numerical representations.

3.2. Composite feature construction

Feature engineering is a core component of credit scoring models, as combining variables can effectively capture non-linear relationships—a critical aspect in accurately predicting borrower default risks for financial risk assessment. Single traditional features often fail to fully reflect a borrower’s creditworthiness and repayment ability, while complex interactions between variables are easily overlooked. Therefore, constructing composite features through feature engineering becomes a critical step to enhance model performance. Based on the business logic of variables, this study generates a total of 10 composite features.

3.3. Feature selection

In loan risk prediction modeling, feature selection aims to construct an efficient and stable prediction model by achieving three key objectives: noise filtering, dimensionality reduction, generalization improvement, and enhanced business interpretability. This study employs a combined screening approach using three methods: chi-square test ($p < 0.05$), Population Stability Index ($PSI < 0.25$), and Information Value ($IV \geq 0.01$). Through this tripartite selection process, 27 features are filtered out, leaving 32 features for model construction. The selection workflow is illustrated in Figure 1.

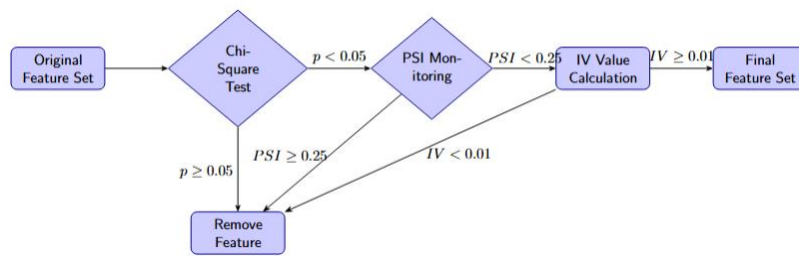


Figure 1. The feature selection flowchart.

IV Selection: IV serves as a core indicator for measuring a feature’s predictive power for the target variable. It quantifies a feature’s ability to distinguish defaults through the Weight of Evidence (WOE), widely used in the risk control domain. The calculation formula is Equation (1):

$$IV = \sum_{i=1}^n \left(\frac{B_i}{B_{total}} - \frac{G_i}{G_{total}} \right) * \ln \left(\frac{B_i/B_{total}}{G_i/G_{total}} \right), \quad (1)$$

where B_i and G_i represent the number of default (bad) and non-default (good) samples in the i -th bin, respectively. Through experiments, a threshold of ($IV > 0.01$) was determined. As shown in Figure 2, this threshold optimizes prediction performance, balancing feature relevance and model complexity. Features with ($IV = 0.01$) are retained, as they demonstrate meaningful discriminatory power for default risks.

