# Research on Credit Risk Warning Model for Enterprise Supply Chain Partners Integrating Ensemble Learning and GAN

**Lingxian Zhu**
School of Marxism, Zhengzhou Railway Vocational & Technical College, Zhengzhou 451460, Henan, China
Corresponding author: Lingxian Zhu (e-mail: cynthiasff@163.com)

**ABSTRACT** In the context of global industrial shifts and heightened market volatility, credit risks among enterprise supply-chain partners exhibit rapid contagion, extensive reach, and deep concealment. This volatility is further compounded by complex procurement, logistics and cooperation networks, which can mask underlying financial instability until it disrupts production continuity. Traditional credit risk warning models are difficult to meet the demand for accurate warning due to data imbalance, insufficient feature extraction, and weak generalization ability of a single algorithm. Ensemble learning improves prediction stability through the collaborative advantage of multiple models, while Generative Adversarial Networks (GANs) can solve sample imbalance problems through data augmentation. The integration of these advanced methodologies provides the technical framework necessary to improve credit risk warning under imbalanced samples and complex supply-chain relationships. By combining synthetic minority samples with ensemble learning, the model supports more accurate risk identification and timely intervention.

**INDEX TERMS** credit risk warning,ensemble learning,generative adversarial network,supply chain partners,industrial supply chain

## I. INTRODUCTION

### A. RESEARCH BACKGROUND

AS the fundamental framework of industrial operations, the seamless flow of fiber sourcing and production directly dictates a firm's market competitiveness and resilience against economic shocks. Ensuring the stability of these upstream and downstream linkages allows manufacturers to maintain consistent fabric quality and output amidst the inherent volatility of the global trade. The 14th Five Year Plan for the Development of Modern Supply Chain clearly proposes to strengthen the construction of the supply chain credit system, build an intelligent credit risk warning mechanism, and enhance the ability to prevent and control supply chain risks. In recent years, factors such as global pandemics, geopolitical conflicts, and fluctuations in raw material prices have intensified supply chain uncertainty, leading to frequent credit default incidents among partners[1-3].

The transmission of credit risk is enhanced, and the default of a single partner may trigger a 'domino effect', leading to supply chain disruptions;

The manifestations of risks are diversified, covering various types such as payment defaults, delivery delays, substandard quality, and broken funding chains;

The traditional warning mechanism is ineffective, and a single evaluation system relying on financial indicators is difficult to capture non-financial risk factors, resulting in significant warning lag;

The problem of data imbalance is prominent, with a high proportion of creditworthy enterprises in the sample (usually over 80%), leading to biased predictions in the model.

### B. RESEARCH SIGNIFICANCE

1) Theoretical Significance

(1) Build a four-dimensional supply chain credit risk evaluation system of "finance operation cooperation environment", improve the theoretical framework of supply chain credit risk, and enrich the dimensions and indicators of credit risk evaluation;

(2) Propose a fusion model architecture of GAN data augmentation+Stacking ensemble learning, clarify the fusion mechanism of generative adversarial networks and ensemble learning, and expand the application boundaries of generative adversarial networks and ensemble learning;

(3) Unveiling the core determinants and transmission mechanisms of credit risk within imbalanced data scenarios provides a robust theoretical foundation for modeling the financial volatility inherent in highprecision manufacturing. This analytical framework enables more accurate predictive modeling where rare default events are often obscured by the

high volume of standard fiber procurement and processing transactions.

#### 2) Practical Significance

(1) Develop a high-precision enterprise supply chain partner credit risk warning model to achieve early identification and accurate warning of credit risks, helping enterprises reduce default losses;

(2) Using GAN technology to solve the problem of imbalanced supply chain credit risk data, improve the model's ability to identify low credit rating enterprises, and avoid the risk of "missed judgment";

(3) Provide data support and decision-making basis for selecting supply chain partners, credit rating, and adjusting cooperation strategies for enterprises, and enhance the scientific level of supply chain management;

(4) Provide replicable technical solutions and practical experience for credit risk warning in other fields such as financial credit and e-commerce platforms, and promote the intelligent development of credit risk management[4].

### C. CURRENT RESEARCH STATUS AT HOME AND ABROAD

#### 1) Current Research Status Abroad

Research on supply chain credit risk warning and machine learning applications started earlier in foreign countries, mainly focusing on three directions:

(1) The research on credit risk evaluation system has constructed a multidimensional evaluation system that includes financial indicators, operational indicators, and market indicators;

(2) Research on a single algorithm warning model, applying algorithms such as logistic regression, support vector machine, and neural network to credit risk prediction;

(3) Research on the application of ensemble learning and data augmentation, attempting to improve predic tion accuracy through ensemble learning and alleviate data imbalance problems using traditional data augmentation methods such as SMOTE.

The advantages of foreign research lie in its solid technological foundation and abundant data resources, but there are also shortcomings:

Traditional methods such as SMOTE are often used for data augmentation, resulting in limited data quality and difficulty in accurately reflecting credit risk characteristics;

The selection and fusion strategy of the base model in ensemble learning lacks specificity and does not fully consider the particularity of supply chain credit risk;

The evaluation system is mostly designed based on the characteristics of Western enterprises, and its adaptability to Chinese supply chain enterprises is insufficient[5].

