Hindawi
Complexity
Volume 2020, Article ID 2394948, 16 pages
https://doi.org/10.1155/2020/2394948

Research Article

Credit Risk Assessment for Small and Microsized Enterprises Using Kernel Feature Selection-Based Multiple Criteria Linear Optimization Classifier: Evidence from China

Yimeng Wang and Yunqi Zhang

Business School, Central University of Finance and Economics, Beijing 100089, China

Correspondence should be addressed to Yimeng Wang; wangym0129@163.com

Received 26 April 2020; Accepted 20 May 2020; Published 8 June 2020

Guest Editor: Guangchen Zhang

Credit risk assessment has gained increasing marked attention in the recent years by researchers, financial institutions, and banks, especially for small and microsized enterprises. Evidence shows that the core of small and microsized enterprises' credit risk assessment is to construct a scientific credit risk indicator system, and the key is to establish an effective credit risk prediction model. Therefore, we analyze the factors that influence the credit risk of Chinese small and microsized enterprises and then construct a comprehensive credit risk indicator system by adding behaviour information, supervision information, and policy information. Furthermore, we improve the multiple criteria linear optimization classifier (MCLOC) by introducing the one-norm kernel feature selection and thereby establish the kernel feature selection-based multiple criteria linear optimization classifier (KFS-MCLOC). As for experiments, we use real business data from a Chinese commercial bank to test the performance of these models. The results show that (1) the proposed KFS-MCLOC has greater advantages in predictive accuracy, interpretability, and stability than other models; (2) the KFS-MCLOC selects 10 features from 53 original features and gives selected features their weight automatically; (3) the features selected by the KFS-MCLOC are further verified and compared by the features selected by the logistic regression model with stepwise parameter, and the indicators of "quick ratio; net operating cash flow; enterprises' abnormal times of water, electricity, and tax fee; overdue days of enterprises' loans; and mortgage and pledge status" are proved to be the most influencing credit risk factors.

1. Introduction

Nowadays, faced with the challenges of economic globalization, credit risk assessment of small and microsized enterprises has attracted more and more considerable attention from financial market investors, financial regulators, and national governments. The main purpose of credit risk assessment is to measure the possibility of enterprises' default through qualitative analysis and quantitative calculation of factors that may lead to credit risk and to provide the basis for credit decision-making and risk prevention of banks. In reality, a small increase in the level of bad credit level of enterprises will lead to huge losses to financial institutions [1]. Thus, if we can accurately evaluate the credit risk of an enterprise, it will not only promote the improvement of the enterprise's own risk management level, but also help the bank to effectively prevent the enterprise's default risk, thereby effectively improving the operation efficiency of the whole capital market.

Small and microsized enterprises have been recognized as the predominant type of business units in most Asian economies. In the recent years, small and microsized enterprises have played an increasingly important role in promoting economic growth, increasing employment opportunities, and creating industries. For example, according to the Ministry of Commerce of China, small and microsized enterprises contribute 60% of GDP, provide 80% of urban employment opportunities, and introduce 75% of new products, accounting for 65% of patents and inventions. Thus, the important role played by Chinese small and microsized enterprises can be seen obviously.

Therefore, it is necessary to establish a credit risk indicator system and a credit risk assessment model, especially for small and microsized enterprises.

However, the construction of the small and microsized enterprises' credit risk indicator system is different from that of the large enterprises. For instance, the financial information of small and microsized enterprises is incomplete and opaque, thereby cannot objectively and truly reflect the comprehensive credit risk status of SMEs. So, it is far from enough to only use financial information, whereas it is mainly used by large enterprises. How to build a comprehensive credit risk indicator system by including some behavior variables, supervision variables, and other novel related variables becomes the main focus in today's SMEs' credit risk assessment. Moreover, the improvements and optimization of models can further increase the predictive accuracy to better help enterprises and financial institutions with their risk management and risk prevention.

Recently, many researchers have paid great attention to improve the algorithm of the credit risk assessment models, including statistical models, intelligent models, and optimization models. Evidence indicates that the models used nowadays are more advanced and complicated than before, and these models are also proved to be more effective than before. For instance, Zhang et al. proposed a sparse multicriteria optimization classifier to deal with credit risk assessment. The results showed that the proposed model is more efficient and has better interpretability as well as generalization power [2]. Zhang et al. presented an improved sequential minimal optimization learning algorithm named FV-SMO by using the credit data from the Chinese Banking Regulatory Commission. The experimental results demonstrated that FV-SMO performed much better in saving the computational cost and increasing predictive accuracy compared with the other five state-of-the-art classification methods in enterprises' credit risk assessment [3]. However, few of them study how to carry out feature selection and classification simultaneously.

In this paper, firstly, we construct a credit risk indicator system especially for Chinese SMEs, which contains six parts, that is, basic information, financial information, actual controllers' information, behavior information, supervision information, and policy information. Secondly, we improve the MCLOC from two aspects. Firstly, the one-norm of feature kernel weight vector is introduced into the objective function of MCLOC for feature selection and data dimensionality reduction. Secondly, the kernel feature selection is introduced, and the importance of each feature is expressed by the weight of the kernel feature. The empirical results show that the KFS-MCLOC not only shows high accuracy in predictive performance, but also has great advantages in the feature selection process. The experimental results are shown as follows. Firstly, the proposed KFS-MCLOC has greater advantages in predictive accuracy, interpretability, and stability than other models. Secondly, the KFS-MCLOC selects 10 features from 53 original features and gives selected features their weight automatically. Thirdly, the features selected by the KFS-MCLOC are further compared with the features selected by the logistic regression model with stepwise parameter, and the comparison results show that the indicators of "quick ratio; net operating cash flow; abnormal times of water, electricity, and tax fee; overdue days of enterprises' loans; and mortgage and pledge status" are proved to be the most influencing risk factors.

The remaining sections are structured as follows. Section 2 provides an introduction to credit risk assessment, feature selection, and sparse learning and classification models, with reviews and comparison of the related literatures. Section 3 proposes a new model—kernel feature selection-based multiple criteria linear optimization classifier (KFS-MCLOC). Section 4 presents the experimental design, including the Chinese small and microsized enterprises' credit risk indicator system, dataset description and preprocessing, parameter setting, and models' evaluation criteria, whilst Section 5 describes and analyzes the experimental results. Finally, Section 6 is devoted to the conclusions as well as future work.

2. Literature Review and Related Works

At present, the research studies on credit risk assessment of small and microsized enterprises in academia and practice mainly focus on two aspects: one is the design of the credit risk indicator system; the other is the construction of the credit risk prediction model. In this section, we review and discuss the credit risk indicator system, feature selection methods, and credit risk assessment models by using examples from the past literature.

2.1. Credit Risk and Credit Risk Assessment. Credit risk assessment is emerging as an important concerning topic nowadays. Recently, it has played a more and more important role in assessing the credit worthiness of individuals and enterprises. Generally, the enterprises' credit risk refers to the risk associated with financing problems [3]. Credit risk can not only lead to creditors' economic losses, but also lead to business failure (or corporate distress or bankruptcy or corporate failure). Therefore, how to avoid the credit risk crisis of small and microsized enterprises has been a main focus in the recent years. The purpose of enterprises' credit risk assessment is to distinguish enterprises from good ones to bad ones by various methods, which is essentially a binary classification problem. Recently, the binary classification problem has become the focus of research in the fields of statistics, machine learning, and optimization algorithms, and various methods including logistic regression, SVM, ANN, and MCLPC have been applied to solve this problem.

2.2. Credit Risk Assessment Models

2.2.1. Statistical Models. Originally, credit risk was evaluated by experts' experience and then evolved into the 5Cs theory. With the development of statistical technology, statistical methods were applied to predict the enterprises' credit risk. In 1936, Fisher first established a discriminant analysis to discriminate between the two groups of applicants, which is a very classical statistical method [4]. Later on, Altman proposed a famous Z-score model, which is based on the discriminate analysis model [5]. After that, Orgler applied

linear regression into credit risk evaluation aiming at differentiating between "good" and "bad" credit applicants for commercial banks in practical credit-scoring applications [6]. However, linear regression has strict linear assumption and many other restrictions. Wiginton proposed the logistic regression model for bankruptcy prediction [7]. Since logistic regression has no linear requirement and is easy to understand and interpret, this method is widely used to solve credit risk assessment problems in real business practice. However, for most statistical methods, the shortcomings are obvious, such as they cannot deal effectively with high-dimensional data, they have some assumptions, and their computation time is too long.

2.2.2. Artificial Intelligence Models. In the recent years, with the development of machine learning and the wide use of big data, more and more sophisticated intelligence approaches emerge which are widely applied to enterprises' credit risk prediction, such as neural networks [8–10], genetic algorithms [11,12], and decision trees [13–16]. The literature shows that the intelligent techniques performed better in credit risk assessment than traditional statistical methods [17], and artificial intelligence methods proved to have higher computation accuracy, less computation time, and lower computation cost. Nevertheless, the higher predictive accuracy of artificial intelligence models is often associated with lower interpretive power and longer training time. So, despite the advantages of using intelligent methods, there are still some challenges. For example, most artificial intelligence methods, such as ANN methods, are "black box" methods, whose output result cannot directly interpret the credit risk evaluation result. However, whether the results can be explained is of great importance in practice since most rejected credit applicants will ask for the reasons for refusal. Lu et al. believed that it is very important to determine the importance of each variable by decision rule generation tools before using a black box for prediction [18].