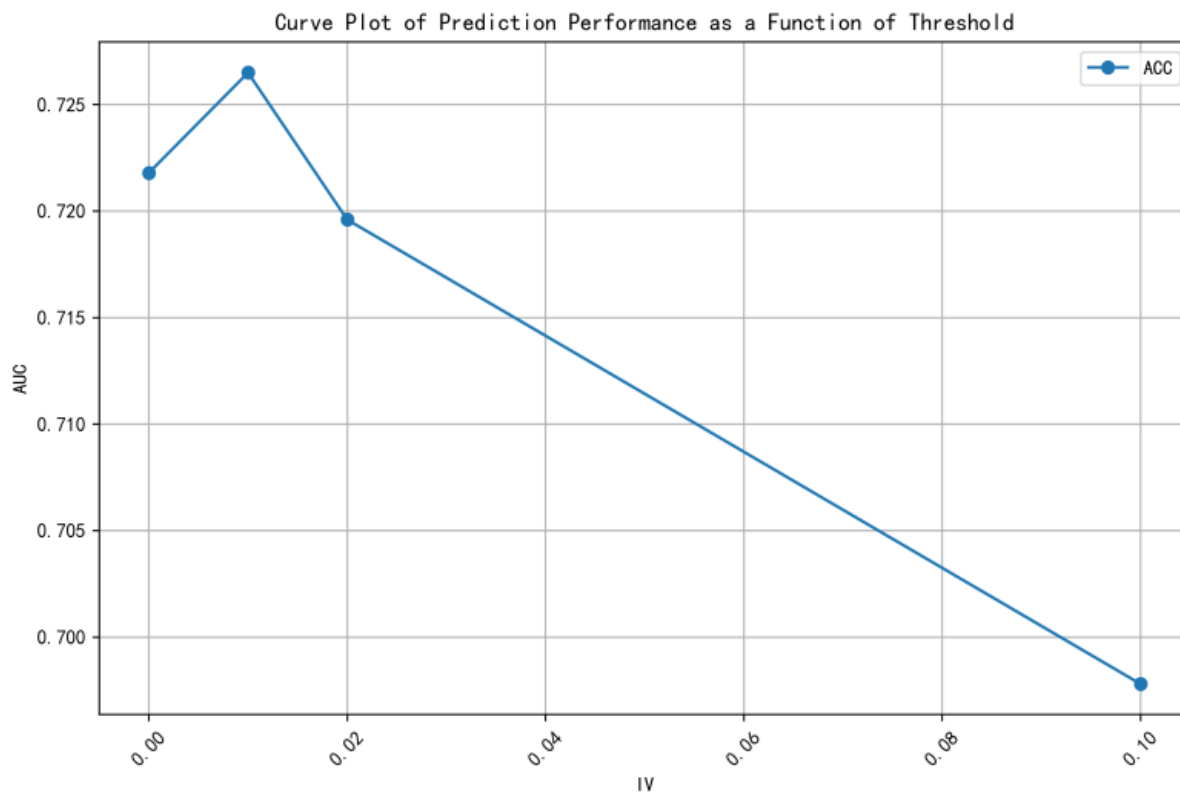


Figure 2. Curve plot of prediction performance.

A total of 34 features were selected through the IV value screening, with their IV values compared in Figure 3. As can be seen, the six most predictive features are loan grade, interest rate, FICO score lower bound at loan issuance, FICO score upper bound at loan issuance, and ratio of income to loan amount—two of which are newly constructed composite features. These results highlight the effectiveness of combining business driven composite features with traditional variables in enhancing risk discrimination.

PSI Selection: The PSI quantifies distributional differences of features across different time points (e.g., between training and test sets) to identify data drift. By calculating the deviation between the actual and expected distributions, it helps identify features with stable long-term distributions. This process ensures consistency between the feature distributions of training data and the target application scenarios. In this study, a PSI threshold of 0.25 was applied, and three features—id, Loan age (date_diff), and earliestCreditLine—exceeded this threshold, indicating unstable distributions and were therefore removed from the feature set.

p -value Selection. The chi-square test quantifies the dependency between features and the target variable by measuring their association, where a larger chi-square value or smaller p value indicates a stronger correlation, helping to rule out random noise. In this study, a threshold of 0.05 for p was applied, resulting in the selection of five features with p values exceeding the threshold. These features demonstrate strong associations with default risks, justifying their retention in the final feature set.