#### 2) Current Research Status in China

In recent years, domestic research has developed rapidly and conducted extensive explorations around supply chain credit risk warning

(1) In terms of constructing an evaluation system, scholars have proposed evaluation indicators that include dimensions such as finance, operations, and cooperative relationships, but the selection of indicators lacks quantitative basis;

(2) In terms of research on warning models, warning models based on a single algorithm or simple ensemble learning have been developed, but the problem of data imbalance has not been effectively solved;

(3) In terms of technology integration applications, some studies have attempted to use GAN technology for credit risk modeling, but there is insufficient research on deep integration with ensemble learning. Prominent gaps in domestic research:

Lack of a fusion model that balances data balance and algorithm performance, making it difficult to simultaneously address issues of data quality and prediction accuracy;

The industry adaptability of the evaluation system is insufficient, and the differences in risk characteristics of different industry supply chains have not been fully considered;

The empirical research sample is small and the industry coverage is narrow, and the generalization ability and practicality of the model need to be verified[6].

## II. CORE CONCEPTS AND THEORETICAL FOUNDATIONS

### A. DEFINITION OF CORE CONCEPTS

#### 1) Credit Risk of Supply Chain Partners

Credit risk of supply chain partners refers to the possibility that upstream and downstream enterprises (suppliers, distributors, logistics providers, etc.) in the supply chain, due to deteriorating financial conditions, poor operational management, decreased willingness to cooperate, external environmental changes, and other factors, fail to fulfill contractual agreements (such as on-time delivery, quality supply, and timely payment), causing economic losses to the cooperating enterprises. Its core features include:

Transmission: The credit risk of a single partner may propagate along the supply chain, triggering a chain reaction;

Concealment: Some risk factors, such as willingness to cooperate and management loopholes, are difficult to capture directly through public data;

Diversity: Risk manifestations include payment defaults, delivery delays, substandard quality, and broken funding chains;

Dynamics: Credit risk dynamically evolves with changes in the company's operating conditions, market environment, and cooperative relationships[7-9].

#### 2) Ensemble Learning

Ensemble learning refers to a machine learning method that improves overall prediction performance by constructing multiple base models and integrating their prediction results. This paper uses Stacking integrated learning framework, takes the prediction results of multiple base models as input, and trains the meta model to generate the final prediction results, which can give full play to the advantages of each base model and reduce the bias and variance of a single model[10].

#### 3) Generative Adversarial Networks (GANs)

Generative Adversarial Network (GAN) is a generative machine learning model composed of a generator and a discriminator. The generator generates synthetic data by learning the distribution of real data, while the discriminator is responsible for distinguishing between real data and synthetic data. The two are continuously optimized through adversarial training, ultimately generating synthetic data with a highly similar distribution to real data, which can be used to solve the problem of sample imbalance[11].

#### 4) Data Imbalance

Data imbalance refers to the phenomenon of significant differences in the sample size of different categories in credit risk warning datasets, usually manifested as a high proportion of creditworthy entities (positive samples) and a low proportion of credit default enterprises (negative samples) (usually less than 20%), resulting in the model leaning towards categories with high predicted proportions and reducing its ability to identify minority class samples[12].

### *B. THEORETICAL BASIS*

#### 1) Credit Risk Theory

The credit risk theory holds that the emergence of credit risk is the result of multiple factors such as a company's financial condition, operational capabilities, market environment, and cooperative relationships. Its core viewpoints include:

Financial condition is the core influencing factor of credit risk, and indicators such as asset liability ratio and current ratio directly reflect a company's debt paying ability;

The operational capability determines the sustained profitability of the enterprise, and indicators such as inventory turnover rate and accounts receivable turnover rate reflect the operational efficiency of the enterprise;

The external environment indirectly affects credit risk by influencing the operational status of enterprises, and industry competition, policy changes, and other factors may all trigger credit defaults;

The history of cooperation can reflect the willingness and ability of enterprises to cooperate, and is an important supplement to credit risk assessment[13].

This theory provides a theoretical basis for the construction of a credit risk evaluation system and clarifies the direction for selecting evaluation indicators.

#### 2) Ensemble Learning Theory

The core of ensemble learning theory is "multi model collaborative optimization", which achieves overall performance improvement by integrating the prediction results of multiple base models. Its core viewpoints include:

The diversity of base models is the key to effective ensemble learning, as different types and parameters of base models can capture different features of data;

The fusion strategy directly affects the integration effect, and commonly used fusion strategies include voting method, weighted voting method, stacking method, etc;

Integrated learning can reduce the risk of overfitting of a single model and improve the generalization ability and stability of the model[14-15].

This theory provides technical support for the construction of Stacking ensemble learning models and clarifies the design principles for base model selection and fusion strategies.

#### 3) Generative Adversarial Network Theory

The theory of generative adversarial networks is based on game theory and deep learning, and achieves the generation of synthetic data through adversarial training of generators and discriminators. Its core viewpoints include:

The generator and discriminator form a zero sum game. The goal of the generator is to generate synthetic data that cannot be distinguished by the discriminator, while the goal of the discriminator is to accurately distinguish between real data and synthetic data;

During adversarial training, the performance of the generator and discriminator mutually promote each other. In this study, the approach toward Nash equilibrium was empirically monitored via the convergence of the Wasserstein distance and the stabilization of the discriminator's loss. Unlike traditional GANs, the improved WGANGP exhibited a steady decline in the gradient penalty-adjusted loss, indicating a stable training process that avoided mode collapse and reached a practical state of distributional alignment;

The generated synthetic data can supplement minority class samples and effectively solve the problem of data imbalance[16].