2.2.3. Optimization Models. Besides conventional statistical techniques and artificial intelligence techniques, more attention has been paid to the collaborative use of optimization methods and data mining methods. For example, the SVM, which was first proposed by Cortes and Vapnik, can achieve a higher generalization power and promising results relative to other classification techniques in credit risk modelling [19]. Subsequently, many researchers used improved SVM based on optimizing theory for enterprises' credit risk assessment. Yao et al. proposed a novel two-stage model which is based on the least square support vector machine, and the results showed that this model yields better performance than that of the other statistical models [20].

Moreover, similar to the idea of SVM, mathematical programming optimization techniques such as linear programming, quadratic programming, integer programming, and multicriteria linear programming, which are also based on the optimizing and data mining methods, are widely used in credit risk assessment. Meanwhile, the literature shows that the mathematical programming techniques are shown to have higher predictive accuracy and better explanatory power. In the early 20th century, Shi et al. proposed a compromise solution-based MCLPC model by using behavior analysis of credit cardholders [21]. Maldonado et al. proposed a mixed-integer linear programming model for simultaneous classification and feature selection. The experimental results showed the effectiveness of this method in terms of predictive performance [22]. In addition, some other researchers also solved the linear and nonlinear problems through optimization methods [23,24].

2.3. Feature Selection and Sparse Learning. Feature selection is a necessary step to select features from a large number of data when using classification algorithms to build credit risk assessment models because the quality of data will significantly affect the performance of almost all algorithms. In most cases, it is highly possible that the real-world data contain many irrelevant and redundant features, whereas an appropriate feature selection method can reduce high feature dimension and remove irrelevant features and redundant features [12]. Ala'raj and Abbod's experimental results proved that choosing an optimal subset of features can improve prediction accuracy of the classifiers when constructing the hybrid model [25]. Zhang et al. also indicated that feature selection process was of great importance in reducing the computation time and increasing predictive accuracy [26].

Generally, the two most commonly used feature selection methods are filter and wrapper [27]. However, these two methods can only improve the predictive accuracy, but cannot automatically find out the most important features. To improve the interpretation ability, some researchers proposed to solve this problem by the sparse method, among which zero-norm and one-norm are more typical data sparse methods. Generally, zero-norm regularization is considered as a good method in theory. However, since zero-norm is a nonconvex discontinuous function, the corresponding mathematical problems are difficult to solve and domain knowledge is needed to control the value of superparameters, so it is not suitable for large-scale high-dimensional data problems. However, because of its convexity, one-norm can be directly used in the case of sparse features. In general, these sparse methods are difficult to integrate into kernel functions of high-dimensional feature space. More recently, many researchers have studied the sparse method; for instance, Sun used sparse nonnegative matrix factorizations for reducing the data dimensionality, and the empirical results showed that the NMF-SVM model has the relatively good predictive performance [28]. Mei proposed a sparse coding with sparse dictionaries (K-SVD method) for enterprises' credit risk prediction, and the empirical results demonstrated that this method shows a superior predictive performance [17].

3. Research Methodology

3.1. Multiple Criteria Linear Optimization Classifier (MCLOC). For a binary classification problem and given dataset $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ with a feature set $X = (X_1, \ldots, X_d)^T$, each input point $\mathbf{x}_i (\mathbf{x}_i \in R^d)$ belongs to

the class label y_i where $y_i \in \{-1, 1\}$, d is the dimensionality of the input space, and n is the sample size.

According to these research works [29–31], two measures can be used to make a separation between the positive class and the negative class for solving a two-class classification problem. One measure is the overlapping degree of deviation from the separating hyperplane, and another is the distance between input points and the separating hyperplane, respectively. Thus, in the case of linearly separable data, a multiple criteria linear optimization classifier (MCLOC) model can be denoted as

$$\begin{aligned} \min_{w,b,\alpha,\beta} \quad & C\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\beta_i \\ \text{s.t.} \quad & y_i\left(\mathbf{w}^T\mathbf{x}_i - b\right) = \beta_i - \alpha_i, \quad \alpha_i \geq 0, \beta_i \geq 0, \\ & i = 1, \ldots, n, \end{aligned} \tag{1}$$

where α_i ($\alpha_i \geq 0$, $\alpha = (\alpha_1, \ldots, \alpha_n)^T$) is the distance at which an input point $\mathbf{x}_i$ deviates from the separating hyperplane $\mathbf{w}^T\mathbf{x}_i = b$, $\beta_i (\beta_i \geq 0, \beta = (\beta_1, \ldots, \beta_n)^T)$ is the distance where the input point $\mathbf{x}_i$ departs from the decision hyperplane, the penalty constant $C(C > 0)$ is used to tradeoff the overlapping degree $\sum_{i=1}^{n}\alpha_i$ and the separation degree $\sum_{i=1}^{n}\beta_i$, the weight vector $\mathbf{w}(\mathbf{w} = (w_1, \ldots, w_d)^T)$ consists of the weights of different features, and the scalar $b(b \in R)$ is the unrestricted variable.

3.2. Multiple Criteria Quadratic Optimization Classifier (MCQOC). For the MCOC model in (1), the function $\|\mathbf{w}\|_2^2$ regarding the weight vector $\mathbf{w}$ with a penalty factor $D(D > 0)$, which determines the complexity of the classifier model, is also added to the objective function of the classification problem. Besides, the linear sum of errors is replaced by the squared sum. Thus, we get a multiple criteria quadratic optimization classifier (MCQOC) with the quadratic objective function and the linear constraints, which can be denoted as

$$\begin{aligned} \min_{w,b,\alpha,\beta} \quad & D\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n}\alpha_i^2 - \sum_{i=1}^{n}\beta_i \\ \text{s.t.} \quad & y_i\left(\mathbf{w}^T\mathbf{x}_i - b\right) = \beta_i - \alpha_i, \quad \alpha_i, \beta_i \geq 0, \\ & i = 1, \ldots, n. \end{aligned} \tag{2}$$

Based on the constraints in the MCOC model in (2), we can calculate the intercept $b(b \in \mathbb{R})$, and the separating hyperplane regarding the weight vector $\mathbf{w}(\mathbf{w} \in \mathbb{R}^d)$ is defined as

$$\mathbf{w}^T\mathbf{x} - b = 0, \tag{3}$$

where $\mathbf{x}$ is any input point from the independent test set.

Thus, for a new input point $\mathbf{x}$, its class label y can be predicted by the following decision function:

$$f(\mathbf{x}) = \operatorname{sign}\left(\mathbf{w}^T\mathbf{x} - b\right). \tag{4}$$

Thus, the input point $\mathbf{x}$ is classified as the positive class ($y = 1$) if $\mathbf{w}^T\mathbf{x} \geq b$. Otherwise, the input point $\mathbf{x}$ is classified as the negative one ($y = -1$).

In general, there are three main characteristics that make the multiple criteria optimization classifier (MCOC) more popular than some other traditional classifier models. Firstly, either the principle of MCOC or the algorithm of MCOC is relatively easy to interpret in practice. Secondly, MCOC gains a perfect generalization power as well as an excellent classification accuracy rate since it can correctly find the best balance between minimizing the overlapping degree and maximizing the total distance departed from the boundary. Thirdly, it is shown that since it is very easy and simple to implement the MCOC classifier and adjust its parameters, the performance of the model can be improved considerably. Besides, for the multiclass classification problem, the above MCOC model can be changed into multiple one versus one and one versus the rest classifiers.

