Research on SMEs' Credit Risk Assessment Based on Blockchain-driven Supply Chain Finance

Ping Xiao*, Li Tan and Mad Ithnin Salleh

Abstract—**The investigation into the significance of credit risk assessment for small and medium-sized enterprises (SMEs) within the context of supply chain finance (SCF) facilitated by blockchain technology presents both opportunities and challenges. The implementation of blockchain has the potential to mitigate information asymmetry. However, it simultaneously complicates the process of credit risk evaluation. To this regard, this study developed a constructed credit risk assessment indicator system and formulated novel forecasting models that employ an imbalanced sampling strategy based on machine learning algorithms, including the classification tree, the bagging, the AdaBoost, the random forest. These models were subsequently applied to predict credit risk of SMEs in China. Additionally, the study examined the selection of characteristic indicators in the prediction model, which were used to calculate and compare the strengths and weaknesses of credit risk assessment models, as well as their predictive capabilities. The empirical result indicate that the AdaBoost algorithm demonstrated the most effective predictive performance. This study addresses the issue of inaccurate credit risk assessment for SMEs and offers valuable management insights for financial institutions assessing the credit risk of SMEs in practice. Furthermore, it provides significant reference value for the development of a blockchain-driven SCF model.**

Index Terms—**Credit risk assessment, SMEs, Blockchain, Supply chain finance, Machine learning method**

I. INTRODUCTION

BLOCKCHAIN was first introduced in China in 2016, facilitating the interconnection of data and enabling the visualization of all transaction information, thereby ensuring transparency among all stakeholders [1]. Concurrently, digital invoices utilizing blockchain technology have the potential to enhance the security and traceability of transactions within the bill market [2].

Manuscript received Jun 20, 2024; revised Feb 18, 2025.

This work was supported in part by the Scientific research project of the Excellent Youth Project of the Hunan Provincial Department of Education (24B0824)/(24B0819) and Natural Science Foundation of Hunan Province (2173) .

Ping Xiao is a lecturer at the College of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi, Hunan, 417000, China. (Corresponding author, email: pingxiao_88@163.com).

Li Tan is a lecturer at the National Taxinstitute of Sta, Changsha, Hunan, 410000, China. (email:710402858@qq.com).

Mad Ithnin Salleh is a professor at the Faculty of Management and Economics, Universiti Pendidikan Sultan Idris, Tanjong Malim, Malaysia. (email: mad.ithnin@fpe.upsi.edu.my).

The implementation of blockchain technology has significantly contributed to addressing the issue of information asymmetry in SCF, alleviating the financing challenges faced by SMEs, and offering a platform for capital providers to manage risk. However, it also presents challenges in the assessment of credit risk [3]. There exists a paucity of research examining the effects of blockchain platform integration on the credit risk of SMEs [4], and there is a lack of assessment systems that adequately align with the characteristics of blockchain-driven SCF operations and effectively identify associated risks [5].

This study proposes a credit risk assessment framework and process, as illustrated in Fig. 1, which is applicable to credit risk evaluation scenarios within the blockchain-driven SCF model. The utilization of blockchain is intended to enhance the precision of credit risk assessments for SMEs and to furnish financial institutions with innovative tools for evaluating credit risk. Furthermore, this research employs actual data for empirical validation, with findings that can be directly applied to data mining practices in the industry and that hold significant relevance for enterprise credit risk management.

This research makes several important contributions to theory and practice. To the best of our knowledge, it is a notable study that identifies operational information, trade relationships, and sustainability as fundamental components for forecasting the credit risk of SMEs within a blockchain-driven SCF context, thereby providing theoretical underpinnings for the development of a credit risk assessment index system. Additionally, this study captures various indicators of critical information from both financial and non-financial perspectives, thereby enriching the credit risk evaluation index system. The conclusions drawn from this research motivate investors to make financing decisions that should consider not only the financial data of SMEs but also their reliability and sustainability. The application of machine learning techniques to construct credit risk assessment models significantly enhances the predictive performance of these models. Practically, this research offers management insights for financial institutions in assessing the credit risk of SMEs and serves as a valuable reference for the advancement of blockchain-driven SCF models.

The structure of the article is arranged as follows. Section 2 reviews the relevant research on credit risk based on blockchain-driven SCF, establishes the credit risk for employing blockchain platform to SCF, and the methods for credit risk assessment. Section 3 proposes the methodology used in this article. Section 4 clarifies the sample collection process. Section 5 presents the empirical results. Section 6 provides the research conclusion and implications.

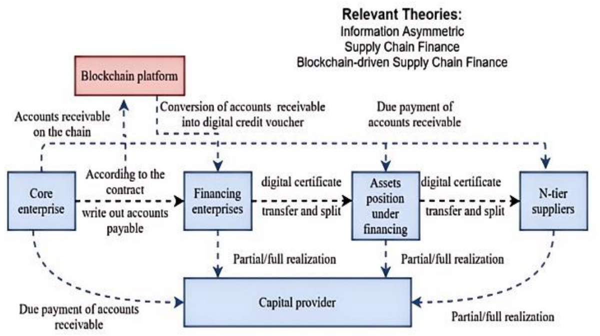

Fig. 1. Framework for Blockchain-driven SCF

II. LITERATURE REVIEW

A. Factors Affecting Credit Risk

The challenges faced by SMEs ability meet legal obligations[6]. Corporate financing risk is closely linked to a company's profitability, with firms that generate higher profits typically credit risk [7]. Wang et al. [8] combined various financial variables to predict the credit risk of SMEs in SCF, focusing particularly on their capital capacity, management proficiency, profitability, growth potential, and solvency. Notably, the profitability, operational efficiency, and solvency of financing enterprises can contribute to credit risk [9].

The credit status of core enterprises is crucial in the credit evaluation of the entire supply chain, as core enterprises typically provide guarantees for the creditworthiness of SMEs. Furthermore, an increase in the number of core enterprises with strong credit ratings will enhance the overall system's immunity to credit risk [10]. Tian, Zhuang, and Zhao [11] conducted regression analysis and tested logistic models, retaining the solvency and profitability factors of core enterprises. Their findings indicate that the credit level of SMEs is significantly influenced by the return on net assets and the operating net interest rate of core enterprises.

The assets position under financing includes the accounts receivable turnover rate and inventory turnover rate [12]. Upstream enterprises obtain short-term credit loans from financial institutions through accounts receivable financing. When SMEs are unable to repay loans, they may transfer assets position under financing to generate cash flow [13]. This approach not only addresses the short-term funding needs of SMEs, and promotes their healthy and stable development, but it also supports the continuous and efficient operation of the entire supply chain while reducing credit risk [14].

Blockchain technology can effectively address risk management issues in the financial market, reshape the credit framework of the SCF market, reduce business costs, and mitigate moral hazard [15]. Within a blockchain platform, the information flow between principals and agents is transparent and accessible, which can significantly reduce credit risk [16]. This technology has successfully resolved the information asymmetry problem that has long plagued SCF [17]. Furthermore, blockchain technology not only facilitates the effective regulation of liquidity within the SCF process but also aids the financial ecosystem in eliminating fraud and defaults [18]. Additionally, the application of blockchain technology in SCF risk management can enhance the stability of relationships among supply chain enterprises and reduce potential risks in the SCF market[19].