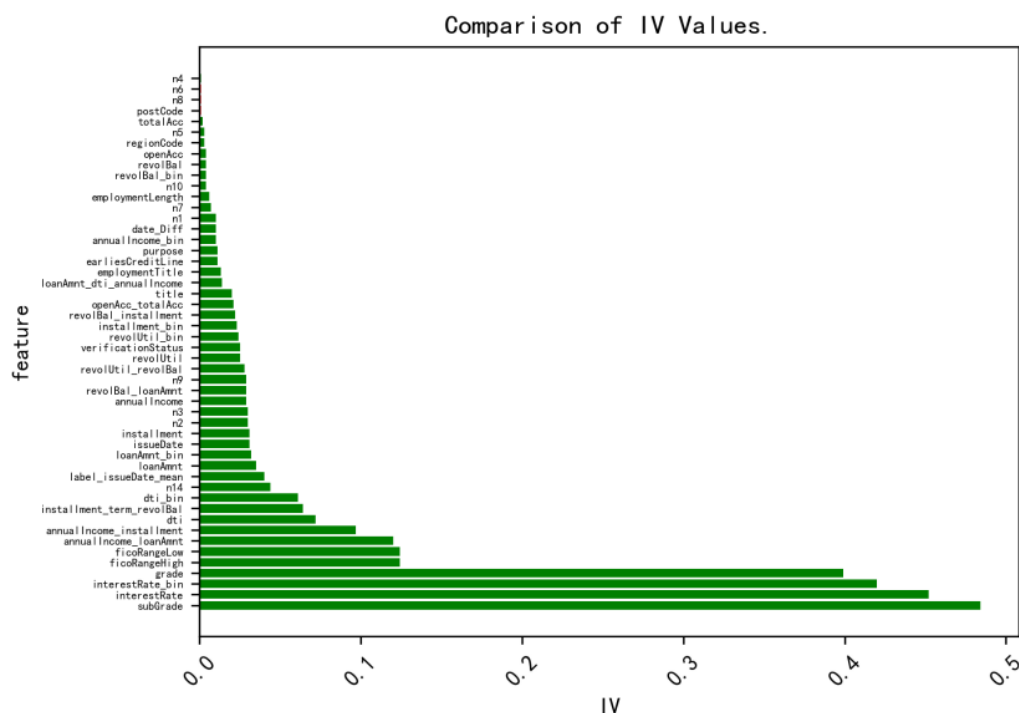


Figure 3. Comparison of IV values. (Zoom in for better view)

3.4. Construction of loan risk prediction model

This paper proposes a comprehensive loan risk model designed to leverage the advantages of different algorithms, enhancing the accuracy and reliability of loan risk prediction. The model consists of three core components: a Feedforward Neural Network (FNN), LightGBM, and CatBoost, which respectively generate probability predictions for loan risks. These predictions are then fused through a soft-voting ensemble method to obtain more stable and accurate probability outputs. Finally, a scorecard model is employed to convert the fused probabilities into intuitive credit scores, facilitating business application and risk quantification. This multi-level, multi-model architecture not only maximizes the extraction of information from data but also provides financial institutions with a comprehensive and actionable loan risk assessment solution through model ensemble and probability-to-score transformation. The overall architecture can be shown in Figure 4.

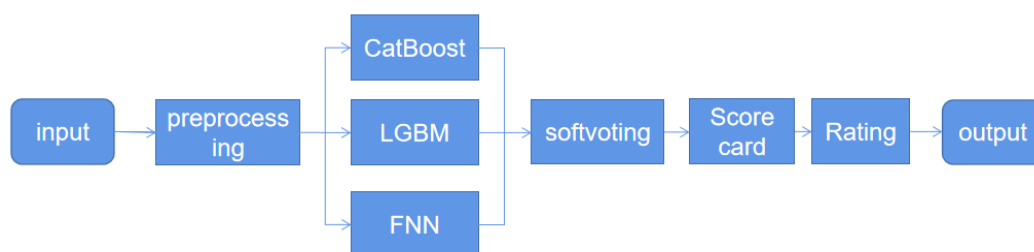


Figure 4. Diagram of the system architecture.

FNN Model: The deep FNN is an artificial neural network suitable for non-linear pattern recognition, classification, and regression tasks. It learns from training data by iteratively updating

weights and biases using an optimizer (e.g., SGD) to minimize the loss function via backpropagation, driving the error to decrease along the negative gradient direction and approaching the expected output.

In this study, a four-layer FNN is designed for the binary classification task of financial risk monitoring. Unlike traditional machine learning models, this model automatically learns feature interactions through cascaded non-linear transformations. The Batch Normalization (BN) layer effectively mitigates the feature scale differences common in financial data, while the Dropout mechanism optimizes for overfitting in modeling rare risk events with small samples. The overall architecture can be expressed as:

$$y = \sigma(W_4 * \text{Dropout}(\text{ReLU}(\text{BN}(W_3 * \text{Dropout}(\text{ReLU}(\text{BN}(W_2 * \text{Dropout}(\text{ReLU}(\text{BN}(W_1 * x + b_1)) + b_2)) + b_3)) + b_4)), \quad (1)$$

where x is the input feature vector; W_1, W_2, W_3 and W_4 are the weight matrices of the four fully connected layers; b_1, b_2, b_3 and b_4 are their corresponding bias vectors; BN denotes the Batch Normalization operation; ReLU is the activation function defined as $\text{ReLU} = \text{MAX}(0, z)$; and σ is the Sigmoid activation function, defined as: $\sigma(z) = 1/(1 + e^{-z})$.