3.3. The KFS-MCLOC Model. Nonlinear separable data often appear in the real business world, especially when classifying the credit status of SMEs. Traditionally, a basis function $\phi(\cdot)$ can be used to transform the nonlinear problem into a linear problem by mapping the input points from the input space to the new high-dimensional feature space, where data are linearly separable. For any input point $\mathbf{x}_j$ from the training set D, the weight vector $\mathbf{w}$ is expressed as a linear combination with respect to the instance coefficient vector $\lambda(\lambda = (\lambda_1, \ldots, \lambda_n)^T)$ and class label $\mathbf{y}$, and we have

$$\mathbf{w} = \sum_{j=1}^{n}\lambda_j y_j \phi\left(x_j\right). \tag{5}$$

Then, for any two input points $\mathbf{x}_i$ and $\mathbf{x}_j$ from the training set D, their dot product $\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$ with respect to the basis function $\phi(\cdot)$ can be replaced with the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$. Therefore, the separating hyperplane is reformulated as

$$\sum_{j=1}^{n}\lambda_j y_j K\left(\mathbf{x}_j, \mathbf{x}\right) = b. \tag{6}$$

Here, for any two input points $\mathbf{x}_i$ and $\mathbf{x}_j$ from the training set D, their dot product $\phi(x_{i,m})^T\phi(x_{j,m})$ with respect to the $m$thdimension is replaced by the mth dimensional kernel $k_m(x_{i,m}, x_{j,m})$. Let the feature kernel weight vector be $\mu(\mu = (\mu_1, \ldots, \mu_d)^T)$, then the total kernel function $K(\mathbf{x}_j, \mathbf{x}_i)$ is further defined as the linear combination of $k_m(x_{i,m}, x_{j,m})$ and the kernel weight of $\mu_m (\mu_m \in R)$. Thus, we have

$$K\left(\mathbf{x}_j, \mathbf{x}_i\right) = \sum_{m=1}^{d}\mu_m k_m\left(x_{j,m}, x_{i,m}\right). \tag{7}$$

For the purpose of dimensionality reduction, the one-norm of the feature kernel weight vector μ with the sparsity factor S can be introduced to the MCLOC model in (1). At the same time, the kernel feature selection is realized by applying the total kernel function to the MCLOC model in (1), so the kernel feature selection-based MCLOC (KFS-MCLOC) model can be written as

$$\begin{aligned} \min_{\mu,b,\alpha,\beta} \quad & S\|\boldsymbol{\mu}\|_1 + C\sum_{y_i=-1}\alpha_i - \sum_{i=1}^{n}\beta_i \\ \text{s.t.} \quad & y_i\left[\sum_{j=1}^{n} y_j \sum_{m=1}^{d} \mu_m k_m\left(x_{j,m}, x_{i,m}\right) - b\right] = \beta_i - \alpha_i, \quad 0 \le \beta_i, 0 \le \alpha_i \le C, \mu_m \in \mathbb{R}, \\ & \qquad\qquad i, j = 1, \ldots, n. \end{aligned} \tag{8}$$

Owing to the discontinuity of the one-norm of the feature kernel weight vector μ in the objective function in the KFS-MCLOC model in (8), let $|\mu_m| \le t_m (t_m \ge 0)$, and we have the final KFS-MCLOC model with the form

$$\begin{aligned} \min_{\mu,b,\alpha,\beta} \quad & S\sum_{m=1}^{d} t_m + C\sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\beta_i \\ \text{s.t.} \quad & y_i\left[\sum_{j=1}^{n} y_j \sum_{m=1}^{d} \mu_m k_m\left(x_{j,m}, x_{i,m}\right) - b\right] = \beta_i - \alpha_i, 0 \le \beta_i, 0 \le \alpha_i \le C, -t_m \le \mu_m \le t_m, 0 \le t_m \le S, \\ & \qquad\qquad i, j = 1, \ldots, n, \\ & \qquad\qquad m = 1, \ldots, d. \end{aligned} \tag{9}$$

As shown in (9), the feature kernel weight μ_m represents the importance of each feature. The larger the μ_m, the higher the importance of the feature in the mth dimension. And then, the mth feature should be held in the feature space. Otherwise, the feature is redundant which can be removed from the feature space. In this way, the high-dimensional feature vector can be transformed to a low-dimensional feature vector by the feature kernel weight vector in the total kernel function, which can make it efficient for the model to make later calculation.

For any input point $\mathbf{x}$, its class label can be determined by the following decision function:

$$f(x) = \operatorname{sign}\left[\sum_{j=1}^{n} y_j \sum_{m=1}^{d} \mu_m k_m\left(x_{j,m}, x_{i,m}\right) - b\right], \tag{10}$$

where for any input point $\mathbf{x}_i$ with $\alpha_i = 0$ and $\beta_i > 0$, the intercept b can be obtained by $b = \sum_{j=1}^{n} y_j \sum_{m=1}^{d} \mu_m k_m(x_{j,m}, x_{i,m})$, based on the optimal feature kernel weight vector μ.

Finally, for two input points $\mathbf{x}_i$ and $\mathbf{x}_j$, the RBF kernel function regarding the mth feature is defined as

$$k_m\left(x_{j,m}, x_{i,m}\right) = \exp\left(-\frac{\left(x_{j,m} - x_{i,m}\right)^2}{2\sigma^2}\right), \tag{11}$$

where the parameter σ is specified by the user.

3.4. Algorithmic Design of KFS-MCLOC. The overall process of the experimental design of KFS-MCLOC is shown in Figure 1.

This process can be divided into several stages as follows:

Stage 1. Data Preprocessing. We will use the Binning method and the min-max normalization method to normalize the original dataset

Stage 2. Data Partitioning. We will divide the dataset into the training subset and the test subset, and we will use the five-fold cross-validation method for the training process

Stage 3. Parameter Setting. We will use grid-search method for parameter setting

Stage 4. Two-Stage KFS-MCLOC Model and Other Models. We will test KFS-MCLOC, logistic regression, SVM, neural networks, MCLOC, and MCQOC with the test set to get the predictive results

Stage 5. Predictive Results and Performance Evaluation of Models. We will use seven performance criteria, including the total accuracy, F_1 score, MCC, KS score, AUC, type-I accuracy, and type-II accuracy, to evaluate six models' predictive performance

Stage 6. Feature Importance Analysis. We will use importance analysis and the reduction rate to make an in-depth analysis of the selected features

4. Empirical Design for Small and Microsized Enterprises' Credit Risk Assessment

In this section, we use the proposed KFS-MCLOC and other classifiers for small and microsized enterprises' credit risk assessment, and this section includes four parts, that is, Chinese small and microsized enterprises' credit risk indicator system, data description and data preprocessing, parameter setting, and performance evaluation criteria.

4.1. Design of Small and Microsized Enterprises' Credit Risk Indicator System. In this experiment, the dependent variable is defined by whether it is in credit risk status. If the enterprise is in credit risk status, then the indicator will be equal to 1; otherwise, the indicator will be equal to 0.

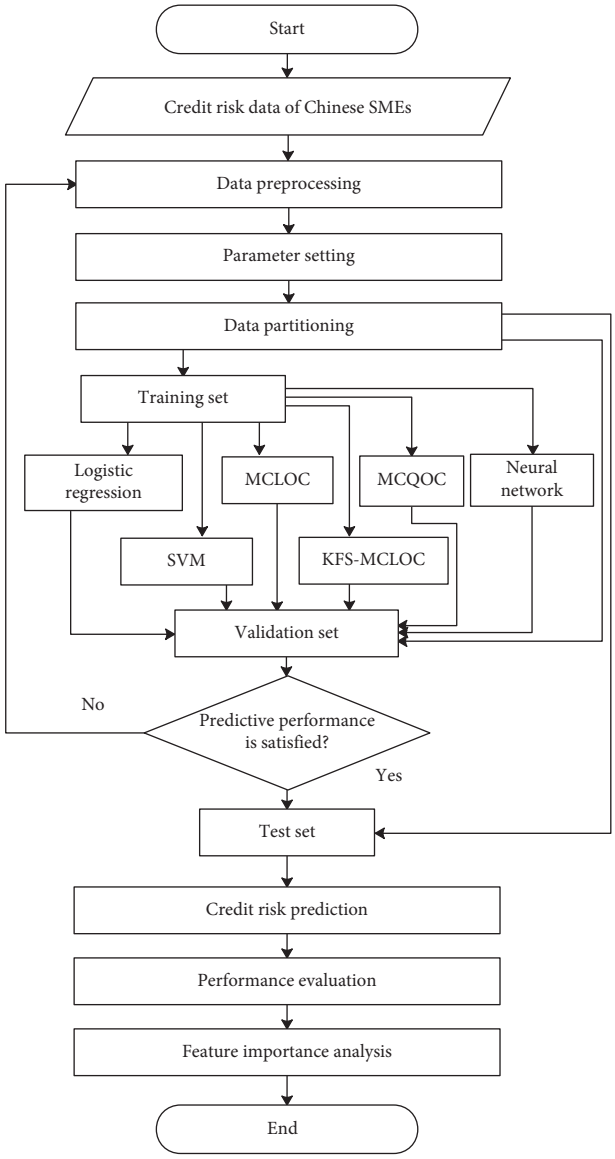

Figure 1: The overall process of the KFS-MCLOC algorithm for small and microsized enterprises' credit risk assessment.

Meanwhile, as for independent variables, a total of 53 credit risk indicators are selected as initial input variables. Based on the characteristics of small and microsized enterprises, we construct a multidimensional and multilevel credit risk indicator system especially for Chinese small and microsized enterprises, and these variables are picked based not only on the financial experts' suggestions but also on the past classical and influential literature [32]. On the whole, the credit risk indicators in this paper can be broadly classified into six overall dimensions: basic information, enterprises' financial information, the actual controller's information, behavior information, supervision and evaluation information, and policy information.