This theory provides a technical basis for improving the construction of the WGANGP model, clarifying the training methods and optimization directions of the model.

#### 4) Data Imbalance Handling Theory

The theory of data imbalance processing focuses on solving the problem of low recognition rate of minority class samples, mainly including processing methods at the data level and algorithm level. Data level methods balance data distribution by adding minority class samples or reducing majority class samples, such as SMOTE, GAN, etc; Algorithm level methods improve the recognition ability of minority class samples by adjusting the model loss function or classification threshold, such as weighted loss function, ensemble learning, etc[17]. This theory provides methodological support for the dual processing approach of "data level (GAN)+algorithm level (ensemble learning)" in this article[18-20].

## III. CONSTRUCTION OF CREDIT RISK EVALUATION SYSTEM FOR SUPPLY CHAIN PARTNERS

### A. CONSTRUCTION PRINCIPLES

#### 1) Principle of Comprehensiveness

Covering the core factors that affect the credit risk of supply chain partners, including financial status, opera tional capabilities, cooperation history, external environment, and other dimensions, to ensure the integrity of the evaluation system.

#### 2) Targeted Principle

Focusing on the particularity of supply chain cooperation scenarios, select indicators closely related to supply chain cooperation (such as on-time delivery rate and cooperation period) to ensure the pertinence of the evaluation system.

#### 3) Principle of Operability

Clear definition of indicators, easy access to data, avoiding the selection of difficult to quantify or unavailable indicators, ensuring the feasibility of data collection and model training.

#### 4) Principle of Scientificity

Based on credit risk theory and empirical research results, combined with expert review, optimize the indicator system to ensure the rationality and relevance of the indicators.

#### 5) Principle of Dynamism

Considering the dynamic changes in the business operation and market environment, the indicator system has adjustability and can adapt to the credit risk evaluation needs at different stages. To move beyond static annual snapshots such as the Asset-Liability Ratio, the system incorporates high-frequency flow data including monthly raw material procurement volumes and weekly subcontracting fulfillment rates. By integrating these real-time operational flows with traditional financial indicators, the model captures the immediate volatility of the supply chain, allowing for the ’agile responses’ required to mitigate sudden shifts in production demands.

### B. EVALUATION SYSTEM CONSTRUCTION PROCESS

#### 1) Preliminary Construction

Based on the four-dimensional framework of "finance operation cooperation environment", combined with literature research and theoretical analysis, a preliminary credit risk evaluation system consisting of 42 indicators is constructed.

#### 2) Expert Review and Optimization

Using the Delphi method for two rounds of expert review:

First round of evaluation: Invite 20 experts to rate the necessity, rationality, and operability of the indicators, delete 7 redundant indicators, and merge 3 duplicate indicators;

Second round of review: Experts proposed supplementary suggestions for the modified system, adding 0 indicators, and ultimately forming a credit risk evaluation system consisting of 4 primary dimensions and 32 secondary indicators.

#### 3) Determination of Indicator Weights

Using Analytic Hierarchy Process (AHP) to determine the weights of each indicator:

1. Build a judgment matrix: Invite experts to compare the relative importance of each dimension and indicator pairwise, and construct a judgment matrix;
2. Consistency check: Calculate the consistency index (CI) and consistency ratio (CR) of the judgment matrix, where CR<0.1 indicates that the judgment matrix meets the consistency requirements;
3. Weight calculation: The weights of each dimension and indicator are calculated using the eigenvalue decomposition method. The total weights of the four primary dimensions are: financial status (0.50), operational capability (0.25), cooperation history (0.15), and external environment (0.10). To address the subjectivity of expert ratings, a quantitative sensitivity analysis was performed by perturbing the primary weights by $\pm 10\%$. The results demonstrated that while the Financial Status weight significantly influences the final score, the model’s classification AUC remained stable (fluctuating by less than 1.5%), confirming that the integrated learning framework is robust against minor variations in subjective indicator weighting.

## IV. DESIGN OF CREDIT RISK WARNING MODEL INTEGRATING ENSEMBLE LEARNING AND GAN

### A. OVERALL MODEL ARCHITECTURE

Construct a five stage model architecture consisting of "data preprocessing GAN data augmentation feature engineering Stacking ensemble learning model evaluation", as follows:

**TABLE 1.** Performance Comparison of Different Models

| Model Type | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Experimental Group: Fusion Model (Improved WGAN-GP + Stacking) | 94.7 | 93.8 | 93.2 | 93.9 |
| Control Group 1: Stacking Ensemble Model | 88.5 | 87.6 | 86.3 | 86.9 |
| Control Group 2: XGBoost | 82.3 | 81.5 | 78.6 | 80.0 |
| Control Group 2: LightGBM | 83.7 | 82.8 | 80.1 | 81.4 |
| Control Group 2: Random Forest | 79.5 | 78.3 | 75.2 | 76.7 |
| Control Group 2: Logistic Regression | 72.6 | 71.8 | 68.3 | 69.9 |