The credit risk of SMEs is influenced not only by internal factors, but also by SCF factors. These include the financial and non-financial status of the enterprise, supply chain operation status, and features of the pledge object [13]. Wicaksana [20] suggested that in addition to economic risks, social and environmental risks should also be considered as emerging risk categories. Wang [21] integrated the LR model with the ML model to identify potential risks. A robust ML-LRA architecture can effectively manage the increasing volume of immutable financial data generated in records. SCF presents a non-linear risk, and the integration of digital supply chain technology with ML-LR technology to enhance human decision-making ability [22].

The existing literature has examined the reasons that affect the credit risk of SCF, mainly analyzing the sources of risk from the perspective of enterprises involved in the blockchain-driven SCF model and the characteristics of this model. This analysis offers a theoretical framework for selecting credit risk assessment indicators for the current study.

B. Index construction of credit risk assessment model

To establish and implement a credit risk assessment index system for SMEs, researchers have conducted extensive studies. Yi and Guo [23] developed a risk assessment indicator system categorized into areas: risk associated with core enterprises, risks related to accounts receivable status, risk concerning financing companies, and risks linked to the operating environment. Zhang et al. [24] identified 19 indicators, including profitability, solvency, growth potential, and operational efficiency, to analyze the factors influencing green credit risk. This study summarized the performance environment and environmental quality, leading to the establishment of a green credit risk assessment indicator system. Kuang et al. [25] created four standard layers by comprehensively analyzing the risks encountered throughout the entire supply chain, which include applicant qualifications, counterparty qualifications, asset positions under financing, and supply chain operations. Additionally, 14 sub-indicators were identified, such as accounts receivable characteristics and performance metrics. Kohler, Bager, and Pizzol [26] examined 16 cases of blockchain-based technology and voluntary sustainability standards, evaluating them against 12 sustainability-related criteria. The findings indicate that the relationship between blockchain-based technologies and sustainability standards can be characterized by coexistence, collaboration, and entagonism. Xia et al. [27] selected risk indicators for manufacturing SMEs from four categories: an overview of financing enterprises, assets positions under financing, core enterprises, and supply chain operations.

The literature presented the sample selection and construction methods for credit risk assessment indicators, providing a technical reference for developing these indicators in this research.

C. Credit risk assessment methods

Some researchers have employed statistical and data mining techniques to assess risks, including logical models, support vector machine models, and neural network models [28]. Numerous publications have examined risk assessment models, such as the KMV model [29]. Wang et al. [8] applied various machine learning techniques in credit risk assessment, including Logistic Regression (LR), Classification and Regression Tree (CART), Neural Network (NN), and Support Vector Machine (SVM). Sang [51] discovered that the classification accuracy of the BP neural network method, when enhanced by genetic algorithms, is relatively higher than that of the support vector method (SVM). Guo [30] showed that the BP neural network algorithm outperforms the LR. Wu [31] found that the GA-BPNN accumulates results more rapidly than traditional BPNN algorithms.

Shen et al. [32] demonstrate that the proposed deep learning integration model is more effective in addressing the issue of unbalanced credit risk assessment, particularly in processing uneven credit data. To enhance the accuracy of credit risk predictions for SMEs, Zhu et al [13] combined two classical integrated ML methods: Random Subspace (RS) and Multi-Boosting. Compared to other techniques, Extreme Gradient Boosting (XGBoost) performs better when the dataset contains missing values [33]. Unlike random forests, XGBoost significantly improves accuracy by iteratively fitting the final value using residuals multiple times [34]. Zhang et al [35] proposed a credit risk assessment method based on RF-SMOTE-XGBoost, which emphasizes the importance of features through the accuracy and robustness of random forests in large-scale data classification challenges. Due to its unique random tree, random forest technology outperforms standard regression trees in mitigating bias, handling missing data, and managing turbulent inputs [36].

Previous studies have primarily employed traditional statistical analysis, data mining, and machine learning techniques to predict credit risk. Generally, data mining and machine learning methods outperform traditional statistical analysis. However, these studies have not thoroughly examined the impact of blockchain integration on the supply chain. Consequently, this research summarizes the factors influencing credit risk within the blockchain-driven SCF model, develops scientific credit risk assessment indicators, and utilizes machine learning techniques to predict credit risk.

III. RESEARCH METHOD

This section introduces four classic machine learning algorithms and factor analysis methods, which process sample indicators through factor analysis, and then compare the four prediction models constructed by the machine learning method.

A. Factor Analysis

Factor analysis is a multivariate statistical method that begins with the examination of the internal relationships among variables. It consolidates several variables with complex interrelationships into a smaller number of underlying factors [11]. The purpose of the Kaiser-Meyer-Olkin (KMO) statistic is to assess the degree of partial correlation among variables and to quantify the suitability of component analysis. The KMO statistic ranges from 0 to 1, evaluating both the strength of simple correlations and the extent of partial correlations that exist between the variables [37]. A KMO value greater than 0.8 indicates an excellent effect; a value greater than 0.6 is considered acceptable; and a value less than 0.5 suggests that factor analysis is not appropriate [11]. The following formula is used to calculate the KMO statistic:

$$KMO = \frac{\sum\sum_{i\neq j} r_{ij}^2}{\sum\sum_{i\neq j} r_{ij}^2 + \sum\sum_{i\neq j} \beta_{ij}^2}$$

Where, r and β are correlation coefficients and partial correlation coefficients respectively.

B. Classification Tree

The classification tree method is a type of classifier that utilizes a tree structure to organize the sample set. It begins at the root and progresses through branches and nodes until it reaches the leaves. This approach is a form of machine learning. As a valid strategy to enhance the prediction accuracy of the maximum likelihood method, researchers are increasingly focusing on optimizing this method, including techniques such as boosting [38].

C. Bagging Algorithm

Bagging, short for bootstrap aggregating, involves generating multiple distinct training sets through bootstrapping sampling. Models are then established using these training sets, resulting in a series of base classifiers [39]. Each classifier exhibits varying prediction performance on the same test set due to their derivation from different training samples. When the training sample set is small and data fluctuations minimally impact the model parameters, the bagging algorithm demonstrates significant improvement. Utilizing bagging techniques for ensemble prediction is particularly beneficial for large datasets [3].

D. AdaBoost Algorithm

Multiple models hold equal status in the prediction voting process, and the variations in prediction accuracy among different models are not taken into account [41]. The AdaBoost algorithm has been refined in these two areas. Compared to the bagging algorithm, the AdaBoost algorithm demonstrates greater sophistication, particularly in scenarios involving unbalanced data distributions [40].

E. Random Forest

The Random Forest algorithm constructs a forest in a stochastic manner, comprising numerous decision trees that exhibit high predictive accuracy and low correlation, or even no correlation, with one another. This ensemble of trees forms a robust predictive model [15]. To reduce the correlation among individual decision trees and enhance classification accuracy, the Random Forest method randomly selects features at each node during the tree-building process. As each tree grows rapidly, the classification process of the Random

Forest is also swift, facilitating easy parallelization [43].

