Hindawi
Complexity
Volume 2022, Article ID 5139562, 18 pages
https://doi.org/10.1155/2022/5139562

Research Article

Default Risk Prediction of Enterprises Based on Convolutional Neural Network in the Age of Big Data: Analysis from the Viewpoint of Different Balance Ratios

Zhe Li,[1] **Zhenhao Jiang**,[2] **and Xianyou Pan**[3]

[1]*School of Economics and Management, Dalian University of Technology, Dalian 116024, China*
[2]*DUT-BSU Joint Institute, Dalian University of Technology, Dalian 116024, China*
[3]*School of Economics and Management, Shanghai University of Electric Power, Shanghai 201306, China*

Correspondence should be addressed to Zhe Li; li_zhe@yeah.net and Zhenhao Jiang; dp_warnel@163.com

Received 17 December 2021; Revised 23 May 2022; Accepted 11 July 2022; Published 1 September 2022

Academic Editor: Daniele Salvati

In the age of big data, machine learning models are globally used to execute default risk prediction. Imbalanced datasets and redundant features are two main problems that can reduce the performance of machine learning models. To address these issues, this study conducts an analysis from the viewpoint of different balance ratios as well as the selection order of feature selection. Accordingly, we first use data rebalancing and feature selection to obtain 32 derived datasets with varying ratios of balance and feature combinations for each dataset. Second, we propose a comprehensive metric model based on multimachine learning algorithms (CMM-MLA) to select the best-derived dataset with the optimal balance ratio and feature combination. Finally, the convolutional neural network (CNN) is trained on the selected derived dataset to evaluate the performance of our approach in terms of type-II error, accuracy, G-mean, and AUC. There are two contributions in this study. First, the optimal balance ratio is found through the classification accuracy, which changes the deficiency of the existing research that samples are imbalanced or the balance ratio is 1 : 1 and ensures the accuracy of the classification model. Second, a comprehensive metric model based on the machine learning algorithm is proposed, which can simultaneously find the best balance ratio and the optimal feature selection. The experimental results show that our method can noticeably improve the performance of CNN, and CNN outperforms the other four commonly used machine learning models in the task of default risk prediction on four benchmark datasets.

1. Introduction

Default risk prediction means the prediction of the repayment ability of enterprise loans. Default risk prediction can effectively avoid default risk and reduce the loss of debtors and investors. In the age of big data, deep learning algorithms, such as convolutional neural networks, are widely applied to various fields. In the area of default risk prediction, there are two main problems: one is the balance ratio of the dataset, and the other is the selection of the best feature combination. To address these two problems, this study proposes a comprehensive metric model based on multimachine learning algorithms to find the best balance ratio and optimal feature combination and to improve the performance of the deep learning model in the default risk prediction.

1.1. The Best Balance Ratio of the Imbalanced Dataset. Due to the objective fact that there are far more nondefault samples than the default samples, the problem of data imbalance is prevalent. The performance of traditional machine learning methods on unbalanced datasets is poor. There are many studies on rebalancing methods for imbalanced datasets, but few studies on the optimal balance ratio. Different balance ratios greatly influence the prediction result of the machine learning model. For the deep learning model, the training time is too long, and it would

take a lot of time to find the optimal balance ratio. Therefore, finding the best balance ratio is a significant problem.

1.2. The Best Feature Combination and Selection Order. Different feature combinations can make the default risk prediction quite different. Feature selection can improve the model performance and save computational costs. Therefore, feature selection is an essential problem in the task of default risk prediction. At the same time, with respect to the orders of rebalancing and feature selection, the datasets with the same balance ratio have different optimal feature combinations. Therefore, the order of feature selection and data rebalancing processing is also a problem worth discussing.

The differences between our study and existing studies are mainly reflected in the following two aspects.

Our study on balance ratio is most related to Hou et al. [1]. Hou et al. applied the method of synthetic minority sample oversampling and minority sample weighting to address the imbalance problem [1]. In this study, from the perspective of searching for the best balance ratio, the score ranking of the comprehensive metric model based on the multimachine learning algorithm in a training set with different balance ratios is used to reverse the optimal balance ratio and improve the performance of the classifier.

Our study on feature selection is most related to Song et al., but the research of Song et al. on feature selection does not consider the influence of data balance processing [2]. In this study, the feature combination selection is conducted before and after the data balancing. Then the optimal feature combination is obtained through the comparative analysis of multiple groups of experiments.

In recent decades, researchers have introduced many resampling strategies and imbalance-proof algorithms to address the data-imbalanced problem for default risk prediction. Unfortunately, the efficiency of the above-mentioned methods almost reaches the limit, which means that we have to find a new facet to push the limit. In this study, the optimal balance ratio is found through the classification accuracy, which changes the deficiency of the existing research that samples are imbalanced or the balance ratio is 1 : 1 and ensures the accuracy of the classification model. A comprehensive metric model based on the machine learning algorithm is proposed, which can simultaneously find the best balance ratio and perform the optimal feature selection. Additionally, our framework is flexible and universal for different resampling methods and machine learning algorithms and obtain a more satisfactory result than the original algorithms. In conclusion, a framework to explore the best balance ratio and feature combination is promising and meaningful. This makes the strategy for the data-imbalanced problem more systematic.

This paper left is organized as follows. Section 2 is the literature review. Section 3 presents the problem statement and solution idea. Section 4 introduces the methodology. Section 5 shows the experimental analysis. Section 6 outlines the conclusion.

2. Literature Review

2.1. Research on Data Imbalance. The solution to data imbalance can be divided into algorithm level, data level, and the combination of algorithm level and data level.

At the algorithm level, researchers improve the performance of models by making the algorithm focus on the minority class. Huang et al. evaluated the ability of neural networks to deal with imbalance problems in the field of default risk prediction. The experimental results show that the proposed methods can obtain consistently high TP prediction rates for each class [3]. Huang et al. used three strategies to construct the hybrid SVM-based credit scoring models. Compared with other tree classifiers, the SVM classifier achieved an identical classificatory accuracy [4]. I. Mues and C. Mues used resampling to process imbalanced credit data and applied LR, DT, and other models to predict default. This empirical study indicates that the random forest and gradient boosting classifiers perform very well in credit scoring [5]. Kvamme et al. used convolutional neural networks to predict mortgage defaults. Experimental results show that the threshold is essential for accuracy [6]. Yu et al. proposed an integrated model based on the deep belief network (DBN) and SVM. The experimental results indicate that the proposed model can be a promising tool for credit risk classification with imbalanced data [7]. Ma et al. applied LightGBM and XGboost to deal with unbalanced data to predict the default risk of loans [8]. Li et al. explored the application of transfer learning in the financial field. Their study highlights the commercial value of the transfer learning concept and provides practitioners and management personnel with a decision basis [9]. M. Lardy and J. P. Lardy offered a simple, global, and transparent CDS structural approximation for the energy industry based on random forest [10]. Zhao et al. constructed a DMDP system that can capture a company's historical performance and use long short-term memory to dynamically consider news from social media and public opinion [11]. Fu et al. utilized BiLSTM to predict the default risk of the platform based on the extracted keywords of investor comments. Experimental results show that the proposed model can better capture semantic features and achieve significant improvement [12].

At the data level, resampling methods are globally used, including undersampling, oversampling, and hybrid sampling. Among them, the oversampling technique recently has recently become a research hotspot. The synthetic minority oversampling technique (SMOTE) is the bedrock of the oversampling method proposed by Chawla et al. [13]. Hernandez et al. presented an empirical study about using oversampling and undersampling methods to improve the accuracy of instance selection methods on imbalanced databases. The experimental results show that using oversampling and undersampling methods significantly improves the accuracy for the minority class [14]. I. Lai-Yuen and S. K. Lai-Yuen presented a novel oversampling approach called A-SUWO which can avoid sample overlapping. The experimental results indicate that the proposed method achieves significantly better than other sampling methods [15]. Liang et al. presented an oversampling

technique that unites k-means and SVM to balance the data, which proves to have a better performance than the SMOTE algorithm [16]. Jiang et al. formulated an OS-CCD method based on the classification contribution degree that can describe the importance of samples. Experimental results show that OS-CCD outperforms six classical oversampling methods on twelve benchmark datasets [17]. Zhang et al. proposed a new hybrid ensemble model with voting-based outlier detection and balanced sampling. The experimental results indicate superior performance and prove its robustness and effectiveness [18].

Several studies improved both the algorithm and data. Hou et al. proposed a new dynamic integrated selection strategy (DES) for default prediction. Before the candidate classifier pool is generated, the training set is balanced by SMOTE, and the importance of minority samples in the evaluation of classifier capability is increased [1]. Shen et al. proposed an improved SMOTE technique to balance the dataset and make use of the integrated deep learning model for default prediction. The experiments show that the proposed model improves the AUC of seven known and popular DES algorithms [19].

A common difference between our study and the abovementioned literature is that different balance ratios are considered in this study.

2.2. Research on Feature Selection. Feature selection includes single feature selection and feature combination selection. Single feature selection only focuses on the function of one feature but ignores the underlying relationship among different features, while feature combination selection considers the influence of multiple features on the results simultaneously.