### B. DATA PREPROCESSING

#### 1) Data Cleaning

Eliminate duplicate data: delete completely identical sample data to avoid data redundancy;

Handling outliers: using box plot method to identify outliers, deleting extreme outliers, and replacing minor outliers with median values;

Fill in missing values: For indicators with a missing rate below 5%, use mean/median filling; For indicators with a missing rate of 5% to 10%, K-nearest neighbors (KNN) are used for filling; For indicators with a missing rate higher than 10%, instead of direct deletion, a ’missingness indicator’ variable is constructed to capture the potential risk signal. In supply chain credit assessment, the absence of data often reflects deliberate ’concealment’ or poor transparency;

therefore, these missing patterns are encoded as distinct categorical features to allow the model to learn the risk implications of non-disclosure.

#### 2) Data Standardization

Using the Zscore standardization method to standardize all indicators, the data follows a normal distribution with a mean of 0 and a variance of 1, eliminating the influence of dimensional and value range differences.

### C. GAN DATA ENHANCEMENT MODULE DESIGN

#### 1) Defects of Traditional WGANGP Model

The traditional WGANGP model has the following shortcomings when generating credit risk data:

The feature distribution of the generated data deviates from the real data, resulting in a decrease in the model's generalization ability;

The training process is unstable and prone to Mode Collapse, resulting in insufficient diversity of generated data;

The generation effect of high-dimensional data is poor, making it difficult to capture the complex correlations between credit risk indicators.

#### 2) Improving WGANGP Model Design

To address the shortcomings of traditional models, improvements are made in three aspects: network structure, loss function, and training strategy

1. Network structure optimization:

Generator: Using a deep neural network (DNN) with three hidden layers, the activation function uses LeakyReLU, and the output layer uses Tanh function to ensure that the value range of the generated data is consistent with the real data;

Discriminator: Adopting Convolutional Neural Network (CNN) to enhance feature extraction capability, con sisting of 2 convolutional layers and 2 fully connected layers, with LeakyReLU as the activation function.

2. Improvement of loss function:

Introduce Feature Matching Loss to ensure that the feature distribution of the generated data is consistent with the real data;

Improve the gradient penalty term by using adaptive gradient penalty coefficients to enhance training stability.

3. Training strategy optimization:

Adopting Minibatch GradientDescent with a batch size set to 64;

The training ratio between the generator and discriminator is set to 1:5 to avoid insufficient training of the generator;

Introduce Early Stopping method, which stops training when the accuracy of the discriminator stabilizes at around 50% to avoid overfitting.

#### 3) Data Enhancement Process

1. Separate negative samples (credit default enterprises) and positive samples (credit good enterprises) from the original dataset;

2. Input the negative samples into the improved WGANGP model generator to generate synthesized negative samples that are consistent with the feature distribution of the real negative samples;

Adopt a diversity-oriented filtering mechanism. While cosine similarity is calculated, the threshold is set to retain samples within a range of 0.70 to 0.80. Samples with similarity $\geq 0.85$ are excluded to prevent data leakage and ensure the synthetic data provides novel boundary cases rather than mere replicas of the training set. This ensures the Stacking ensemble learns generalized decision boundaries rather than over-fitting to minor variations of existing default cases;

4. Merge the filtered synthesized negative samples with the original data to form a balanced dataset (with a positive to negative sample ratio of approximately 1:1).

### D. FEATURE ENGINEERING

#### 1) Feature Selection

Using random forest feature importance score to screen core features:

1. Train a random forest model and calculate the feature importance scores for each indicator;

2. Sort in descending order of importance scores and select the top 20 core features (cumulative importance score $\geq 85\%$) as input features to reduce the impact of redundant features.

#### 2) Feature Dimensionality Reduction

Use principal component analysis (PCA) to reduce the dimensionality of the selected features:

1. Calculate the covariance matrix of features, extract eigenvalues and eigenvectors;

2. Select principal components with a cumulative variance contribution rate of $\geq 90\%$ as the final input features to reduce the computational complexity of the model.

#### 3) Feature Construction

Construct three new features based on business logic:

Financial Health Index: Weighted calculation based on financial indicators such as asset liability ratio and current ratio;

Operational Efficiency Index: Weighted calculation based on operational indicators such as on-time delivery rate and inventory turnover rate;

Cooperative Credit Index: Weighted calculation based on cooperative indicators such as years of cooperation and default records.

