

Research Article

Research on Financial Risk Prediction Based on Improved Random Subspace

Yinghui Li 

Henan Finance University, Zhengzhou 450046, China

Correspondence should be addressed to Yinghui Li; liyingshui@hafu.edu.cn

Received 8 November 2021; Revised 9 December 2021; Accepted 7 January 2022; Published 4 March 2022

Academic Editor: Man Fai Leung

Copyright © 2022 Yinghui Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to provide timely and effective information and decision support for financial market entities, combined with random subspace and weight fused Lasso, this paper constructs a financial risk prediction model based on the improved random subspace method. Firstly, the basic principles of random subspace and SVM algorithm are introduced. Then, WFL and AI methods are introduced to improve random subspace, so as to reduce the dimension of multisource heterogeneous data and realize the adaptive fusion of features. Then, a financial risk prediction model based on weighted fusion adaptive random subspace is constructed, in which SVM is used as the basic classifier and the output strategy of result integration is introduced. Finally, based on the data of some listed companies, the improved random subspace method is compared with other methods. The results show that the improved random subspace method has a higher prediction value, which indicates that the method is reasonable and effective in financial risk prediction. In the improved random subspace method, combined feature $F1 + F2 + F3$ is better than other methods in $T-3$, $T-4$, and $T-5$, and the prediction value is more than 95%, which fully demonstrates the rationality of the improved random subspace method in financial risk prediction. The area under the ROC curve (AUC) predicted by weight fused adaptive integration-based random subspace (FAIB_RS) method is about 95% in $T-3$, 93% in $T-4$, and 95.5% in $T-5$, which is obviously higher than that of the other eight methods.

1. Introduction

With the rapid development of the financial industry and the continuous popularization of the internet financial model, the financial market is facing more severe financial risks. In order to improve the ability of financial market subjects to obtain financial risk early warning information, domestic experts and scholars have done a lot of research on financial risk and put forward a variety of financial risk prediction methods. For example, literature [1] used RBF network model to establish Jiangsu Province's financial risk early warning model and predicted the regional total risk in 2019 with the sample set in 2018, so as to obtain the prevention and treatment suggestions of Jiangsu Province's financial risk. Literature [2] constructed the regional financial risk index; evaluated and predicted the financial risk of 31 provinces, cities, and autonomous regions (excluding Hong Kong, Macao, and Taiwan) in China; and concluded that the overall financial risk of China and the 6

regions showed a synchronous trend of change and got the enlightenment of formulating preventive measures. According to the current situation of financial development, literature [3] proposed a multifactor international supply chain financial risk prediction model including external environmental risk factors, obtained the structural characteristics and formation mechanism of international supply chain financial risk system, and put forward countermeasures and suggestions. Most of the above studies are based on a single data source, and the prediction accuracy is low; the effect is not good. Therefore, combined with the financial risk prediction problem, this paper constructs a new weighted fusion adaptive stochastic subspace financial risk prediction method and verifies the rationality and effectiveness of the method. Literature [4] shows that the data of compliant financial activities are of high quality and quantity in all aspects. On the contrary, the data of informal financial activities are of low quality. Therefore, the machine learning method based on

single-source data can only detect the risks of formal financial activities, while the effect of using multisource data to detect financial risks is not good. Then, TSAIB_RS method is proposed to integrate various data adaptively. Literature [5] makes quarterly observations from 225,813 company samples. After investigation, it is found that financial system risks play a great role in predicting the failure of an enterprise when its internal financial sector fluctuates greatly; its scale is small; and its debts are large. Although the integration of classifiers in reference [6] has been applied in the financial industry, there are also some wrong factors that hinder the prediction performance, such as irrelevant features, inclined categories, and so on. In the event of such an error, the cost of the wrong classification is far greater than the cost associated with the non-default or non-insolvency (negative) category. In the future, we need to deeply study the potential relationship between classifier ensemble and positive sample type. In today's "big data" era [7], big data has been gradually integrated with finance and become the core of the financial industry. How to make good use of big data to effectively predict and prevent is of great significance to the financial industry, which is also the essence of financial management, that is, risk management and control. Combined with the advantages of big data prevention and control and prediction, this paper summarizes feasible and effective financial risk management and control countermeasures.