F. Model Evaluation Criterion

It is crucial to utilize appropriate standards for assessing the performance of the model. In credit risk management, accurately and effectively identifying potential defaulting customers is a fundamental aspect of model evaluation. The performance evaluation metrics for the classifier primarily include the confusion matrix, accuracy, sensitivity, specificity, AUC, and KS [44]. Consequently, this research selected the following indicators for a comprehensive evaluation of the model. In classification problems, a confusion matrix is a widely used tool for evaluating model performance. It is constructed by calculating the number of samples accurately classified by the model, divided by the total number of samples. This metric indicates the percentage of samples correctly classified by the model out of all samples [13]. In this context, when the true value of a sample is positive, and the model correctly identifies it as a true positive (TP). Conversely, if the true value of the sample is positive but the model incorrectly classifies it as negative, this is known as a false negative (FN), which is statistically recognized as a type II error. On the other hand, when the true value of a sample is negative, but the model incorrectly classifies it as positive, this is termed a false positive (FP), which is statistically referred to as a type I error.

TABLE I
CONFUSION MATRIX

	Forecast category (Positive)	Forecast category (Negative)
Real category (Positive)	TP	FN
Real category (Negative)	FP	TN

To facilitate the discrimination of credit risk of SMEs, this research improved the confusion matrix according to the credit risk assessment model, as shown in Table II.

TABLE II
CONFUSION MATRIX

	Forecast category (risk)	Forecast category (non-risk)
Real category (risk)	TR	FR
Real category (non-risk)	FN	TN

True Risk (TR) represents actual credit risk, while Non-True Risk (N_{TR}) indicates the number of samples that exhibit both actual credit risk and predicted credit risk. False Risk (FR) refers to instance without credit risk, and Non-False Risk (N_{FR}) denotes the number of samples that possess credit risk but are incorrectly predicted as non-credit risk. False Negative (FN) signifies instance of credit risk that are misclassified, and Non-False Negative (N_{FN}) represents the number of samples that are inaccurately predicted to have credit risk when they do not. True Negative (TN) indicates instances of non-credit risk, and Non-True Negative (NTN) refers to the number of samples that are accurately identified as non-credit risk.

To evaluate the predictive performance of the four machine learning methods, we applied eight assessment criteria based on the table: Accuracy, Sensitivity, Recall Rate, F1 Score, Type I Error Rate, Type II Error Rate, Area Under the Curve (AUC), and Kolmogorov-Smirnov (KS) statistic, as follows:

$$P = \frac{N_{TR} + N_{TN}}{N_{TR} + N_{FN} + N_{FR} + N_{TN}} \times 100\% \tag{1}$$

$$S_e = \frac{N_{TR}}{N_{TR} + N_{FR}} \times 100\% \tag{2}$$

$$S_p = \frac{N_{TN}}{N_{FN} + N_{TN}} \times 100\% \tag{3}$$

$$\text{Type I error} = FR/(TR+FR) \tag{4}$$

$$\text{Type II error} = FN/(FN + TN) \tag{5}$$

$$F1 = \frac{2 \times TR}{2 \times TR + FN + FR} \tag{6}$$

$$TPR = \frac{TR}{TR + FR} \tag{7}$$

$$FPR = \frac{TN}{FN + TN} \tag{8}$$

$$AUC = \frac{1}{2}\left(\frac{TR}{TR+FR} + \frac{TN}{FN+TN}\right) \tag{9}$$

$$KS = \max(TPR - FPR) \tag{10}$$

Accuracy is the evaluation of the overall classification and prediction ability of the classifier [48], which is given in Eq.(1). Sensitivity S_e represents the accuracy of the model in predicting positive samples, given in Eq. (2). The higher the index value, the better the prediction ability. The research represents the accuracy of predicting enterprises with credit risk. Specificity S_p represents the prediction accuracy of the model for negative samples [49], given in Eq. (3). In this research, it represented the accuracy of enterprises without credit risk. The higher the index value, the better the prediction ability. This research also adopts the Type I error given in Eq. (4) represents the ratio of positive applicants being incorrectly forecasted as negative. Type II error given in Eq. (5) represents the ratio of negative applicants being incorrectly forecasted as positive [37]. F1 value is the amicable mean of accuracy rate and recall rate, which is given in Eq.(6). In this research, the curve is abscissa and ordinate, which are false positive rate (FPR) given in Eq. (7) and true positive rate (TPR) given in Eq. (8). The AUC value represents classifier's ability to avoid false classification [50], which given in Eq. (9). Kolmogorov Smirnov (KS) index is usually used to evaluate the discrimination of the model and is generally used for non-parametric test to test whether the data obey a certain distribution or whether the overall distribution of two samples is consistent. When calculating KS, the samples were sorted from small to large according to the predicted risky probability, and then the cumulative proportion of risky SMEs and non-risky SMEs under each default probability [47]. The maximum difference between the cumulative proportion of non-risky samples and risky samples is KS, given in Eq. (10).

When evaluating credit risk, the key indicators of the model vary across different business scenarios, or multiple indicators need to be taken into consideration. When comparing various models, if only paying attention to one of the indicators, it is easy to make the selected model not applicable in other business scenarios. Therefore, this research constructed the evaluation index system of the credit risk evaluation model to appraise the advantages and disadvantages of the model, as shown in Table III.

TABLE III
EVALUATION OF CREDIT RISK ASSESSMENT MODEL

Primary Index	Secondary Index	Index Description
Accuracy	S_e	The preciseness of the model in predicting enterprises with credit risk
	S_p	The preciseness of the model in predicting enterprises without credit risk
	F1	The indicators of accuracy and recall
Discrimination	KS	The model distinguishes between enterprises with credit risk and enterprises without credit risk.
Stability	AUC value	Classifier's ability to avoid false classification
Interpretability	Importance of each variable	The order of importance of each feature and the odds ratio of default events

The above credit risk assessment model regards the performance and interpretability of the model and can evaluate the model in all aspects. Through this system, models appropriate for different scenes can be selected.

G. The Whole Experimental Step

Based on the analysis presented above, the forecasting process of SMEs in blockchain-driven SCF, utilizing all 27 potential predictors can be divided into four steps.

Step 1: Building a Knowledge Database for forecasting SMEs' credit risk in blockchain-driven SCF, We collected data on 27 potential predictors across four categories: financial information, operational information, relationship strength within the supply chain, macroeconomic environment, and sustainability.

Step 2: Select key predictors of credit risk for SMEs in blockchain-driven SCF. We compiled 27 potential predictors and applied four established machine learning algorithms to predict SMEs' credit risk within the context of blockchain-driven SCF. The predicted results were analyzed to determine the key predictors.

Step 3: Utilize factor analysis to extract common factors from the 27 indicators. Having too many indicators can lead to multicollinearity, which will provide a source of indicators for establishing a credit risk assessment model.