First of all, the basic information of the enterprise directly reflects the most basic situation of the enterprise and reflects the basic quality of the enterprise itself. In general, the higher the quality of the enterprise itself, the less the credit risk of the enterprise will be. Experience shows that the size and the ownership structure of the enterprises are very essential for credit risk assessment. Moreover, the industry to which the enterprise belongs also has a great influence on whether the enterprise will go bankrupt in the future. This is because some industries are largely affected by macroeconomic factors such as national policies, which may lead to unstable operating conditions to enterprise. The description of enterprises' basic information is shown in Table 1.

In the second level, the financial information is an indispensable link in the construction of the credit risk indicator system. The literature shows that the financial information of the enterprise must be fully considered when constructing the enterprise credit evaluation index system. Hence, we select commonly used financial indicators (solvency, profitability, operating capacity, and growth capacity) to reflect the financial situation of the enterprise and select the cash flow-related indicators to reflect the capital turnover of the enterprise. The description of small and microsized enterprises' financial information is shown in Table 2.

As for the third level, we focus on the information about small and microsized enterprises' actual controllers. It is of great importance since the management right and ownership of small and microsized enterprises are usually highly concentrated in the hands of the actual controller, which means that the actual controller has the absolute power to influence the decision-making and the future development direction of the enterprise. Simon found that managerial ability plays a much higher role in predicting credit rating performance than other indicators [33]. Karabag found the great impact of top management actions and the quality of human resources about enterprises' failures [34]. In terms of the existing literature, we find that some indicators are frequently chosen, for instance, marital status, educational background, age, gender, daily behavior, and their credit status. In this paper, the actual controller's information includes the basic information, behaviour information, and supervision information. The description of the actual controller's information is shown in Table 3.

In terms of the fourth level, we focus on the small and microsized enterprises' behaviour information. Under the "Internet Plus" and "Big Data Era," enterprises' behaviour records are increasingly being concerned by the government, financial institutions, and counterparts. Since the influence of enterprises' behaviour becomes more and more extensive, researchers pay more attention to the enterprises' daily behaviour. In fact, nowadays, the relationship between the expected behaviour of enterprises and their past social relations as well as their past behaviour rules is far more important than that of their financial information. For instance, Wang et al. proved that besides conventional hard information, soft information like behaviour information also enters into the lending decision process [35]. Therefore, we add the behaviour information to supplement the content of the credit risk indicator system. The description of behaviour information is shown in Table 4.

As for the fifth level, since supervision information reflects the past production, operation, and credit status of the enterprise, it can help the government to better supervise enterprises and help banks to better prevent the default risk. Therefore, supervision information is necessary. The

Table 1: Small and microsized enterprises' basic information.

Variable	Variable name	Variable definition
X_1	Registered capital	The amount of registered capital
X_2	The size of enterprises	Micro 1, small 2
X_3	The size of employees	The number of the enterprises' employees
X_4	Years of establishment	1 plus years since the enterprises' started to incorporate
X_5	The nature of enterprises	Private 1, collective 2, foreign commercial 3, Hong Kong, Macao, and Taiwan 4, state owned 5.
X_6	Industry of enterprise	Manufacturing 1, wholesale and retail 2, electricity and thermal 3, transportation 4, water conservancy 5, agriculture and forestry 6, construction 7, resident service 8, and mining 9

Table 2: Small and microsized enterprises' financial information.

Variable	Category	Variable name
X_7	Solvency	Asset-liability ratio
X_8		Current ratio
X_9		Quick ratio
X_{10}		Interest coverage ratio
X_{11}		Operating cash flow liability ratio
X_{12}	Profitability	Return on total asset
X_{13}		Gross profit margin
X_{14}		Return on sales
X_{15}		Return on total assets
X_{16}		Return on equity
X_{17}	Operating capacity	Inventory turnover ratio
X_{18}		Account receivable turnover
X_{19}		Total asset turnover
X_{20}	Growth capacity	Sales growth rate
X_{21}		Profit growth rate
X_{22}		The growth rate of total assets
X_{23}	Cash flow information	Capital accumulation rate
X_{24}		Net cash flow from operating activities
X_{25}		Net cash flow

Table 3: Actual controller's information.

Variable	Category	Variable name	Variable definition
X_{26}	Actual controller's basic information	Family status	Married 1, unmarried 2, divorce 3
X_{27}		Educational background	High school 1, secondary school 2, college 3, university 4
X_{28}		Gender	Male 1, female 2
X_{29}		Age	The age of the actual controller
X_{30}	Actual controller's behaviour information	Managerial experience	Number of year that the actual controller engaged in management
X_{31}		Number of self-owned enterprises	Number of enterprises owned by the actual controller
X_{32}		Accumulated overdue repayment	Accumulated overdue repayment times of the actual controller in the bank in the past two years
X_{33}		Financial situation of actual controller	Amount of deposit owned by the actual controller
X_{34}	Actual controller's supervision information	Whether there are bad tax records and illegal behaviours	Yes 1, no 0
X_{35}		Positive comments	Number of positive comments of actual controller
X_{36}		Negative comments	Number of negative comments of actual controller

description of supervision and evaluation information is shown in Table 5.

Finally, in the macrolevel, policy information can largely influence the future development direction of the enterprises. In general, enterprises which are in line with national industrial policies or bank's own policy are more likely to get credit support from banks; on the contrary, banks will be more cautious. The basic description of policy information is shown in Table 6.

Table 4: Small and microsized enterprises' behaviour information.

Variable	Variable name	Variable definition
X_{37}	Whether the wages of employees in the past 12 months can be paid on time	Yes 1, no 0
X_{38}	Employees provided with insurance	The proportion of employees provided with insurance
X_{39}	Whether water, electricity, and tax fees are abnormal in the past 12 months	Abnormal times of water, electricity, and tax in the past 12 months
X_{40}	Change of manager in the past three years	The number of the changes of the general manager in the past three years
X_{41}	Whether has credit business with the bank	Yes 1, no 0
X_{42}	Overdue enterprise loans	Overdue days of enterprises' loans
X_{43}	Whether affiliated enterprises is abnormal	Yes 1, no 0
X_{44}	Mortgage and pledge status	Whether the mortgage and pledge of enterprises is abnormal
X_{45}	Whether the guarantee enterprise is abnormal	Yes 1, no 0

Table 5: Supervision and evaluation information.

Variable	Category	Variable name	Variable definition
X_{46}	Abnormal information of business operation	Business exceptions	Number of business exceptions
X_{47}		Administrative sanction	Number of administrative sanction
X_{48}	Judicial litigation information	Civil judgment document	Number of civil judgment document
X_{49}		Abnormal tax payment	Number of abnormal tax payment
X_{50}	Evaluation information	Positive comments	Number of positive comments
X_{51}		Negative comments	Number of negative comments

Table 6: Policy information.

Variable	Variable name	Variable definition
X_{52}	Whether the enterprise complies with the preferential policies of the bank	Yes 1, no 0
X_{53}	Whether the enterprise conforms to the industry preferential policies	Yes 1, no 0

4.2. Data Description and Data Preprocessing

4.2.1. Data Description. The data we use in this paper come from a Chinese commercial bank. This dataset comprises 188 small and microsized enterprises, in which 130 enterprises are regarded as "not in credit risk status (normal)" while 58 of them are regarded as "in credit risk status (default)." The enterprises we choose cover various industries, such as industrial enterprises, agricultural enterprises, and marine enterprises. Moreover, the nature of enterprises also covers many types, such as private, joint venture, and foreign capital. In terms of data type, this dataset includes various types of variables, such as binary data (male/female), character data (bank's historical credit rating), and numerical data (financial ratios). Therefore, the enterprises' data chosen in this paper are relatively representative.

4.2.2. Data Preprocessing. In reality, irrelevant and redundant features will not only reduce the predictive performance of a classification model but also increase the computational complexity [36]. Generally, data will be firstly preprocessed before going to the model by converting the initial data to standard form data. In this paper, the "Binning" and "min-max normalization" are used as data preprocessing methods. The process of using the min-max normalization method is shown as follows:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{12}$$

where $X_{\min}$ represents the minimum value of X and $X_{\max}$ represents the maximum value of X.

In order to ensure the typicality of sample data extraction, we adopt the stratified random sampling method to collect samples from the training set. Through the stratified random sampling method, 70 normal enterprises and 30 default enterprises are selected from the total samples as training samples. The remaining 60 normal enterprises and 28 default enterprises are used as test samples. The distribution of the sample is shown in Table 7.

In addition, Stone proved that the cross-validation method is an effective method to test the strength of the predictive power of models [37]. Since the K-fold cross-validation is applied due to its properties of being simple, easy, and using all data for training and validation [38], we use a five-fold cross-validation for each of the aforementioned classifiers on the training subset in each iterative process. After that, we can obtain the final predictive accuracy by using the average of the five groups' results. Furthermore, an iterative process on the predefined

Table 7: Distribution of samples.