### E. STACKING INTEGRATED LEARNING MODULE DESIGN

#### 1) Basic Model Selection

Three high-performance machine learning algorithms, XGBoost, LightGBM, and Random Forest, were selected as the base models. The core parameters of the three models are as follows:

**TABLE 2.** Performance Comparison of Different Models

| Stability Test Method | Model Type | Average Accuracy (%) | Accuracy Standard Deviation (%) | Coefficient of Variation (%) |
|---|---|---|---|---|
| 5-Fold Cross Validation | Experimental Group | 94.3 | 0.87 | 0.92 |
| | Control Group 1 | 88.1 | 1.23 | 1.40 |
| | Control Group 2 (XGBoost) | 81.9 | 1.56 | 1.90 |
| Dataset Split Ratio 6:4 | Experimental Group | 94.1 | 0.92 | 0.98 |
| | Control Group 1 | 87.8 | 1.31 | 1.49 |
| | Control Group 2 (XGBoost) | 81.5 | 1.62 | 1.99 |
| Dataset Split Ratio 8:2 | Experimental Group | 95.0 | 0.78 | 0.82 |
| | Control Group 1 | 88.7 | 1.15 | 1.30 |
| | Control Group 2 (XGBoost) | 82.7 | 1.48 | 1.79 |

#### 2) Meta Model Selection

Choosing logistic regression as the meta model is based on comparative testing against non-linear alternatives like Neural Networks. While the base models (XGBoost, LightGBM) capture complex non-linear interactions, a linear meta-model effectively reduces the risk of 'meta-overfitting.' By acting as a weighted regularizer of the base learners' outputs, Logistic Regression provides a more stable integration than non-linear meta-learners, which were found to amplify noise in high-dimensional industry datasets.

#### 3) Stacking Fusion Process

1. Divide the balanced dataset into a training set and a validation set in a 7:3 ratio;

2. Train three base models using 5-fold cross validation to obtain the prediction results of the training set (first stage prediction);

3. Train a logistic regression meta model by using the predicted results of the three base models as input features for the meta model;

4. Use the trained base model to predict the validation set, input the prediction results into the meta model, and obtain the final prediction result (second stage prediction).

#### 4) Model Evaluation Indicators

Select the following indicators to comprehensively evaluate the performance of the model:

Accuracy: The proportion of correctly predicted samples to the total sample size, reflecting the overall predictive ability of the model;

Precision: The proportion of correctly predicted negative samples to all predicted negative samples, reflecting the risk of "misjudgment" in the model;

Recall: The proportion of correctly predicted negative samples to all true negative samples, reflecting the risk of "missed judgments" in the model;

F1 score: the harmonic average of precision and recall, which comprehensively reflects the classification performance of the model;

ROC curve and AUC value: The ROC curve has false positive case rate (FPR) as the horizontal axis and true case rate (TPR) as the vertical axis, and AUC value is the area under the ROC curve. The larger the AUC value, the stronger the model's discriminative ability.

## V. EMPIRICAL RESEARCH DESIGN AND RESULT ANALYSIS

### *A. EMPIRICAL RESEARCH DESIGN*

Data sources and sample selection

Collect sample data using a multi-source data collection method:

Financial data: sourced from financial reports of listed companies, Wind database, and Tonghuashun database;

Operational data: sourced from enterprise supply chain management systems and industry statistical reports;

Collaboration data: sourced from supply chain collaboration records and questionnaire surveys of core enterprises;

Environmental data: sourced from statistical data released by the National Bureau of Statistics and industry associations.

1864 supply chain enterprises from the manufacturing, retail, and logistics industries were selected as research samples, including 1523 credit worthy enterprises (positive sample) (accounting for 81.7%) and 341 credit default enterprises (negative sample) (accounting for 18.3%). The industry distribution of the samples is as follows:

**TABLE 3.** Model Performance in Different Industries

| Industry Type | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Manufacturing (n=1026) | 95.2 | 94.3 | 93.8 | 94.0 |
| Retail (n=518) | 94.4 | 93.5 | 92.7 | 93.1 |
| Logistics Industry (n=320) | 93.8 | 92.9 | 91.5 | 92.2 |

#### 1) Experimental Design

Set up three sets of comparative experiments to verify the superiority of the fusion model:

Experimental group: Ensemble learning and GAN fusion model (improved WGANGP+Stacking ensemble learning);

Comparison group 1: Ensemble learning model without GAN fusion (Stacking Ensemble Learning);

Control group 2: traditional single model (XGBoost, LightGBM, random forest, logistic regression).

Experimental environment: Intel Core i912900K processor, 64GB memory NVIDIA RTX3090 GPU, Implement model training and evaluation using Python 3.9, TensorFlow 2.10, and Scikitlearn 1.2.2.

Experimental steps

1. Data preprocessing: cleaning, missing value filling, outlier handling, and standardization of raw data;

2. Data augmentation: Using an improved WGANGP model to generate synthetic negative samples and construct a balanced dataset;

3. Feature engineering: perform feature screening, dimensionality reduction, and construction to obtain the final input features;

4. Model training: Train the experimental group and control group models separately, optimize model parameters;

5. Model evaluation: Evaluate the performance of the model using indicators such as accuracy, precision, recall, F1 score, AUC value, etc;

6. Stability test: Use 5-fold cross validation and different dataset partitioning ratios (6:4, 7:3, 8:2) to test the stability of the model;

**TABLE 4.** Model Performance in Different Industries

| Rank | Feature Name | Importance Score | Belonging Dimension |
|---|---|---|---|
| 1 | Asset-Liability Ratio | 0.126 | Financial Status |
| 2 | Cash Flow Ratio | 0.108 | Financial Status |
| 3 | Number of Default Records | 0.095 | Cooperation History |
| 4 | On-Time Delivery Rate | 0.087 | Operational Capability |
| 5 | Current Ratio | 0.076 | Financial Status |
| 6 | Qualified Rate of Products | 0.068 | Operational Capability |
| 7 | Net Profit Margin on Sales | 0.059 | Financial Status |
| 8 | Cooperation Duration | 0.053 | Cooperation History |
| 9 | Accounts Receivable Turnover | 0.048 | Financial Status |
| 10 | Industry Competition Intensity | 0.039 | External Environment |

7. Industry adaptability test: Verify the performance of the model in samples from the manufacturing, retail, and logistics industries.