Step 4: Forecasting models that utilize an imbalance sampling strategy within a machine learning algorithm to predict the credit risk based on 11 factors. We then compare these results with the original predictions, evaluating them based on accuracy, discrimination, stability, and interpretability.

IV. DATASET INFORMATION AND PREDICTORS

A. Data Resource

In this study, as blockchain technology began to gain popularity in China in 2016, we selected Chinese A-share listed companies from 2017 to 2021 as our sample. This included non-financial listed companies that were designated as Special Treatment (ST) due to financial reasons, representing the sample with credit risk, and non-ST enterprises, which served as the sample without credit risk [6]. Furthermore, we study employed statistical analysis and machine learning methods in our research, ensuring that the findings are both quantitative and based on secondary data.

The research initially focused on the manufacturing industry, followed by an examination of blockchain businesses involved in the supply chain. Subsequently, cooperative SMEs and core enterprises from the blockchain platform, and the sample size was calculated. This constituted the fundamental process for determining the sample size. Consequently, the data for this study were sourced from the CSMAR Database and the financial statements of companies that are publicly traded. traded companies. Ultimately, 90 sample sizes were selected, comprising 56 SMEs, 8 core enterprises, and 26 blockchain platforms, covering the period form 2017 to 2021.

B. Data Collection Procedure

This research focused on the manufacturing sector to more effectively examine the factors that contribute to credit risk and to develop a model for assessing it. By concentrating on this sector, the researchers were able to mitigate errors arising from disparities present in other industries. The following is the approach that should be taken when selecting samples:

Selection of blockchain platform Samples: The report indicates that 211 listed companies engaged in blockchain development possess high overall qualifications, with medium and large-scale enterprises comprising over 90% of this group. Of more than three years, representing 92.42% of the total; among them, 50.24% have been publicly listed for more than 90%. 195 enterprises have been listed for more over ten years. These 211 companies span various industries, with the largest representation in information transmission, software, and information technology services (99 enterprises), manufacturing (57 enterprises), and finance (19 enterprises). Approximately 38% of theses listed companies utilize their current blockchain capabilities primarily to address internal business needs or to diversify their business lines. About 27% of the enterprises focus mainly on exporting technology, while another 35% export their technical capabilities externally while also fulfilling their own business requirements.

The blockchain platforms examined in this research primarily offered information technology services. Consequently, this study initially identified 99 blockchain enterprises that provide such services. From these 99 enterprises, we then identified those participating in the supply chain, revealing that a total of 26 blockchain platforms are involved with manufacturing enterprises in the supply chain.

The white paper on the development of blockchain financial applications resulted in the identification of 26 blockchain platform companies. These companies often engage in business with SMEs related to manufacturing sector. The SCF mode is supported by these companies. By reviewing the official websites of the 26 blockchain platform companies and collecting information on the SMEs with which they collaborate, a total of 56 SMEs were identified.

(1) Screening of SMEs. Out of the 56 SMEs, 11 were identified as socially targeted as ST enterprises, while the remaining 45 were classified as non-ST companies. Consequently, the research established a set of risky enterprises comprising 11 sample groups and a set of non-risky enterprises consisting of 45 sample groups.

(2) Screening of core enterprises. This research utilized financial reports from official websites and data from the CSMAR database to identify and screen the upstream and downstream enterprises that engage in business transactions with selected SMEs. Large enterprises that interact with the SMEs within the supply chain were designated as core enterprises, resulting in the selection of eight main board enterprises. This approach was adopted to more accurately reflect the characteristics of the supply chain. Each of the 56 SMEs engages in genuine commercial transactions with one of the eight large businesses, these eight large businesses exhibit characteristics that are typical of core firms within the supply chain.

C. *Tables of Variables*

This research validated 27 credit risk sub-assessment indices, which will be used as explanatory variables in the model construction. These indices were developed through the creation of a comprehensive credit risk assessment index. The methodology categorized each indicator into one of five distinct groups.

TABLE IV
INDEPENDENT VARIABLES INVOLVED IN CREDIT RISK ASSESSMENT

Primary indicators	Secondary indicators	Symbol	Tertiary indicators
Financing enterprises	Profitability	X_1	Net interest rate on total assets
		X_2	Operating profit ratio
	Growth Ability	X_3	Growth rate of total assets
		X_4	Net profit growth rate
	Operation Ability	X_5	Current asset turnover
		X_6	Turnover of assets
	Debt paying ability	X_7	Current ratio
		X_8	Quick ratio
		X_9	Cash ratio
		X_{10}	Asset–liability ratio
Core enterprises	Profitability	X_{11}	Net interest rate on total assets
		X_{12}	Operating profit ratio
	Operation ability	X_{13}	Turnover of assets
		X_{14}	Credit status
	Debt paying ability	X_{15}	Quick ratio
		X_{16}	Asset–liability ratio
Assets position under financing		X_{17}	Receivable turnover ratio
		X_{18}	Inventory turnover ratio
Blockchain platform	Profitability	X_{19}	Operating profit ratio
		X_{20}	Main revenue growth rate
	Operation Ability	X_{21}	Turnover of assets
		X_{22}	Credit status
	Debt-paying Ability	X_{23}	Quick ratio
		X_{24}	Asset–liability ratio
Supply chain operation	Strength	X_{25}	Relationship strength of supply chain
	Macro	X_{26}	Macro environment
	Sustainability	X_{27}	Sustainability

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. *The Result of Factor Analysis*

This research employed factor analysis to identify underlying factors from the 27 indicators presented in Table 4, as an excessive number of indicators may result in multicollinearity. This approach aims to furnish a foundational source of indicators for the development of the credit risk assessment model.

B. *Basic Principle of Factor Analysis*

The study utilized factor analysis to identify the key components among 27 distinct variables that underwent principal component analysis. Through this process, the 27 indicators were consolidated into several primary factors, which exhibited no correlation with one another.

TABLE V
KMO AND BARTLET

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.554
Bartlett's Test of Sphericity	Approx. Chi-Square	3080.656
	df	351
	Sig.	.000

The KMO statistic presented in Table 5 is 0.554, indicating that factor analysis is appropriate. The original hypothesis is rejected due to a significance level of less than 0.05, suggesting a significant correlation among the variables. This finding implies that the selected data are suitable for Principal Component Analysis (PCA).

C. Identification Common Factors

This study employed the principal component analysis technique, which facilitates the conversion of pertinent variables into non-relevant variables via coordinate transformation, thereby achieving the objective of dimensionality reduction. To enhance the clarity of the primary factor, the maximum variance method was utilized for the rotation process.