Sample set	Number of total enterprise	Number of normal enterprise	Number of default enterprise	Proportion of normal enterprise (%)	Proportion of default enterprise (%)
Total sample set	188	130	58	69.15	30.85
Training set	100	70	30	70	30.00
Test set	88	60	28	68.18	31.82

parametric set will be used to find the best parameter. At last, the predictive performance of each classifier will be reported for comparison in order to get the best classification accuracy.

4.3. Parameter Setting. Based on the given training set, we employ the 5-fold cross-validation method to train KFS-MCLOC, SVM, neural networks, MCLOC, and MCQOC, and then they are tested by using the independent test. In the process of training these classifiers, some discrete sets corresponding with different parameters are predefined. As for neural networks, we define a two-dimensional grid corresponding with various numbers of hidden layers and nodes in each hidden layer to search the optimal network structure. That is, the number of hidden layers took values from 1 to 3, while the number of nodes in each hidden layer varies from 10 to 50 with the stride 2. For the SVM classifier, the optimal penalty factor C is selected from the set {1, 2, 5, 10, 20, 50, 100, 200, 500, 1000}, whereas the range of the parameter σ is set to {0.01, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100} for the RBF kernel function. For MCLOC, MCQOC, and KFS-MCLOC, the penalty factor C and the shrinkage factor S for feature selection are uniformly defined as the set {1, 2, 5, 10, 20, 50, 100, 200, 500, 1000}. Meanwhile, the domain of the parameter σ for the RBF kernel function is set to the set {0.01, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100}. Furthermore, all of the experiments are carried out using the MATLAB 8.1 platform.

4.4. Performance Evaluation Criteria. In this paper, we use seven criteria to evaluate the performance of six classifiers, which includes total accuracy, type-I accuracy, type-II accuracy, F_1 score, MCC, and KS score.

4.4.1. Total Accuracy. Total accuracy is one of the most popular performance evaluation criteria which is defined as the correct predicted samples divided by the total sample, and it is computed as

$$\text{total accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{13}$$

where TP refers to the number of the correct classification of the good creditor as the good creditor; FN refers to the number of the wrong classification of the poor creditor as the good creditor; TN refers to the number of the correct classification of the poor creditor as the poor creditor; and FP refers to the number of the wrong classification of the good creditor as the poor creditor.

4.4.2. Type-I Accuracy or Type-I Error. Type-I error is known as the false negative rate, and type-I accuracy is known as the true positive rate, which are, respectively, computed as

$$\begin{aligned} \text{type} - \text{I error} &= \frac{\text{FN}}{\text{TP} + \text{FN}}, \\ \text{type} - \text{I accuracy} &= 1 - \text{type} - \text{II error}. \end{aligned} \tag{14}$$

4.4.3. Type-II Accuracy or Type-II Error. Type-II error is known as the false positive rate, and type-II accuracy is known as the true negative rate, which are, respectively, computed as

$$\begin{aligned} \text{type} - \text{II error} &= \frac{\text{FN}}{\text{TP} + \text{FN}}, \\ \text{type} - \text{II accuracy} &= 1 - \text{type} - \text{II error}. \end{aligned} \tag{15}$$

4.4.4. F_1 Score. F_1 score is commonly used to measure the predictive accuracy of the binary classification model in statistics, and it is computed as

$$F_1 \text{ score} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FN} + \text{FP}}. \tag{16}$$

4.4.5. Matthew's Correlation Coefficient (MCC). MCC is usually used to judge the correlation relationship between two groups of data, and it is computed as

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{[(\text{TP} + \text{FN}) \times (\text{TP} + \text{FP}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})]^{1/2}}. \tag{17}$$

4.4.6. Kolmogorov–Smirnov Score (KS Score). KS score is widely used in evaluating the discriminatory ability of the model between positive and negative samples because it is sensitive to the difference of position and shape between two kinds of empirical cumulative distribution functions. The higher the KS score, the better the performance of the model. The KS score is computed as

$$\text{KS score} = \max(|\text{TP} - \text{FP}|). \tag{18}$$

4.4.7. AUC. AUC is regarded as a widely used measure for model performance evaluation. As a numerical value which ranges from 0 to 1, AUC can evaluate the classifier

intuitively. The evaluation criterion is that the larger the AUC value, the better the evaluation performance of the model.

5. Results and Discussion

5.1. Optimal Parameter Setting. In this paper, we study the effect of different parameters on the classification performance of SVM, MCLOC, MCQOC, KFS-MCLOC, and ANN, respectively. By running the grid search and 5-fold cross-validation against the training data, the best parameters are found for SVM with $C = 100$ and $\sigma = 10$, MCLOC with $C = 5$, and MCQOC with $C = 10$. As for KFS-MCLOC, its optimal parameters are, respectively, 20 for the shrinkage factor S for feature selection, 200 for the penalty factor C, and 0.2 for the bandwidth σ of the RBF kernel function. For the neural network classifier, the optimal network topological structure is composed of 1 hidden layer with 44 nodes in addition to input and output layers.

5.2. Model Evaluation of Predictive Results. As for Chinese small and microsized enterprises' credit risk assessment, we use the five-fold cross-validation method to train the training subsets of the proposed KFS-MCLOC and the other five models, respectively, and then provide each predictive result. In this paper, we totally use seven evaluation criteria, including total accuracy (TA), type-I accuracy, type-II accuracy, F_1 score, MCC, KS score, and AUC to measure the performance of six models.

As shown in Table 8, we can find that the performance of KFS-MCLOC is significantly better than that of the other classifiers. The total accuracy of KFS-MCLOC is 93.63%, which is the highest, indicating that KFS-MCLOC has the best overall predictive performance. As for the F_1 score, only KFS-MCLOC's F_1 score exceeds 0.9, which is 0.91. In comparison, the other five models' F_1 scores are all below 0.9, i.e., SVM (0.82), neural networks (0.77), logistic regression (0.75), MCQOC (0.74), and MCLOC (0.72). In terms of MCC, KFS-MCLOC also has an absolute advantage (0.87), which is much higher than the other classifiers.

In order to compare the predictive ability of each model more clearly, we draw an ROC curve. Figure 2 shows that the ROC curve of KFS-MCLOC is on the far left, which means that it is far better than MCLOC and MCQOC. This indicates that the predictive performance of the MCLOC can be largely improved by introducing the one-norm kernel feature selection. In addition, the ROC curves of neural networks and SVM almost overlap, and both are slightly better than logistic regression in most cases.

Moreover, compared with the ROC curve, since AUC has a specific value, it can intuitively show the classification performance of the model. Therefore, we make a supplementary comparison of the results through AUC. As shown in Table 8, the AUC of KFS-MCLOC is the highest, which is 0.93, and MCLOC performed the worst, which is 0.79.

In addition, experience shows that when the negative sample has a greater influence on the judgment of results, the KS score can better reflect the distinguishing ability of the model than AUC. Figure 3 shows the comparison of KS score among six classifiers. We can see in Table 8 that the KS score of KFS-MCLOC is also the largest (0.85), while the KS score of logistic regression is the lowest (0.61). This result fully demonstrates that unbalanced data have a negative impact on the classification performance of traditional classifiers, which also proves the conclusion of Galar et al.'s research, in which they said that when faced with unbalanced distribution of classes, the performance of traditional classifiers was often disappointing [39].

However, in real business practice, the banks are more concentrated on the type-II error since the cost of the type-II error is much higher than the cost of the type-I error. Table 9 shows the type-I accuracy, type-I error, type-II error accuracy, and type-II error of the six models.

In order to display the results more intuitively, we draw the histograms of type-I accuracy and type-II accuracy of six classifiers, respectively (Figure 4). From the comparison results of type-I accuracy, SVM performs the best (94.64%). However, from the comparison results of type-II accuracy, KFS-MCLOC performs the best (98.57%), followed by logistic regression (85.71%); and the worst is MCLOC (68.57%), which is almost 30% lower than KFS-MCLOC.

Table 8: Predictive results of six models based on five evaluation criteria.

Model	TA (%)	F_1 score	AUC	KS score	MCC
Logistic regression	82.05	0.75	0.82	0.61	0.63
Neural networks	85.68	0.77	0.83	0.72	0.67
MCLOC	82.73	0.72	0.79	0.75	0.6
MCQOC	84.32	0.74	0.8	0.68	0.63
SVM	89.09	0.82	0.87	0.8	0.74
KFS-MCLOC	93.63	0.91	0.93	0.85	0.87

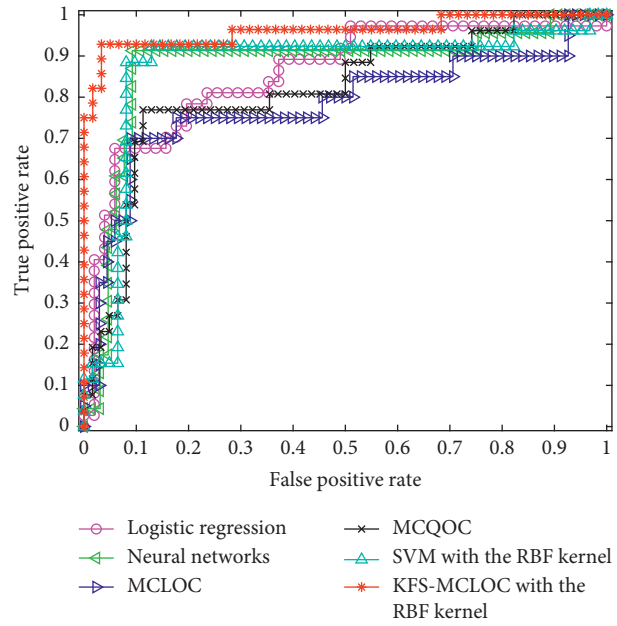

Figure 2: ROC curve of six models.