### B. EMPIRICAL RESULTS AND ANALYSIS

#### 1) Analysis of Data Enhancement Effect

The improved WGANGP model generated 1182 synthesized negative samples, which were fused with the original 341 negative samples to obtain 1523 negative samples, consistent with the number of positive samples, forming a balanced dataset. Use the following indicators to evaluate the quality of generated data:

Similarity of feature distribution: The difference in mean and variance between the generated data and the real data is less than 5%, indicating a high degree of consistency in feature distribution;

Visual analysis: Through tSNE dimensionality reduction visualization, the clustering distribution overlap between generated data and real data reaches 89.3%;

Classification accuracy: After mixing the generated data with real data and using the XGBoost model for classification, the accuracy is only 52.7%, indicating that the generated data is difficult to distinguish from real data.

The results indicate that the synthesized data generated by the improved WGANGP model has excellent quality, can effectively supplement negative samples, and balance data distribution.

Comparative analysis of model performance

The performance evaluation results of the three models are shown in the following table 4:

Result analysis:

The performance indicators of the experimental group model were significantly better than those of the control group model, with an accuracy of 94.7%, F1 value of 93.9%, and AUC value of 0.976, indicating that the fusion of ensemble learning and GAN can effectively improve the accuracy of credit risk warning; The performance of comparison group 1 (Stacking integrated model) is better than that of comparison group 2 (traditional single model), which verifies the collaborative optimization effect of integrated learning; The recall rate of the experimental group model reached 93.2%, significantly higher than the 86.3% of the control group 1 and the 68.3% 80.1% of the control group 2, indicating that GAN data augmentation can effectively improve the model's ability to identify negative samples and reduce the risk of "missed judgments".

#### 2) Results of Model Stability Test

The stability of the model was tested using 5-fold cross validation and different dataset partitioning ratios. The results are as follows:

**TABLE 5.** Model Stability Test Results

| Stability Test Method | Model Type | Average Accuracy (%) | Accuracy Standard Deviation (%) |
|---|---|---|---|
| 5-Fold Cross Validation | Experimental Group | 94.3 | 0.87 |
| | Control Group 1 | 88.1 | 1.23 |
| | Control Group 2 (XGBoost) | 81.9 | 1.56 |
| Dataset Split Ratio 6:4 | Experimental Group | 94.1 | 0.92 |
| | Control Group 1 | 87.8 | 1.31 |
| | Control Group 2 (XGBoost) | 81.5 | 1.62 |
| Dataset Split Ratio 8:2 | Experimental Group | 95.0 | 0.78 |
| | Control Group 1 | 88.7 | 1.15 |
| | Control Group 2 (XGBoost) | 82.7 | 1.48 |

Result analysis: The accuracy standard deviation and coefficient of variation of the experimental group model were significantly lower than those of the control group model, indicating that the ensemble learning and GAN fusion model has stronger stability and generalization ability, and is less affected by the data partitioning ratio.

### C. EMPIRICAL CONCLUSION

The empirical research results indicate that:

1. Improving the WGANGP model can generate high-quality synthesized negative samples, with feature distributions highly consistent with real data, effectively solving the problem of imbalanced supply chain credit risk data;

2. The credit risk early warning performance of the integrated learning and GAN integration model is significantly better than that of the non integrated GAN integration model and the traditional single model, with the accuracy rate of 94.7%, recall rate of 93.2%, and F1 value of 93.9%;

3. The model has good stability and generalization ability, and exhibits excellent performance in different industries and data partitioning ratios;

4. Asset liability ratio, cash flow ratio, number of default records, and on-time delivery rate are the core characteristics that affect the credit risk of supply chain partners;

5. The practical application effect of the model is good, which can help enterprises accurately identify high-risk partners and reduce default losses.

## VI. MODEL OPTIMIZATION AND APPLICATION STRATEGIES

### A. MODEL OPTIMIZATION STRATEGY

Data level optimization

1. Expand data coverage: Increase the industry coverage of the sample, including supply chain enterprises in more industries such as agriculture and services, to enhance the generalization ability of the model;

2. Improve data timeliness: Establish a real-time data collection mechanism, update the financial, operational, and cooperation data of the enterprise in a timely manner, and ensure the dynamic warning capability of the model;

3. Enrich data dimensions: Introduce emerging dimensions such as corporate social responsibility and public opinion information to further improve the credit risk evaluation system.

Technical optimization

1. Model structure optimization: Attempt to use more advanced generative models (such as StyleGAN, VAEGAN) to replace WGANGP and improve the quality of synthesized data;

2. Algorithm fusion innovation: Integrating deep learning algorithms (such as CNN, LSTM) as the base model to capture the temporal and spatial features of data;

3. Parameter adaptive adjustment: Introducing reinforcement learning techniques to achieve adaptive adjustment of model parameters and enhance the model's adaptability to different scenarios;

4. Model lightweighting: Pruning and optimizing the model to reduce computational complexity and improve real-time response speed.