TABLE VI
TOTAL VARIANCE EXPLAINED

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.302	12.229	12.229	3.302	12.229	12.229	3.178	11.771	11.771
2	2.990	11.074	23.303	2.990	11.074	23.303	2.420	8.964	20.735
3	2.392	8.860	32.163	2.392	8.860	32.163	2.359	8.736	29.471
4	2.040	7.554	39.718	2.040	7.554	39.718	1.850	6.852	36.323
5	1.774	6.569	46.286	1.774	6.569	46.286	1.731	6.412	42.735
6	1.358	5.029	51.315	1.358	5.029	51.315	1.634	6.053	48.789
7	1.264	4.682	55.997	1.264	4.682	55.997	1.387	5.138	53.927
8	1.169	4.328	60.325	1.169	4.328	60.325	1.266	4.688	58.615
9	1.132	4.191	64.516	1.132	4.191	64.516	1.254	4.644	63.259
10	1.034	3.831	68.347	1.034	3.831	68.347	1.220	4.517	67.776
11	1.013	3.754	72.100	1.013	3.754	72.100	1.168	4.324	72.100
12	.947	3.507	75.608						

Table VI presents the common factor variance for each component. The eigenvalue for the first component is 3.302, while the second component has an eigenvalue of 2.99, and the eleventh component has an eigenvalue of 1.013. The "Variance%" column indicates the proportion of total variance accounted for by each component, calculated as the percentage of each factor's eigenvalue relative to the total sum of eigenvalues. The "Cumulative%" column reflects the cumulative percentage of variance attributed to each factor, arranged in descending order. Collectively, the eigenvalues of the top eleven factors account for over 72.1% of the total variance, suggesting that these eleven factors can explain 72.1% of the variation observed in the original 27 variables.

A. Establishment of the Factor Load Matrix

The application of orthogonal rotation transformation enhances the explanatory power of common factors concerning the variables, thereby facilitating the identification of these common factors. Consequently, the factors can be designated based on their correlation with the original 27 indices following the orthogonal rotation of the eleven factors. In this study, the maximum variance method is employed for the factor rotation process.

TABLE VII
ROTATED COMPONENT MATRIX

	Component										
	1	2	3	4	5	6	7	8	9	10	11
X_1	.072	-.015	.020	.031	-.038	.016	-.025	.076	.034	.824	-.111
X_2	.097	-.048	-.043	.332	.184	-.090	.325	.372	.073	.164	-.059
X_3	.061	-.046	-.077	-.048	-.102	.085	.037	.830	-.014	.176	.025
X_4	.080	-.055	-.010	.003	.037	-.030	-.055	.100	.013	-.097	.808
X_5	-.004	.013	-.113	.025	.896	.109	.049	-.044	-.029	-.008	.035
X_6	-.051	-.008	-.073	-.025	.861	.101	-.185	-.079	.126	-.045	.004
X_7	.952	-.008	-.003	.042	-.017	.009	.031	-.008	-.006	.008	.007
X_8	.961	-.017	.049	.018	.007	.006	.001	.015	.050	-.005	-.003
X_9	.887	.013	.013	.008	.011	.085	-.009	.094	.074	-.058	-.031
X_{10}	-.679	.034	.044	.064	.105	.246	-.138	.099	.090	-.345	-.191
X_{11}	-.005	.966	.008	-.037	.016	-.054	-.015	-.005	.019	-.018	.004
X_{12}	-.037	.717	.179	-.097	-.046	.377	-.039	-.108	-.002	.073	.101
X_{13}	-.008	.636	-.234	.075	.031	-.483	-.015	.082	.023	-.029	-.007
X_{14}	.011	-.251	.642	-.131	.050	-.257	.027	.125	-.113	-.231	-.217
X_{15}	.078	.105	.828	.014	-.184	-.038	.014	-.144	.080	.012	.098
X_{16}	.006	-.180	-.857	.065	.072	.065	.017	-.011	.049	-.037	.077
X_{17}	.123	.021	-.287	.103	.133	.704	.078	-.076	-.191	.016	.005
X_{18}	-.131	-.063	-.062	.010	.156	.741	-.181	.164	.232	-.023	-.032
X_{19}	.064	-.024	.077	-.059	-.129	-.060	.814	.081	.006	-.055	-.056
X_{20}	.023	-.020	-.236	.008	.011	.039	.307	-.156	.734	-.041	.057
X_{21}	-.071	-.009	-.162	-.162	-.126	-.001	.343	-.169	-.693	-.153	.032
X_{22}	-.043	.022	-.080	.452	-.013	-.024	-.081	.085	.114	.386	.295
X_{23}	-.041	-.047	.013	.782	-.021	.047	-.446	.031	-.066	-.052	-.019
X_{24}	-.066	-.001	.069	-.875	-.008	-.063	-.178	-.039	-.103	-.023	.037
X_{25}	-.101	-.084	.431	.094	.027	.085	.254	.055	-.223	.240	.310
X_{26}	-.012	-.636	-.178	-.064	-.005	.108	-.017	-.134	.064	.073	.401
X_{27}	-.067	.179	.084	.236	-.092	-.040	.022	.468	-.024	-.209	.182

Table VII presents the factor loading matrix subsequent to rotation. The principal component analysis method was employed to extract the factors, while the maximum variance method facilitated the rotation, achieving convergence after nine iterations. The values displayed in the table indicate the degree of correlation between 27 indices and 11 factors, with values ranging from 0 to 1, higher values signify stronger correlations.

The empirical findings reveal a significant alteration in the correlation coefficients following the rotation process. The first principal factor exhibits a substantial absolute correlation coefficient with variables X_7, X_8, X_9, and X_{10}, which pertain to the debt repayment capacity of financing enterprises. Consequently, this factor is designated as the "financing enterprises debt-paying ability factor" (F1). The second factor (F2) is characterized by larger absolute correlation coefficients for variables X_{11}, X_{12}, X_{13}, and X_{26}, with the first three variables reflecting the profitability and operational capacity of core enterprises, while X_{26} pertains to the net profit margin of total assets within the industry. Thus, F2 is referred to as the "joint influence factor of core enterprises and supply chain operations."

The third factor (F3) comprises variables related to the operational capacity and debt repayment ability of core enterprises (X_{14}, X_{15}, X_{16}) alongside the strength of supply chain operations (X_{25}), and is termed the "joint impact factor of core enterprises and supply chain operations." The fourth factor (F4) is associated with the operational capacity and debt repayment ability of the blockchain platform (X_{22}, X_{23}, X_{24}), hence it is labeled the "operational debt-paying ability factor of the blockchain platform." The fifth factor (F5) reflects the operational capabilities of financing enterprises (X_5, X_6) and is designated as the "financing enterprise's operational capability factor."

The sixth factor (F6) pertains to the asset position under finance (X_{17}, X_{18}) and is referred to as the "assets position under financing factor." The seventh factor (F7) demonstrates a strong correlation solely with the profitability of the blockchain platform (X_{19}), thus it is named the "blockchain platform profitability factor." The eighth factor (F8) encompasses financing enterprises (X_2, X_3) and the sustainability of supply chain operations (X_{27}), leading to its designation as the "sustainable development factor." The ninth factor (F9) is characterized by the absolute values of the correlation coefficients for X_{20} and X_{21}, which represent the profitability and operational capacity of the blockchain platform, and is termed the "blockchain platform operating profitability factor." The tenth factor (F10) is exclusively associated with the profitability of financing enterprises (X_1), hence it is referred to as the "profitability factor of financing enterprises." Finally, the eleventh factor (F11) is solely related to the growth capacity of financing enterprises (X_4), and is designated as the "financing enterprises growth ability factor." Overall, these 11 factors encapsulate various dimensions of the indicators and serve as variable sources for the subsequent development of risk assessment models.