The reasoning method based on belief rule in reference [8] is widely used in risk assessment of research and development (R&D) projects. Because there are many risk factors in the performance evaluation of R&D projects, the BRB method will produce a rule base. Therefore, the stochastic subspace BRB model has been experimented and applied as a new RS-BRB model. It constructs several subspaces from sampling; then develops BRB subsystems according to the subspaces, thus obtaining results; and finally carries out combination mode according to different subsystem results. The traditional forecasting model can no longer meet the current forecasting needs [9], and then researchers put forward a method called RS-multiple boosting to improve the accuracy of forecasting credit risk. This method is a combination of two classical integrated ML methods: random subspace (RS) and multiple boosting. There are many methods for risk assessment and prediction in the financial market, but most of the above studies are based on a single data source, with low prediction accuracy and poor results. The random subspace method can be used for portfolio selection in different data sets, which shows that it is essentially superior to the traditional bagging-based resampling portfolio. In order to solve the problem of single information source, less data application, low prediction accuracy, and poor prediction accuracy of financial risk prediction at present. In this study, a financial risk prediction method based on the adaptive fusion of multisource heterogeneous data is constructed for the financial risk of listed companies and the default risk of individual borrowers, and its effectiveness is verified by using various types of real data sets collected from online platforms. It has a good effect on the prediction effect and accuracy and can solve the problems existing in the financial system. In view of the current

financial data with heterogeneous, redundant, and other issues, the current prediction model in the multisource heterogeneous data prediction is not very satisfactory. This paper proposes a financial risk prediction method based on the adaptive fusion of multisource heterogeneous data, which can effectively predict the multisource data of financial companies and improve the prediction accuracy. In this paper, the problems of financial risk prediction are introduced in detail; then, a financial risk prediction method WFAIB_RS based on the weighted fusion of adaptive random subspace is constructed; finally, through the experimental comparison of the data sets of listed companies, WFAIB_RS has a better prediction effect.

2. Introduction to Basic Methods

2.1. Random Subspace Brief Introduction of Algorithm. Random subspace [10, 11] (RS) is a kind of ensemble learning. Random subspace trains each classifier by using random partial features instead of all features to reduce the correlation between each classifier. Therefore, this method is very suitable for learning tasks with high feature dimensions. The RS steps are as follows:

Step1: according to the feature dimension of data samples, data samples are randomly selected to form data subsets of similar sizes. The subspace ratio parameter r is used to adjust the size of the data subset.

Step2: the sampled data subsets are input into the base classifier and trained.

Step3: finally, according to the training results, the sample results are fused.

2.2. SVM Classification Principle. Using SVM [12] as the base classifier can be solved according to the following objective function:

$$\begin{aligned} s.t. & y_i(\omega^T \varphi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, n, \\ & \zeta_i, \quad (1) \\ \min_{\omega, b} & = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \end{aligned}$$

where C is the normal vector, C is the penalty coefficient, B is the displacement term, and C is the non-negative relaxation factor.

The above problem is transformed into a dual problem:

$$\begin{aligned} s.t. & \sum_{i=1}^n a_i y_i = 0, 0 < a_i < C, \quad i = 1, 2, \dots, n, \\ & (2) \\ \max_a & \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j \varphi(x_i)^T \varphi(x_j). \end{aligned}$$

In the calculation process of dual problem, through the calculation of $\varphi(x_i)^T \varphi(x_j)$, which is mainly the inner product operation after the sum of pairs. Then the inner product operation is input into the kernel function. The expression of the kernel function is as follows:

$$k(x_i, x_j) = \varphi(x_i)^T \varphi(x_j). \quad (3)$$

The final decision function is obtained:

$$y(x) = \text{sign} \left(\sum_{i=1}^n a_i y_i k(x_i, x_j) \right) + b, \quad (4)$$

where a_i and b are constant real numbers; $a_i > 0$. The kernel function $k(i, j)$ using radial basis function kernel function is defined as follows:

$$k(x_i, x_j) = \exp \left(-\gamma \|x_i - x_j\|^2 \right). \quad (5)$$

$$s_{ij} = \text{sgn}(p_{ij}) = \begin{cases} +1, & p_{ij} > 0 \\ -1, & p_{ij} < 0 \end{cases}$$

$$a_{ij} = \frac{p_{ij}}{1 - p_{ij}},$$

$$\beta^* = \arg_{\beta} \min \left\{ \frac{1}{2} \left\| y - \sum_{i=1}^p x_i \beta_i \right\|_2^2 + \lambda |\beta_i| + \frac{\lambda_2}{p} \sum_{i < j} a_{ij} (\beta_i - s_{ij} \beta_j)^2 \right\}. \quad (6)$$