Application level optimization

1. Personalized warning threshold: Based on the risk preferences and cooperation strategies of the enterprise, set personalized warning thresholds to balance the risks of "misjudgment" and "missed judgment";

2. Risk level classification: Divide credit risk into three levels: high, medium, and low, and provide differentiated risk response recommendations for different levels;

3. Visual interface development: Develop an intuitive visual warning interface that displays risk levels, core influencing factors, warning criteria, and improves the usability of the model.

#### 1) Model Application Strategy

Scenario specific application strategy

1. Partner selection scenario: Apply the model to new partner admission evaluation, prioritize low-risk enterprises to establish cooperative relationships, and reduce initial cooperation risks;

2. Collaboration process monitoring scenario: Real time credit risk monitoring of existing partners, regular generation of risk assessment reports, and timely detection of risk hazards;

3. Scenario for adjusting cooperation strategy: Based on the model warning results, measures such as reducing cooperation scale, strengthening payment constraints, and increasing quality inspections are taken for high-risk partners to reduce cooperation risks;

4. Emergency response scenario: When the model issues a high-risk warning, the emergency response mechanism is activated to quickly find alternative partners and avoid supply chain disruptions.

#### 2) Scale based Enterprise Application Strategy

1. Large enterprises: comprehensively promote the application of models, integrate supply chain management systems and model warning functions, and achieve full process and automated management of credit risks;

2. Small and medium-sized enterprises: adopting a lightweight version of the model, focusing on credit risk warning for core partners, and reducing application costs;

3. Start up enterprises: Prioritize the application of models to evaluate the credit risk of key suppliers and customers, ensuring the stable operation of core businesses.

Implementation of safeguard measures

1. Organizational support: Establish a dedicated supply chain credit risk management team responsible for the application, maintenance, and optimization of the model;

2. Technical support: invest necessary technical resources, build a data collection and model operation platform, and ensure the stable operation of the model;

3. Personnel training: Conduct specialized training on model application to enhance the model usage and risk identification abilities of supply chain management personnel;

4. Institutional guarantee: Establish rules and regulations for supply chain credit risk management, clarify the model application process, division of responsibilities, and incentive mechanisms.

## VII. RESEARCH CONCLUSION

This article focuses on developing a credit risk warning model for manufacturing entities by integrating ensemble learning and GAN to address the data imbalances inherent in processing and industrial sectors. By leveraging this hybrid framework, the research captures the nuanced financial volatility of upstream and downstream partners, ensuring more resilient risk detection across the broader production landscape. Through a series of studies such as theoretical review, system construction, model design, and empirical verification, the following conclusions are mainly drawn:

(1) A supply chain credit risk evaluation system has been constructed, which includes four primary dimensions of "financial condition operational capability cooperation history external environment" and 32 secondary indicators. The financial condition dimension has the highest weight (0.50), and the core indicators include asset liability ratio, cash flow ratio, current ratio, etc., providing scientific indicator support for credit risk warning;

(2) A fusion model architecture of "Improved WGANGP Data Enhancement+Stacking Ensemble Learning" was proposed. The improved WGANGP model optimizes the network structure, loss function, and training strategy to generate high-quality synthesized data, effectively solving the problem of data imbalance;

(3) The empirical results show that the accuracy rate of credit risk early warning of the integrated model is 94.7%, the recall rate is 93.2%, and the F1 value is 93.9%, which is significantly better than the integrated model without GAN and the traditional single model, and has good stability and industry adaptability;

(4) The asset liability ratio, cash flow ratio, number of default records, and on-time delivery rate are the core

characteristics that affect the credit risk of supply chain partners. The debt paying ability and cooperation history of enterprises are the key factors of credit risk warning;

(5) The practical application effect of the model is good, which can help enterprises accurately identify high-risk partners, reduce default losses by 37.6%, and provide operational technical solutions for supply chain credit risk management.

## AUTHOR CONTRIBUTIONS

Lingxian Zhu designed, collected and analyzed the data, and drafted the manuscript. Lingxian Zhu conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Lingxian Zhu participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## CONFLICTS OF INTEREST

The author declares no conflict of interest.

## FUNDING

This research received no external funding.

## ACKNOWLEDGEMENTS

Not applicable.

## REFERENCES

[1] M. Baraheni, H. B. Soudmand, S. Amini, et al., “Stacked generalization ensemble learning strategy for multi-variate prediction of delamination and maximum thrust force in composite drilling,” Journal of Composite Materials, vol. 58, no. 30, pp. 3113–3138, 2024, doi: 10.1177/00219983241289494.

[2] M. Atemkeng, V. Osanyindoro, R. Rockefeller, et al., “Ensemble learning and deep learning-based defect detection in power generation plants,” Journal of Intelligent Systems, pp. 33(1), 2024, doi: 10.1515/JISYS-2023-0283.

[3] K. Singh, S. Dhawan, and N. Bali, “An Ensemble Learning Hybrid Recommendation System Using Content-Based, Collaborative Filtering, Supervised Learning and Boosting Algorithms,” Automatic Control and Computer Sciences, vol. 58, no. 5, pp. 491–505, 2024, doi: 10.3103/S0146411624700615.

