Credit Evaluation of Supply Chain Partners Using Ensemble Learning and Conditional GANs

Dan Cui
School of Finance and Business, Dalian Vocational & Technical College, Dalian, 116035, China
E-mail: cuidancd1@outlook.com

Keywords: supply chain finance, credit rating, ensemble learning, GAN, Bayesian optimization

Received: January 6, 2026

Driven by the deep integration of the industrial internet and digital finance, supply chain risk control is shifting towards dynamic data intelligence. However, extreme sample class imbalances and data feature heterogeneity in real business scenarios severely restrict the accuracy of credit assessment. To overcome the identification blind spots of sparse risk samples, the study proposes an enterprise supply chain partner credit evaluation model combining Conditional Generative Adversarial Networks (CGAN) and Ensemble Learning (EL). First, an improved CGAN reconstructs the true distribution of default samples to alleviate data skew. Then, a Tree-structured Parzen Estimator (TPE) based Bayesian Optimization (BO) algorithm adaptively tunes a two-layer Stacking ensemble framework (including XGBoost, LightGBM, and CatBoost) for deep heterogeneous feature mining. Results show that the model exhibits excellent comprehensive classification performance, with an F1 score of 0.949 and a Kolmokolov-Smirnov statistic of 0.62, significantly outperforming benchmark models such as Attentive Interpretable Tabular Learning (TabNet). In credit simulation, the model achieves a 94.5% default rejection rate while maintaining high credit coverage, avoiding huge bad debt losses. The study effectively solves the problem of capturing minute risks in high-dimensional heterogeneous spaces, providing a scientific basis with high economic value for financial institutions to optimize credit resource allocation and reduce decision-making uncertainty.

Povzetek: Študija predstavlja napreden model za ocenjevanje kreditnega tveganja v dobavnih verigah, ki z uporabo umetne inteligence natančneje prepoznava tvegane partnerje tudi pri neuravnoteženih in raznolikih podatkih.

1 Introduction

1.1 Motivation for the proposed research

As the industrial internet and digital finance become deeply integrated, the risk management approach in supply chain finance is experiencing a major transformation, shifting from static financial auditing to dynamic data intelligence. Cloud-based ecosystems provide multimodal streaming data for credit profiling [1], yet transforming this into accurate risk insight remains computationally challenging [2]. In real business scenarios, high-quality enterprises occupy an absolute dominant position, while default behavior exhibits extreme sparsity and concealment. Severe class imbalance makes it difficult to capture risk characteristics in a large number of normal samples. At the same time, data such as financial ratios and enterprise attributes have feature heterogeneity, which further increases the nonlinear complexity of data distribution [3]. Traditional credit assessment mainly relies on expert scoring cards or basic statistical models, which reveals subjectivity and lag when dealing with high-dimensional dynamic data [4]. Heuristic oversampling often ignores financial indicator distributions, generating synthetic samples lacking business logic near decision boundaries [5]. This data distortion causes spurious correlations, hindering conventional classification models from identifying minority risks. Therefore, how to reconstruct the true distribution of scarce risk samples in a heterogeneous feature space and overcome the blind spots in risk identification of traditional models has become a key theoretical problem that urgently needs to be solved in current supply chain credit evaluation.

1.2 Related works

Extensive research has explored supply chain credit evaluation via index systems and intelligent algorithms. To address variable screening and class imbalance, L. Wang et al. proposed an unbalanced sampling strategy for machine learning, confirming that multi-source information fusion improves SME credit prediction [6]. To address specific industry scenarios and misjudgment risks, researchers have deployed hybrid ensemble algorithms and heterogeneous evaluation standards integrated with three-way decision mechanisms [7, 8]. Furthermore, Generative Adversarial Networks (GANs) have demonstrated exceptional data reconstruction capabilities through adversarial learning [9]. This generative advantage has been widely validated in resolving data scarcity and imbalance across diverse domains, such as satellite imagery [10], smart home

time-series [11], and vehicle controller networks [12]. These cross-domain successes confirm GAN's potential in reconstructing complex distributions, providing methodological support for addressing extreme scarcity and heterogeneity in supply chain credit evaluation. The comparison of related works on supply chain credit evaluation and generative data augmentation is shown in Table 1.

Table 1: Comparison of related works on supply chain credit evaluation and generative data augmentation

Research source	Core methodology	Key contributions	Limitations (research gap)
Wang et al. [6]	Machine learning with sampling strategy	Confirmed multi-source information fusion improves SME credit prediction accuracy.	Heuristic sampling ignores the joint distribution of financial indicators, causing potential data distortion.
Belhadi et al. [7]	Hybrid ensemble (Rotating Forest + Logistic Boosting)	Identified current ratio and financial leverage as core discriminators for agricultural SMEs.	Relies on static datasets and lacks mechanisms to handle extreme class imbalance in real-time scenarios.
Liang et al. [8]	Heterogeneous information + Three-way decision	Collaboratively handled misjudgment risks and improved decision-making refinement.	Focused on decision logic rather than reconstructing the underlying sparse data manifold.
Alzahem et al. [10]	GAN-based augmentation + Vision Transformers	Successfully addressed high-dimensional data imbalance in satellite imagery.	The model is designed for spatial image data and cannot be directly applied to tabular credit features.
Chougule et al. [12]	GAN-based synthetic dataset generation	Proved that synthetic data can outperform original data in specific security tasks.	Lacks adaptive hyperparameter optimization for complex, multi-model ensemble systems.
This Study (EL-CGAN)	Improved CGAN + BO-driven Stacking Ensemble	Reconstructs high-fidelity default samples while adaptively optimizing heterogeneous model fusion.	Overcomes both distribution distortion and manual parameter tuning inefficiencies (Current Focus).

1.3 Research objective and highlights

Existing research has made some progress in the construction of supply chain credit evaluation index systems and the application of intelligent algorithms, confirming the effectiveness of multi-source information fusion. However, traditional heuristic oversampling distorts synthetic sample distributions, and single models struggle with high-dimensional interpretability and complex nonlinear risk boundaries. To address these limitations, this research proposes a Credit Evaluation Model Using Ensemble Learning and Conditional GANs (EL-CGAN). Specifically, to address the research questions (RQ1) how to generate high-fidelity default samples without feature distortion, and (RQ2) how to adaptively integrate heterogeneous models for optimal decision boundaries, this model uses an improved CGAN to reconstruct default distributions and introduces a Tree-structured Parzen Estimator (TPE) based Stacking strategy to integrate base learners.

This research aims to enhance minority sample identification in high-dimensional heterogeneous data, providing scientific decision-making tools for financial institutions. The main innovations relative to existing work are: (1) Constructing a conditional generation mechanism with gradient penalty for mixed-attribute data, achieving high-fidelity sample generation conforming to business logic. (2) Designing a BO-driven two-layer Stacking framework to dynamically optimize hyperparameters, significantly enhancing risk identification efficiency.

2 Methods and materials

2.1 Supply chain sample distribution reconstruction based on conditional GAN

Given the multimodal heterogeneity and extreme scarcity of default samples in supply chain data, we propose a Conditional Generative Adversarial Network (CGAN) framework to synthesize high-fidelity risk samples. This method embeds conditional vectors into the generator to constrain the mapping space, ensuring financial logic completeness while filling class gaps. Its core generative model architecture is shown in Figure 1.