Here, $\lambda_2/p \sum_{i < j} a_{ij} (\beta_i - s_{ij} \beta_j)^2$ is the penalty term added based on the Lasso model, p_{ij} is the correlation coefficient between any two features x_i and x_j , and λ_2 represents the regular penalty parameter, which mainly adjusts the penalty intensity for feature correlation.

2.3. Random Subspace Algorithm Improvement. In order to ensure the accuracy of financial risk prediction and the problem of multisource heterogeneous financial data, the random subspace algorithm needs to be improved first. Therefore, there is a need for multisource heterogeneous data fusion in financial risk prediction data. Based on the Lasso model, the weighted fusion adaptive stochastic subspace model (WFAIB_RS) is introduced for the adaptive fusion of multisource heterogeneous data features. The Lasso model takes the form of

The WFL model can effectively solve the multicollinearity problem between features and improve the stability of the model. After the Lasso model is introduced into WFL, the model form is as follows:

$$\beta^* = \arg_{\beta} \min \left\{ \frac{1}{2} \left\| y - \sum_{i=1}^p x_i \beta_i \right\|_2^2 + \lambda w_i^{(1)} |\beta_i| + \frac{\lambda_2}{p} \sum_{i < j} a_{ij} (\beta_i - s_{ij} \beta_j)^2 \right\}. \quad (7)$$

Here, $w_i^{(1)}$ represents the adaptive weight, which can be added to the WFAL model to obtain a more accurate feature subset; $1/\sqrt{n}$ denotes the smoothing item. λ with the increasing of parameters, more feature weights are 0. As the regular penalty parameter λ_2 becomes larger, more features will be regarded as related features.

The adaptive feature weights $W = (w_1, w_2, \dots, w_p)^T \in R_+^p$ based on weighted fusion are obtained by WFAL model estimation, and then the features are sampled according to the weights, and the data subsets $\{D_{\text{sub}}^1, D_{\text{sub}}^2, \dots, D_{\text{sub}}^M\}$ and $D_{\text{sub}}^i = \{(x_1^i, y_1^i), \dots, (x_j^i, y_j^i), \dots, (x_{p_i}^i, y_{p_i}^i)\}$ that can be used for base classifier training are obtained. The sampling process is mainly adjusted by subspace ratio parameters. The larger the r , the higher the feature dimension of the obtained sample subset.

3. Construction of Financial Risk Prediction Model

3.1. Financial Risk Forecasting Process. Set $D_{\text{train}} = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\}^T$ and test set

$D_{\text{test}} = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\}$. Firstly, Lasso estimation is carried out on the data to obtain the adaptive weight, specifically as follows:

$$w_i^{(1)} = \frac{1}{|\beta_{i\text{Lasso}} + 1/\sqrt{n}|}. \quad (8)$$

Secondly, the adaptive weight is used to estimate the data by WFAL to obtain the feature weight vector $W = (w_1, w_2, \dots, w_p)^T \in R_+^p$.

Thirdly, the feature weight is taken as the sampling probability, and the sampling $\{D_{\text{sub}}^1, D_{\text{sub}}^2, \dots, D_{\text{sub}}^M\}$ of data subset is carried out under the adjustment of subspace ratio r .

Fourthly, the base classifier is trained according to the sampled data subset. Fifthly, according to the rules of evidential reasoning, the prediction results of the base classifier are synthesized to get the final prediction results.

3.2. Base Classifier. Support vector machine (SVM) is selected as the base classifier [13–15]. This method is a typical

classification algorithm, and its characteristic is that it can solve high-dimensional and non-linear classification problems by using the principle of structural risk minimization and can efficiently classify data samples with few samples and high feature dimensions [16, 17]. Therefore, SVM is chosen as the base classifier of the financial risk prediction model.