In line with the results of the above factor analysis, this research extracted 11 factors as independent variables, credit risk as the dependent variable, and established a credit risk assessment model by using the machine learning method. Then, predicted the credit risk of SMEs, and selected the model with the best prediction accuracy and discrimination effect.

A. Outcomes of Machine Learning Techniques

This study employed a Classification Tree, Bagging algorithm, AdaBoost algorithm, and Random Forest models to assess the credit risk associated with SMEs. The findings derived from these four credit risk evaluation models are presented and compared in the following section.

TABLE VIII
COMPARISON OF CREDIT RISK ASSESSMENT MODEL PERFORMANCE

Evaluation Criteria	Classification Tree	Bagging	AdaBoost	Random Forest
P	90%	91.10%	97.22%	86.40%
Se	72.70%	96.88%	100%	50.90%
Sp	94%	90.32%	93.49%	99.1%
F1	74%	71.34%	74.51%	65.88%
KS	0.67	0.87	0.93	0.5
Error I	27.30%	3.13%	4.42%	49%
Error II	5.78%	9.68%	0	0.89%
AUC value	0.8335	0.936	0.985	0.713
Interpretability	F7,F1,F10,F8,F9	F1,F7,F5,F4,F10	F7,F1,F10,F5,F3	F7,F1,F9,F5,F4

As illustrated in Table VIII, the AdaBoost algorithm demonstrates the highest predictive accuracy, achieving commendable values for precision (P), sensitivity (Se), specificity (Sp), F1 score, and area under the curve (AUC). The Bagging algorithm follows closely in terms of performance. Notably, the prediction accuracy of the Bagging algorithm (P=91.1%) is nearly equivalent to that of the Classification tree models. However, the Bagging algorithm exhibits superior sensitivity in predicting SMEs with credit risk (Se=96.88%) compared to the Classification tree model (Se=72.4%). Conversely, the Classification tree model outperforms the Bagging algorithm in predicting enterprises without credit risk, achieving a specificity of 94%. In terms of interpretability, factors F1 and F7 are identified as significant contributors across all four machine learning algorithms. The findings indicate that the debt repayment capacity of financing enterprises and the profitability of blockchain platforms are critical elements in the assessment of credit risk.

TABLE IX
PERFORMANCE OF THE MODEL IN VARIOUS INDICATORS

Primary index	Two-level index	Model performance
Accuracy	P	AdaBoost>Bagging>Classification Tree>Random Forest
	Se	AdaBoost>Bagging>Classification Tree>Random Forest
	Sp	Random Forest>Classification Tree>AdaBoost>Bagging
	F1	AdaBoost>Classification Tree >Bagging>Random Forest
Discrimination	KS	AdaBoost>Bagging>Classification Tree>Random Forest
stability	AUC interval	AdaBoost>Bagging>Classification Tree>Random Forest
Interpretation	Importance of each variable	F1, F3, F4, F5, F7, F8, F9, F10

In terms of predictive accuracy, the AdaBoost algorithm demonstrates superior performance, followed by the bagging algorithm, while the random forest model exhibits the lowest predictive accuracy. This trend is consistent when evaluating the prediction rates for enterprises with credit risk. Conversely, when assessing enterprises without credit risk, the random forest model achieves the highest accuracy, whereas the bagging algorithm shows the least predictive effectiveness. The AdaBoost algorithm model also ranks highest in terms of overall accuracy and recall rate, followed by the classification tree model. Furthermore, the AdaBoost algorithm exhibits the best discrimination capabilities, with the bagging algorithm following closely behind. In terms of stability, the AdaBoost algorithm model performs commendably.

The interpretation of the results indicates that the predictive and discriminatory outcomes are primarily influenced by factors F1, F3, F4, F5, F7, F8, F9, and F10. The first factor (F1) pertains to the debt repayment capacity of financing enterprises. The second factor (F2) is characterized by the significant absolute values of the correlation coefficients associated with variables X_{11}, X_{12}, X_{13}, and X_{26}. The third factor (F3) reflects the combined impact of core enterprises and supply chain operations. Factor F4 encompasses the operational capacity of the blockchain platform alongside its debt repayment ability. Factor F5 denotes the operational capabilities of financing enterprises. Factor F7 exhibits a strong correlation solely with the profitability of blockchain platforms. Factor F8 integrates the financing enterprises and the sustainability of supply chain operations, serving as a deputy factor for sustainable development. Factor F9 combines the profitability and operational capacity of the blockchain platform, representing the operational profitability of the blockchain platform. Finally, factor F10 is exclusively related to the profitability of financing enterprises.

VI. Discussion and Conclusions

This study examines a sample of 56 publicly listed SMEs, 8 core enterprises, and 26 blockchain enterprises in China over the period from 2017 to 2021. We developed a novel credit risk assessment index system and four distinct credit risk assessment models for SMEs utilizing machine learning algorithms.

The findings indicate that the predictive outcomes are predominantly influenced by the operational capacity, profitability, and debt repayment capabilities of the financing enterprises. The core enterprises provide insights into the predictive results through their collaborative roles within the supply chain. Furthermore, the operational capacity and debt repayment ability of blockchain platforms significantly contribute to the predictive explanations. Additionally, sustainable development emerges as a critical input variable influencing the prediction results.

Upon evaluating the strengths and weaknesses of the aforementioned models and their applicable scenarios, it is evident that each of the four models selected for this research possesses unique advantages and limitations, making them suitable for different contexts. Overall, the AdaBoost algorithm demonstrates superior performance, exhibiting the highest capability to differentiate between enterprises with and without credit risk, alongside notable stability. This model is particularly adept at handling extensive, high-dimensional, nonlinear credit risk data characterized by significant noise.

In terms of differentiation, the AdaBoost algorithm excels, effectively distinguishing SMEs with credit risk from those without. Regarding stability, the AdaBoost algorithm outperforms others, with the Bagging algorithm following closely. This finding suggests that the accuracy of these models remains relatively consistent when applied to other large datasets for credit risk evaluation. Both algorithms also provide interpretability by enabling the calculation of the importance of individual features.

A. Theoretical Implication

This study contributes to the theoretical framework for establishing a credit risk assessment index system specifically tailored for SMEs within a blockchain-driven SCF model. Employing a range of innovative hybrid machine learning techniques, the research developed a credit risk assessment model adept at addressing both imbalanced datasets and nonlinear relationships inherent in SCF. Given that the dataset comprises panel data spanning from 2017 to 2021, the baseline machine learning model was trained on each individual dataset. The empirical findings suggest that the AdaBoost algorithm emerges as the most effective solution, demonstrating superior robustness and accuracy.