LightGBM Model: LightGBM is an efficient machine learning algorithm based on gradient boosting decision trees. In loan risk assessment, LightGBM effectively processes high-dimensional data and categorical features, learning the relationships between features and loan risks by constructing decision trees. Its hierarchical feature splitting and regularization mechanisms enable it to capture non-linear interactions and prevent overfitting, making it well suited for imbalanced financial datasets where rare default events need to be accurately identified.

CatBoost Model: For the categorical variables commonly present in loan data, CatBoost achieves unbiased estimation through its ordered boosting strategy and symmetric tree structure. Its built-in target encoding mechanism avoids the dimensionality explosion issue caused by traditional one-hot encoding, making it particularly suitable for handling high cardinality categorical features. This characteristic is critical in loan risk modeling, where variables like employmentLength or purpose often have numerous distinct categories but carry significant predictive value for default risks.

3.5. Soft-voting ensemble model

The soft-voting layer acts as a coordinator for the predictions of base models, combining their predicted probabilities through weighted averaging to generate the final prediction probability. Compared with hard voting, soft-voting utilizes the probabilistic information of models, more comprehensively reflecting the uncertainty of the model. In this loan risk model, integrating the predicted probabilities of FNN, LightGBM, and CatBoost via soft-voting leverages the strengths of these three models, enhancing prediction accuracy and stability.

In this study, a classification soft-voting model was built using Python: the expected probabilities of the three models were weighted and averaged. Through exhaustive experiments, the optimal weights were determined as 0.1, 0.1, and 0.8. As shown in Figure 5, when fixing the LightGBM weight, the AUC curve varies with the FNN weight—indicating the best performance is achieved when the LightGBM weight is 0.1 and the FNN weight is 0.1.

$$\text{Probability} = 0.1 * P_{\text{FNN}} + 0.1 * P_{\text{lightGBM}} + 0.8 * P_{\text{CatBoost}}, \quad (2)$$

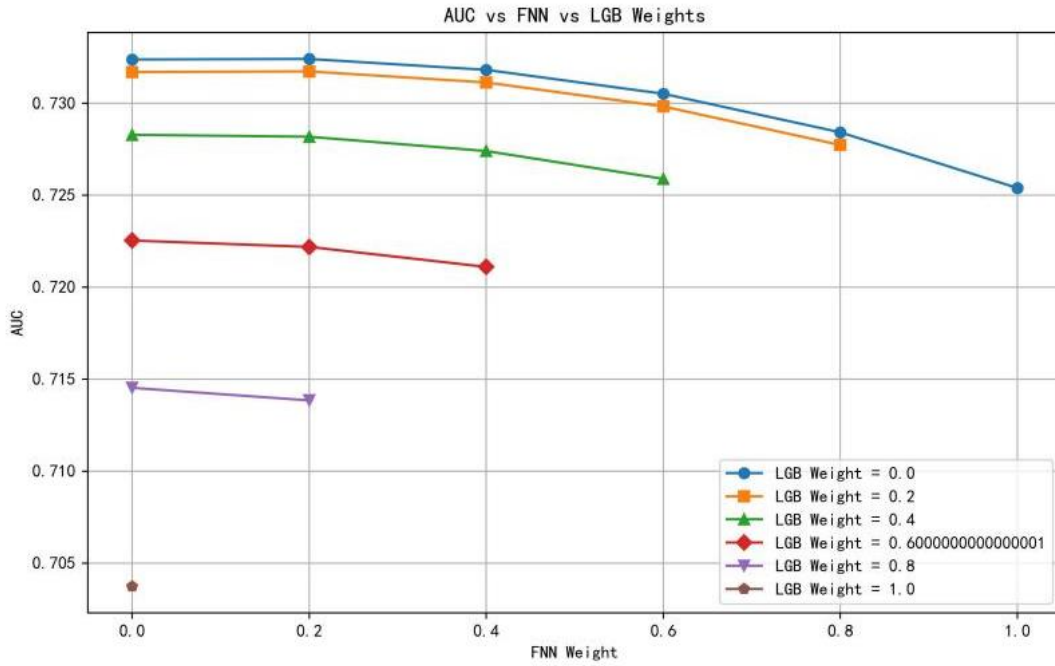


Figure 5. AUC comparisons under different model weights.