Research on single feature selection. Chen and Li proposed four approaches combined with the support vector machine classifier for feature selection that retains enough information for classification purposes. The result suggests that the hybrid credit scoring approach is robust and effective in finding optimal subsets [20]. S. Oreski and G. Oreski proposed a feature selection method based on genetic algorithms. The preset threshold eliminates the unqualified features in terms of FDAF-score. Experimental results show that the proposed classifier is promising for feature selection and classification [21]. Song et al. calculated the FDAF-score of each feature to evaluate the feature's contribution to classification. Experiments demonstrate that the proposed FDAF-score algorithm can obtain good results and deal with the classification problem with noises [2]. Sheikhpour et al. studied the semisupervised feature selection method to reduce the cost of dataset annotation and simplify data collection. Based on the literature review, it was observed that most of the semi-supervised feature selection methods had been presented for classification problems [22]. Urbanowicz et al. placed Relief-based algorithms in the context of other feature selection methods and provided an in-depth introduction to the Relief-algorithm concept [23]. Hu et al. conducted feature selection by calculating the correlation between features and classification results. The classification results show that the proposed model performs better than the other five methods [24]. Kozodoi et al. proposed a profit-driven multiobjective feature selection method. Experiments demonstrate that the proposed approach can yield a higher expected profit using fewer features [25].

The above is a single feature selection method, and many scholars have significantly contributed to feature combination selection. Luo et al. presented an adaptive unsupervised feature selection method that can generate a picture of the original feature space and output a reliable feature combination [26]. Zhang et al. applied the multiobjective particle swarm optimization to a multiclass dataset that verifies that the proposed algorithm is a helpful approach to feature selection for multi-label classification problems [27]. Mafarja et al. presented a wrapper feature selection method based on the binary dragonfly algorithm. The results show the ability to search the most informative features for classification tasks [28]. Gu et al. presented a competitive swarm optimizer that deals with high-dimensional feature selection. Experiments demonstrate that the proposed feature selection algorithm can select a much smaller number of features with better classification performance [29]. M. Mirjalili and S. Mirjalili modified wrapper feature selection by whale algorithm proved efficient in searching for the optimal feature subsets [30]. Abualigah et al. used the particle swarm optimization algorithm to perform feature selection which can eliminate the uninformative features of the text [31]. Sayed et al. proposed a chaotic crow search algorithm to improve the convergence rate and find the optimal feature combination [32]. Zhang et al. proposed a two-archive multiobjective artificial bee colony algorithm that is proved to be an efficient and robust optimization method for solving cost-sensitive feature selection problems [33]. Ghosh et al. proposed a wrapper-filter combination of ant colony optimization in feature selection. The experimental results clearly show that our method outperforms most state-of-the-art algorithms used for feature selection [34]. Zhang et al. proposed a self-learning and multiobjective algorithm to balance local and global searching [35].

The difference between the research on feature selection in this study and the abovementioned studies is that our study first uses the feature combination selection method to select the datasets with different balanced ratios and then selects the optimal feature combination from the selected feature combinations. In addition, this paper studies the selection order.

2.3. Research on Default Risk Prediction Methods. There is extensive literature on various approaches to credit scoring, default risk prediction, and fraud detection. Credit risk prediction indicates the level of the risk in investing with the company. It represents the likelihood that the company pays its financial obligations on time.

Credit risk prediction models can be divided into qualitative and quantitative analysis, machine learning models, and deep learning models. Standard and Poor's, Moody's, and Fitch use a twofold analysis (qualitative and

quantitative) to assign a credit score to a company. Qualitative analysis is based on different factors such as company strategy and economic market outlook, while quantitative analysis is based only on financial statements. However, how these analyses lead to the final credit score is still unclear [36].

Many machine learning methods have been extensively employed to forecast corporate default risk. Linear default risk prediction models determine which category the new individual sample belongs to by summarizing a classification rule with a large number of samples. The representative models include linear discriminant analysis [37], multivariate discriminant analysis [38], logistic regression [39], and probit models [40], among others. Although the linear models have the advantages of simplicity and ease of use, they cannot effectively deal with the nonlinear relationship between variables. Thus there is a significant difference between the classification results and the actual default status. With the advent of the big data era, artificial intelligence-based default risk prediction methods are increasingly advantageous. Machine learning models involve support vector machines [41], decision trees[42], k-nearest neighbors [43], BP neural networks [44], and so on. Compared with the linear models, which require strict assumptions and are sensitive to noise data, the artificial intelligence models effectively overcome linear models' limitations with strong robustness and unstructured characteristics.

In addition to standard neural networks, some researchers focus on deep learning models in many financial fields. Compared with an artificial neural network (ANN), a deep neural network (DNN) is a model with more than one hidden layer between the input and output layers [45]. Deep learning models such as Convolutional Neural Networks (CNN) [46], Generative Adversarial Networks (GAN) [47], and Recurrent Neural Networks (RNN) [48] have been proven to significantly improve the accuracy of classification in various financial problems [49]. Previous applications of CNN include their use in image processing, sequence, and time-series [50, 51], while in financial problems, they are mainly used in the stock market analysis [52]. Tsantekidis et al. used CNN to predict mid-price movements of the limit order book; their empirical evidence reveals that CNN can obtain more accurate results than the multilayer perceptron model [53]. Chung and Shin applied one of the representative deep learning techniques, multichannel convolutional neural networks, to predict the fluctuation of the stock index. The experimental results show that CNN outperforms the comparative models, which demonstrates the effectiveness of CNN [54]. Chen et al. proposed a novel method for stock trend prediction using a graph convolutional feature based convolutional neural network model, in which both stock market information and individual stock information are considered [55]. Some studies have tried applying CNN to the default risk prediction field. Kvamme et al. predict mortgage default by applying CNN to consumer transaction data [6]. Carrasco and Sicilia-Urbán used DNN to measure their ability to detect false positives by processing alerts triggered by a fraud detection system [56]. J. I. Z. Lai and K. L. Lai proposed the deep convolution neural network scheme based on a financial fraud detection scheme using a deep learning algorithm. Over a time duration of 45 s, the detection accuracy of 99% was obtained using the proposed model, as observed in the experimental results [57].

In our study, CNN is used to predict the default risk of the enterprises, and the experimental results show that CNN outperforms KNN, DT, SVM, and LR. As a deep learning model, CNN can accurately predict enterprises' credit risk.

3. Problem Description and Solution Ideas

3.1. The Determination of the Best Balance Ratio of Data. In order to solve the problem that the default samples are far fewer than the nondefault samples, the dataset is often rebalanced before constructing the default risk prediction model. As an excellent sampling method, the synthetic minority oversampling technique (SMOTE) is widely used to process imbalanced datasets. In this study, the SMOTE algorithm is used to balance the dataset. In the oversampling process, the difference in balance ratios between default and nondefault samples will affect the model performance.

In this study, the approach to solve the problem is to use default and nondefault samples with different balance ratios to establish a prediction model. The best balance ratio is obtained through the comprehensive ranking of multiple metrics of the prediction model.

The approach to constructing the balance ratio is as follows. Assume that A represents the majority (nondefault) samples in the training set and B represents the minority (default) samples. It is clear that the proportion of the majority class and minority class of the training set is $A:B$. The SMOTE algorithm is used to perform the oversampling operation on default samples. It is assumed that every round of sampling is performed on all minority samples, i.e., repeated B times. Then, the maximum sampling round $N_{\max}$ is:

$$n_{\max} = \lceil \frac{A-B}{B} \rceil. \tag{1}$$

The abovementioned formula means round-up, which means the number of times required to apply the SMOTE algorithm to the minority samples when the dataset is balanced to 1 : 1. Then, the sampling round N_α required to achieve is shown as follows.

$$n_\alpha = \frac{A/\alpha - B}{B}. \tag{2}$$

The abovementioned formula indicates the times required to apply the SMOTE algorithm when the dataset is balanced to $\alpha:1$, where α is the balanced ratio and set to 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10. The datasets with different balance ratios are put into support vector machine (SVM), logical regression (LR), decision tree (DT), and k-nearest neighbor (KNN). The optimal balance ratio can be obtained by comparing the comprehensive ranking of multiple metrics of the four models with different balance ratios.

3.2. The Determination of the Best Feature Combination and the Selection Order. In the task of enterprise default risk prediction, enterprise data often have many features, which creates the problem of feature redundancy. Selecting the best feature combination can optimize the performance of machine learning models and reduce the calculation cost.

Different feature combinations significantly influence the results, and there are always several feature combinations that can make the models have higher prediction accuracy. Feature selection can be divided into single feature selection and feature combination selection. Feature combination selection refers to the selection of multiple different feature combinations simultaneously. In this study, feature selection technology based on the genetic algorithm is a feature combination selection method. In the feature selection process, we often face the problem of the order of feature selection treatment and rebalancing treatment. The selection order is also studied in this work.