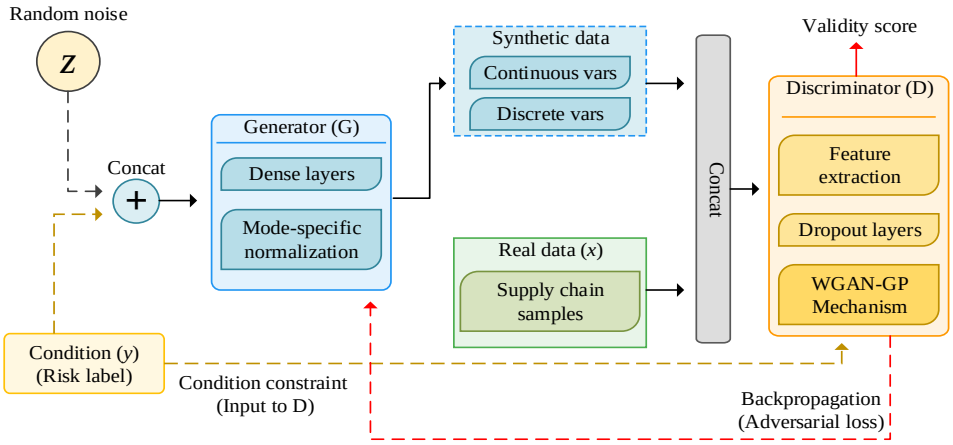


Figure 1: Architecture diagram of the credit data generation model based on CGAN

Figure 1 illustrates an adversarial closed-loop system where the generator concatenates random noise and risk labels, processing them via mode-specific normalization to handle feature heterogeneity [13]. The condition vector and sample generation formulas for the generator are defined in equations (1) and (2).

$$c = \text{Concat}(z, y) \tag{1}$$

$$\begin{cases} \hat{X} = G(c;\theta_G) = \text{BN}(W_c \cdot c + b_c) \\ \theta_G = \{W_c, b_c\} \end{cases} \tag{2}$$

Where the condition constraint vector (c) is formed by concatenating a uniformly distributed random noise vector (z) and a risk label vector (y). The synthetic sample ($\hat{X}$) is generated by the network ($G(\cdot;\theta_G)$) utilizing batch normalization ($\text{BN}(\cdot)$) over linearly transformed conditions (with weight $W_c \cdot c + b_c$) to directionally constrain specific risk attributes.To optimize parameters and alleviate gradient vanishing in high-dimensional spaces, a gradient penalty mechanism is introduced during training [14], as calculated in equation (3).

$$\begin{cases} L_{gp} = \lambda \cdot \text{E}_{\hat{X}_{\grave{o}} \sim P_{\grave{o}}} \left[\left(\| \nabla_{\hat{X}_{\grave{o}}} D(\hat{X}_{\grave{o}}, c) \|_2 - 1 \right)^2 \right] \\ \hat{X}_{\grave{o}} = \grave{o} \cdot X + (1 - \grave{o}) \cdot \hat{X}, \grave{o} \sim U[0,1] \end{cases} \tag{3}$$

Where the gradient penalty term (L_{gp}) applies a penalty coefficient (λ =10, conforming to WGAN-GP) to penalize the degree to which the gradient norm ($\| \nabla_{\hat{X}_{\grave{o}}} D(\hat{X}_{\grave{o}}, c) \|_2$) of the interpolated sample ($\hat{X}_{\grave{o}}$) deviates from 1, thereby enforcing the 1-Lipschitz continuity constraint. The specific process of adversarial training and generation of mixed attribute data is shown in Figure 2.

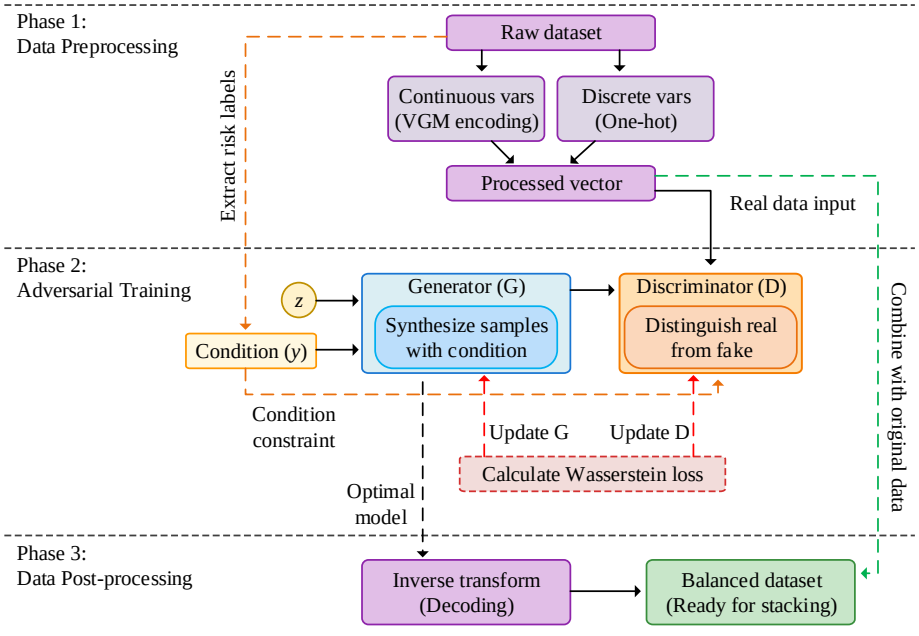


Figure 2: Flowchart of adversarial training and generation of mixed attribute data

Figure 2 details three collaborative stages: preprocessing original data, adversarial synthesis, and inverse transformation for dataset balancing. In the preprocessing stage, continuous financial indicators are encoded via a variational Gaussian mixture model [15], while discrete credit features are one-hot encoded [16], as defined in equations (4) and (5).

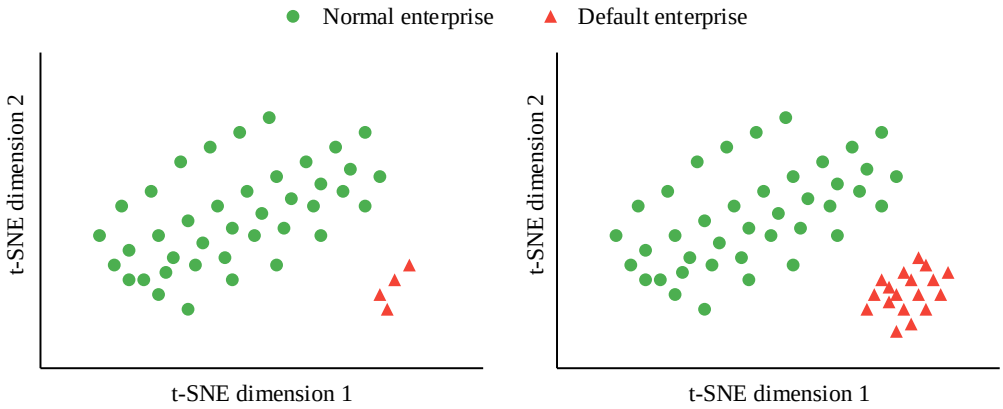
$$\begin{cases} p(X_c \mid \phi) = \sum_{k=1}^{K} \pi_k \cdot \mathrm{N}\,(X_c; \mu_k, \sigma_k^2) \\ \phi = \{\pi_k, \mu_k, \sigma_k^2\}_{k=1}^{K}. \end{cases} \tag{4}$$

$$O_{i,j} = \mathrm{I}\left(X_{d,i} = c_j\right) \tag{5}$$

Where the continuous indicator distribution ($p(X_c \mid \phi)$) is fitted using K mixture components with corresponding weights (π_k), means (μ_k), and variances (σ_k^2); simultaneously, the discrete feature matrix (O) normalizes textual categories into numerical vectors via an indicator function ($\mathrm{I}(\cdot)$) mapping the original feature ($X_{d,i}$) to its specific category (c_j). During the core adversarial stage, the generator synthesizes specific attributes [17], and the discriminator evaluates real and fake samples using the Wasserstein distance (WD) to approximate the Nash equilibrium [18], as shown in equation (6).

$$WD(P_r, P_g) = \sup_{\|D\|_L \le 1} \mathrm{E}_{X \sim P_r}[D(X, c)] - \mathrm{E}_{\hat{X} \sim P_g}[D(\hat{X}, c)] \tag{6}$$

Where $WD(P_r, P_g)$ quantifies the distributional distance by calculating the expectation difference between the conditionally constrained discriminator outputs for real supply chain data (P_r) and synthesized data (P_g). Crucially, to explicitly avoid data leakage, synthetic-sample generation is strictly executed only on the training folds during cross-validation. Furthermore, to ensure full reproducibility, the exact mathematical formulations of the generative mechanism, alongside the specific experimental environment and dataset configurations, are comprehensively documented. In the final post-processing stage, the optimal generator's output is inverse-transformed and fused to construct a class-balanced dataset. The reconstruction effect of this mechanism on the data manifold is shown in Figure 3.



(a) Original imbalanced data (b) GAN augmented data

Figure 3: Comparison of feature space distribution between real samples and generated samples

Figure 3 utilizes t-SNE to intuitively demonstrate that generated defaulting samples effectively fill the sparse regions while maintaining original characteristics, without deviating from the real data manifold structure [19].