This weighted fusion mechanism balances the non-linear feature extraction capability of FNN, the efficiency of LightGBM in dealing with high-dimensional data, and CatBoost's in categorical feature modeling, significantly improving the ensemble model's robustness against data noise and distribution shifts.

3.6. Scorecard conversion based on ensemble model

To meet the transparency requirements of financial regulation, a logistic scorecard is used to map probabilities into integer scaled scores. By defining a linear relationship between the odds ratio and the score scale, the model output can directly interface with traditional risk control systems.

The dataset in this study covers key dimensions such as loan amount, term, credit grade, and first credit line date, reflecting the real-time status of loans—highly consistent with the application scenario of in-loan behavioral scorecards (B-cards). Therefore, this paper focuses on constructing and validating a B-card model. The scorecard formula is:

$$\text{score} = A - B * \ln(\text{odds}), \quad (3)$$

where A and B are constants, and odds is the probability ratio of user risk, calculated as:

$$\text{odds} = \frac{p}{1 - p}, \quad (4)$$

The base score and the points to double the odds (PDO) are two critical parameters: The base score is the initial score corresponding to a reference odds ratio (typical values: 500/600/650). PDO represents the score change when the odds ratio doubles (e.g., PDO = 50 means scores decrease by 50 when odds double). Calibrating A and B using predefined base score and PDO:

$$\begin{aligned} P_0 &= A - B * \ln(\theta_0), \\ P_0 - \text{PDO} &= A - B * \ln(2\theta_0). \end{aligned} \quad (5)$$

Substituting A and B into the score formula converts the ensemble model's probability output into interpretable behavioral scores, where higher risk corresponds to lower scores.

4. Experimental results

4.1. Integrated model evaluation based on soft-voting

This study integrates three base models—FNN, LightGBM, and CatBoost—using soft-voting with weights of 0.1:0.1:0.8. After training, a confusion matrix is generated via *sklearn.confusion_matrix*, as shown in Table 1.

Table 1. Confusion matrix of the integrated model.

	Predicted Positive	Predicted Negative	Total
Actual Positive	94296	1763	96214
Actual Negative	21437	2504	23786
Total	115733	4267	120000

Using the confusion matrix, the key metrics are calculated and compared with the base models in Table 2.

Table 2. Model comparison results.

	Predicted Positive	Predicted Negative	Total
Actual Positive	94296	1763	96214
Actual Negative	21437	2504	23786
Total	115733	4267	120000

The ROC curve of the model can be plotted using the recall rate and the false positive rate (FPR). Calculations show that the area under the curve (AUC) is 0.73, which falls within the range of 0.7 to 0.8, further demonstrating the model's accuracy. The ROC curve is shown in Figure 6 (a)–(d) below:

4.2. Scorecard

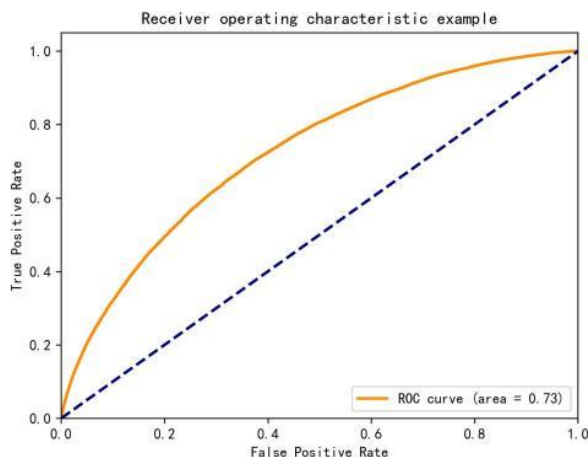
The scorecard conversion formula is defined as follows. With a predefined base score of 600, PDO of 30, and initial *odds* ratio of 0.5, the constants A and B are calculated as:

$$\begin{aligned} B &= \frac{\text{PDO}}{\ln(2)} = \frac{30}{\ln(2)} \approx 43, \\ A &= P_0 + B * \ln(\theta_0) = 600 + \frac{30}{\ln(2)} * \ln\left(\frac{1}{2}\right) = 570. \end{aligned} \quad (6)$$

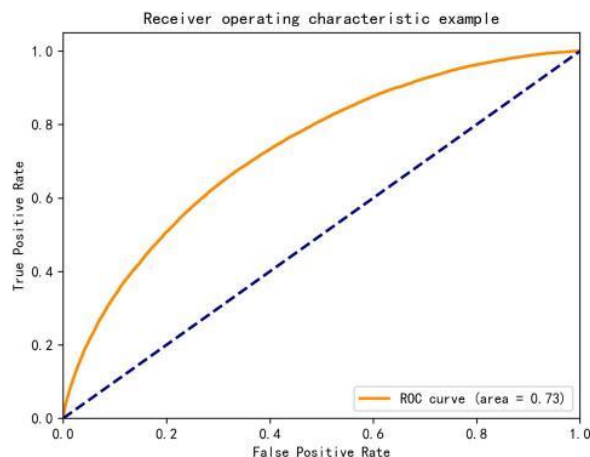
Substituting A and B into the score formula yields:

$$\text{score} = A - B * \ln(p/(1 - p)) = 570 - 43 * \ln(p/(1 - p)), \quad (7)$$

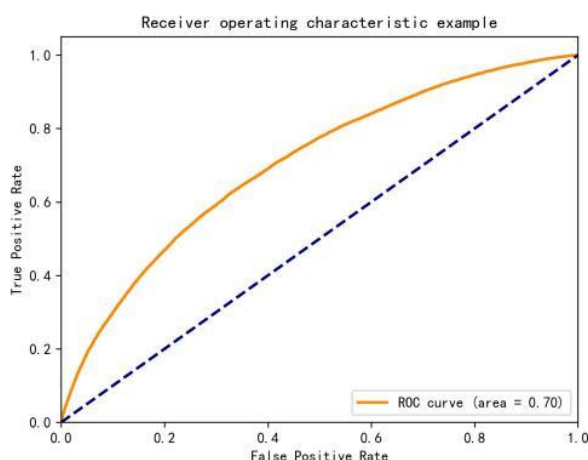
where p is the predicted default probability from the integrated model. This formula directly maps probabilities to an intuitive score scale.



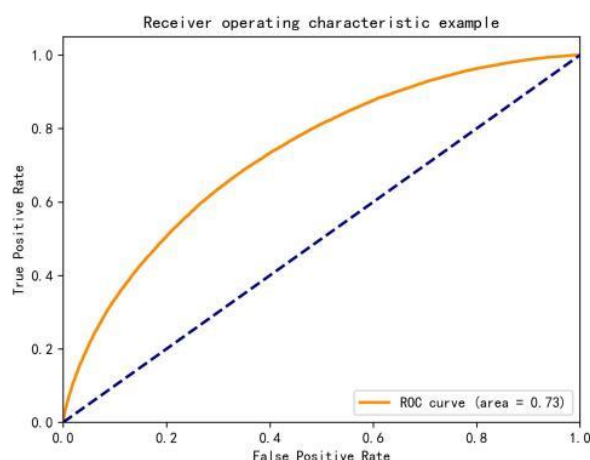
(a) FNN ROC curve and AUC.



(b) CatBoost ROC curve and AUC.



(c) LGBM ROC curve and AUC.



(d) Soft-voting ROC curve and AUC.

Figure 6. The ROC curve and AUC.

