From Data to Risk: An AI-Integrated Financial Risk Early Warning Framework for Bank Credit Operations

Weijia Li

Business and Tourism School, Sichuan Agricultural University, Chengdu, 611830, China
18113181571@163.com

Abstract. Amid accelerating digital transformation, commercial banks face challenges in credit risk management due to multi-source, unstructured data. Traditional models relying on human experience and static indicators are insufficient for real-time risk identification. Taking a city commercial bank in western China as a case study, this paper addresses three core issues—inefficient data collection, lack of early warning mechanisms, and delayed post-lending management—by designing an AI-integrated framework. The framework uses OCR and NLP to automate financial report extraction, and constructs a two-tier risk model: Gaussian process regression for window-dressing detection and logistic regression for financial performance assessment. Back-testing on the bank's credit clients validates its effectiveness, offering a replicable pathway for intelligent risk control in similar banks.

Keywords: Artificial Intelligence; Financial Risk Early Warning; Credit Risk Management; Intelligent Risk Control; Framework Construction

1 Introduction

Against a backdrop of global macroeconomic fluctuations and financial reforms, commercial banks—particularly city banks serving local SMEs—face heightened credit risks. Credit operations are a core profit source, and their stability directly affects bank soundness. For city banks, information asymmetry and weak financial compliance make accurate risk assessment from financial statements difficult.

Although fintech has advanced, many small and medium-sized banks still struggle with three interrelated deficiencies: inefficient manual data collection, reliance on basic tools like Excel, and absence of dynamic post-lending monitoring. As shown in Fig. 1, the traditional risk control model suffers from these three deficiencies.

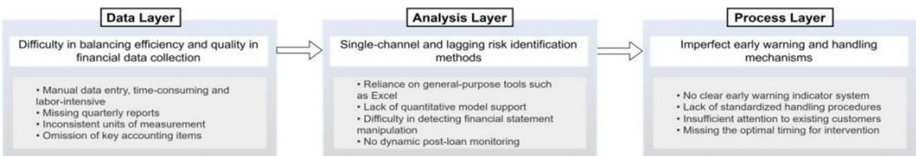

Fig. 1. Analysis of Issues in the Traditional Risk Control Model

N. Malik et al. (eds.), *Proceedings of the 6th International Conference on Public Management and Intelligent Society (PMIS 2026)*, Advances in Engineering Research 301,
https://doi.org/10.2991/978-94-6239-721-7_22

AI offers alternatives. OCR structures scanned statements, NLP extracts risk information, and machine learning supports forward-looking assessments[1][2]. Yet existing AI applications are still fragmented—handling individual tasks rather than combining data collection, risk identification, and post-lending monitoring, which weakens their practical value[3].

A city commercial bank in western China provides a relevant case. From 2020 to 2022, its corporate NPL ratio rose steadily and remained elevated in 2023, while provision coverage dropped sharply—reflecting broader risk management challenges. Drawing on this case, this paper proposes an AI-based financial risk early warning framework. Its core is two complementary models: GPR for window-dressing detection and logistic regression for financial performance assessment. Integrated with OCR and NLP, the framework addresses the three deficiencies in a unified manner. Back-testing on the bank's clients validates the approach, offering a replicable pathway for similar institutions. The architecture of the proposed framework is illustrated in Fig. 2.

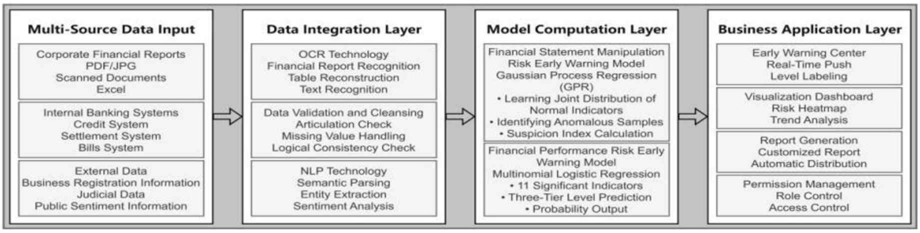

Fig. 2. Architecture Diagram of the AI-Based Financial Risk Early Warning Framework

2 Case Bank Profile and Data Overview

2.1 Bank Background and Credit Operations

The case bank is a city commercial bank in western China. Corporate lending contributes over 60% of operating revenue, with corporate loan interest accounting for nearly 40% of total interest income. Its customer base is concentrated among local SMEs—a segment with high information asymmetry and less standardized reporting.