2.2 Weighted stacking integration model based on BO

Although CGAN has effectively corrected the skewed distribution of samples, the reconstructed supply chain data still retains complex, high-dimensional, heterogeneous properties. A single classifier is highly susceptible to getting trapped in local optima when facing such non-convex decision boundaries, leading to risk identification failure. To address this, the study introduces a Stacking strategy to capture nonlinear correlations through algorithmic complementarity. Furthermore, to overcome hyperparameter explosion and manual tuning inefficiencies, a Bayesian Optimization (BO) mechanism is employed to adaptively lock the global optimal solution via a probabilistic surrogate model. This constructs a highly robust two-layer cascaded evaluation system. The specific integrated framework is shown in Figure 4.

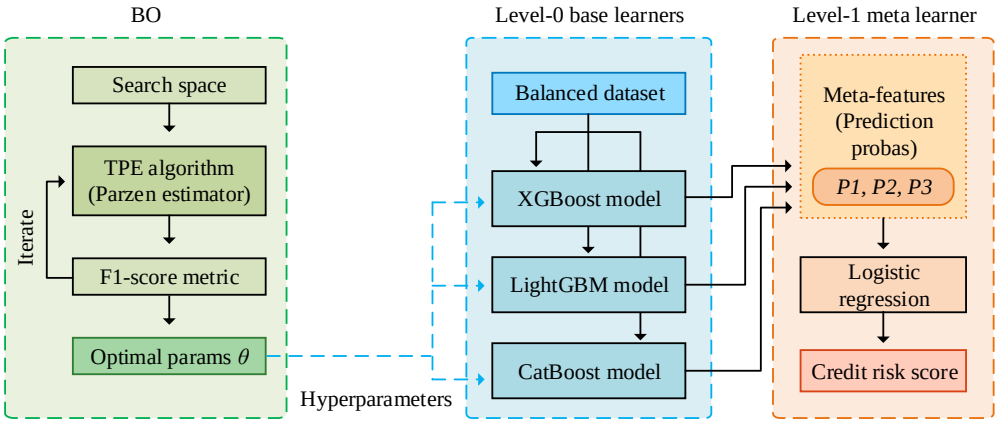


Figure 4: Two-layer Stacking integration framework based on BO

Figure 4 illustrates the BO-driven two-layer Stacking system, utilizing a TPE to dynamically optimize hyperparameters for three parallel heterogeneous base learners (XGBoost, LightGBM, and CatBoost) [20 – 23]. The prediction probability output of these base learners and the construction of the higher-order meta-feature matrix are defined in equations (7) and (8).

$$p_{i,t} = H_t(D_i; \lambda_t) \tag{7}$$

$$M = [p_{1,1}, p_{1,2}, p_{1,3}; p_{2,1}, p_{2,2}, p_{2,3}; \ldots; p_{N,1}, p_{N,2}, p_{N,3}]_{N\times 3} \tag{8}$$

Where $p_{i,t}$ represents the predicted default probability (range [0,1]) derived from the t th base learner $H_t(\cdot)$ (XGBoost, LightGBM, or CatBoost) for the i th CGAN-reconstructed sample vector D_i under specific hyperparameters λ_t . These individual predictions populate the $N \times 3$ meta-feature matrix M, realizing the structured integration of multi-dimensional prediction information for subsequent nonlinear correction by the logistic regression meta-model [24]. To handle complex discrete attributes within the supply chain data, CatBoost employs a unique sorting and encoding strategy, as shown in Figure 5.

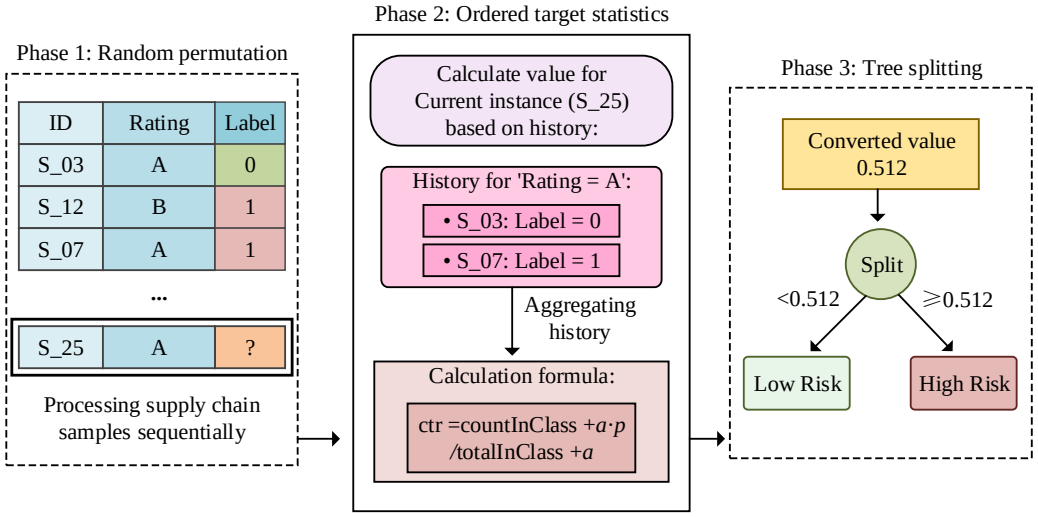


Figure 5: CatBoost's ranking boosting mechanism for categorical features.

Figure 5 demonstrates how CatBoost randomly rearranges sample sequences to prevent overfitting, then maps text labels to continuous values using historical target statistics and smoothing calculations, thereby enhancing decision tree splitting efficiency [25, 26]. To overcome the high computational cost and local optima limitations of manual parameter tuning, a hyperparameter adaptive optimization closed loop based on the TPE is established, as shown in Figure 6.

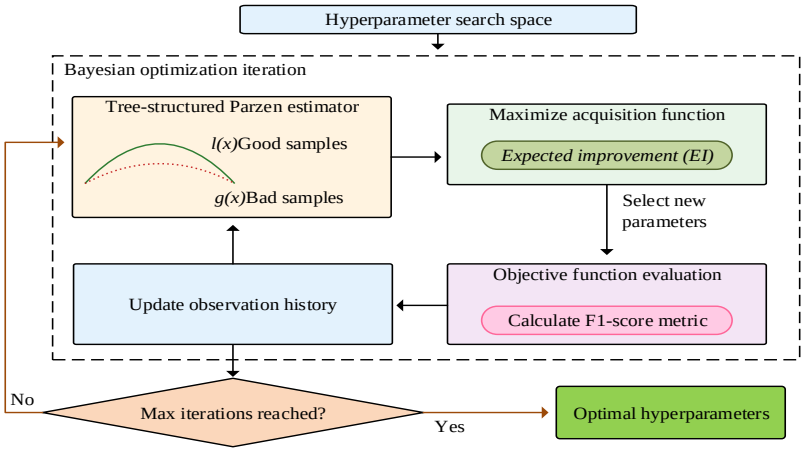


Figure 6: Flowchart of TPE Bayesian hyperparameter adaptive optimization

Figure 6 details the TPE BO process, which models posterior distributions from historical data to balance global exploration and local development. The Expected Improvement (EI) acquisition function and the optimal hyperparameter selection process are calculated in equations (9) and (10).

$$EI(\lambda) = \mathrm{E}\left[\max(0, F1^* - F1(\lambda))\right] = (F1^* - \mu(\lambda))\Phi\left(\frac{F1^* - \mu(\lambda)}{\sigma(\lambda)}\right) + \sigma(\lambda)\phi\left(\frac{F1^* - \mu(\lambda)}{\sigma(\lambda)}\right) \tag{9}$$

$$\begin{cases} \lambda^* = \arg\max_{\lambda \in \Lambda} EI(\lambda) \\ \lambda^* = [\lambda_1^*, \lambda_2^*, \lambda_3^*] \end{cases} \tag{10}$$

Where the acquisition function $EI(\lambda)$ evaluates a potential hyperparameter combination λ using the currently optimal F1 score ($F1^*$) alongside the TPE surrogate model's predicted mean ($\mu(\lambda)$) and standard deviation ($\sigma(\lambda)$) [27], mapped via the standard normal cumulative distribution ($\Phi(\cdot)$) and probability density ($\phi(\cdot)$). The operator $\arg\max(\cdot)$ ultimately selects the optimal combination (λ^*) from the preset search space (Λ), automatically locking the global optimal solution to ensure maximum generalization performance in complex feature spaces.