Data shows that 15.7% of clients have scores below 600, and the score distribution histogram highlights that most clients (78.2%) fall within the 570–750 score range. As shown in Table 3, there is a clear negative correlation between credit scores and default rates: higher scores correspond to lower within group default proportions, validating the scorecard's risk discriminative power.

Table 3. Prediction performance of the scorecard model.

Score	Range	Quantity	Predicted	Non-Default Samples Proportion	Probability	Within Group Default Proportion	Cumulative Default Samples
(300, 510]	14	0.829	9	5	0.643	9	0.001
(510, 570]	4253	0.573	2495	1758	0.587	2504	0.105
(570, 600]	14562	0.401	5919	8643	0.406	8423	0.352
(600, 630]	29353	0.257	7652	21701	0.261	16075	0.671
(630, 690]	56534	0.124	7282	49252	0.129	23357	0.976
(690, 750]	14598	0.039	578	14020	0.040	23935	1.000
(750, 850]	686	0.012	6	680	0.009	23941	1.000

4.3. User credit score grading and strategies

Based on the credit scores derived from predicted default probabilities, users are classified into four risk tiers (A to D). The corresponding lending strategies are formulated according to the default rates within each tier, as outlined in Table 4.

Table 4. User credit score grading and strategies.

Tier	Score Range	Predicted default rate	Default rate within the grade	Lending Strategy
A	630 – 850	$0 \leq p < 0.2$	0.129	Give priority to lending
B	600 – 630	$0.2 \leq p < 0.4$	0.352	Allow loans
C	570 – 600	$0.4 \leq p < 0.6$	0.594	Reduce the credit limit
D	189 – 570	$0.6 \leq p \leq 1$	0.7	Take out loans carefully

Tier A: High-quality clients with a very low default rate (3%). Offer diversified loan products and moderately reduce interest rates or increase credit limits to enhance customer loyalty and long-term relationships.

Tier B: Clients with manageable loan risks (default rate: 25%). After loan approval, send repayment reminders via SMS or other channels several days before the due date to minimize overdue risks from forgetfulness.

Tier C: Higher default risk (40%). Reduce loan amounts and intensify post-loan repayment monitoring. Clients with large overdue amounts or long-term defaults should be blacklisted, with future loan applications automatically rejected.

Tier D: High-risk clients with a default rate of 60%. In principle, reject loan applications. If exceptions are made, they require strict manual review and sufficient collateral.

5. Conclusion

This study addresses the issue of intelligent loan risk early warning by constructing a multi-model fusion framework based on soft-voting, integrating deep learning and machine learning algorithms to achieve precise assessment of loan default risks.

The main conclusions are as follows. The multi-model fusion approach integrating FNN, LightGBM, and CatBoost via soft-voting enhances prediction stability and accuracy. The ensemble model achieves an AUC of 0.73 on high-dimensional financial data, outperforming single models. In data preprocessing and feature engineering, binning, standardization, and composite feature construction are employed to address nonlinearity. Joint feature selection using IV, PSI, and chi-square retains 32 key features, with composite features significantly improving default prediction performance. The scorecard model, which converts model probabilities into credit scores, enhances interpretability and bridges the technical and business domains. Notably, there is a strong negative correlation between credit scores and default rates, facilitating actionable risk assessment. Clients are stratified into tiers A–D based on their scores, enabling the formulation of differentiated lending strategies to optimize risk management.

Acknowledgments

This work was supported in part by the Scientific Research Program of Anhui Provincial Ministry of Education (No. 2024AH051686 and 2025AHGXZK31128); the Laboratory Opening Project of CHNU (No. 2025xjsyskf039), the University Synergy Innovation Program of Anhui Province (No. GXXT-2023-108); and the National College Student Innovation and Entrepreneurship Training Program (No. 202510373003 and 202510373004).