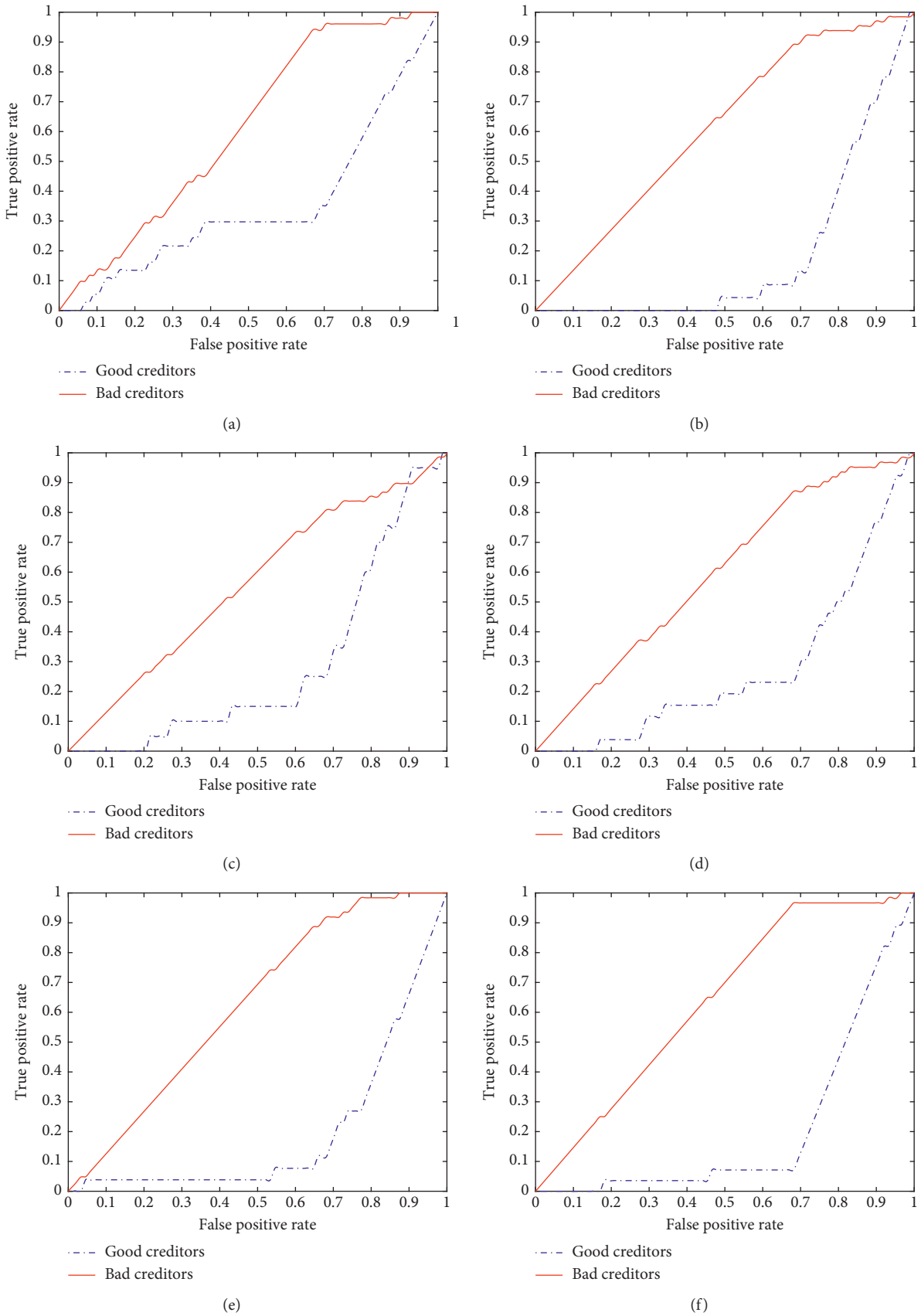

Figure 3: KS curves of six models. (a) Logistic regression. (b) Neural networks. (c) MCLOC. (d) MCQOC. (e) SVM. (f) KFS-MCLOC.

Table 9: Predictive results of six models based on four evaluation criteria.

Model	Type-I accuracy (%)	Type-I error (%)	Type-II accuracy (%)	Type-II error (%)
Logistic regression	80.33	19.67	85.71	14.29
Neural networks	91.67	8.33	72.86	27.14
MCLOC	89.33	10.67	68.57	31.43
MCQOC	90.67	9.33	70.71	29.29
SVM	94.67	5.33	77.14	22.86
KFS-MCLOC	91.33	8.67	98.57	1.43

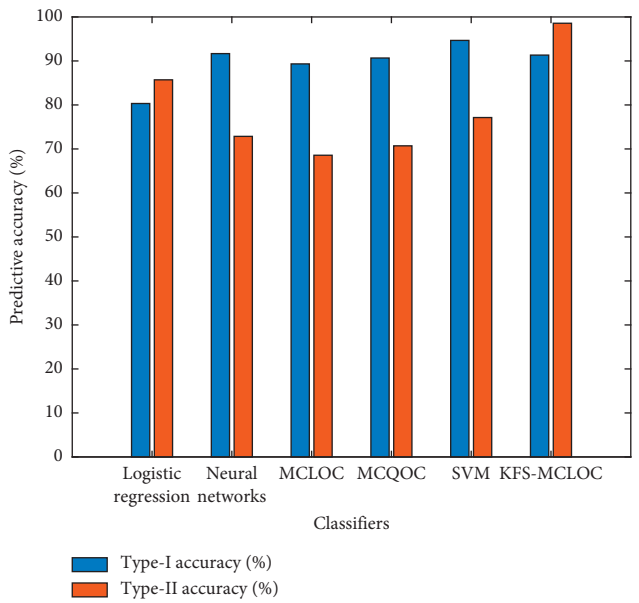

Figure 4: Type-I accuracy and type-II accuracy of six models.

In conclusion, there are three main findings: firstly, in the comparison of the results of total accuracy, type-II accuracy, AUC, KS score, F_1 score, and MCC, KFS-MCLOC's performance is significantly better than that of the other models, i.e., logistic regression, SVM, neural networks, MCLOC, and MCQOC, except for type-I accuracy, in which KFS-MCLOC's performance is slightly worse than that of SVM. Secondly, SVM has almost the second-best predictive performance among the six models. Thirdly, the predictive performance of MCLOC is the worst, which fully shows that it is very necessary to improve the algorithm of MCLOC.

5.3. The Analysis and Discussion of Feature Selection Results

5.3.1. Reduction Rate Analysis. Generally, the purpose of dimension reduction or attribute reduction is to solve the problem of too many features in the original feature set. Obviously, the computational complexity of the model will be greatly reduced and the calculation efficiency of the model will be greatly improved, by reducing the number of original features in the feature space. Furthermore, a proper feature reduction can ensure that the final predictive results from the reduced data are basically consistent with the predictive results obtained from the original data set.

In this paper, the sparse kernel method is used to filter the original feature set, which can reduce the attributes and ensure the accuracy of the predictive results simultaneously. Among them, the reduction rate is an important indicator, i.e.,

$$\text{reduction rate} = \frac{\text{total features} - \text{used features}}{\text{total features}} \times 100\%. \tag{19}$$

In general, the less the number of features in credit risk indicators is, the more the explainable power of small and microsized enterprises' credit risk assessment will be. The comparison of reduction rates of six classifiers is listed in Table 10.

Normally, the fewer the selected features, the higher the reduction rate and the stronger the explanatory power of the selected features. As shown in Table 10, KFS-MCLOC selects 10 most important features from 53 original features, with a reduction rate of 81.13%; while other models (logistic regression, SVM, neural networks, MCLOC, and MCQOC) are with a reduction rate of 0%.

5.3.2. Feature Importance Analysis. The 10 selected features and their weights are shown in Figure 5 and are described in brief as follows: X_9(−2.04%), X_{24}(−19.76%), X_{30}(−2.52%), X_{34}(7.95%), X_{37}(−7.21%), X_{39}(21.42%), X_{42}(10.14%), X_{44}(−13.09%), X_{45}(10.06%), and X_{47}(5.81%).

In order to make a more specific analysis, we combine and classify the name, category, positive/negative impact, and weight proportion of the features selected by KFS-MCLOC, which is shown in Table 11 in detail.