4. Methodology

4.1. Synthetic Minority Oversampling Technique. Chawla et al. proposed the synthetic minority samples oversampling technique (SMOTE), which is a type of oversampling method to solve the problem of data imbalance [13]. The basic principle is based on the distance between the samples in the sample space, using the existing minority samples to synthesize new samples. Given an imbalanced dataset, for each sample x_i of the minority class, we calculate the Euclidean distance between x_i and other minority samples. SMOTE finds K neighbors of x_i, selects a random point x_j from these K points, and then synthesizes a new sample point x_{new}. The calculation formula of the new sample is as follows.

$$x_{\text{new}} = x_i + \left(x_i - x_j\right) \times \delta, \tag{3}$$

where δ is a random number between 0 and 1. The abovementioned formula indicates that a new sample is synthesized between two minority samples, and the coordinate value of the new sample point is calculated.

In Figure 1, the circles represent the minority samples, the rectangles represent the majority samples, and the crosses represent the synthesized samples.

4.2. Genetic Algorithm. In this study, the genetic algorithm is used to implement the operation of feature selection, which is a kind of feature combination selection. The introduction of the genetic algorithm is as follows.

A genetic algorithm is an optimization algorithm that simulates natural evolution. In nature, organisms evolve through natural selection. In genetic algorithms, computer-generated "creatures" are selected and evolved by fitness functions. The basic process of the genetic algorithm is shown in Figure 2.

(1) The population is generated randomly, and numerical arrays represent the individuals. Additionally, an individual represents a combination of features. The elements in the array are called genes, which are values of 0 or 1. In the feature selection task, 0 represents that the feature corresponding to the position is removed, and 1 represents that the feature corresponding to the position is selected.

(2) The fitness function is used to calculate the fitness of individuals. In our study, the fitness function is the average $F1$-score of the fivefold cross-validation results of the SVM on the training set. The details of the SVM are described in Section 4.3.3. The $F1$-score is introduced in Section 4.5.

(3) According to the preset threshold, the unqualified individuals in terms of fitness are removed.

(4) The remaining individuals are "hybrid" to produce a new individual, and the offspring's genes are the binary sum of the same locus of its parents' genes. For example, an individual with a genome of [1, 0, 1, 1, 0] crosses with that of [0, 1, 1, 1, 1] to obtain [1, 1, 0, 0, 1]. Then, some individuals are randomly selected for gene mutation at a random locus; the gene is changed from 0 to 1 or from 1 to 0. Then, the population is updated.

(5) The fitness function is used to calculate the fitness of the individuals in the new population, and the unqualified individuals are eliminated.

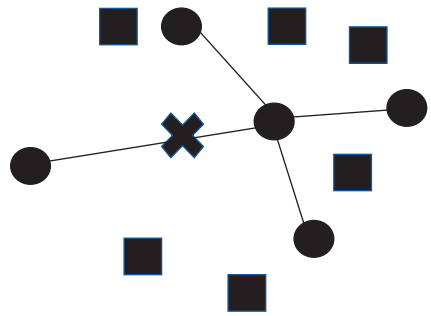

Figure 1: Schematic diagram of SMOTE.

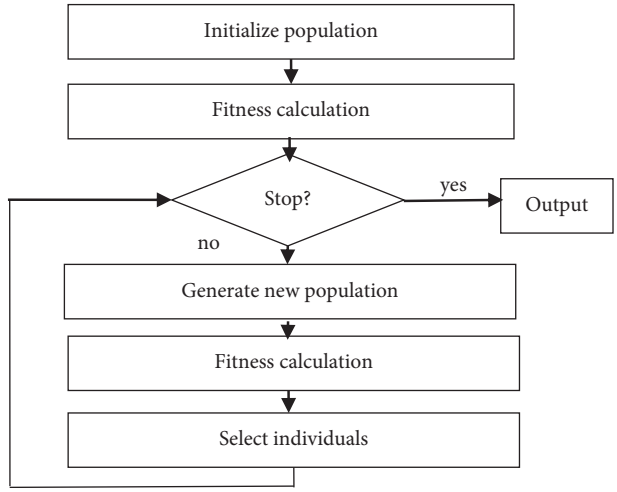

Figure 2: Flowchart of genetic algorithm.

(6) Judge whether the stop condition is met. If it is, the individual with the highest fitness in the evolutionary process is output. If not, it is returned to step (4).

The genetic algorithm is a heuristic search and optimization technique that mimics the process of natural evolution. The algorithm is easy to understand, good for noisy environments, and robust with respect to local maxima/minima. However, the genetic algorithm may fall into the local optimum when dealing with a complex optimization problem, and the result of the solution strongly depends on the initial value.

The genetic algorithm jumping out of local optimization to obtain global optimization is based on crossover and mutation in the algorithm. Compared with the standard genetic algorithm, this study makes the following two improvements.

4.2.1. The Improved Crossover. First, according to the matching principle, the parents are queued, i.e., sorted by the fitness function. Individuals with small objective function values are paired, and individuals with large objective function values are paired. Then, the position of the crossover point is determined using the logistic chaotic sequence $x(n+1) = 4x(n)\ (1 - x(n))$. Finally, the determined crossover is performed.

For example, paired individuals (Ω_1, Ω_2), $\Omega_1 = w_1^1\ w_2^1, \ldots, w_k^1$, $\Omega_2 = w_1^2\ w_2^2, \ldots, w_k^2$. First, an initial value between 0 and 1 is taken, and $x(n+1) = 4x(n)\ (1 - x(n))$ is used to generate one chaotic value on (0, 1) at a time; multiply this value by k, and finally round it.

4.2.2. The Improved Mutation. The mutation is also a means to achieve group diversity and is an essential guarantee for jumping out of local optimization. The improved mutation in this study is designed as follows. According to the given mutation rate, two integers between 2 and k are randomly selected to mutate the genes at the corresponding positions of these two numbers. The mutation takes the current gene value as the initial value, and the chaotic sequence $x(n+1) = 4x(n)\ (1 - x(n))$ is used to perform a number of iterations to obtain the new gene after mutation; thereby, a new chromosome is obtained.

4.3. Machine Learning Model. The following four well-known machine learning models are not the focus of this article; thus, they are briefly described.

4.3.1. K-Nearest Neighbor. K-nearest neighbor (KNN) is a classification algorithm based on the distance between samples in space. The basic idea of the KNN algorithm is as follows. Given a test sample, k training samples closest to it are obtained, and the prediction based on the information of these k neighbors is made. Generally, the voting method can be used for the prediction. The prediction result is the class label that appears most in the k samples. The schematic diagram of KNN algorithm is shown in Figure 3.

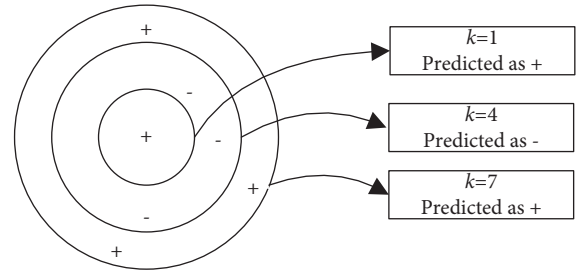

Figure 3: Schematic diagram of K-nearest neighbor algorithm.

The steps of KNN are as follows.

(1) The distances between every test sample and training samples are calculated.

(2) The distances are sorted by ascending order.

(3) k training samples with the smallest distances are selected.

(4) The frequencies of the k samples in different classes are obtained.

(5) The class with the highest frequency is the predicted class of the test sample.

4.3.2. Decision Tree. Decision tree is a tree-like structure that can classify a sample through multiple levels of nodes. Among them, each internal node represents the judgment of a feature, and the classification result is finally output by the leaf node. The decision tree usually divides the sample attributes according to the purity of the dataset. The decision tree used in our paper uses the Gini index to measure the purity of the dataset. The Gini index formula is as follows.

$$\text{Gini}(X) = \sum_{k=1}^{n} \sum_{i \neq k} p_k p_i, \tag{4}$$

where Gini(X) is the Gini index of dataset X, n is the number of samples in the dataset, and P_k is the probability that the sample belongs to class K. The above formula means the sum of the probabilities of two samples randomly selected from the dataset in different classes. The smaller the Gini index, the higher the purity of the dataset.

The generation process of a decision tree is mainly divided into the following three parts.

(1) Feature selection

A feature is chosen from many features in the training data as the split criterion of the current node. Different feature selection methods derive different decision tree algorithms. This article uses information gain to divide nodes.

(2) Decision tree generation

According to the selected feature evaluation criteria, child nodes are generated recursively from top to bottom, and the decision tree stops growing until the data set is indivisible.

(3) Pruning

The decision tree has been over-fitting, and it is necessary to reduce the size of the tree structure through pruning. There are two commonly used pre-pruning and backward pruning.

Figure 4 is a schematic diagram of a decision tree. Rectangles represent the internal nodes, and ellipses represent the leaf nodes.