Furthermore, the credit risk assessment model incorporates a diverse array of indicators, encompassing both qualitative and financial metrics. These include factors such as credit status, the strength of supply chain relationships, and sustainability considerations. The research identifies various information sources pertinent to SMEs, which serve as effective predictors of credit risk within the SCF context. Additionally, the selection of key variables highlights the importance of multi-source information encompassing finance, operational management, sustainability, and adverse events, marking a departure from previous research findings.

Lastly, the credit risk model facilitates a comprehensive understanding of the credit risk landscape for SMEs. Consequently, this research advances the study of credit risk prediction methodologies for SMEs by concurrently addressing the challenges posed by imbalanced datasets and nonlinear relationships that arise from temporal dynamics in blockchain-driven SCF.

B. Practical Implication

This study demonstrates that the primary motivation for investors in making financing decisions should extend beyond the financial information of SMEs to encompass the reliability and sustainability of these entities. Consequently, managers must continuously adapt their organizational models in response to diverse sources of credit risk information to align with stakeholder interests. Furthermore, to enhance investor interest in SMEs, it is imperative for managers to prioritize factors such as profitability, market potential, and corporate reputation.

Additionally, the findings indicate that the predictive outcomes are predominantly influenced by the operational capacity, profitability, and debt repayment capabilities of financing enterprises. Core enterprises interpret these predictive results through the collaborative dynamics of supply chain operations. The operational capacity and debt repayment ability of blockchain platforms serve as significant factors in these predictions. Moreover, sustainable development emerges as a critical input variable influencing the predictive outcomes.

Lastly, this study identifies various indicators of essential information pertaining to both financial and non-financial dimensions of SMEs' credit risk, thereby enriching the credit risk evaluation index system. The data concerning non-financial indicators are sourced from multiple channels. These insights not only offer valuable research directions for the establishment of a credit risk assessment indicator system but also provide practical guidance for the effective implementation of such a system.

C. Limitations and Future Research

The dataset utilized in this study exhibited a significant imbalance between credit-risky enterprises and non-credit-risk enterprises, with credit-risk samples constituting only 20% of the total dataset. This disparity may result in a low precision rate for credit-risky samples while yielding a high precision rate for non-credit-risk samples, thereby leading to a skewed evaluation. The availability of data was limited, and there were constraints regarding data sources. Consequently, this research focused exclusively on the data from listed enterprises within the matching group when conducting a one-to-one sample match between SMEs and core enterprises. Future research endeavors could expand the sample scope by incorporating interviews and investigations of unlisted suppliers. Additionally, subsequent studies could further examine the internal relationship between the financial distress of SMEs and the credit risk associated with their downstream customers, thereby enhancing the research framework and system concerning the external impacts of SMEs' financial distress.

While the proposed machine learning methodology has demonstrated potential benefits for assessing the credit risk of SMEs within blockchain-driven SCF, further investigation is warranted. Future research can be expanded in three key areas. First, this study categorized credit risk indicators into only two groups: enterprises with credit risk and those without. Future investigations should consider a more nuanced classification of credit risk indicators, such as high, medium, and low risk, which would provide additional perspectives and practical significance for various SME categories.

Second, the current research exclusively examined SMEs within the manufacturing sector, which may limit the applicability of the findings to other industries. Future studies should encompass SMEs from diverse industrial backgrounds to enhance the generalizability of the results.

Third, existing scholarly work on the credit risk of SMEs in the context of blockchain-driven SCF primarily focuses on two domains: one aimed at improving or optimizing the SCF model to mitigate credit risk for SMEs, and the other centered on predicting potential credit risk for SMEs based on the blockchain-driven SCF framework. It is evident that this research has predominantly concentrated on the latter aspect. Therefore, future research should prioritize strategies for reducing the credit risk of SMEs through the enhancement and optimization of the blockchain-driven SCF model.

REFERENCES

[1] R. Adams, B. Kewell, and G. Parry, " Blockchain for good? Digital ledger technology and sustainable development goals, " *Handbook of sustainability and social science research*, pp.127-140, 2018.

[2] R.Wang and Y. J. Wu, "Application of Blockchain Technology in Supply Chain Finance of Beibu Gulf Region," *Mathematical Problems in Engineering,* vol.3, pp.10, 2021.

[3] M. Xue, "Risk assessment and analysis in supply chain finance based on blockchain technology," *Journal of Sensors,* pp. 11, 2022.

[4] F. Zhang and W. Y. Song, "Sustainability risk assessment of blockchain adoption in the sustainable supply chain: An integrated method," *Computers & Industrial Engineering,* vol. 171, pp. 108378, 2022.

[5] U. Mia, M. Rahman, A. Elbeltagi, Abdullah-Al-Mahbub, H. M.Gitika Sharma, T.Islam, Subodh Chandra Pal, Romulus Costache, Abu Reza, Md.Towfiqul Islam, Md Monirul Islam, Ningsheng Chen, Edris Alam, and Rana Muhammad Ali Washakh, "Sustainable flood risk assessment using deep learning-based algorithms with a blockchain technology," *Geocarto International,* vol.23, no.8, 2022.

[6] Y. Zhu, "Research on credit risk assessment of small and medium-sized enterprises under supply chain financial environment," *Hunan University,* 2016

[7] Q. X. Rao and W. B. Li, "Risk evaluation and forecast behavior analysis of supply chain financing based on blockchain," *Wireless Communications and Mobile Computing,* no.11, pp. 7668474, 2022.

[8] K.Wang, M. X. Li, and J. Y. Cheng, "Research on personal credit risk evaluation based on XGBoost," *Procedia Computer Science,* vol. 199, pp. 1128-1135, 2022.

[9] Y. Yang, X. Chu, R. Pang, F. Liu, and P. Yang, "Identifying and Predicting the Credit Risk of Small and Medium-Sized Enterprises in Sustainable Supply Chain Finance: Evidence from China," *Sustainability,* vol.13, no. 10, 2021.

[10] R. F. Zhang and X. Q. Li, "Research on measurement of financial credit risk contagion in Supply Chain Based on SWN-SEIRS model," *Financial Theory and Practice,* no. 03, pp. 21-25, 2021.

[11] K. Tian, X. T. Zhuang, and W. T. Zhao, "Credit risk assessment of small and medium-sized enterprises under the mode of Supply Chain Finance," *Industrial technology economy*, vol. 331, no.5, pp.15-18, 2021.

[12] C. Yin, C. Jiang, H. K. Jain, and Z. Wang, "Evaluating the credit risk of SMEs using legal judgments," *Decision Support Systems*, vol. 136, 2020.

[13] Y. Zhu, L. Zhou, C. Xie, G. J. Wang, and T. V. Nguyen, "Forecasting SMEs ' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach," *International Journal of Production Economics*, vol. 211, pp. 22–33, 2019.

[14] W. A. Abbasi, Z. Wang, and A. Alsakarneh, "Overcoming SMEs financing and supply chain obstacles by introducing supply chain finance," *International Journal of Business Management*, vol. 13, no. 6, pp.165-173, 2018.

[15] H. Y. Dong, Y. D. Wang, and L. H. Li, "Summary of Stochastic Forest optimization algorithms," *Information and Computer*, vol. 17, pp. 34-38, 2021.