Specifically, from the perspective of the categories of the selected features, they cover the financial information, basic information of the actual controller's information, behavioural information, and supervision information. Among them, the number of features belonging to behavioural information is the largest, accounting for more than 50% of all indicators. In terms of contribution degree, the sum of the three features ("abnormal times of water, electricity, and tax in the past 12 months," "net operation cash flow," and "mortgage and pledge status") importance weights exceeds 50% and plays a decisive role. Among which, two features belong to behaviour information and one feature belongs to financial information. As for the influence direction of selected features, we can see that there are 5 selected features that have positive influence and five selected features that have negative influence.

5.3.3. Further Verification of Selected Features. In order to further verify the rationality of the important features

Table 10: Reduction rate of six models.

Model	Selected features	Number of selected features	Reduction rate (%)
Logistic regression	$X_1, \ldots, X_{53}$	53	0
SVM	$X_1, \ldots, X_{53}$	53	0
Neural networks	$X_1, \ldots, X_{53}$	53	0
MCLOC	$X_1, \ldots, X_{53}$	53	0
MCQOC	$X_1, \ldots, X_{53}$	53	0
KFS-MCLOC	$X_9, X_{24}, X_{30}, X_{34}, X_{37}, X_{39}, X_{42}, X_{44}, X_{45}, X_{47}$	10	81.13

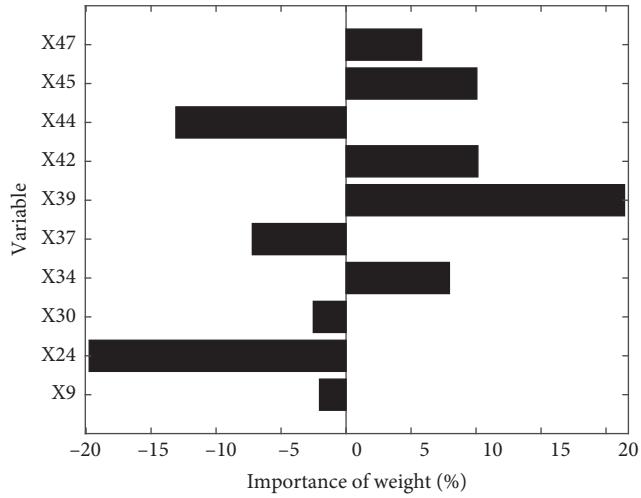

Figure 5: The 10 features selected by KFS-MCLOC and their importance weights.

Table 11: Feature selection results by KFS-MCLOC.

Variable	Variable name	Category	Positive/ negative	Weight
X_9	Quick ratio	Enterprises' financial information	Negative	−2.04%
X_{24}	Operating net cash flow	Enterprises' financial information	Negative	−19.76%
X_{30}	Actual controller's management experience	Actual controller's information	Negative	−2.52%
X_{34}	Whether the actual controller has any record of tax default and illegal behaviour	Actual controller's information	Positive	7.95%
X_{37}	Whether the wages of employees in the past 12 months can be paid on time	Enterprises' behaviour information	Negative	−7.21%
X_{39}	Abnormal times of water, electricity, and tax in the past 12 months	Enterprises' behaviour information	Positive	21.42%
X_{42}	Overdue days of enterprises' loans	Enterprises' behaviour information	Positive	10.14%
X_{44}	Mortgage and pledge status	Enterprises' behaviour information	Negative	−13.09%
X_{45}	Whether the guarantee enterprise is abnormal	Enterprises' behaviour information	Positive	10.06%
X_{47}	Number of administrative punishments	Enterprises' supervision information	Positive	5.81%

selected by KFS-MCLOC, we choose the logistic regression model with stepwise parameter to screen the original 53 features. At the significance level of 5%, logistic regression with stepwise parameter selects 8 important features which are shown in Table 12.

As we can see from Table 12, according to the categories of selected features, most of the selected features belong to small and microsized enterprises' behaviour information, which again proves that the behaviour information plays an important role in SME's credit risk assessment. In addition,

Table 12: Feature selection results by logistic regression with stepwise parameter.

Variable	Variable name	Category	Coefficient	p value
X_9	Quick ratio	Enterprises' financial information	−0.3168	0.0014
X_{20}	Interest coverage ratio	Enterprises' financial information	−0.2924	0.0076
X_{24}	Operating net cash flow	Enterprises' financial information	−0.2401	0.0198
X_{32}	Accumulated overdue repayment of actual controller	Actual controller's information	−0.3426	≤ 0.001
X_{39}	Abnormal times of water, electricity, and tax in the past 12 months	Enterprises' behaviour information	1.2816	≤ 0.001
X_{42}	Overdue days of enterprises' loans	Enterprises' behaviour information	0.2164	0.0248
X_{44}	Mortgage and pledge status	Enterprises' behaviour information	−0.271	≤ 0.001
X_{46}	Number of administrative punishments	Enterprises' supervision information	0.345	0.0006

by comparing the features selected by KFS-MCLOC and features selected by the logistic regression model, it is found that there is no significant difference between them. It can be shown that five indicators including "quick ratio, net operating cash flow, abnormal times of water, electricity, and taxes in the past 12 months, overdue days of enterprises' loans, and mortgage and pledge status" are selected by both KFS-MCLOC and the logistic regression model. To some extent, this result further proves the effectiveness and correctness of the KFS-MCLOC model in feature selection. Furthermore, it should be worth noting that although both models can select important features, KFS-MCLOC can automatically give importance weight to each selected feature, whereas the logistic regression model is relatively weak in this respect. From this point of view, KFS-MCLOC has more advantages in feature selection process.

In addition, combined with the results of Tables 11 and 12, we make a more in-depth discussion on the indicators selected by both KFS-MCLOC and logistic regression with stepwise parameters, as follows.

Firstly, "quick ratio" is considered to be the ability of quick assets to pay current liabilities and is an important index to measure the short-term solvency of enterprises. The higher the "quick ratio" is, the stronger the short-term solvency is and the less likely the enterprise is to have credit risk.

Secondly, "net operating cash flow" is proved to be one of the most important factors, and its importance weight is relatively high. This is probably because small and microsized enterprises have a small scale of assets and a single type of production and operation, which lessen their external financing channels. Once they face short-term financing pressure or fall into production difficulties, it is usually difficult to get the support of external funds. In this case, the internal capital generated by the production and operation of small and microsized enterprises is particularly important. Therefore, through operating cash flow, we can analyze the rationality of enterprise capital operation and judge the ability of enterprise to repay loans, and thereby assess the enterprises' credit risk status.

Thirdly, "abnormal times of water, electricity, and tax fees in the past 12 months" can well illustrate the daily operating status of enterprises' credit risk. Because, if the enterprise cannot pay the water, electricity, and tax fees on time, it means that there may be a high probability of problems in the operation of the enterprise at this stage. Therefore, it can be preliminarily judged that the probability of enterprises' credit risk will increase, which needs to be paid attention.

Fourthly, "overdue days of enterprises' loans" is a very important signal to judge small and microsized enterprises' risk status. The longer the overdue days of enterprises' loans are, the greater the risk of operation of the enterprise is, the less the cash flow is, and the greater the credit risk of the enterprise is. Moreover, "overdue days of enterprises' loans" is the precursor index for the determination of the high risk of enterprises.

Finally, "mortgage and pledge status" is also a very significant indicator. In general, the higher the value of the mortgage, the less the risk the enterprise will have. At present, since the financial information and operation information between banks and small and microsized enterprises are seriously asymmetric, the "mortgage and pledge status" is shown to be a particularly important factor. By increasing mortgage and pledge, credit risk can be effectively mitigated, thus reducing the occurrence of nonperforming loans.

6. Conclusions

Credit risk assessment has always been an important research topic in the fields of accounting, finance, and business. At the same time, it has become a hot research field of statistical learning, artificial intelligence, and optimization algorithm in the recent years. Nowadays, enterprises' credit risk analysis is gradually forming its own theoretical system and research framework. A good credit risk assessment model for enterprises has important practical significance on improving the awareness of credit risk, preventing the credit crisis, and avoiding the bankruptcy liquidation.

Based on the credit data of small and microsized enterprises of a Chinese commercial bank, we design a credit risk indicator system, especially for small and microsized enterprises, including basic information, financial information, actual controllers' information, behaviour information, supervision information, and policy information. As for model construction, we improve the MCLOC by introducing the one-norm kernel feature selection and thereby establish the KFS-MCLOC. In order to test the effectiveness of the KFS-MCLOC, we use total accuracy, F_1 score, MCC, KS score, and AUC to compare models' predictive performance. The empirical result shows that the KFS-MCLOC

model performs better than the other models in almost all aspects by using a real-world credit dataset from a Chinese commercial bank. Secondly, the KFS-MCLOC selects 10 features from 53 original features and gives selected features their weight automatically. Thirdly, the features selected by KFS-MCLOC are further verified and compared by the features selected by logistic regression with stepwise parameter, and the indicators of "quick ratio; net operating cash flow; enterprises' abnormal times of water, electricity, and taxes fee; overdue days of enterprises' loans; and mortgage and pledge status" are proved to be the most influencing credit risk factors. This finding is meaningful for banks and regulatory institutions because these key indicators can be regarded as important credit risk factors and should be paid more attention in practice in the future. In theory, this study provides a useful idea and reference for enriching and developing the credit risk indicator system for Chinese small and microsized enterprises. In practice, this paper also has practical contribution since the effectiveness of the KFS-MCLOC model has been validated by a real-world credit dataset from a Chinese commercial bank.