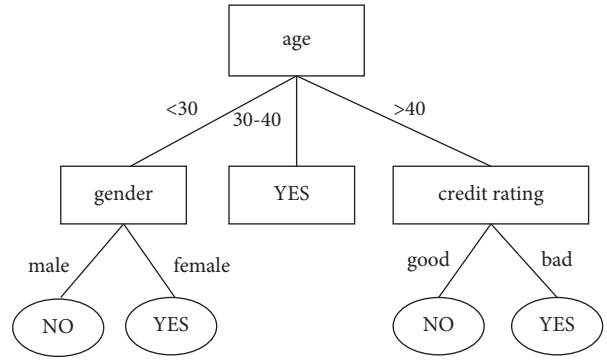

Figure 4: Schematic diagram of decision tree.

4.3.3. Support Vector Machine. Support vector machine (SVM) is an algorithm that can classify samples by maximally spaced hyperplanes. Its goal is to find a support vector that is the most distant from the samples. The maximal interval hyperplane is found by solving the optimization problem in (5). The objective function of the support vector machine is shown as follows.

$$\begin{aligned} &\min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}\xi_i, \\ &s.t.\ y_i\left[(w \cdot x_i + b)\right] \geq 1 - \xi_i, \end{aligned} \tag{5}$$

where w is the weight vector, b is the displacement term, ξ is the slack variable, and $C > 0$ is the penalty factor. In (4), the first expression is the optimization goal, and the second expression is the constraint condition. (4) means minimizing the square of the modulus of the weight matrix under constraint conditions. Figure 5 is a schematic diagram of SVM.

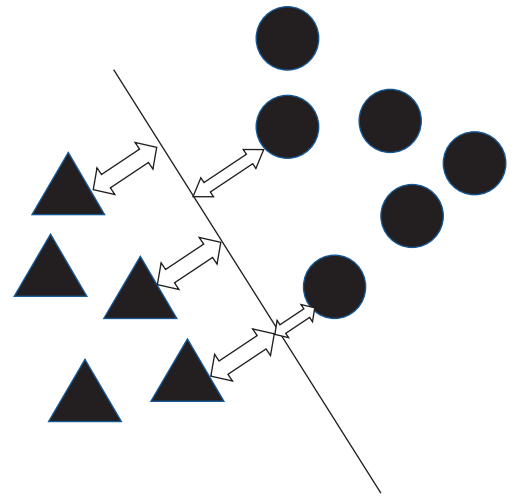

Figure 5: Schematic diagram of SVM.

4.3.4. Logistic Regression. The basic idea of logistic regression (LR) is to judge the classification by finding the relationship between the classification probability and the input vector. Firstly, it is assumed that the data obey a certain distribution, and then the maximum likelihood estimation is used for parameter estimation.

$$\ln\frac{p}{1-p} = \alpha + \beta_1 x_1 + \cdots + \beta_h x_h. \tag{6}$$

Where p is the default probability, $1-p$ is the nondefault probability, α is the constant in the model, β_h is the logistic regression coefficient, x_h is the hth independent variable.

The maximum likelihood method is used to estimate the parameters α and β. Given the data set $\{(x_i, y_i)\}_{i=1}^{N}$, the logarithmic likelihood function can be expressed as

$$\ln\ L(\beta) = \sum_{i=1}^{N}\left(y_i \ln\ p_i\right) + \left(1 - y_i\right)\ln\left(1 - p_i\right)). \tag{7}$$

The maximum likelihood function of (5) is equivalent to the following minimum cost function:

$$J(\beta) = \mathrm{In}L(\beta) - \frac{1}{N}\sum_{i=1}^{N}\left(y_i \ln\ s_i\right) + \left(1 - y_i\right)\ln\left(1 - s_i\right)). \tag{8}$$

In this paper, the logistic regression parameters are estimated using gradient descent. The main goal of parameter optimization is to find a direction in which the value of the cost function can be reduced after the parameters move in this direction.

4.4. Convolutional Neural Network. To explore the application of deep learning algorithms to default risk prediction, we choose the convolutional neural network (CNN), a well-known model in computer vision, to compare it with the abovementioned four traditional machine learning algorithms.

As a representative algorithm of deep learning, CNN is widely used in the field of image and audio in the era of big data. In recent years, convolutional neural networks have also produced many research results in the area of default prediction. The CNN constructed in this study includes the convolutional layer (CONV), linear rectified function (ReLU), and full-connection layer (FC) for extracting features, as shown in Figure 6.

4.4.1. Convolutional Layer. The Convolutional layer is a unique structure of CNN, which can be used to extract data features. The convolution layer contains multiple convolution kernels. Since the object of study in this paper is one-dimensional data, the one-dimensional convolution kernel is used. The convolution layer uses the convolution kernel to

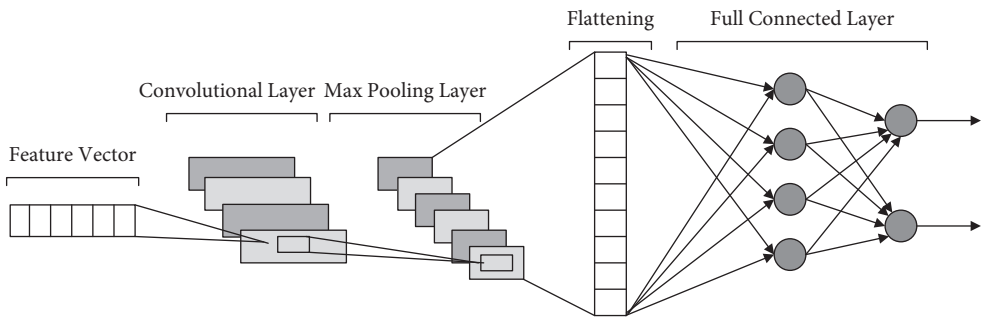

Figure 6: Convolutional neural network architecture.

perform the convolution operation on the input matrix and convert the input matrix to the output matrix. The convolution operation process shown in Figure 7 is illustrated by taking one-dimensional convolution as an example.

The weight matrix of the convolution kernel starts from the initial position of the input matrix, calculates the sum of the product of the weight matrix and the corresponding position elements of the input matrix, and fills the result into the output matrix. The model self-adjusted the weight matrix through the backpropagation (BP) algorithm. After each calculation, the convolution kernel would move on to calculate the value of the next element of the output matrix until all calculations are completed.

A 10-dimensional vector shown in Figure 8 represents the input matrix. Moreover, the weight matrix of the convolution kernel is represented by a three-dimensional vector shown in Figure 9, and the size of the convolution kernel is 1×3. Then the first element of the output matrix is as follows.

$$3 \times 7 + 6 \times 4 + 9 \times 2 = 63. \tag{9}$$

The second element of the output matrix is

$$6 \times 7 + 9 \times 4 + 10 \times 2 = 98. \tag{10}$$

The output matrix is finally obtained by analogy, as shown in Figure 10.

Let the dimension of the output vector of the convolutional layer be n_{out} (in the one-dimensional case), then the size of the output matrix of the convolutional layer can be calculated by the following formula.

$$n_{\text{out}} = \frac{n_{\text{in}} + 2p - k}{s} + 1, \tag{11}$$

where n_{in} is the dimension of the input vector, p is the padding, k is the size of the convolution kernel, and s is the step. Padding = 0 and $s = 1$ in this paper.

4.4.2. Linear Rectifying Function. The linear rectifier function is a commonly used activation function, which is the slope function in this paper, and its expression is

$$f(x) = \max(0, x). \tag{12}$$

Where $f(x)$ is a linear rectifying function and x is an input matrix. The meaning of the above equation is that for each element value in the input matrix, if greater than 0, it remains unchanged; otherwise, it becomes 0.

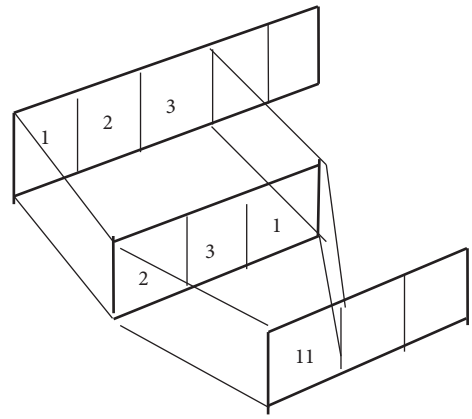

Figure 7: Schematic diagram of the convolution operation.

3	6	9	10	1	2	4	7	3	5

Figure 8: Input vector.

7	4	2

Figure 9: Weight matrix.

63	98	105	78	23	44	62	71

Figure 10: Output matrix.

4.4.3. Full-Connected Layer. The full-connected layer lies behind the convolution layer and acts as a feed-forward neural network. After the output matrix of the convolution layer is input to FC, the Softmax function can calculate the default probability. The expression of the Softmax function is

$$P(y = k|x) \frac{\exp\left(x^T w_k\right)}{\sum_{i=1}^{m} \exp\left(x^T w_i\right)}. \tag{13}$$

Where x^T is the transpose of the input matrix, P is the probability that the sample belongs to class k, and w_k is the weight vector under class k. The above equation means that for each input x, Softmax calculates the probability P of its being divided into each classification k. When the probability of default exceeds the threshold, the sample would be predicted as a default sample.