3.3. Integration Strategy. In the result fusion, evidential reasoning is used as a new ensemble strategy to synthesize the results produced by different base classifiers. Evidential

reasoning is to treat the classification results of different base classifiers as evidence and the accuracy as evidence reliability and initial weight when fusing the results of base classifiers and finally fuse the results by continuous optimization [18, 19].

Firstly, all the classification results are regarded as a set of mutually exclusive identification frames, which can constitute a complete set, and are designated as Θ . The results produced by the base classifier s can then be converted into the following evidence:

$$e_s = \left\{ (y_j, p_{y_j, s}) \mid j = 1, 2, \dots, N, \sum_{j=1}^n p_{y_j, s} = 1 \right\}, \quad s = 1, 2, \dots, S, \quad (9)$$

where e_s represents the evidence of transformation from the classification results obtained from the s -th base classifier and $p_{y_j, s}$ denotes the probability that the classification result goes to y_j .

In the process of evidential reasoning, weight W and reliability R are often used to define the reliability distribution function, which is mainly to ensure that the result information of the base classifier does not conflict. The expression of the reliability distribution function is as follows:

$$m_s = \{(y_j, \tilde{m}_{j, s}); (P(\otimes), \tilde{m}_{P(\otimes), s})\}. \quad (10)$$

Here, $\tilde{m}_{j, s}$ represents the holding degree of evidence e_s considering reliability and weight to y_j , which is defined as follows:

$$\tilde{m}_{y_j, s} = \begin{cases} 0, & y_j = \Phi, \\ c_{r\omega, s} m_{y_j, s}, & y_j \in \Theta, y_j \neq \Phi, \\ c_{r\omega, s} (1 - r_s), & y_j \in P(\Theta), \end{cases} \quad (11)$$

where $m_{y_j, s} = \omega_s p_{y_j, s}$, $c_{r\omega, s} = 1/(1 + \omega_s - r_s)$ is the normalization factor.

Secondly, by fusing S pieces of evidence e_1, e_2, \dots, e_n provided by different base classifiers, the reliability function $p_{y_j, e(s)}$ of S pieces of evidence jointly supporting y_j is obtained, and the expression of reliability function $p_{y_j, e(s)}$ is as follows:

$$\begin{aligned} \hat{m}_{y_j, e(s)} &= \left[(1 - r_s) \hat{m}_{y_j, e(s-1)} + m_{P(\Theta), e(s-1)} m_{y_j, s} \right] \\ &+ \sum_{B \cap C = D_j} m_{B, e(s-1)} m_{C, s}, \quad \forall y_j \in \Theta, \\ P_{y_j, e(s)} &= \begin{cases} 0, & y_j = \Phi \\ \frac{\hat{m}_{y_j, e(s)}}{\sum_{y_j \in \Theta} \hat{m}_{y_j, e(s)}}, & y_j \in \Theta, y_j \neq \Phi \end{cases} \end{aligned} \quad (12)$$

After evidential reasoning, the fusion result of the base classifier is the category corresponding to the maximum $p_{y_j, e(s)}$ value of the final classification result of $\{(y_j, p_{y_j, e(s)}), j = 1, 2, \dots, N\}$ model.

Finally, this paper takes different classification results as evidence, takes accuracy as the reliability and initial weight of evidence, and then obtains the optimization weight of the base classifier through repeated optimization.

The training model is as follows:

$$\begin{cases} \min \varepsilon = \frac{1}{m} \sum_{i=1}^m d_E(u_i - v_i)^2, & 0 \leq \omega_s \leq 1, \quad s = 1, 2, \dots, S, \end{cases} \quad (13)$$

where m represents the amount of data in the training set; u_i represents the distribution of true classification results, and v_i represents the probability distribution of base classifier results on different categories after synthesis.

4. Experimental Verification

In this paper, the improved random subspace method is used to predict the financial risk of listed companies. The experimental flow chart is shown in Figure 1.