2.3 Construction of enterprise supply chain partner credit evaluation model

Based on the aforementioned modules, the overall calculation process of the enterprise credit risk dynamic assessment model is shown in Figure 7..

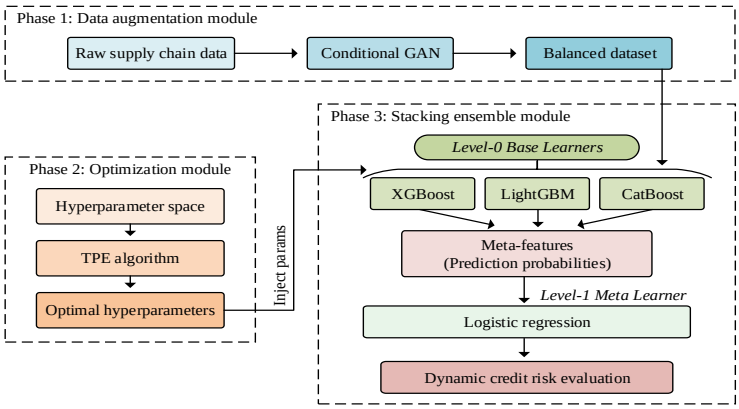


Figure 7: Overall calculation chart of enterprise credit risk dynamic assessment model

Figure 7 illustrates the end-to-end computational closed loop integrating three core modules: CGAN-based data augmentation for overcoming class skew, TPE-driven intelligent parameter optimization, and the parallel heterogeneous Stacking ensemble for meta-feature extraction. In the final integration stage, the logistic regression meta-learner performs nonlinear fusion and secondary correction on the prediction probabilities to output accurate dynamic assessment results [28], as calculated in equation (11).

$$S_i = \sigma\left(\omega_1 p_{i,1}^* + \omega_2 p_{i,2}^* + \omega_3 p_{i,3}^* + b\right) = \frac{1}{1+\exp\left(-(\omega_1 p_{i,1}^* + \omega_2 p_{i,2}^* + \omega_3 p_{i,3}^* + b)\right)} \quad (11)$$

Where the final credit risk score (S_i with values [0,1]) for the i th enterprise is derived by mapping the linearly weighted sum of the prediction probabilities ($p_{i,1}^*$, $p_{i,2}^*$, $p_{i,3}^*$) from the three optimal base learners —using their respective importance weights (ω_1, ω_2, ω_3) and a secondary correction bias (b)—through a Sigmoid activation function ($\sigma(\cdot)$), thereby achieving a panoramic insight into default probabilities.

3 Results

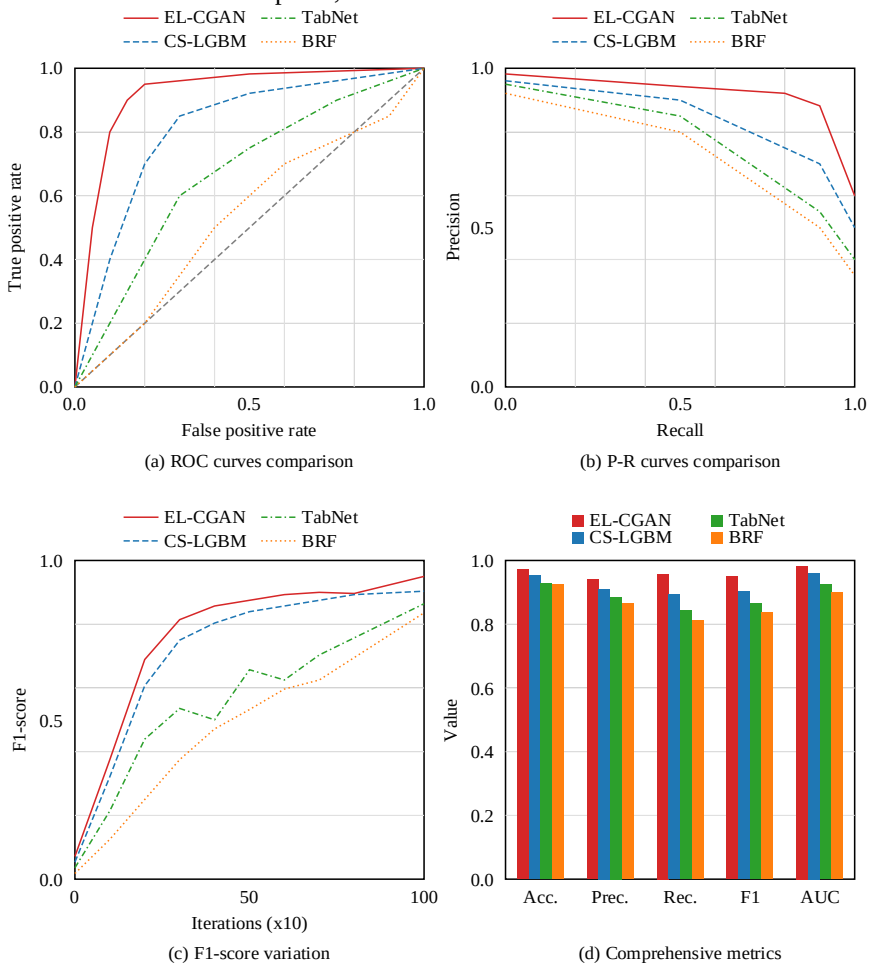
3.1 Performance validation of the EL-CGAN model

To comprehensively evaluate the risk identification accuracy of the EL-CGAN model under extreme class imbalance and feature heterogeneity, experiments utilized the UCI Taiwan Bankruptcy Prediction Dataset. Specifically, the dataset comprises 6,819 enterprise samples (6,599 normal, 220 default) with 95 financial features. Missing values were median-imputed, and a chronological split (70% train, 30% test) was strictly applied to simulate real-world temporal forecasting.

To thoroughly verify methodological superiority, benchmarks were expanded to include standard imbalance treatments (Class Weights, SMOTE, CTGAN) and strong standalone tabular learners (XGBoost, LightGBM, CatBoost), alongside the originally selected Cost-Sensitive LightGBM (CS-LightGBM), Attentive Interpretable Tabular Learning (TabNet), and Balanced Random Forest (BRF). Furthermore, to ensure robust evaluation and address decision-centric requirements, Brier Score and Expected Calibration Error (ECE) were incorporated to assess probability calibration, alongside SHAP analysis for feature interpretability. All experiments were uniformly executed in a Python 3.8 environment utilizing PyTorch and Scikit-learn frameworks.

3.1.1 Comparison of the overall classification performance of the EL-CGAN model

The study first conducted a comparative experiment on the overall classification performance of the models to visually present the differences between the models in the core indicators. The specific results are shown in Figure 8.



(a) ROC curves comparison (b) P-R curves comparison

(c) F1-score variation (d) Comprehensive metrics

Figure 8: Comparison of the overall classification performance of EL-CGAN and the baseline model

Figure 8 confirms EL-CGAN possessed the optimal generalization boundary, with its Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves completely enclosing these initial benchmarks, alongside rapid, non-oscillating training convergence (Fig. 8a-c). Furthermore, to comprehensively address state-of-the-art imbalance treatments and robust standalone models under strict probability calibration, an extended quantitative evaluation is presented in Table 2.