4.5. Model Testing. The model evaluations adopted in this paper are as follows.

(1) $F1$-score is the harmonic mean of precision and recall and is shown in.

$$F1 - \text{Score} = \frac{2 \times TP/(TP + FP) \times TP/(TP + FN)}{TP/(TP + FP) + TP/(TP + FN)}. \quad (14)$$

(2) The ratio of default samples wrongly predicted to the nondefault sample, namely Type-II error is

$$\text{Type} - II \text{ error} = \frac{FN}{TP + FN}. \quad (15)$$

(3) Accuracy is the ratio of the number of correctly classified samples to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}. \quad (16)$$

(4) The G-mean is the geometric mean of the correct rates of nondefault and default samples and is shown as follows.

$$G - \text{mean} = \left(\frac{TN}{FP + TN} - \frac{TP}{TP + FN}\right)^{1/2}. \quad (17)$$

(5) The AUC value is equivalent to the probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example[58].

The confusion matrix of prediction results is shown in Table 1.

4.6. Comprehensive Metric Model Based on Multi-Machine Learning Algorithms. The training of CNN is a time-consuming task. Compared with the convolutional neural network, the four machine learning models mentioned above can be easily trained and are simpler to debug. Therefore, before training the CNN, we first use the machine learning models to evaluate the datasets with different balance ratios and feature combinations. The average ranking of multiple metrics determines the best balance ratio and feature combination. This result is applied to CNN to improve the performance of the deep learning model. The basic process of our approach is as follows.

(1) An unprocessed copy of the original training set is denoted by T. The copy of the original training set processed by feature selection is denoted as T_0, and its corresponding feature combination is denoted as F_0.

(2) Rebalancing T, the balance ratios are set to 1 : 1, 2 : 1, . . ., 10 : 1, and ten training sets are denoted as T_1, T_2, . . ., T_{10}.

(3) The ten training sets are processed by feature selection to obtain T_{11}, T_{12}, . . ., T_{20} and their corresponding feature combinations are F_1, F_2, . . ., F_{10}.

(4) Applying the same rebalancing process as above to T_0, the ten training sets obtained are denoted as T_{21}, T_{22}, . . ., T_{30}. Ultimately, we get 32 different training sets and 11 different feature combinations. Namely, T is a dataset without rebalancing and feature selection (a copy of the original dataset), T_0 is a feature selected dataset without rebalancing, T_1–T_{10} are balanced in different balance ratios, but no feature section, T_{11}–T_{20} are rebalanced followed by feature section, T_{21}–T_{30} are feature selected followed by rebalancing. In addition, we call these 32 training sets derivative datasets of the original dataset.

(5) We input the 32 derived datasets into KNN, DT, SVM, and LR and calculate the accuracy, G-mean, Type-II error, and AUC of the four models tested on the same test set after training the derived datasets.

(6) The four metrics of a machine learning algorithm in different training sets are ranked, and the comprehensive ranking of the performance of the machine learning algorithm in different training sets is obtained by averaging the rankings of the four evaluations.

(7) Averaging the comprehensive ranking of the four machine learning algorithms on the derived training set can obtain the score of the comprehensive metric model of multimachine learning algorithms. The derivative datasets with the highest rank can be found, and their balanced ratio and feature combination are the best.

(8) Training the CNN with the selected derivative dataset and testing it on the same test set.

The flowchart of our approach is shown in Figure 11.

Table 1: Confusion matrix.

Actual default status	Prediction		Total
	1 (default)	0 (nondefault)	
1 (default)	True positive (TP)	False negative (FN)	TP + FN
0 (non-default)	False positive (FP)	Ture negative (TN)	TN + FP
Total	TP + FP	FN + TN	T

5. Empirical Analysis

5.1. The Empirical Research. In this section, the dataset (China Dataset) is collected from the credit database of a regional commercial bank in China. The data contain 80 features, which can be divided into three levels (i.e., internal financial, nonfinancial, and external macro factors) and nine second-level criterion layers, including solvency,

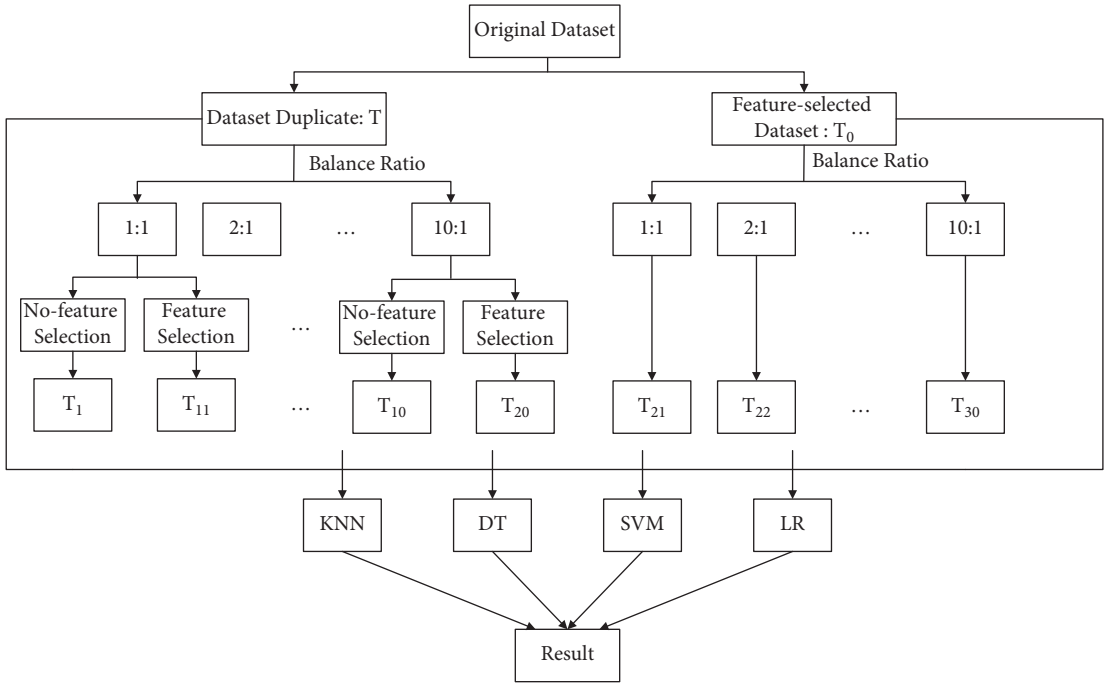

Figure 11: Flowchart of the comprehensive metric model based on multimachine learning algorithms.

profitability, and operating capability, among others. This dataset contains 2,995 nondefault enterprises and 50 default enterprises, with a default rate of 0.0164. The details of the China Dataset are shown in Table 2.

Table 3 shows the parameters we used for the algorithms. In this study, the random search method [59] is used to select the best hyperparameters of CNN.

The original training set of the China Dataset is denoted by C. C_0 is obtained by feature selection of C. After the balance treatment of C, ten training sets are obtained, which are denoted as C_1, C_2, …, C_{10}. By performing feature selection on the ten datasets, we can obtain C_{11}, C_{12}, …, C_{20}. Then, the ten datasets obtained by balancing C_0 are denoted as C_{21}, C_{22}, …, C_{30}. These 32 derived datasets are input into our proposed approach. Table 4 shows the rank of the comprehensive metric model based on multimachine learning algorithms on the 32 derived datasets.

Table 4 shows that through the results of our proposed method on different datasets, we can conclude that the optimal balance ratio of default data of small enterprises is 2 : 1, and the optimal feature combination includes 31 features shown in Table 5.

The research results in Table 5 show that, for the default prediction of Chinese small enterprises, the best feature combination includes 31 features. Next, we train the CNN with the C_{12} dataset. In this experiment, the CNN has two convolutional layers and four fully connected layers. The first convolution layer consists of two one-by-five convolution kernels. The second convolution layer is the same as the first one. Because the C_{12} dataset contains only 35 features, the padding for the convolution operation is 0, and the step size is 1, which means that the size of the output matrix is 1×27 after processing by two convolution layers. The number of input neurons in the first FC is 27, and the number of output neurons is 15. The number of output neurons in the second FC is 9. Additionally, the third and fourth output neurons are 3 and 1, respectively. The performance of the CNN and its ranking with the other four models are shown in Table 6.

It can be concluded from Table 6 that it is feasible to use our method to find the best balance ratio and the optimal feature combination. The bold value means that the method ranks first in some evaluation criterion. CNN ranks first in terms of G-mean, accuracy, and AUC, which indicates that its comprehensive ranking also ranks first. CNN shows excellent predictive ability on the selected dataset.

5.2. Robustness Test

5.2.1. The Comparison of Different Datasets. In this section, three datasets are collected from the UCI Machine Learning Repository to evaluate the robustness of our method. Table 7 shows the details of the datasets.

To verify the generalization ability of our method, we apply it to the Japan Dataset, Australia Dataset, and Chile Dataset to perform research. The three datasets are denoted as J, A, and CH, respectively. We pretreated the abovementioned three datasets in the same way as in Section 5.1. The experimental results are shown in Table 8.