4.1. Experimental Data Set and Model Evaluation Indicators.

In this study, 1,726 listed companies in China were used as experimental samples, and 1,597 normal samples and 129 risk samples were obtained according to ST markers. The time panel for collecting experimental data of risk samples is divided into 3 years ahead of schedule, 4 years ahead of schedule, and 5 years ahead of schedule, which are expressed by $T-3$, $T-4$, and $T-5$, and the time span is 5 years from 2016 to 2020. The specific distribution is 14 in 2016, 27 in 2017, 22 in 2018, and 32 in 2019.

The experimental data set consists of 39 financial features, 12 emotional features, and qualitative text features, which are represented by $F1$, $F2$, and $F3$, respectively.

In the experimental process of financial risk prediction of listed companies, we mainly adopt four evaluation criteria: average accuracy (AA), type I error, type II error, and

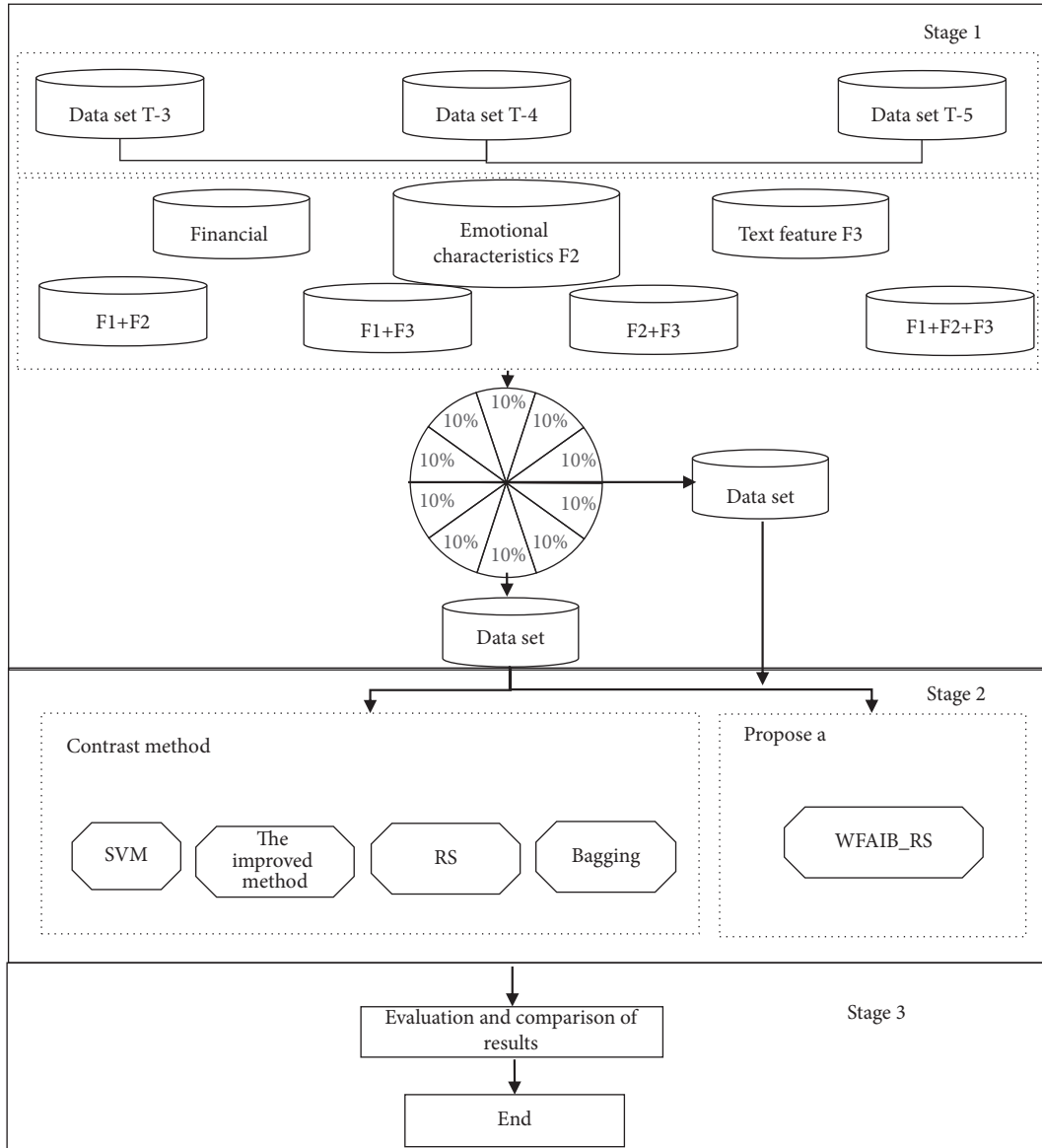


FIGURE 1: Experimental flow chart.