[4] C. Fu, Y. Jiang, and A. Tian, “Quantitative prediction of soil chromium content using laboratory-based visible and near-infrared spectroscopy with different ensemble learning models,” Advances in Space Research, vol. 74, no. 10, pp. 4705–4720, 2024, doi: 10.1016/J.ASR.2024.07.056.

[5] P. C and A. Y P V, “Deep hybrid architecture with stacked ensemble learning for binary classification of retinal disease,” Results in Engineering, vol. 24, Art. no. 103219, 2024, doi: 10.1016/J.RINENG.2024.103219.

[6] H. Lv, Z. Wu, X. Zheng, et al., “Multi-attribute diagnosis of urban flood-bearing bodies based on integrated learning with Stacking-GPR-QPSO coupling,” Journal of Hydrology, vol. 645, no. PA, Art. no. 132222, 2024, doi: 10.1016/J.JHYDROL.2024.132222.

[7] X. Li, G. Liang, L. Wang, et al., “Identifying the spatial pattern and driving factors of nitrate in groundwater using a novel framework of interpretable stacking ensemble learning,” Environmental Geochemistry and Health, vol. 46, no. 11, pp. 482, 2024, doi: 10.1007/S10653-024-02201-1.

[8] A. Singh P, B. Chaurasia K, and A. Tripathi, “Stacking Enabled Ensemble Learning Based Intrusion Detection Scheme (SELIDS) for IoV,” SN Computer Science, vol. 5, no. 8, pp. 1000, 2024, doi: 10.1007/S42979-024-03376-1.

[9] M. Tahmouresi S, M. Niksokhan H, and A. Ehsani H, “Enhancing spatial resolution of satellite soil moisture data through stacking ensemble learning techniques,” Scientific Reports, vol. 14, no. 1, Art. no. 25454, 2024, doi: 10.1038/S41598-024-77050-0.

[10] H. Liu, W. Zhang, Z. Cheng, et al., “Investigating the contributing factors of urban crash levels: A novel stacking integrated learning framework,” Applied Geography, vol. 173, Art. no. 103440, 2024, doi: 10.1016/J.APGEOG.2024.103440.

[11] F. Zhang, Y. Liu, X. Yu, et al., “Towards facial micro-expression detection and classification using modified multimodal ensemble learning approach,” Information Fusion, vol. 115, Art. no. 102735, 2025, doi: 10.1016/J.INFFUS.2024.102735.

[12] M. Saraswat and A. Dubey K, “Early stage brain tumor prediction using dilated and Attention-based ensemble learning with enhanced Artificial rabbit optimization for brain data,” Biomedical Signal Processing and Control, vol. 100, no. PC, Art. no. 107033, 2025, doi: 10.1016/J.BSPC.2024.107033.

[13] H. Zhang, K. Kang, C. Wang, et al., “Cross-variety seed vigor detection using new spectral analysis techniques and ensemble learning methods,” Journal of Food Composition and Analysis, vol. 136, Art. no. 106845, 2024, doi: 10.1016/J.JFCA.2024.106845.

[14] D. Qi and T. Liu, “VotePLMs-AFP: Identification of antifreeze proteins using transformer-embedding features and ensemble learning,” BBA - General Subjects, vol. 1868, no. 12, Art. no. 130721, 2024, doi: 10.1016/J.BBAGEN.2024.130721.

[15] S. Kamro, M. Rafiee, and S. Mirjalili, “Metaheuristic-driven space partitioning and ensemble learning for im-balanced classification,” Applied Soft Computing, vol. 167, no. PB, Art. no. 112278, 2024, doi: 10.1016/J.ASOC.2024.112278.

[16] F. Mengistu, B. Hailu T, T. Abera A, et al., “Improving the accuracy of honey bee forage class mapping using ensemble learning and multi-source satellite data in Google Earth Engine,” Scientific African, vol. 26, Art. no. e02415, 2024, doi: 10.1016/J.SCIAF.2024.E02415.

[17] J. Fu, X. Sun, Z. Ma, et al., “Efficient Optimization: Unveiling the Application of Ensemble Learning Combined with the CMA-ES Algorithm in Hydraulic Fracturing Design,” Processes, vol. 12, no. 10, pp. 2299, 2024, doi: 10.3390/PR12102299.

[18] F. Xing, N. Li, B. Zhao, et al., “An Investigation into the Susceptibility to Landslides Using Integrated Learning and Bayesian Optimization: A Case Study of Xichang City,” Sustainability, vol. 16, no. 20, pp. 9085, 2024, doi: 10.3390/SU16209085.

[19] Y. Jia, H. Wang, Z. Yuan, et al., “Biomedical relation extraction method based on ensemble learning and attention mechanism,” BMC Bioinformatics, vol. 25, no. 1, pp. 333, 2024, doi: 10.1186/S12859-024-05951-Y.

[20] B. Çetinus, S. Oyucu, A. Aksöz, et al., “The Role of Machine Learning in Enhancing Battery Management for Drone Operations: A Focus on SoH Prediction Using Ensemble Learning Techniques,” Batteries, vol. 10, no. 10, pp. 371, 2024, doi: 10.3390/BATTERIES10100371.