The abovementioned results show that the best balance ratio of the Japanese dataset is 3 : 1, the Australian dataset is 1 : 1, and the Chilean dataset is 3 : 1. Next, we bring the experimental results into the CNN. Because these three datasets have few features in this experiment, the CNN only contains one convolutional kernel and three fully connected layers. Due to the different numbers of features in the three datasets, the number of neurons in the FC is slightly different, but the structures of the three networks for the three datasets are the same. The experimental results are shown in Tables 9–11.

Table 2: The feature data of China dataset.

Serial number	First-level criterion layer	Second-level criterion layer	Feature	3,111 original data			3,045 standardization data		
				Sample 1	...	Sample 3,111	Sample 1	...	Sample 3,045
1	Financial factors	Solvency	Asset-liability ratio	6.923	...	0.556	0.328	...	0.419
...			...	...	...	...	...	...	...
9		Profitability	Return on equity	0	...	0.003	0.053	...	0.003
...			...	...	...	...	...	...	...
35		Operation ability	Inventory turnover rate	3	...	8.580	0.005	...	0.01462
...			...	...	...	...	...	...	...
45		Growth ability	Profit growth rate	0	...	1.695	0.494	...	0.530
...			...	...	...	...	...	...	...
52	Nonfinancial factors	Nonfinancial factors within the enterprise	Patent status	NA	...	NA	0	...	0
...			...	...	...	...	...	...	...
58		Basic information of legal representative	Education background	Bachelor	...	Bachelor	0	...	0.663
...			...	...	...	...	...	...	...
69		Basic credit status of the enterprise	Type of registered capital in place	Capital type registered capital	...	Capital type registered capital	0.685	...	1
...			...	...	...	...	...	...	...
71		Business reputation of the enterprise	Corporate tax records	No record of tax arrears	...	No record of tax arrears	0.669	...	0.669
...			...	...	...	...	...	...	...
75	External macro factors	External macro conditions of enterprises	Industry climate index	127.960	...	139.500	0.626	...	0.742
...			...	...	...	...	...	...	...
80			Engel coefficient	39.400	...	37.000	0.651	...	0.755

Table 3: The parameters of algorithms.

Algorithm	Parameter	Value
KNN	K-neighbors	5
DT	Max depth	12
	Split criterion	Gini
SVM	Kernel function	Rbf
	Kernel scale	2
LR	Objective function minimization	SGD
	Learning rate	10^{-3}
	Iteration	10^{4}
	Epoch	10–100
CNN	Initial learning rate(AdaGrad)	2×10^{-6}
	Batch size	32–256
	Padding mode	Same
	Kernel size	1×7
SMOTE	K-neighbors	5

For the Japanese dataset, the CNN's *G*-mean, Type II error, and AUC rank first. Although DT obtains the highest accuracy, Type-II error is 3.3 times higher than that of CNN. Although the accuracy of CNN is not the first for the Australian dataset, it is close to first place, while the other three metrics are all ranked first. With regard to the Chilean dataset, the *G*-mean of SVM is 0; although Type-II error ranks first, it is worthless. The AUC of the CNN is very close to first place, while the *G*-mean and accuracy rank first. It can be considered that the comprehensive performance of the convolutional neural network is the best. We can conclude that the convolutional neural network performs well on the three public datasets processed by our method and is superior to the other three machine learning models. Thus far, we have discovered that our proposed comprehensive metric model based on multimachine learning algorithms has good robustness.

5.2.2. The Comparison of Different Resampling Methods. To evaluate the performance of our proposed method, we perform comparison experiments on the China Dataset in terms of undersampling and oversampling [14]. In this section, the SMOTE method is replaced by two methods to evaluate the robustness of our method.

For undersampling, the CNN's accuracy and G-mean rank first, DT's AUC ranks first, and SVM's Type-II error ranks first. For oversampling, the CNN's G-mean and AUC rank first, DT's Type-II error ranks first, and SVM's accuracy ranks first. It can be seen from the comparison in Table 12 that CNN has better classification performance than the other four models in both undersampling and oversampling, which verifies the robustness of the model proposed in this study from the aspect of sample processing.

5.2.3. The Comparison of Different Feature Selection Methods. To evaluate the performance of our proposed method, we perform comparison experiments on the China Dataset in terms of FDAF-score [2], correlation coefficient

Table 4: The rank of four classifiers on different datasets (China dataset)

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Dataset	C	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	C_{27}	C_{28}	C_{29}	C_{30}
Rank	31	32	21	15	9	5	8	7	16	19	17	19	6	**1**	11	3	12	10	14	2	4	12	25	22	27	23	28	17	25	24	30	29

Table 5: Optimal feature combination system for small business default risk prediction.

1	2	3	4	5	6	7	8
Asset liability ratio	Quick ratio	Current ratio	Equity ratio	Super-quick ratio	Net assets to year-end loan balance ratio	Cash ratio	Long-term asset suitability ratio
9	10	11	12	13	14	15	16
Outstanding loans to total asset ratio	Net cash ratio f operating non-current liability	Gross profit rate	Cost margin	Net cash flow from operating activities	Current assets turnover speed	Working capital ratio	Return on investment
17	18	19	20	21	22	23	24
Account payable turnover rate	Cash cycle	Revenue growth rate	Capital accumulation ratio	Years of working in related industries	Patent status	Account opening status with the bank	Credit card records of the legal representative
25	26	27	28	29	30	31	
Marital status	Dwelling condition	Duration of holding the position	Type of registered capital	Enterprise credit granting in recent three years	GDP growth rate	Per capita disposable income of urban residents	

The bold value means that the method ranks first in some evaluation criterion. For example, CNN ranks first in terms of G-mean, accuracy, and AUC, which indicates that its comprehensive ranking also ranks first.

Table 6: The rank of performance of CNN and other four models on C_{12}.

Method	*G*-mean	Accuracy	Type-II error	AUC
CNN	**0.917**	**0.931**	0.101	**0.917**
KNN	0.853	0.822	0.179	0.853
DT	0.853	0.822	0.182	0.853
SVM	0.903	0.874	0.063	0.904
LR	0.819	0.722	**0.044**	0.829

Table 7: Information of datasets.

Dataset	Non-default	Default	Attribute	Default rate
Japan dataset	383	307	15	0.4449
Australia dataset	383	307	14	0.4449
Chile dataset	8000	2000	21	0.2

Table 8: The rank of four models on different datasets.

ID	Dataset	Rank	Dataset	Rank	Dataset	Rank
1	J	31	A	32	Ch	27
2	J_0	32	A_0	29	Ch_0	31
3	J_1	5	A_1	15	Ch_1	6
4	J_2	10	A_2	18	Ch_2	9
5	J_3	11	A_3	20	Ch_3	6
6	J_4	20	A_4	16	Ch_4	19
7	J_5	18	A_5	26	Ch_5	15
8	J_6	28	A_6	28	Ch_6	11
9	J_7	22	A_7	24	Ch_7	22
10	J_8	26	A_8	31	Ch_8	15
11	J_9	26	A_9	27	Ch_9	13
12	J_{10}	29	A_{10}	29	Ch_{10}	17
13	J_{11}	16	$\mathbf{A_{11}}$	**1**	Ch_{11}	4
14	J_{12}	4	A_{12}	7	Ch_{12}	10
15	$\mathbf{J_{13}}$	**1**	A_{13}	4	$\mathbf{Ch_{13}}$	**1**
16	J_{14}	12	A_{14}	11	Ch_{14}	6
17	J_{15}	7	A_{15}	2	Ch_{15}	3
18	J_{16}	15	A_{16}	9	Ch_{16}	2
19	J_{17}	23	A_{17}	22	Ch_{17}	13
20	J_{18}	12	A_{18}	22	Ch_{18}	20
21	J_{19}	18	A_{19}	13	Ch_{19}	21

Table 8: Continued.

ID	Dataset	Rank	Dataset	Rank	Dataset	Rank
22	J_{20}	6	A_{20}	16	Ch_{20}	24
23	J_{21}	9	A_{21}	13	Ch_{21}	11
24	J_{22}	17	A_{22}	10	Ch_{22}	24
25	J_{23}	3	A_{23}	7	Ch_{23}	4
26	J_{24}	2	A_{24}	5	Ch_{24}	18
27	J_{25}	12	A_{25}	3	Ch_{25}	26
28	J_{26}	23	A_{26}	12	Ch_{26}	23
29	J_{27}	21	A_{27}	21	Ch_{27}	27
30	J_{28}	8	A_{28}	19	Ch_{28}	29
31	J_{29}	25	A_{29}	6	Ch_{29}	30
32	J_{30}	30	A_{30}	25	Ch_{30}	31

The bold value means that the method ranks first in some evaluation criterion. For example, CNN ranks first in terms of G-mean, accuracy, and AUC, which indicates that its comprehensive ranking also ranks first.

Table 9: The performance of CNN and other four models on J_{13}.

Method	*G*-mean	Accuracy	Type-II error	AUC
CNN	**0.775**	0.727	**0.143**	**0.779**
KNN	0.663	0.785	0.476	0.682
DT	0.679	**0.818**	0.476	0.702
SVM	0.741	0.678	**0.143**	0.749
LR	0.735	0.719	0.238	0.736

The bold value means that the method ranks first in some evaluation criterion. For example, CNN ranks first in terms of G-mean, accuracy, and AUC, which indicates that its comprehensive ranking also ranks first.