[16] J. Wu and J. Yu, "Blockchain ' s impact on platform supply chains: transaction cost and information transparency perspectives," *International Journal of Production Research*, pp. 1-14, 2022.

[17] T. K. Agrawal, J. Angelis, W. A. Khilji, R. Kalaiarasan, and M. Wiktorsson, "Demonstration of a blockchain-based framework using smart contracts for supply chain collaboration," *International Journal of Production Research*, vol. 61, no. 5, pp. 1497-1516, 2022.

[18] C. H. Tsai, "Supply chain financing scheme based on blockchain technology from a business application perspective," *Annals of Operations Research*, vol. 22, no. 10, 2022.

[19] C. L. Dang, F. Wang, Z. M. Yang, H. X. Zhang, and Y. F. Qian, "Evaluating and forecasting the risks of small to medium-sized enterprises in the supply chain finance market using blockchain technology and deep learning model," *Operations Management Research*, vol. 15, pp. 662-675, 2022.

[20] A. Wicaksana, W. Ho, S. Talluri, and A. Dolgui, "A decade of progress in supply chain risk management: risk typology, emerging topics, and research collaborators," *International Journal of Production Research,* vol. 60, no. 24, pp. 7155-7177, 2022.

[21] Y. J. Wang, "Research on supply chain financial risk assessment based on blockchain and fuzzy neural networks," *Wireless Communications and Mobile Computing*, no. 8, pp. 5565980, 2021.

[22] Y. Q. Wei, "A Machine Learning Algorithm for Supply Credit Risk Assessment Based on Supply Chain Management," *International Transactions on Electrical Energy Systems*, vol. 11, pp. 4766597, 2022.

[23] H. H. Yi and F. Guo, "Selection and application of enterprise credit risk evaluation index system from the perspective of Supply Chain Finance," *Economic Research*, vol. 31, pp. 103-106, 2021.

[24] H. Zhang, Y. Shi, X. Yang, R. Zhou, "A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance," *J. R. i. I. B., & Finance*, vol. 58, pp. 101482, 2021.

[25] H. B. Kuang, H. Du, and H. Y. Feng, "Construction of credit risk index system of small and medium-sized enterprises under supply chain finance," *Scientific research management*, vol. 21, no. 4, pp. 210-215, 2020.

[26] S. Köhler, S. Bager, and M. Pizzol, "Sustainability standards and blockchain in agro-food supply chains: Synergies and conflicts," *Technological Forecasting and Social Change*, vol. 185, pp. 122094, 2022.

[27] Y. Xia, T. Xu, M. X. Wei, Z. K. Wei, L. J. Tang, "Predicting Chain's Manufacturing SME Credit Risk in Supply Chain Finance Based on Machine Learning Methods," *Sustainability*, vol. 15, pp. 1087, 2023.

[28] S. Shi, R. Tse, W. M. Luo, Stefano, and Giovanni, "Machine learning-driven credit risk: a systemic review," *Neural Computing and Applications*, vol. 34, pp. 14327-14339, 2022.

[29] G. R. Li, J. X. Yang, and Y. Huang, "Supply chain finance credit risk evolving intelligent analysis system based on system dynamic model," *J. Intell. Fuzzy Syst*, vol. 38, pp. 7837–7847, 2020.
[30] Y. Guo, "Credit risk assessment of P2P lending platform towards big data based on BP neural network," *Journal of Visual Communication and Image Representation*, vol. 71, 2020.
[31] Y. L. Wu, X. Li, Q. Q. Liu, and G. J. Tong, "The analysis of credit risk in agricultural supply chain finance assessment model based on genetic algorithm and backpropagation neural network," *Computational Economics*, vol. 60, pp. 1269-1292, 2022.
[32] F. Shen, X. Zhao, G. Kou, and F. E. Alsaadi, "A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique," *Applied Soft Computing*, vol. 98, 2021.
[33] S. Mirzaei, M. Vafakhah, B. Pradhan, and S. J. Alavi. "Flood susceptibility assessment using extreme gradient boosting (EGB)," *Iran. Earth Sci Inform*, vol. 14, no. 1, pp. 51–67, 2021.
[34] M. H. Wang and X. C. Liang, "Personal Credit Evaluation Based on CPSO XGboost," *Computer Engineering and Design,* vol. 40, no. 7, pp. 1891-1895, 2019.
[35] L. Zhang, J. Q.Wang, Z.Y. Fei, S. Luo, and J. Q. Sui, "A personal credit risk assessment model for bank users based on RF-SMOTE-XGboost," *Modern Electronic Technology*, vol. 43, no. 16, pp. 77-81, 2022.
[36] M. Rahman, N. Chen, A. Elbeltagi, M. M. Islam, M. Alam, H. R. Pourghasemi, W. Tao, J. Zhang, T. Shufeng, and H. Faiz, "Application of stacking hybrid machine learning algorithms in delineating multitype flooding in Bangladesh," *J Environ Manage*, vol. 295, pp. 113086, 2021.
[37] W. L. Gao and L. S. Cheng, "Difference analysis of China's tourism development based on factor analysis and spatial econometric model," *Mathematical statistics and management*, vol. 33, no. 1, pp. 138-146, 2014.
[38] C. X. Jin, F. C. Li, S. J. Ma, and Y.Wang, "Sampling scheme-based classification rule mining method using decision tree in big data environment," *Knowledge-Based Systems*, vol. 244, pp. 108522, 2022.
[39] H. He, and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, 2009.
[40] W. Huang and Z. L. Wang, "Data mining: R language practice," *Publishing House of Electronics Industry*, vol. 208, 2016.
[41] W.Y. Qian and H. N. Zhang, "Research on supply chain financial credit risk evaluation based on AdaBoost DPSO SVM model," *Industrial technical economy*, vol. 341, no. 3, pp. 72-80, 2022.
[42] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.
[43] D. Wu, B. Zhang, and O. Baron, "A Trade Credit Model with Asymmetric Competing Retailers," *Production and Operations Management*, vol. 28, no. 1, pp. 206-231, 2019.
[44] Z. W. Ren, Y. D. Wang, and T. Chen, "A new combined classification algorithm based on deep random forest," *Communication technology*, vol. 54, no. 12, pp. 2614-2618, 2021.
[45] H. Abdou and J. Pointon, "Credit scoring, statistical techniques and evaluation criteria: a review of the literature," *Intelligent Systems in Accounting, Finance and Management*, vol. 18, pp. 59-88, 2011.
[46] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection," *Proceedings of the Fourteenth International Conference on Machine Learning*, pp. 179-186, 1997.
[47] M. Rishehchi Fayyaz, M. R. Rasouli, and B. Amiri, "A data-driven and network-aware approach for credit risk prediction in supply chain finance," *Industrial Management & Data Systems*, vol. 121, no. 4, pp. 785-808, 2021.
[48] B. Sang, " Application of genetic algorithm and BP neural network in supply chain finance under information sharing," *Journal of Computational and Applied Mathematics*, vol. 384, no. 11, pp. 113170, 2021.