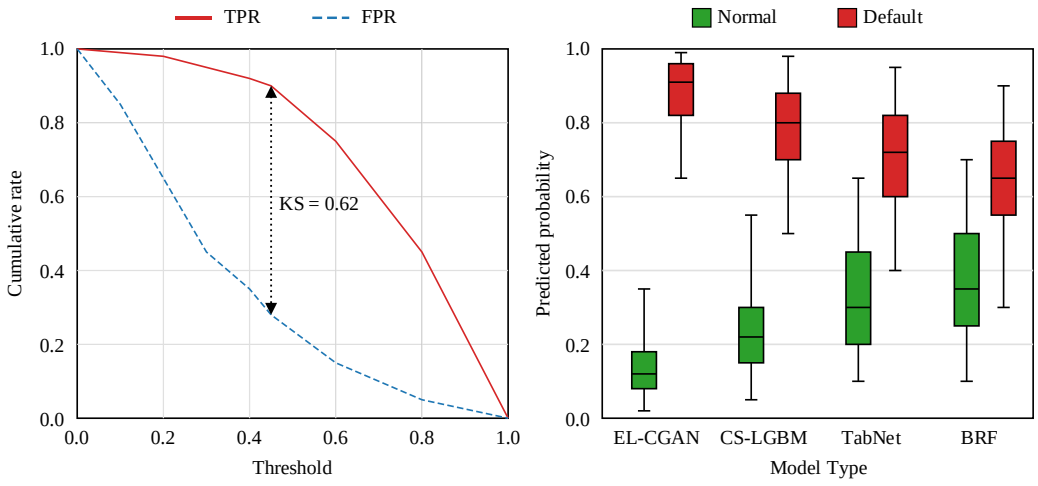
Table 2: Extended quantitative comparison of classification performance and probability calibration

Model category	Specific model	F1 score (↑)	KS statistic (↑)	Brier score (↓)	ECE (↓)
Proposed model	EL-CGAN	0.949 ± 0.012	0.620 ± 0.015	0.045 ± 0.003	0.021 ± 0.002
Original baselines	CS-LightGBM	0.903 ± 0.018	0.581 ± 0.020	0.062 ± 0.005	0.045 ± 0.004
	TabNet	0.863 ± 0.025	0.512 ± 0.022	0.081 ± 0.007	0.063 ± 0.005
	Balanced RF (BRF)	0.845 ± 0.015	0.498 ± 0.018	0.088 ± 0.006	0.071 ± 0.004
Imbalance treatments	Class Weights (w/ XGB)	0.821 ± 0.020	0.485 ± 0.021	0.095 ± 0.008	0.082 ± 0.007
	SMOTE (w/ XGB)	0.875 ± 0.016	0.530 ± 0.019	0.072 ± 0.005	0.055 ± 0.006
	CTGAN (w/ XGB)	0.912 ± 0.022	0.575 ± 0.024	0.058 ± 0.007	0.038 ± 0.005
Standalone learners	XGBoost	0.895 ± 0.014	0.560 ± 0.016	0.065 ± 0.004	0.050 ± 0.005
	LightGBM	0.888 ± 0.015	0.555 ± 0.017	0.068 ± 0.005	0.052 ± 0.004
	CatBoost	0.905 ± 0.012	0.570 ± 0.014	0.055 ± 0.003	0.035 ± 0.003

Table 2 details the comprehensive performance and calibration metrics across all evaluated models. While advanced resampling (e.g., CTGAN) and strong standalone tree models (e.g., CatBoost) improved upon basic methods, EL-CGAN maintained strict dominance. Specifically, it achieved the peak F1 score (0.949) and Kolmogorov-Smirnov statistic (0.620), while uniquely securing the lowest Expected Calibration Error (ECE) and Brier Score. Traditional heuristic methods like SMOTE suffered from feature distortion in heterogeneous spaces, generating synthetic samples that lacked business logic near decision boundaries, thereby hitting a performance bottleneck (F1 of 0.875). Conversely, standalone tree models greedily pushed predicted probabilities to extremes to minimize cross-entropy, leading to severe probability overconfidence (ECE > 0.035). EL-CGAN successfully mitigated these issues through adversarial distribution reconstruction and logistic regression meta-correction, providing superior probability calibration and risk identification accuracy.

3.1.2 Validation of the credit risk discrimination ability of the EL-CGAN model

Although the EL-CGAN model was proven to perform well on overall metrics, its ability to accurately isolate defaulting and normal samples was even more crucial in supply chain finance risk control practices. Therefore, this study further introduced the Kolmogorov-Smirnov Curve (KS Curve) and predicted probability distribution analysis to verify the model's ability to effectively differentiate high-risk enterprises from a large sample size and its ranking stability. The specific results are shown in Figure 9.



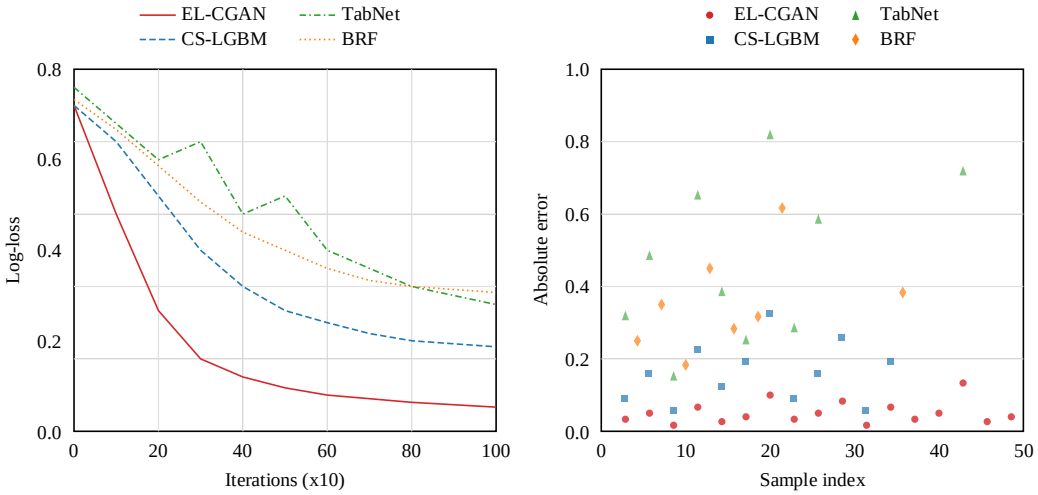
(a) KS statistic analysis (EL-CGAN) (b) Predicted probability distribution

Figure 9: Comparison of credit risk identification ability and probability distribution

Figure 9 illustrates the clearest separation of predicted probability distributions by EL-CGAN. With minimal box overlap compared to the significant confusion in intermediate intervals seen in baselines, EL-CGAN significantly reduced decision uncertainty by precisely targeting high-risk enterprises (Fig. 9b).

3.1.3 Stability and error analysis of EL-CGAN model training

After verifying the classification accuracy and discriminative ability of the EL-CGAN model, this study further analyzed the learning stability and generalization boundary of EL-CGAN under the BO mechanism by recording the loss change trajectory during the training cycle and the prediction bias distribution of the test samples. The specific results are shown in Figure 10.



(a) Training loss convergence (b) Prediction error distribution

Figure 10: Comparison of training stability and error between EL-CGAN model and baseline model

Figure 10 compares the training stability and prediction error distribution. The EL-CGAN training loss demonstrated the fastest convergence and lowest steady-state error (logarithmic loss < 0.05), proving the TPE-BO strategy effectively avoided local optima (Fig. 10a). In randomly sampled predictions, EL-CGAN's absolute error converged tightly near the zero axis, demonstrating superior robustness and generalization accuracy over models like TabNet which exhibit wide error distributions (Fig. 10b).

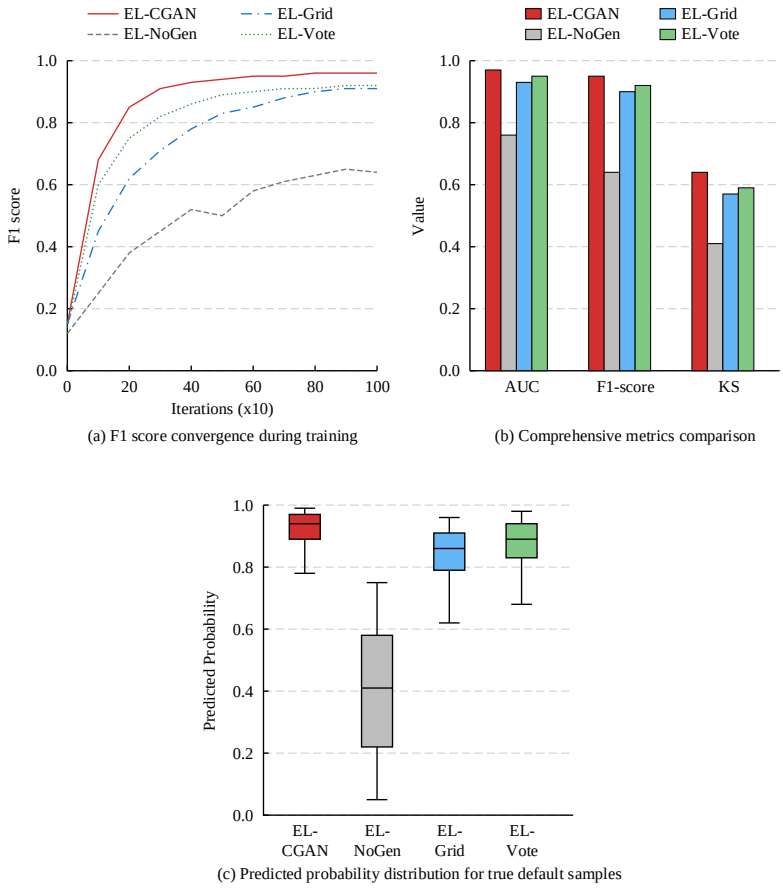
3.2 Ablation experimental analysis of the EL-CGAN model

To isolate and quantify the specific contributions of the CGAN, TPE-BO, and Stacking ensemble strategies in addressing imbalanced and heterogeneous features, a systematic ablation experiment was conducted using the previously established dataset and environment. The specific ablation variants (EL-NoGen, EL-Grid, EL-Vote) and their comparison scheme are detailed in Table 3.