Table 10: The performance of CNN and other four models on A_{11}.

Method	*G*-mean	Accuracy	Type-II error	AUC
CNN	**0.889**	0.883	**0.113**	**0.890**
KNN	0.849	0.883	0.221	0.851
DT	0.858	**0.900**	0.231	0.863
SVM	0.880	0.867	**0.110**	0.881
LR	0.880	0.867	**0.110**	0.881

The bold value means that the method ranks first in some evaluation criterion. For example, CNN ranks first in terms of G-mean, accuracy, and AUC, which indicates that its comprehensive ranking also ranks first.

Table 11: The performance of CNN and other four models on CH_{13}.

Method	*G*-mean	Accuracy	Type-II error	AUC
CNN	**0.655**	**0.507**	0.082	0.693
KNN	0.553	0.376	0.043	0.639
DT	0.411	0.245	1.000	0.576
SVM	0.000	0.091	**0.000**	0.500
LR	0.652	0.496	0.061	**0.696**

The bold value means that the method ranks first in some evaluation criterion. For example, CNN ranks first in terms of G-mean, accuracy, and AUC, which indicates that its comprehensive ranking also ranks first.

Table 12: The comparison experiments of undersampling and oversampling on China dataset.

Method	G-mean	Accuracy	Type-II error	AUC
Undersampling				
CNN	**0.935**	**0.924**	0.171	0.910
KNN	0.875	0.867	0.224	0.893
DT	0.884	0.872	0.195	**0.911**
SVM	0.932	0.905	**0.108**	0.907
LR	0.872	0.731	0.151	0.836
Oversampling				
CNN	**0.912**	0.935	0.117	**0.920**

Table 12: Continued.

Method	G-mean	Accuracy	Type-II error	AUC
KNN	0.873	0.842	0.195	0.913
DT	0.903	0.892	**0.107**	0.885
SVM	0.914	**0.940**	0.119	0.912
LR	0.819	0.722	0.118	0.893

The bold value means that the method ranks first in some evaluation criterion. For example, CNN ranks first in terms of G-mean, accuracy, and AUC, which indicates that its comprehensive ranking also ranks first.

Table 13: The comparison experiments of feature selection methods on China dataset.

Method	*G*-mean	Accuracy	Type-II error	AUC
FDAF-score				
CNN	**0.904**	**0.927**	0.124	**0.921**
KNN	0.868	0.845	0.177	0.891
DT	0.895	0.900	0.121	0.899
SVM	0.900	0.914	**0.116**	0.904
LR	0.857	0.876	0.145	0.884
Correlation coefficient				
CNN	**0.920**	**0.912**	0.102	**0.931**
KNN	0.894	0.882	0.178	0.921
DT	0.903	0.904	0.131	0.864
SVM	0.923	0.900	0.115	0.922
LR	0.891	0.895	**0.097**	0.924
Particle swarm optimization algorithm				
CNN	**0.929**	**0.935**	0.101	**0.935**
KNN	0.867	0.852	0.203	0.908
DT	0.907	0.912	0.147	0.879
SVM	0.924	0.934	0.119	0.921
LR	0.921	0.927	**0.093**	0.893
Artificial bee colony algorithm				
CNN	**0.941**	0.925	**0.067**	0.935
KNN	0.888	**0.929**	0.178	0.894
DT	0.915	0.878	0.187	0.917
SVM	0.921	0.930	0.075	**0.942**
LR	0.845	0.884	0.238	0.876

The bold value means that the method ranks first in some evaluation criterion. For example, CNN ranks first in terms of G-mean, accuracy, and AUC, which indicates that its comprehensive ranking also ranks first.

[24], particle swarm optimization algorithm [31], and artificial bee colony algorithm [33]. In this section, the GA method is replaced by four methods to evaluate the robustness of our method.

In this study, four feature selection methods (i.e., the FDAF-score, correlation coefficient, particle swarm optimization algorithm, and artificial bee colony algorithm) are used to verify the robustness of the model proposed in this study from the aspect of feature selection. For the FDAF-score, correlation coefficient, and particle swarm optimization algorithm, the CNN ranks first in terms of G-mean, accuracy, and AUC. For the artificial bee colony algorithm, the CNN ranks first in terms of G-mean, Type-II error. The empirical results in Table 13 show that CNN performs better than the other four classification models.

6. Conclusion and Future Research

The main conclusions of this study are as follows.

(1) The performance of models with different balance ratios and feature combinations is different for an enterprise credit dataset. There is always the best balance ratio and its corresponding optimal feature combination for each dataset. Based on our proposed method, we can find the best results among the 32 derived datasets.

(2) Overall, the dataset whose selection order is data rebalancing followed by feature selection always achieves a better result, which means that it is the best selection order.

(3) The best balance ratio and our method's corresponding optimal feature combination are also suitable for the convolutional neural network. In addition, the performance of CNN is better than that of traditional machine learning models in the task of default risk prediction.

The primary contributions of this study are as follows.

(1) Different balance ratios will result in different classification accuracies, and there is bound to be an optimal balance ratio. In this study, the optimal

balance ratio is found through the classification accuracy, which changes the deficiency of the existing research that samples are imbalanced or the balance ratio is 1 : 1 and ensures the accuracy of the classification model.

(2) The order of data balance and feature selection affects the model accuracy. This study proposes a comprehensive metric model based on the machine learning algorithm, which can simultaneously find the best balance ratio and perform the optimal feature selection.

Further studies may include the use of a new feature selection method. Although we make two improvements to the GA to avoid falling into the local optimization, it cannot guarantee that the results are globally optimal. As a heuristic algorithm, GA relies on the initial condition that makes it unstable. Thus, a robust and effective optimization method is our next goal. In the future, feature selection methods such as principal component analysis, rough set, or lasso regression may be an option for redundant features.

Data Availability

The Chinese data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge financial support from the Key Programs of National Natural Science Foundation of China (71731003), the General Programs of National Natural Science Foundation of China (72071026), the General Program of National Social Science Foundation of China (21BJY065), the Youth Programs of National Natural Science Foundation of China (71902077), the University Level Training Program for Innovation and Entrepreneurship (2020101410601041378).

References

[1] W. H. Hou, X. K. Wang, H. Y. Zhang, J. Q. Li, and L. Li, "A novel dynamic ensemble selection classifier for an imbalanced data set: an application for credit risk assessment," *Knowledge-Based Systems*, vol. 208, Article ID 106462, 2020.

[2] Q. Song, H. Jiang, and J. Liu, "Feature selection based on FDA and F-score for multi-class classification," *Expert Systems with Applications*, vol. 81, pp. 22–27, 2017.

[3] Y. M. Huang, C. M. Jiau, and H. C. Jiau, "Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem," *Nonlinear Analysis: Real World Applications*, vol. 7, no. 4, pp. 720–747, 2006.

[4] C. L. Huang, M. C. Wang, and C. J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications*, vol. 33, no. 4, pp. 847–856, 2007.

[5] I. Mues and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3446–3453, 2012.

[6] H. Kvamme, N. Sellereite, K. Sjursen, and S. Sjursen, "Predicting mortgage default using convolutional neural networks," *Expert Systems with Applications*, vol. 102, pp. 207–217, 2018.

[7] L. Yu, R. Zhou, L. Chen, and R. Chen, "A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data," *Applied Soft Computing*, vol. 69, pp. 192–202, 2018.

[8] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Niu, and X. Niu, "Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning," *Electronic Commerce Research and Applications*, vol. 31, pp. 24–39, 2018.

[9] W. Li, S. Ding, Y. Chen, H. Yang, and S. Yang, "Transfer learning-based default prediction model for consumer credit in China," *The Journal of Supercomputing*, vol. 75, no. 2, pp. 862–884, 2019.

[10] M. Lardy and J. P. Lardy, "Credit spread approximation and improvement using random forest regression," *European Journal of Operational Research*, vol. 277, no. 1, pp. 351–365, 2019.

[11] Y. Zhao, Y. Huang, and Y. Huang, "Dmdp: a dynamic multi-source default probability prediction framework," *Data Science and Engineering*, vol. 4, no. 1, pp. 3–13, 2019.

[12] X. Fu, T. Ouyang, J. Luo, and X. Luo, "Listening to the investors: a novel framework for online lending default prediction using deep learning neural networks," *Information Processing & Management*, vol. 57, no. 4, Article ID 102236, 2020.

[13] N. V. Chawla, K. W. Bowyer, L. O. Kegelmeyer, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[14] J. Hernandez, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "An empirical study of oversampling and undersampling for instance selection methods on imbalance datasets," in *Proceedings of the Pattern Recognition, Image Analysis, Computer Vision, and Applications(CIARP)*, pp. 262–269, Havana, Cuba, November 2013.

[15] I. Lai-Yuen and S. K. Lai-Yuen, "Adaptive semi-unsupervised weighted oversampling (A-SUWO) for imbalanced datasets," *Expert Systems with Applications*, vol. 46, pp. 405–416, 2016.