AUC [20, 21] (the area under the ROC curve). The main classification results are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The calculation formula of specific indicators is as follows.

$$\begin{aligned} \text{Type II Error} &= \frac{FN}{TP + FN}, \\ \text{Type I Error} &= \frac{FP}{FP + TN}, \\ \text{Average Accuracy} &= \frac{TP + TN}{TP + FP + FN + TN}. \end{aligned} \quad (14)$$

AUC is the area below the ROC curve, which is usually between 0 and 1. ROC is a curve in two-dimensional coordinates, and its horizontal axis is false positive case rate, and its vertical axis is true case rate.

4.2. Comparison Methods. In this paper, the base classifier, three ensemble learning methods, and unbalanced classification methods are used as comparison methods. Therefore, the comparison methods in this paper include SVM, bagging, RS, and the improved random subspace method.

In this experiment, the stability of the experimental results is verified by the cross method of tenfolds and ten times.

The specific steps are as follows.

- (1) Dividing the data set into 10 data with the same size and scale on average; 1 of them is used as the test set, and the other 9 are used as the training set. In the process of training, each fold should be repeatedly trained 10 times. The average value of 100 experimental results obtained after training is calculated, and the final result is obtained.

TABLE 1: Experimental results under different characteristics, different methods, and different time panels for evaluation indexes AA and AUC.

Characteristic Method	F1		F2		F3		F1 + F2		F1 + F3		F2 + F3		E1 + F2 + F3	
	AA	AUC	AA	AUC	AA	AUC	AA	AUC	AA	AUC	AA	AUC	AA	AUC
SVM	76.33	81.40	62.60	60.12	74.50	74.62	79.55	85.18	77.64	84.27	77.15	81.67	80.80	84.83
Bagging	92.87	92.13	89.65	71.32	91.39	81.61	93.22	93.31	93.79	92.84	93.20	83.80	95.27	92.99
RS	92.71	91.83	90.69	69.17	91.43	80.78	93.76	93.85	93.13	92.90	92.86	89.39	95.10	93.02
WEAIB_RS	92.79	82.94	90.69	68.76	93.61	93.09	92.99	84.22	94.78	93.70	95.05	93.93	96.39	95.24
SVM	76.72	82.53	62.83	60.33	74.61	77.04	77.67	84.91	77.76	85.63	77.49	81.70	80.75	86.47
Bagging	92.83	87.95	89.79	76.80	91.50	82.33	92.91	91.59	93.29	92.70	94.40	85.77	95.91	93.04
RS	92.60	85.27	90.84	75.24	91.58	81.33	93.03	91.33	94.89	93.91	93.40	87.67	95.59	94.04
WFAIB_RS	92.92	83.22	90.78	74.74	94.53	84.91	92.92	85.28	95.59	94.12	94.37	88.53	96.77	94.30
SVM	75.58	8008	62.66	60.29	75.83	78.87	76.55	83.21	77.91	86.00	78.27	83.06	81.46	88.24
Bagging	92.73	87.86	89.70	74.58	92.06	82.95	93.87	92.64	95.13	94.22	94.42	87.11	96.67	94.36
RS	92.65	88.45	90.79	73.64	91.75	82.62	93.29	92.61	95.85	95.05	95.37	92.25	96.55	95.26
WFAIB_RS	92.66	83.16	90.84	75.04	95.79	93.96	93.69	86.34	96.17	95.45	96.03	94.30	97.67	95.91

During the experiment, the ratio of random subspace increased from 0.1 to 0.9 according to the increase of 0.1. The setting ratios of misdivision cost for positive and negative cases are 1,726/129 and 1,726/1,597, respectively. The regularization parameters are 0.001, 0.01, 0.1, 1 and 10, respectively. The regularization parameter λ_2 is optimized by cross-validation.