Table 1 summarizes asset quality trends from 2020 to 2023. The corporate NPL ratio rose from 1.31% to 1.73%, then stabilized at 1.72%, while provision coverage fell from 305.87% to 208.46%, approaching the 150% regulatory threshold. These trends indicate delayed risk identification under traditional practices.

Table 1. Key Asset Quality Indicators of the Case Bank (2020–2023)

Year	Corporate Loan Balance (RMB billion)	Corporate NPL Ratio	Provision Coverage Ratio
2020	1,856.23	1.31%	305.87%
2021	1,925.46	1.70%	268.34%
2022	2,087.38	1.73%	211.19%
2023	2,215.68	1.72%	208.46%

This case bank is representative of many city commercial banks in China: heavy reliance on corporate lending, customer base dominated by local SMEs, and limited technical infrastructure compared to large state-owned banks. A review of its credit operations confirmed the inefficiencies mentioned in the introduction—manual data entry leading to missing quarterly reports, inconsistent units, and omitted key accounts[4]. According to industry research, over 80% of small and medium-sized credit approvals still rely primarily on manual judgment, and interviews with client managers at the bank confirmed that financial analysis is conducted using basic tools such as Excel, with no automated systems for data capture or anomaly detection. These documented deficiencies make the bank a suitable case for designing an AI-integrated early warning framework, and findings from this bank are likely relevant for similar regional institutions.

2.2 Data Background for Model Validation

To support empirical validation, a dataset was extracted from the bank's credit client records for 2020–2023. This four-year window captures both the pre-pandemic stable period and subsequent deterioration. The sample consists of 380 corporate credit clients, selected by stratified random sampling by loan size (small, medium, large). Based on repayment status, they are classified as 170 normal (no overdue), 130 with overdue interest, and 80 non-performing.

3 Design of an AI-Based Financial Risk Early Warning Framework

3.1 Technical Implementation Details

Data Sources and Sample Selection.
Two independent datasets were used for model development and validation, corresponding to the two risk assessment models. Table 2 summarizes the key characteristics of each dataset.

Table 2. Summary of Datasets for Model Development

Model	Data Source	Period	Sample Size	Composition	Selection Criteria
GPR (Window-dressing)	Shanghai & Shenzhen A-share listed companies	2020–2023	1,200 firms	200 negative (regulatory penalties), 1,000 positive	Complete annual reports; balanced industry
Logistic (Performance)	Case bank's corporate clients	2020–2023	380 clients	80 non-performing, 130 overdue, 170 normal	Stratified random by loan size

To account for industry heterogeneity in the window-dressing detection model, the 1,200 firms were stratified into six industry categories based on their operational and financial reporting characteristics: (1) manufacturing and industrial, (2) construction and real estate, (3) wholesale, retail, and services, (4) technology and information, (5) agriculture and resources, and (6) diversified conglomerates. Separate sub-models were constructed for each category to ensure that anomaly detection thresholds reflect industry-specific financial patterns[5].

For the financial performance model, the 380 clients were selected using stratified random sampling stratified by loan size, ensuring proportional representation across small (loan amount < RMB 10 million), medium (RMB 10–50 million), and large (> RMB 50 million) loan segments.

Variable Selection and Processing.

For window-dressing detection, indicators were organized into a three-tier framework. All 41 minor indicators derive from financial statement items; continuous variables were standardized. The detailed hierarchical structure of the 41 financial indicators is provided in Table 3.

Table 3. Hierarchical Structure of Financial Window-Dressing Indicators

Manipulation Motivation	Major Categories (10)	Minor Indicators (41)
Profit Fabrication	Revenue, earnings quality, cost/expense anomalies	15 (e.g., revenue-cost deviation, abnormal gross margin)
Asset Structure Falsification	Asset quality, related-party, investment anomalies	14 (e.g., abnormal receivables, inventory anomalies)
Cash Flow Concealment	Cash flow quality, financing, investment, financial distress	12 (e.g., cash flow-revenue divergence)

For the financial performance model, 23 candidate indicators across five dimensions were reduced to 8 significant ones ($p < 0.05$) using K-S test (normality), ANOVA (normal), or Kruskal-Wallis H test (non-normal). Table 4 summarizes the selection.