Table 3: Ablation experiment setup and comparison scheme

Group	Model name	Data augmentation	Optimization strategy	Ensemble strategy	Validation goal
A	EL-CGAN	Conditional GAN (CGAN)	TPE (Bayesian Opt)	Stacking (Logistic Regression)	Validate the optimal performance of the full architecture
B	EL-NoGen	None (Raw Imbalanced)	TPE (Bayesian Opt)	Stacking (Logistic Regression)	Validate the necessity of generative augmentation for class skew
C	EL-Grid	Conditional GAN (CGAN)	Grid Search	Stacking (Logistic Regression)	Validate the efficiency and accuracy improvement of BO
D	EL-Vote	Conditional GAN (CGAN)	TPE (Bayesian Opt)	Soft Voting	Validate the bias-correction capability of Stacking meta-learner

Table 3 outlines the control variable framework, sequentially stripping away core modules to quantify their marginal contributions to generative augmentation, optimization efficiency, and bias correction. The multi-dimensional performance differences across these variants are visualized in Figure 11.



(a) F1 score convergence during training

(b) Comprehensive metrics comparison

(c) Predicted probability distribution for true default samples

Figure 11: Comparison of ablation experimental results of EL-CGAN model

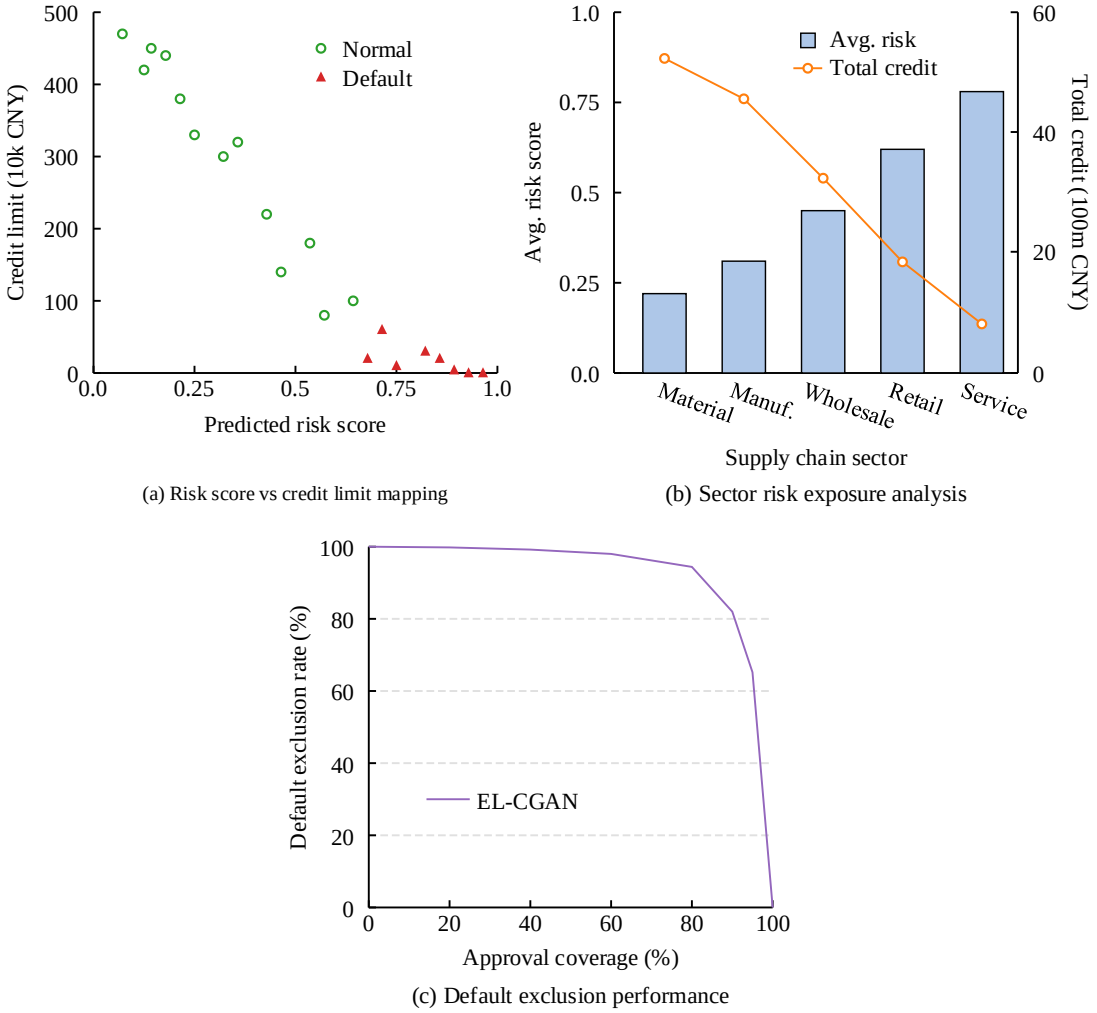
Figure 11 demonstrates the ablation results across training convergence, classification metrics, and predicted probability distributions. Driven by the TPE-BO mechanism, the complete EL-CGAN model achieved the fastest training convergence with a peak steady-state F1 score of 0.96, drastically outperforming the EL-NoGen variant (0.64) (Fig. 11a). Furthermore, EL-CGAN secured an outstanding KS score of 0.64 compared to EL-NoGen's 0.41, confirming that generative adversarial reconstruction was critical for mitigating class bias (Fig. 11b). Finally, EL-CGAN yielded a highly compact predicted probability distribution for default samples with a median of 0.94, whereas all variant models exhibited severe predictive divergence (Fig. 11c). Overall, the absence of any single core component significantly degraded risk identification accuracy, proving the indispensable synergy of the proposed architecture.

3.3 Analysis of the simulation application effect of the EL-CGAN model

To evaluate the model's practical economic and decision-making value, simulation experiments based on real supply chain finance business logic were conducted using the previously established dataset and environment.

3.3.1 Dynamic credit mapping distribution based on risk scoring

The study first constructed an inverse proportional credit granting function based on risk scoring to simulate differentiated credit resource allocation strategies. By analyzing the mapping relationship between the score distribution of the test set samples and the suggested credit limit, the aim was to verify the rationality of the business logic of guiding funds to high-quality enterprises and reducing high-risk exposure. The specific results are shown in Figure 12.



(a) Risk score vs credit limit mapping

(b) Sector risk exposure analysis

(c) Default exclusion performance

Figure 12: Simulation results of dynamic credit granting strategy and risk distribution based on EL-CGAN model

Figure 12 illustrates the dynamic credit mapping distribution and risk hedging capabilities. The model constructed a strict inverse-proportional boundary, accurately confining all default samples to the high-risk, low-limit zone (score > 0.65) (Fig. 12a). Furthermore, it dynamically hedged against industry risks by intelligently minimizing credit exposure (to 820 million yuan) for the service sector, which exhibited the highest average risk score (~0.78) (Fig. 12b). Ultimately, the model demonstrated significant early risk identification, eliminating 94.5% of potential defaults while maintaining an 80% credit coverage rate (Fig. 12c).

3.3.2 Simulation of the economic benefits of credit portfolio

To translate the performance metrics of the EL-CGAN model into intuitive economic value, this study assumed a principal of 1 million yuan per loan, an annualized interest rate of 6%, and a default loss rate of 100%. Under this assumption, the net profit of the credit portfolio corresponding to different risk control thresholds was calculated, aiming to find the optimal decision point that balances interest income and bad debt losses. The specific results are shown in Table 4.