4.3. Experimental Results. It works with different characteristics, different methods, and different time panels, and the experimental results are shown in Table 1. The bold data in Table 1 is the highest value of this feature.

It can be seen from the bold data in Table 1 that the improved random subspace method has achieved excellent results compared with other methods in this experiment. In the AA index, the improved random subspace method achieves good results on higher dimensional feature sets $F3$, $F1 + F3$, $F2 + F3$, and $F1 + F2 + F3$. In $T - 3$, the values of the improved random subspace method are 93.61%, 94.78%, and 96.39%. In $T - 4$, the values of this method are 94.53%, 95.59%, and 96.77%. In $T - 5$, the values of this method are 95.79%, 96.17%, and 96.67%. The comparison analysis shows that the improved random subspace method has achieved the highest results in financial risk prediction among all methods, especially in the feature set of $F1 + F2 + F3$ under the time panel $T - 5$, which has achieved an average rate of 97.67%.

In the AUC index, the improved random subspace method also achieves good results, especially in the feature set $F1 + F2 + F3$; for example, the value of $T - 3$ is 95.24%. The value in $T - 4$ was 94.3%. The value of $T - 5$ was 95.91%, which was higher than that of other eight methods. Therefore, the improved random subspace method is very suitable for financial risk prediction.

From the bold data in Table 2, it can be seen that the improved random subspace method can balance the two types of error rates of type I error and type II error, which shows that this method can effectively deal with both high-dimensional problems and unbalanced problems in financial risk prediction. Among them, the error rate of type II error is

higher than that of type I error as a whole. The main reason is that the data distribution is uneven, and there are few samples in a few classes, which leads to insufficient training. Therefore, the data of a few samples are easy to be misclassified in prediction.

4.4. Analysis of Experimental Results

4.4.1. Analysis of Prediction Results of Different Feature Sets. According to the above experimental results, this paper will compare and analyze the prediction results of different features and different time panels. The comparison results of features under $T - 3$, $T - 4$, and $T - 5$ time panels are shown in Figure 2.

It can be seen from the figure that in a single feature, the performance of $F1$ is relatively stable, the prediction performance of $F2$ is relatively weak, and the prediction effect of $F3$ is obviously better than that of $F2$. This result shows that financial features play the strongest role in the financial risk prediction of listed companies and are in a leading position. Compared with the single features $F1$, $F2$, and $F3$ and the combination features $F1 + F2$, $F1 + F3$, $F2 + F3$, and $F1 + F2 + F3$, the combination feature $F1 + F2 + F3$ is better than the single feature. In the improved random subspace method, combined feature $F1 + F2 + F3$ is better than other methods in $T - 3$, $T - 4$, and $T - 5$, and the prediction value is more than 95%, which fully demonstrates the rationality of the improved random subspace method in financial risk prediction.

4.4.2. Analysis of Prediction Results by Different Methods. According to the prediction results of AUC by different methods, the effectiveness of the improved random subspace method in financial risk prediction is shown in Figure 3 for the prediction results of AUC of $T - 3$, $T - 4$, and $T - 5$.

As can be seen from Figure 3, among the combined features $F1 + F2 + F3$, the AUC predicted by the improved random subspace method is the highest, and the AUC predicted by FAIB_RS method is about 95% in $T - 3$, 93% in $T - 4$, and 95.5% in $T - 5$, which is obviously higher than that

TABLE 2: Experimental results of type I error and type II error under different characteristics, different methods, and different time panels.