Table 4. Variable Selection for Financial Performance Model

Dimension	Initial (23) (examples)	Final (8)	Selection Method
Debt Repayment Capacity	Asset-liability ratio, net debt ratio, current ratio, quick ratio	Asset-liability ratio (X_1), current ratio (X_3), quick ratio (X_4)	K-S + Kruskal-Wallis
Profitability	Gross profit margin, period expense ratio, net profit margin	Net profit margin (X_{22}), return on equity (X_{23})	K-S + ANOVA
Growth Potential	Total asset growth, core revenue growth, main profit growth	Total asset growth (X_{12}), core revenue growth (X_{13}), net profit growth (X_{15})	K-S + Kruskal-Wallis

Operational Efficiency	Receivables turnover, inventory turnover, payables turnover	None (insignificant)	—
Cash Flow	Sales cash ratio, operating cash flow to total debt	None (insignificant)	—

Model Specifications and Parameter Settings.

The GPR model for window-dressing detection assumes a multivariate Gaussian distribution with a squared exponential covariance function. The prior distribution is estimated via maximum likelihood using the training sample of normal firms. Warning thresholds are set at the 85th (standard) and 95th (severe) percentiles of the joint probability density.

The logistic regression model for financial performance assessment uses three ordinal categories: Y=0 (sound), Y=1 (average), Y=2 (poor), with Y=2 as the reference category. Parameters are estimated via maximum likelihood at a 5% significance level. Model validation employs five-fold cross-validation.

Validation Approach.

Both models were validated using back-testing on the case bank's historical data. For the logistic regression model, five-fold cross-validation was employed, where the 380 samples were randomly divided into five subsets, with four used for training and one for testing in each iteration. This process was repeated five times to ensure classification stability. For the Gaussian process regression model, the trained prior distribution was applied to the case bank's client data to identify enterprises flagged as suspicious.

3.2 Financial Window-Dressing Risk Early Warning Model

The financial window-dressing detection model is built on the GPR framework. GPR assumes that the relationship between input financial indicators follows a multivariate Gaussian distribution, which naturally captures non-linear correlations among variables[6]. For a given set of financial indicators x=(x1,x2,…,xd), the joint probability density function is:

$$P(x|\mu,\Sigma)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

where μ is the mean vector and Σ is the covariance matrix. The squared exponential covariance function is adopted to model non-linear dependencies. The prior distribution for normal enterprises is estimated via maximum likelihood using the training sample of 1,000 healthy firms. Thresholds for flagging suspicious enterprises are set at the 85th (standard warning) and 95th (severe warning) percentiles of the joint probability density, based on expert judgment from external auditors.

Back-testing was performed on the case bank's historical data of 380 credit clients. Fig.3 (ROC curve, AUC = 0.91) demonstrates strong discriminatory power of the model. The ROC curve is reproduced from the original analysis.

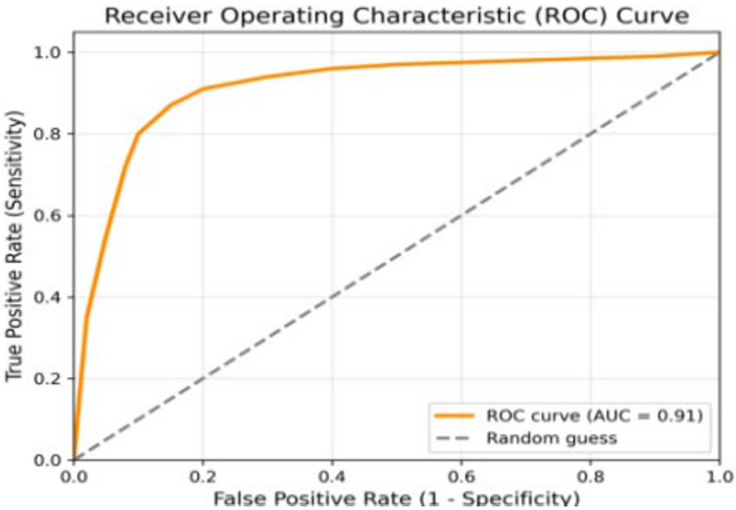

Fig. 3. ROC curve of the financial window-dressing detection model

To provide a more comprehensive evaluation, the optimal classification threshold was determined using the Youden index (sensitivity + specificity − 1), which corresponds to the 87th percentile of the joint probability density. Table 5 reports key performance metrics at this threshold.