Authors' contribution

Conceptualization, C.Q. and L.L.; methodology, C.Q. and Q.W.; software, C.Q. and H.W.; validation, S.W.; formal analysis, K.Z.; investigation, L.S.; writing—original draft preparation, C.Q.; writing—review and editing, L.L.; visualization, D.S. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

- [1] People's Bank of China. Financial Technology Development Plan (2022–2025). 2021. Available: <https://law.esnai.com/do.aspx?action=show&lawid=206239> (accessed on 15 March 2025).
- [2] China Banking and Insurance Regulatory Commission. Guidelines on the Digital Transformation of the Banking and Insurance Industry. 2022. Available: <https://china-cba.net/Index/show/catid/322/id/40501.html> (accessed on 15 March 2025).
- [3] European Commission. Artificial Intelligence Act. 2023. Available: <https://ec-europa-eu.libguides.com/ai-and-ethics/websites> (accessed on 15 March 2025).
- [4] Markov A, Seleznyova Z, Lapshin V. Credit scoring methods: Latest trends and points to consider. *J. Finance Data Sci.* 2022, 8:180–201.
- [5] iResearch. 2023 China Financial Technology Industry Insight Report. 2023. Available: <https://www.idigital.com.cn/report/detail?id=4136> (accessed on 16 March 2025).
- [6] KPMG. Report on the Development of China's Banking Industry. 2023. Available: <https://kpmg.com/cn/zh/insights/2023/05/china-banking-industry-survey-report-2023.html> (accessed on 16 March 2025).
- [7] China Banking Regulatory Commission. Notice on Issuing the Guidelines for Stress Testing of Commercial Banks (2014 Revision). 2014. Available: https://wenku.baidu.com/view/4a9afc18970590c69ec3d5bbfd0a79563d1ed40d.html?_wkts_=1773376550326 (accessed on 16 March 2025).
- [8] Guo L, Li Z, Qian K, Ding W, Chen Z. Bank credit risk early warning model based on machine learning decision trees. *J. Econ. Theory Bus. Manag.* 2024, 1(3):24–30.
- [9] Liu T, Yang L. Financial risk early warning model for listed companies using BP neural network and rough set theory. *IEEE Access* 2024, 12:27456–27464.
- [10] Zhang Y. Research on risk management of personal housing loan business in Zhengzhou Branch of H Bank. Master's Thesis, Shenyang University, 2024.

- [11] Yang G. Risk prevention, control and early warning model of enterprises' loan in the Yangtze River Delta. *J. Coast. Res.* 2020, 115(SI):614–617.
- [12] Zhou H, Qiu Y, Wu Y. An early warning system for loan risk assessment based on rare event simulation. In *Proceedings of the 2007 Asian Simulation Conference*, Seoul, Republic of Korea, October 10–12, 2007, pp. 85–94.
- [13] Li X, Long Z, Fu H, Sun J. Comparison of P2P default prediction algorithms based on machine learning: taking “renrendai” as an example. *Stat. Manag.* 2019, 6:104–109.
- [14] Lu Y. Research on loan default prediction based on machine learning classification algorithms. Master's Thesis, Lanzhou University, 2021.
- [15] Ge L, Ai S, Zhao H, Yu J, Yang J. Research on the construction of financial supervision information system based on machine learning. *Wireless Commun. Mobile Comput.* 2022, 2022(1):9986095.
- [16] Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge: MIT Press, 2016.
- [17] Yang D. Research on personal credit risk assessment based on deep learning. Master's Thesis, Central University of Finance and Economics, 2019.
- [18] Chen S. Research on personal loan credit risk prediction based on deep learning. Master's Thesis, Heilongjiang University, 2023.
- [19] Li Y. Research on personal loan default risk of small and medium commercial banks based on neural networks. Master's Thesis, Southwestern University of Finance and Economics, 2019.
- [20] Jagtiani J, Kolari J, Lemieux C, Shin GH. Early warning models for bank supervision: Simpler could be better. *Econ. Perspect.* 2003, 27(3):49–59.
- [21] Zhuang Y, Wei H. Early warning model and prevention of regional financial risk integrated into legal system. *PLoS One* 2023, 18(6):e0286685.
- [22] Chen S. Research on personal loan credit risk prediction based on deep learning. Master's Thesis, Heilongjiang University, 2023.
- [23] Li Y. Research on personal loan default risk of small and medium commercial banks based on neural networks. Master's Thesis, Southwestern University of Finance and Economics, 2019.