Table 4: Simulation results of the economic benefits of credit portfolios based on the EL-CGAN model

Decision threshold	Approved count	Approval rate (%)	Portfolio default rate (%)	Est. interest income (10k CNY)	Est. bad debt loss (10k CNY)	net portfolio profit (10k CNY)
0.1	520	26.00%	0.00%	3,120	0	3,120
0.2	890	44.50%	0.11%	5,340	98	5,242
0.3 (Optimal)	1,240	62.00%	0.56%	7,440	694	6,746
0.4	1,480	74.00%	1.62%	8,880	2,398	6,482
0.5	1,650	82.50%	3.03%	9,900	4,999	4,901

0.6	1,760	88.00%	4.55%	10,560	8,008	2,552
0.7	1,840	92.00%	5.43%	11,040	9,991	1,049
0.8	1,910	95.50%	6.02%	11,460	11,498	-38
0.9	1,960	98.00%	6.38%	11,760	12,505	-745
Baseline (None)	2,000	100.00%	6.60%	12,000	13,200	-1,200

Table 4 details the net profit of the credit portfolio across different risk control thresholds to identify the optimal balance between interest income and bad debt. As entry thresholds relaxed, linear interest income growth was offset by exponentially rising bad debt losses. However, at the optimal threshold of 0.3, EL-CGAN perfectly balanced scale expansion and risk control, achieving a peak net profit of 67.46 million yuan. In stark contrast, a baseline scenario without risk control (full entry) suffered a 6.60% default rate, resulting in a severe 12-million-yuan economic loss. This confirmed the model's capacity to precisely identify hidden risks and maximize portfolio returns through scientifically optimized boundaries.

3.3.3 Individual enterprise risk assessment and decision backtesting

To verify the interpretability and accuracy of the model's decisions at a micro level, this study selected typical "highly concealed defaulting firms" and "highly risky but easily misjudged high-quality firms" from the test set for backtesting analysis. By comparing the model scores, suggested credit strategies, and the firms' actual financial situations and default outcomes, the aim was to demonstrate the EL-CGAN model's keen insight in handling complex feature heterogeneity. Specific results are shown in Table 5.

Table 5: Detailed backtesting of credit assessment and decision-making for typical supply chain enterprises using the EL-CGAN model.

Sample ID	Debt ratio	ROE	Risk score	Decision	Suggested limit (10k CNY)	Actual label
S_2045	0.18 (Low)	12.5% (High)	0.042	Priority	500	Normal
S_0312	0.45 (Med)	4.2% (Med)	0.215	Prudent	200	Normal
S_1108	0.82 (High)	-1.5% (Neg)	0.892	Reject	0	Default
S_4562*	0.65 (High)	8.1% (High)	0.354	Observation	100	Normal
S_0059*	0.35 (Low)	2.1% (Low)	0.781	Early Warning	0	Default
S_1233	0.22 (Low)	15.3% (High)	0.033	Priority	450	Normal
S_3881	0.91 (Crit)	-12.4% (Neg)	0.995	Blacklist	0	Default
S_2210	0.55 (Med)	5.6% (Med)	0.288	Standard	250	Normal
S_0992*	0.41 (Med)	1.8% (Low)	0.655	High Risk	0	Default
S_1776	0.29 (Low)	9.8% (High)	0.102	Priority	400	Normal

Note: Samples marked with an asterisk (*) are misleading "highly concealed" or "highly leveraged" samples.

Table 5 presents the detailed backtesting of credit assessment and decision-making for these highly deceptive enterprise samples. For standard extreme cases (e.g., S_2045, S_3881), the model's judgments perfectly aligned with actual outcomes. Crucially, for highly deceptive "hidden defaults" (e.g., S_0059 with a seemingly safe debt ratio of 0.35), EL-CGAN still issued a high-risk warning (0.781) to successfully avoid bad debts; SHAP analysis logic suggested this was driven by the model capturing nonlinear anomalies in secondary heterogeneous features. Conversely, for high-quality yet highly leveraged enterprises (e.g., S_4562), the model avoided naive misjudgments by comprehensively weighting strong profitability (ROE) against the high debt ratio. This fully demonstrated the architecture's deep insight, interpretability, and robust decision-making in complex heterogeneous spaces.

4 Discussion

The proposed EL-CGAN successfully addressed extreme class imbalance and data heterogeneity, outperforming benchmarks with a peak F1 score of 0.949 and a KS statistic of 0.62. Traditional linear oversampling destroyed the inherent logic of financial indicators by ignoring the "hollowing out" of high-dimensional risk manifolds. Conversely, the proposed CGAN mechanism reconstructed the true joint distribution of features via adversarial games. While GANs have demonstrated universal applicability in resolving data scarcity across diverse domains like satellite imagery [10], smart home energy [11], and vehicle networks [12], this study explicitly extended their paradigm to supply chain finance. The study proved that even in tabular heterogeneous data lacking spatial or temporal correlation, generative manifold reconstruction effectively alleviated identification blind spots, establishing profound application value for generative risk control.

Furthermore, the TPE-BO driven Stacking strategy maximized the utility of augmented data by capturing complex nonlinear boundaries. Compared to existing

studies validating multi-source fusion [6] and ensemble learning for key variable identification [7], EL-CGAN explicitly overcame their lingering distribution distortion limitations. Addressing the misjudgment risks highlighted by Liang et al. [8], this study advanced beyond theoretical prediction by determining an optimal balance threshold (0.3) in simulation experiments. This leap from accurate classification to maximized decision benefit provides financial institutions with a highly operational algorithmic framework.

Despite these advancements, the study exhibits specific limitations defining future research trajectories. First, relying on static cross-sectional data assumes sample independence, neglecting the risk contagion effect inherent in supply chain networks. Future research should integrate Knowledge Graphs (KG) and Graph Neural Networks (GNNs), which have proven effective in capturing semantic associations between borrowers [29], facilitating a transition from isolated "point assessment" to holistic "network assessment." Second, the model currently relies exclusively on structured financial data, ignoring unstructured soft information. Leveraging Large Language Models (LLMs) to extract sentiment and logical reasoning from financial texts, as advocated by Q. Chen [30], represents a promising avenue to construct a comprehensive multimodal risk control system.

5 Conclusion

To address extreme class imbalance and heterogeneous data in supply chain credit evaluation, this study proposed the EL-CGAN model. It utilized an improved CGAN to reconstruct sparse default distributions and introduces a TPE driven BO Stacking strategy for accurate high-dimensional classification. Results validated EL-CGAN's superior risk identification and economic viability. Specifically, it significantly outperformed benchmark algorithms with a peak F1 score of 0.949, a KS statistic of 0.62, and highly robust training convergence (logarithmic loss < 0.05). In simulated applications, the model achieved a 94.5% default rejection rate while maintaining 80% credit coverage. Crucially, it maximized the credit portfolio net profit to 67.46 million yuan at an optimal 0.3 decision threshold, effectively averting the massive 12-million-yuan loss seen in unmitigated baselines. While this validated the generative and integrated decision-making paradigm for supply chain risk control, current reliance on static, structured cross-sectional data limits the capture of network contagion and unstructured soft information. Future research will integrate Knowledge Graphs (KGs), Graph Neural Networks (GNNs), and Large Language Models (LLMs) to construct a dynamic, multimodal intelligent risk control system capable of holistic network assessment.

Funding

The research is supported by 2025 Liaoning Provincial Social Science Planning Fund Project: Research on Empowering High-Quality Development of the Private Economy through Supply Chain Finance in the Era of Digital Intelligence (L25BJY011).

References

[1] U. O. Nnaji, L. B. Benjamin, N. L. Eyo-Udo, and E. A. Etukudoh, "Advanced risk management models for supply chain finance," World J. Adv. Res. Rev., vol. 22, no. 2, pp. 612-618, May, 2024, DOI: 10.30574/wjarr.2024.22.2.1444.

[2] S. S. Goswami, S. Mondal, S. Sarkar, K. K. Gupta, S. K. Sahoo, and R. Halder, "Artificial intelligence-enabled supply chain management: Unlocking new opportunities and challenges," Artif. Intell. Appl., vol. 3, no. 1, pp. 110-121, September, 2024, DOI: 10.47852/bonviewAIA42021814.

[3] S. O. Khalifaturofi'ah, "Cost efficiency, innovation and financial performance of banks in Indonesia," J. Econ. Adm. Sci., vol. 39, no. 1, pp. 100-116, February, 2023, DOI: 10.1108/JEAS-07-2020-0124.