Table 5. Performance metrics of the GPR model at the optimal threshold

Metric	Value	Definition
AUC	0.91	Area under the ROC curve
Optimal threshold	87th percentile	Based on Youden index
Precision	0.85	TP / (TP + FP)
Recall (sensitivity)	0.82	TP / (TP + FN)
Specificity	0.88	TN / (TN + FP)
F1-score	0.83	2 × (Precision × Recall) / (Precision + Recall)

Note: TP = true positives (suspicious firms correctly identified), FP = false positives (normal firms incorrectly flagged), TN = true negatives, FN = false negatives (suspicious firms missed).

To benchmark the proposed GPR model against conventional approaches, two baseline models were implemented for comparison. The single-indicator threshold method uses the asset-liability ratio as the sole predictor, with a fixed threshold of 80% (firms above 80% are flagged as suspicious). The traditional discriminant analysis applies linear discriminant analysis to the same 41 financial indicators used in the GPR model, with equal prior probabilities. Both baselines were trained on the 1,200-firm dataset and evaluated on the same 380-client back-testing set following the procedure. Table 6 compares the performance of the GPR model against these benchmark models.

Table 6. Comparison of the GPR model with benchmark models

Model	AUC	Precision	Recall	F1-score
Single-indicator threshold	0.67	0.52	0.48	0.50
Traditional discriminant analysis	0.79	0.68	0.71	0.69
GPR model	0.91	0.85	0.82	0.83

3.3 Financial Performance Risk Early Warning Model

This model uses multi-class logistic regression to assess a firm's debt repayment capacity[7]. Financial performance is classified into three ordinal categories: Y=0 (sound, no overdue), Y=1 (average, with overdue interest), and Y=2 (poor, non-performing), with Y=2 as the reference category:

$$log\left(\frac{P(Y{=}j)}{P(Y{=}2)}\right)=\beta_{j0}+\beta_{j1}X_1+...+\beta_{jk}X_k\ ,j{=}0,1$$

where X_1, X_3, X_4 and so on are the 8 significant indicators identified in Section 3.1.2 (Table 4). Parameters are estimated via maximum likelihood using the 380-client sample. Table 7 presents the parameter estimates for indicators with significance level below 5%.

Table 7. Parameter estimates of the logistic regression model

Category / Variable	Estimated coefficient	Significance level
Intercept (j=0, Good)	1.15	—
Intercept (j=1, Average)	8.80	—
Asset-liability ratio (X_1)	–6.85	0.001
Current ratio (X_3)	1.45	0.005
Quick ratio (X_4)	11.80	0.002
Total asset growth rate (X_{12})	–3.50	0.045
Core business revenue growth rate (X_{13})	2.15	0.006
Net profit growth rate (X_{15})	7.20	0.002
Net profit margin (X_{22})	2.50	0.000
Return on equity (X_{23})	0.28	0.018

Substituting the coefficients gives the log-odds equations:

$$ln\left(\frac{P(Y{=}0)}{P(Y{=}2)}\right)=1.15-6.85X_1+1.45X_3+11.80X_4-3.50X_{12}+2.15X_{13}+7.20X_{15}+2.50X_{22}+0.28X_{23}$$

$$ln\left(\frac{P(Y{=}1)}{P(Y{=}2)}\right)=8.80-6.85X_1+1.45X_3+11.80X_4-3.50X_{12}+2.15X_{13}+7.20X_{15}+2.50X_{22}+0.28X_{23}$$

The three probabilities P(Y=0),P(Y=1),P(Y=2) are obtained by solving the system; the category with the highest probability is the prediction. The model was validated using five-fold cross-validation. Table 8 presents the confusion matrix on the full sample of 380 clients.

Table 8. Confusion matrix of the financial performance risk early warning model

Actual \ Predicted	Good	Average	Poor	Total
Good (Y=0)	158	9	3	170
Average (Y=1)	8	114	8	130
Poor (Y=2)	2	6	72	80
Total	168	129	83	380

Table 9 reports precision, recall, and F1 score for each category. The overall classification accuracy is 90.53% (344 out of 380), and the macro-averaged precision, recall, and F1 score are all 0.90, indicating balanced performance across the three risk levels.