[16] X. W. Liang, A. P. Jiang, T. Li, Y. Y. Xue, and G. T. Wang, "LR-SMOTE—an improved unbalanced data set oversampling based on K-means and SVM," *Knowledge-Based Systems*, vol. 196, Article ID 105845, 2020.

[17] Z. Jiang, T. Pan, C. Yang, and J. Yang, "A new oversampling method based on the classification contribution degree," *Symmetry*, vol. 13, no. 2, p. 194, 2021.

[18] W. Zhang, D. Zhang, and S. Zhang, "A new hybrid ensemble model with voting-based outlier detection and balanced sampling for credit scoring," *Expert Systems with Applications*, vol. 174, Article ID 114744, 2021.

[19] F. Shen, X. Zhao, G. Alsaadi, 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, Article ID 106852, 2021.

[20] F. L. Chen and F. C. Li, "Combination of feature selection approaches with SVM in credit scoring," *Expert Systems with Applications*, vol. 37, no. 7, pp. 4902–4909, 2010.

[21] S. Oreski and G. Oreski, "Genetic algorithm-based heuristic for feature selection in credit risk assessment," *Expert Systems with Applications*, vol. 41, no. 4, pp. 2052–2064, 2014.

[22] R. Sheikhpour, M. A. Sarram, S. Chahooki, and M. A. Z. Chahooki, "A survey on semi-supervised feature selection methods," *Pattern Recognition*, vol. 64, pp. 141–158, 2017.

[23] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Moore, and J. H. Moore, "Relief-based feature selection: introduction and review," *Journal of Biomedical Informatics*, vol. 85, pp. 189–203, 2018.

[24] L. Hu, W. Gao, K. Zhao, P. Wang, and F. Wang, "Feature selection considering two types of feature relevancy and feature interdependency," *Expert Systems with Applications*, vol. 93, pp. 423–434, 2018.

[25] N. Kozodoi, S. Lessmann, K. Papakonstantinou, Y. Baesens, and B. Baesens, "A multi-objective approach for profit-driven feature selection in credit scoring," *Decision Support Systems*, vol. 120, pp. 106–117, 2019.

[26] M. Luo, F. Nie, X. Chang, Y. Yang, A. G. Zheng, and Q. Zheng, "Adaptive unsupervised feature selection with structure regularization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 944–956, 2018.

[27] Y. Zhang, D. W. Gong, X. Y. Guo, and Y. N. Guo, "A PSO-based multi-objective multi-label feature selection method in classification," *Scientific Reports*, vol. 7, no. 1, p. 376, 2017.

[28] M. M. Mafarja, D. Eleyan, I. Jaber, A. Hammouri, and S. Mirjalili, "Binary dragonfly algorithm for feature selection," in *Proceedings of the 2017 International Conference on New Trends in Computing Sciences (ICTCS)*, pp. 12–17, Amman, Jordan, January 2018.

[29] S. Gu, R. Jin, and Y. Jin, "Feature selection for high-dimensional classification using a competitive swarm optimizer," *Soft Computing*, vol. 22, no. 3, pp. 811–822, 2018.

[30] M. Mirjalili and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," *Applied Soft Computing*, vol. 62, pp. 441–453, 2018.

[31] L. M. Abualigah, A. T. Hanandeh, and E. S. Hanandeh, "A new feature selection method to improve the document clustering using particle swarm optimization algorithm," *Journal of Computational Science*, vol. 25, pp. 456–466, 2018.

[32] G. I. Sayed, A. E. Azar, and A. T. Azar, "Feature selection via a novel chaotic crow search algorithm," *Neural Computing & Applications*, vol. 31, no. 1, pp. 171–188, 2019.

[33] Y. Zhang, S. Cheng, Y. Shi, D. W. Zhao, and X. Zhao, "Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm," *Expert Systems with Applications*, vol. 137, pp. 46–58, 2019.

[34] M. Ghosh, R. Guha, R. Abraham, and A. Abraham, "A wrapper-filter feature selection technique based on ant colony optimization," *Neural Computing & Applications*, vol. 32, no. 12, pp. 7839–7857, 2020.

[35] Y. Zhang, D. W. Gong, X. Z. Gao, T. Sun, and X. Y. Sun, "Binary differential evolution with self-learning for multi-objective feature selection," *Information Sciences*, vol. 507, pp. 67–85, 2020.

[36] F. Zakaria and M. Zakaria, "Credit rating as a mechanism for capital structure optimization: empirical evidence from panel data analysis," *International Journal of Financial Studies*, vol. 6, no. 1, p. 13, 2018.

[37] K. S. Gyamfi, J. Brusey, A. Gaura, and E. Gaura, "Linear classifier design under heteroscedasticity in linear discriminant analysis," *Expert Systems with Applications*, vol. 79, pp. 44–52, 2017.

[38] L. Tang, F. Ouyang, and Y. Ouyang, "Applying a nonparametric random forest algorithm to assess the credit risk of the energy industry in China," *Technological Forecasting and Social Change*, vol. 144, pp. 563–572, 2019.

[39] E. Dumitrescu, S. Hué, C. Tokpavi, and S. Tokpavi, "Machine learning for credit scoring: improving logistic regression with non-linear decision-tree effects," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1178–1192, 2022.

[40] J. W. Sohn and S. Y. Sohn, "Evaluating borrowers' default risk with a spatial probit model reflecting the distance in their relational network," *PLoS One*, vol. 16, no. 12, Article ID e0261737, 2021.

[41] S. Maldonado, J. Bravo, and C. Bravo, "Cost-based feature selection for Support Vector Machines: an application in credit scoring," *European Journal of Operational Research*, vol. 261, no. 2, pp. 656–665, 2017.

[42] Y. Xia, C. Liu, Y. Liu, 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.

[43] S. Shajari and M. Shajari, "Cost-sensitive payment card fraud detection based on dynamic random forest and k -nearest neighbors," *Expert Systems with Applications*, vol. 110, pp. 381–392, 2018.

[44] Z. Ma, W. Zhang, and D. Zhang, "A credit risk assessment model of borrowers in P2P lending based on BP neural network," *PLoS One*, vol. 16, no. 8, Article ID e0255216, 2021.

[45] M. Mahbobi, S. Vasudevan, and M. Vasudevan, "Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks," *Annals of Operations Research*, vol. 12, 2021.

[46] X. Zhang, Y. Han, W. Wang, and Q. Wang, "HOBA: a novel feature engineering methodology for credit card fraud detection with a deep learning architecture," *Information Sciences*, vol. 557, pp. 302–316, 2021.

[47] U. Fiore, A. De Santis, F. Perla, P. Palmieri, and F. Palmieri, "Using generative adversarial networks for improving classification effectiveness in credit card fraud detection," *Information Sciences*, vol. 479, pp. 448–455, 2019.

[48] J. Momtazi and S. Momtazi, "Ensemble of deep sequential models for credit card fraud detection," *Applied Soft Computing*, vol. 99, Article ID 106883, 2021.

[49] S. Liu and T. Liu, "Performance prediction for higher education students using deep learning," *Complexity*, vol. 2021, pp. 1–10, Article ID 9958203, 2021.

[50] Y. Yu, C. Wang, X. Li, and J. Li, "A novel deep learning-based method for damage identification of smart building structures," *Structural Health Monitoring*, vol. 18, no. 1, pp. 143–163, 2019.

[51] Y. Wang, Y. Han, Y. Wu, E. Korkina, Z. Gagarin, and V. Gagarin, "An occupant-centric adaptive façade based on real-time and contactless glare and thermal discomfort estimation using deep learning algorithm," *Building and Environment*, vol. 214, Article ID 108907, 2022.

[52] Y. Liu, Q. Zeng, J. Ordieres Meré, and H. Yang, "Anticipating stock market of the renowned companies: a knowledge graph approach," *Complexity*, vol. 2019, pp. 1–15, Article ID 920245, 2019.

[53] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Forecasting stock prices from

the limit order book using convolutional neural networks," in *Proceedings of the IEEE 19th Conference on Business Informatics (CBI)*, pp. 7–12, Thessaloniki, Greece, July 2017.

[54] H. Shin and K. S. Shin, "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction," *Neural Computing & Applications*, vol. 32, no. 12, pp. 7897–7914, 2020.

[55] W. Chen, M. Jiang, W. G. Chen, and Z. Chen, "A novel graph convolutional feature based convolutional neural network for stock trend prediction," *Information Sciences*, vol. 556, pp. 67–94, 2021.

[56] R. S. M. Carrasco and M. Á. Sicilia-Urban, "Evaluation of deep neural networks for reduction of credit card fraud alerts," *IEEE Access*, vol. 8, pp. 186421–186432, 2020.

[57] J. I. Z. Lai and K. L. Lai, "Deep convolution neural network model for credit-card fraud detection and alert," *June 2021*, vol. 3, no. 2, pp. 101–112, 2021.

[58] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[59] N. R. Huber, A. D. Missert, H. Gong et al., "Random search as a neural network optimization strategy for Convolutional-Neural-Network (CNN)-based noise reduction in CT," in *Proceedings of the Medical Imaging 2021: Image Processing (SPIE)*, p. 115961U, California United States, February 2021.