[4] F. Ahmed, K. Nizam, Z. Sajid, and S. Qamar, "Striking a balance: Evaluating credit risk with traditional and machine learning models," Bull. Bus. Econ., vol. 13, no. 3, pp. 30-35, August, 2024, DOI: 10.61506/01.00425.

[5] J. Feng, Q. Xiao, S. Lu, and H. Zhang, "A double remote magnetic field synthesis method for reducing high-speed MFL signal distortion caused by velocity effect," IEEE Trans. Ind. Electron., vol. 71, no. 1, pp. 1049-1059, February, 2023, DOI: 10.1109/TIE.2023.3239857.

[6] L. Wang, F. Jia, L. Chen, and Q. Xu, "Forecasting SMEs' credit risk in supply chain finance with a sampling strategy based on machine learning techniques," Ann. Oper. Res., vol. 331, no. 1, pp. 1-33, January, 2022, DOI: 10.1007/s10479-022-04518-5.

[7] A. Belhadi, S. S. Kamble, V. Mani, I. Benkhati, and F. E. Touriki, "An ensemble machine learning approach for forecasting credit risk of agricultural SMEs' investments in agriculture 4.0 through supply chain finance," Ann. Oper. Res., vol. 345, no. 2, pp. 779-807, November, 2021, DOI: 10.1007/s10479-021-04366-9.

[8] D. Liang, W. Cao, and M. Wang, "Credit rating of sustainable agricultural supply chain finance by integrating heterogeneous evaluation information and misclassification risk," Ann. Oper. Res., vol. 331, no. 1, pp. 189-219, November, 2021, DOI: 10.1007/s10479-021-04453-x.

[9] J. P. Kozak, R. Zhang, M. Porter, Q. Song, J. Liu, B. Wang, et al., "Stability, reliability, and robustness of GaN power devices: A review," IEEE Trans. Power Electron., vol. 38, no. 7, pp. 8442-8471, April, 2023, DOI: 10.1109/TPEL.2023.3266365.

[10] A. Alzahem, W. Boulila, A. Koubaa, Z. Khan, and I. Alturki, "Improving satellite image classification accuracy using GAN-based data augmentation and vision transformers," Earth Sci. Inf., vol. 16, no. 4, pp. 4169-4186, November, 2023, DOI:

10.1007/s12145-023-01153-x.
[11] M. Razghandi, H. Zhou, M. Erol-Kantarci, and D. Turgut, "Smart home energy management: VAE-GAN synthetic dataset generator and Q-learning," IEEE Trans. Smart Grid, vol. 15, no. 2, pp. 1562-1573, June, 2023, DOI: 10.1109/TSG.2023.3288824.
[12] A. Chougule, K. Agrawal, and V. Chamola, "Scan-gan: Generative adversarial network based synthetic data generation technique for controller area network," IEEE Internet Things Mag., vol. 6, no. 3, pp. 126-130, September, 2023, DOI: 10.1109/IOTM.001.2300013.
[13] S. Zhang, A. Wijesinghe, and Z. Ding, "RME-GAN: A learning framework for radio map estimation based on conditional generative adversarial network," IEEE Internet Things J., vol. 10, no. 20, pp. 18016-18027, May, 2023, DOI: 10.1109/JIOT.2023.3278235.
[14] P. Shen, L. Zhang, M. Wang, and G. Yin, "Deeper super-resolution generative adversarial network with gradient penalty for sonar image enhancement," Multimedia Tools Appl., vol. 80, no. 18, pp. 28087-28107, May, 2021, DOI: 10.1007/s11042-021-10888-y.
[15] E. M. Campos, A. Gonzalez-Vidal, J. L. Hernandez-Ramos, and A. Skarmeta, "Federated learning for misbehaviour detection with variational autoencoders and Gaussian mixture models," Int. J. Inf. Secur., vol. 24, no. 2, pp. 1-16, March, 2025, DOI: 10.1007/s10207-025-01000-8.
[16] T. Li, G. Kou, Y. Peng, and P. S. Yu, "Feature selection and grouping effect analysis for credit evaluation via regularized diagonal distance metric learning," INFORMS J. Comput., vol. 37, no. 5, pp. 1391-1412, December, 2024, DOI: 10.1287/ijoc.2023.0322.
[17] Z. Yao, S. Xia, Y. Li, and G. Wu, "Cooperative task offloading and service caching for digital twin edge networks: A graph attention multi-agent reinforcement learning approach," IEEE J. Sel. Areas Commun., vol. 41, no. 11, pp. 3401-3413, August, 2023, DOI: 10.1109/JSAC.2023.3310080.
[18] Y. Jiang, W. Wang, L. Zou, and Y. Cao, "Regional landslide susceptibility assessment based on improved semi-supervised clustering and deep learning," Acta Geotech., vol. 19, no. 1, pp. 509-529, June, 2023, DOI: 10.1007/s11440-023-01950-0.
[19] Z. Chen, L. Zhang, J. Sun, R. Meng, S. Yin, and Q. Zhao, "DCAMCP: A deep learning model based on capsule network and attention mechanism for molecular carcinogenicity prediction," J. Cell. Mol. Med., vol. 27, no. 20, pp. 3117-3126, July, 2023, DOI: 10.1111/jcmm.17889.
[20] Y. Jiang, S. Xie, X. Xie, Y. Cui, and H. Tang, "Emotion recognition via multiscale feature fusion network and attention mechanism," IEEE Sens. J., vol. 23, no. 10, pp. 10790-10800, April, 2023, DOI: 10.1109/JSEN.2023.3265688.
[21] Y. Xu, Y. Lyu, G. Xiong, S. Wang, W. Wu, H. Cui, and J. Luo, "Adaptive feature fusion networks for origin-destination passenger flow prediction in metro systems," IEEE Trans. Intell. Transp. Syst., vol. 24, no. 5, pp. 5296-5312, January, 2023, DOI: 10.1109/TITS.2023.3239101.
[22] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," Multimedia Tools Appl., vol. 82, no. 3, pp. 3713-3744, July, 2022, DOI: 10.1007/s11042-022-13428-4.
[23] B. Hou, Z. Cai, K. Wu, T. Yang, and T. Zhou, "6scan: A high-efficiency dynamic internet-wide ipv6 scanner with regional encoding," IEEE/ACM Trans. Netw., vol. 31, no. 4, pp. 1870-1885, January, 2023, DOI: 10.1109/TNET.2023.3233953.
[24] Y. Yuan, Z. Li, and B. Zhao, "A survey of multimodal learning: Methods, applications, and future," ACM Comput. Surv., vol. 57, no. 7, pp. 1-34, February, 2025, DOI: 10.1145/3713070.
[25] M. Liu, D. Li, Y. Li, X. Song, and L. Nie, "Audio-semantic enhanced pose-driven talking head generation," IEEE Trans. Circuits Syst. Video Technol., vol. 34, no. 11, pp. 11056-11069, June, 2024, DOI: 10.1109/TCSVT.2024.3414412.
[26] Z. Fan, J. Gou, and S. Weng, "A feature importance-based multi-layer catboost for student performance prediction," IEEE Trans. Knowl. Data Eng., vol. 36, no. 11, pp. 5495-5507, May, 2024, DOI: 10.1109/TKDE.2024.3393472.
[27] D. Shan, K. Yao, and X. Zhang, "Sequential learning network with residual blocks: Incorporating temporal convolutional information into recurrent neural networks," IEEE Trans. Cogn. Dev. Syst., vol. 16, no. 1, pp. 396-401, October, 2023, DOI: 10.1109/TCDS.2023.3325358.
[28] H. Ghaderi Zefrehi, G. Sheikhi, and H. Altınçay, "Threshold prediction for detecting rare positive samples using a meta-learner," Pattern Anal. Appl., vol. 26, no. 1, pp. 289-306, September, 2022, DOI: 10.1007/s10044-022-01103-1.
[29] Y. Zhu and D. Wu, "P2P credit risk management with KG-GNN: a knowledge graph and graph neural network-based approach," J. Oper. Res. Soc., vol. 76, no. 5, pp. 866-880, September, 2024, DOI: 10.1080/01605682.2024.2398762.
[30] Q. Chen, "Explore the use of prompt-based LLM for credit risk classification," J. Comput. Commun., vol. 13, no. 6, pp. 33-46, January, 2025, DOI: 10.1007/s00521-024-10495-6.