6.1. Contribution. In general, the contributions of this paper are as follows. Firstly, we construct a comprehensive multi-dimensional credit risk indicator system especially for small and microsized enterprises by adding enterprises' behaviour information, supervision information, and policy information. Secondly, we test the evaluation performance of the model based on the real credit dataset of Chinese small and microsized enterprises. The empirical results show that the KFS-MCLOC model has great advantages in predictive accuracy and stability, which means that the model is suitable for evaluating the credit risk of small and microsized enterprises in the real business world. Thirdly, in the financial field, all credit decisions are required to be interpretable. The proposed KFS-MCLOC model can automatically select the most important indicators and determine their importance weights, which is very effective in solving the "black box" problem, so as to help the credit personnel make effective understandable and traceable decisions.

6.2. Limitations and Future Works. In this section, we conclude the limitations of the proposed model, and put forward the corresponding future works. At first, since the sample size in this paper is relatively small, in the future, a large dataset with a more complex data structure should be explored to further validate the proposed model. In addition, dynamic data changing is a relatively new research problem, and more attention should be paid to it in the future. Finally, although the KFS-MCLOC is shown to be relatively effective in small and microsized enterprises' credit risk assessment, "expert technology" could also be added to the model for a higher predictive accuracy.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest in connection with the work submitted.

References

[1] N. Gulsoy and S. Kulluk, "A data mining application in credit scoring processes of small and medium enterprises commercial corporate customers," *Wiley Interdisciplinary Reviews-data Mining And Knowledge Discovery*, vol. 9, no. 3, p. e1299, 2019.

[2] Z. Zhang, J. He, G. Gao, and Y. Tian, "Sparse multi-criteria optimization classifier for credit risk evaluation," *Soft Computing*, vol. 23, no. 9, pp. 3053–3066, 2019.

[3] Q. Zhang, J. Wang, A. Lu, S. Wang, and J. Ma, "An improved SMO algorithm for financial credit risk assessment - evidence from China's banking," *Neurocomputing*, vol. 272, pp. 314–325, 2018.

[4] A. R. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[5] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[6] Y. E. Orgler, "A credit scoring model for commercial loans," *Journal of Money, Credit and Banking*, vol. 2, no. 4, pp. 435–445, 1970.

[7] J. C. Wiginton, "A note on the comparison of logit and discriminant models of consumer credit behavior," *The Journal of Financial and Quantitative Analysis*, vol. 15, no. 3, pp. 757–770, 1980.

[8] A. Zakaryazad and E. Duman, "A profit-driven Artificial Neural Network (ANN) with applications to fraud detection and direct marketing," *Neurocomputing*, vol. 175, pp. 121–131, 2016.

[9] Z. Zhao, S. Xu, B. H. Kang, M. M. J. Kabir, Y. Liu, and R. Wasinger, "Investigation and improvement of multi-layer perceptron neural networks for credit scoring," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3508–3516, 2015.

[10] X. Huang, X. Liu, and Y. Ren, "Enterprise credit risk evaluation based on neural network algorithm," *Cognitive Systems Research*, vol. 52, pp. 317–324, 2018.

[11] V. Kozeny, "Genetic algorithms for credit scoring: alternative fitness function performance comparison," *Expert Systems with Applications*, vol. 42, no. 6, pp. 2998–3004, 2015.

[12] S. Oreski, D. Oreski, and G. Oreski, "Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment," *Expert Systems with Applications*, vol. 39, no. 16, pp. 12605–12617, 2012.

[13] Y. Xia, C. Liu, Y. Li, and N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," *Expert Systems with Applications*, vol. 78, pp. 225–241, 2017.

[14] C. Qiu, L. Jiang, and C. Li, "Randomly selected decision tree for test-cost sensitive learning," *Applied Soft Computing*, vol. 53, pp. 27–33, 2017.

[15] Y. Li, G. Chi, and Z. Zhang, "Decision tree for credit scoring and discovery of significant features: an empirical analysis based on Chinese microfinance for farmers," *Filomat*, vol. 32, no. 5, pp. 1513–1521, 2018.

[16] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: decision tree ensemble based on SMOTE and bagging with differentiated sampling rates," *Information Sciences*, vol. 425, pp. 76–91, 2018.

[17] X. Mei, "Sparse coding with sparse dictionaries for credit risk classification," in *Proceedings of the International Conference on Progress in Informatics and Computing*, pp. 23–26, Shanghai, China, December 2016.

[18] W. Lu, Z. Li, and J. Chu, "Adaptive ensemble undersampling-boost: a novel learning framework for imbalanced data," *Journal of Systems and Software*, vol. 132, pp. 272–282, 2017.

[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[20] X. Yao, J. Crook, and G. Andreeva, "Enhancing two-stage modelling methodology for loss given default with support vector machines," *European Journal Of Operational Research*, vol. 263, no. 2, pp. 679–689, 2017.

[21] Y. Shi, M. Wise, M. Luo, and Y. Lin, "Data mining in credit card portfolio management: a multiple criteria decision-making approach," in *Proceedings of the Multiple Criteria Decision Making in the New Millennium*, pp. 427–436, Berlin, Germany, July 2001.

[22] S. Maldonado, C. Bravo, J. Pérez, and J. Perez, "Integrated framework for profit-based feature selection and SVM classification in credit scoring," *Decision Support Systems*, vol. 104, pp. 113–121, 2017.

[23] S. He, H. Fang, M. Zhang, F. Liu, and Z. Ding, "Adaptive optimal control for a class of nonlinear systems: the online policy iteration approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 2, pp. 549–558, 2020.

[24] S. He, H. Fang, M. Zhang, F. Liu, X. Luan, and Z. Ding, "Online policy iterative-based H∞ optimization algorithm for a class of nonlinear systems," *Information Sciences*, vol. 495, pp. 1–13, 2019.

[25] M. Ala'raj and M. F. Abbod, "A new hybrid ensemble credit scoring model based on classifiers consensus system approach," *Expert Systems with Applications*, vol. 64, pp. 36–55, 2016.

[26] W. Zhang, H. He, and S. Zhang, "A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: an application in credit scoring," *Expert Systems with Applications*, vol. 121, pp. 221–232, 2019.

[27] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153–158, 1997.

[28] H. Sun, Z. Chen, and J. Chen, "Credit risk analysis using sparse non-negative matrix factorizations," in *Proceeding of the Information Science and Control Engineering*, pp. 181–184, Shanghai, China, April 2015.

[29] N. Freed and F. Glover, "Simple but powerful goal programming models for discriminant problems," *European Journal of Operational Research*, vol. 7, no. 1, pp. 44–60, 1981.

[30] Y. Shi, "Multiple criteria optimization-based data mining methods and applications: a systematic survey," *Knowledge and Information Systems*, vol. 24, no. 3, pp. 369–391, 2010.

[31] Z. Zhang, G. Gao, J. Yue, Y. Duan, and Y. Shi, "Multi-criteria optimization classifier using fuzzification, kernel and penalty factors for predicting protein interaction hot spots," *Applied Soft Computing*, vol. 18, pp. 115–125, 2014.

[32] F. Barboza, H. Kimura, and E. Altman, "Machine learning models and bankruptcy prediction," *Expert Systems with Applications*, vol. 83, pp. 405–417, 2017.

[33] S. Cornée, "The relevance of soft information for predicting small business credit default: evidence from a social bank," *Journal of Small Business Management*, vol. 57, no. 3, pp. 699–719, 2019.

[34] S. F. Karabag, "Factors impacting firm failure and technological development: a study of three emerging-economy firms," *Journal of Business Research*, vol. 98, pp. 462–474, 2019.

[35] Z. Wang, C. Jiang, H. Zhao, and Y. Ding, "Mining semantic soft factors for credit risk evaluation in Peer-to-Peer lending," *Journal of Management Information Systems*, vol. 37, no. 1, pp. 282–308, 2020.

[36] H. He, W. Zhang, and S. Zhang, "A novel ensemble method for credit scoring: adaption of different imbalance ratios," *Expert Systems with Applications*, vol. 98, pp. 105–117, 2018.

[37] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[38] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," *Knowledge-Based Systems*, vol. 64, pp. 22–31, 2014.

[39] D. D. Galar, R. Villarejo, C. A. Johansson, U. Kumar, and L. F. B. Muro, "Hybrid models for PHM deployment techniques in railway," in *Proceedings of the Tenth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, pp. 1047–1056, Krakóv, Poland, June 2013.