Table 9. Performance metrics by category

Financial performance classification	Precision	Recall	F1-score	Support
Good (Y=0)	0.94	0.93	0.93	170
Average (Y=1)	0.88	0.88	0.88	130
Poor (Y=2)	0.87	0.90	0.88	80
Overall (macro average)	**0.90**	**0.90**	**0.90**	**380**

Finally, to demonstrate the superiority of the logistic regression approach, a benchmark comparison was performed with a traditional discriminant analysis model. Table 10 shows that the proposed model achieves higher overall accuracy and better balanced performance, particularly in identifying enterprises with overdue interest (average category) and non-performing loans (poor category). The traditional discriminant analysis benchmark was implemented using the same 8 significant indicators and the same 380-client dataset. Linear discriminant analysis was applied with equal prior probabilities and without any feature selection beyond the 8 indicators. Its performance was evaluated using five-fold cross-validation (identical to the logistic regression model), and the reported metrics are the averages across the five folds.

Table 10. Comparison with the benchmark discriminant analysis model

Model	Overall accuracy	Macro-avg precision	Macro-avg recall	Macro-avg F1
Traditional discriminant analysis	82.37%	0.81	0.80	0.80
Logistic regression	90.53%	0.90	0.90	0.90

The two-tier model—GPR for window-dressing detection and logistic regression for financial performance assessment—is well suited to the case bank's business environment. Together, they provide an end-to-end solution for credit risk early warning, and the approach can be adapted to other city commercial banks by recalibrating parameters based on their historical data.

4 Conclusions and Outlook

This paper develops an AI-integrated financial risk early warning framework for bank credit operations—combining OCR, NLP, Gaussian process regression, and logistic regression—to address inefficiencies in data collection, risk identification, and post-lending monitoring, validated through back-testing on a city commercial bank in western China. The main contribution is the deep integration of AI into the risk control practices of urban commercial banks, providing a comprehensive framework from problem diagnosis to implementation assurance. This offers a referenceable technical pathway for intelligent risk control transformation in similar institutions. The findings show that systemic restructuring of data collection, risk identification, and early warning response enhances banks' credit lifecycle management capabilities, shifting from reactive to proactive management[8].

Nevertheless, the framework has several limitations. First, its effectiveness relies heavily on the quality and completeness of corporate financial statements, which can be problematic for smaller enterprises with less standardized reporting. Second, implementation requires a certain level of technical infrastructure and expertise, which may pose scalability challenges for smaller banks with limited IT budgets[9]. Third, the current models are calibrated using data from a single bank and a specific time period; their generalizability to other regions or economic cycles requires further validation. These limitations suggest that the framework should be viewed as a starting point rather than a one-size-fits-all solution.

Future research may explore advanced algorithms like graph neural networks to handle complex corporate relationships and unstructured data, while integrating macroeconomic data to construct a more comprehensive risk view, driving frameworks toward greater intelligence and precision.

References

1. Jiang, J.Q.: Application of Artificial Intelligence in Credit Risk Management. New Finance 2024(8), 4–7 (2024).
2. Altman, E.I.: Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance 23(4), 589–609 (1968).
3. Ohlson, J.A.: Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research 18(1), 109–131 (1980).
4. Lessmann, S., Baesens, B., Seow, H.V., Thomas, L.C.: Benchmarking state-of-the-art classification algorithms for credit scoring: A ten-year update. European Journal of Operational Research 247(1), 124–136 (2015).
5. Qu, Z.F., Wu, H.H., Li, F.J.: Credit Risk Assessment Based on Logistic Regression and Credit Strategy Optimization Modeling for Small and Medium-Sized Enterprises. Journal of Shandong University (Natural Science) 59(8), 94–102 (2024).
6. Zhao, Y.T., Lin, D.Q.: Prediction of Micro- and Small-Sized Enterprise Default Risk Based on a Logistic Model: Evidence from a Bank of China. Sustainability 15(5), 4097 (2023).
7. Horak, J., Sabek, A.: Gaussian Process Regression's Hyperparameters Optimization to Predict Financial Distress. Retos: Revista de Ciencias de la Administración y Economía 13(26), 123–136 (2023).
8. Li, X.D.: Application of Artificial Intelligence in Corporate Financial Risk Prevention and Control: Based on Big Data Environment. Communication of Finance and Accounting 2021(22), 137–142 (2021).
9. Zhao, L.Q., Nie, L.Q.: Research on the Application of Artificial Intelligence in Corporate Financial Risk Early Warning. Management & Technology of SME 2023(20), 134–136 (2023).