Characteristic Method	F1		F2		F3		F1 + F2		F2 + F3		F2 + F3		F1 + F2 + F3	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II
SVM	22.67	36.05	36.00	54.73	24.51	37.76	19.38	33.70	21.66	31.03	22.15	31.52	18.26	30.84
Bagging	6.64	13.20	8.13	37.83	7.25	25.69	6.03	16.06	4.17	18.22	6.37	12.12	3.72	17.23
RS	6.31	19.42	8.09	24.41	7.33	23.92	6.01	9.09	5.16	28.04	6.04	20.76	4.01	15.92
WFAIB_RS	6.28	18.72	8.09	24.41	5.21	21.00	6.16	17.53	3.84	22.30	3.73	20.05	5.17	9.06
SVM	22.71	30.34	35.81	54.01	24.28	39.13	2136	34.34	2164	29.67	21.92	29.81	18.62	27.05
Bagging	6.71	12.86	8.01	37.45	6.90	28.31	6.23	17.74	5.82	17.73	5.26	9.81	2.86	1932
RS	6.46	19.04	7.92	24.51	7.14	24.27	6.42	13.78	4.39	14.02	5.86	15.76	3.58	1469
WFAIB_RS	6.24	17.48	8.01	24.20	4.96	11.78	6.51	14.14	11.84	4.76	4.76	13.72	2.79	8.68
SVM	23.41	36.92	35.99	54.05	23.59	31.35	22.69	32.86	21.55	28.78	20.92	31.76	18.03	24.85
Bagging	6.51	16.68	8.02	38.53	6.73	22.92	5.34	15.91	3.90	16.88	5.24	9.79	2.68	11.38
RS	6.45	18.49	7.98	24.44	7.08	22.73	6.26	12.28	3.61	10.84	4.33	8.34	3.02	8.77
WFAIB_RS	6.43	18.61	7.86	25.25	3.92	7.80	5.47	16.71	3.32	10.14	3.64	8.06	1.92	7.41

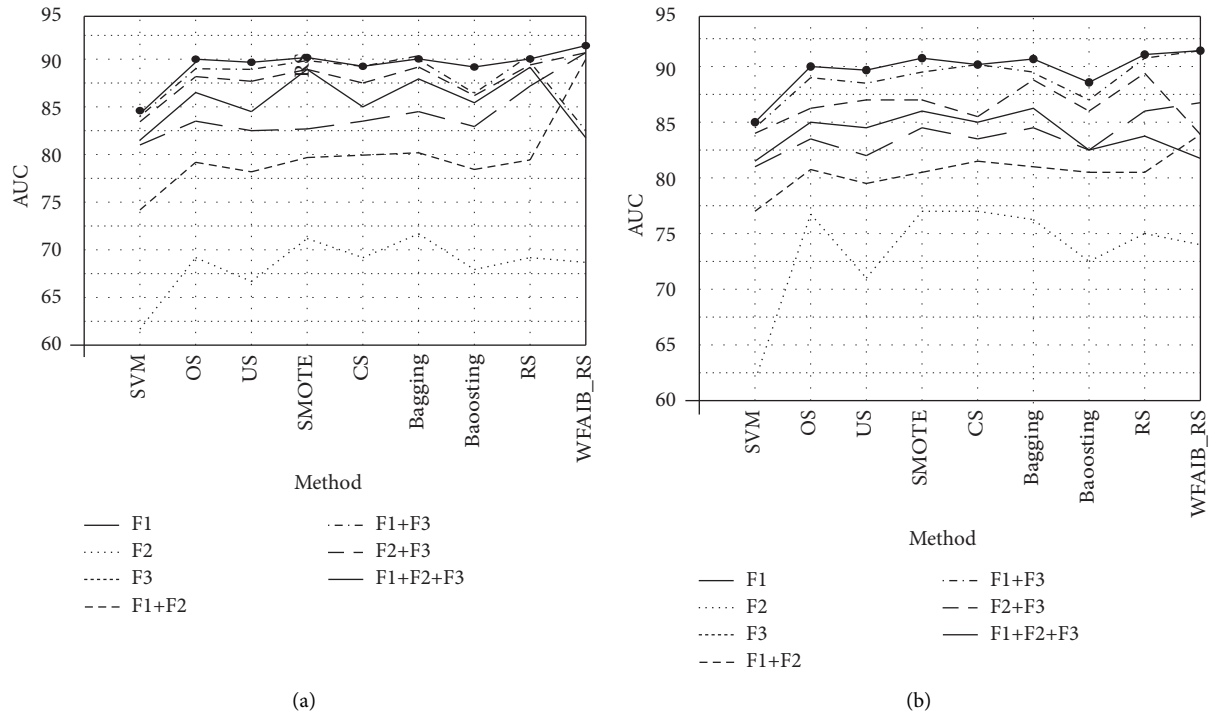


FIGURE 2: Continued.

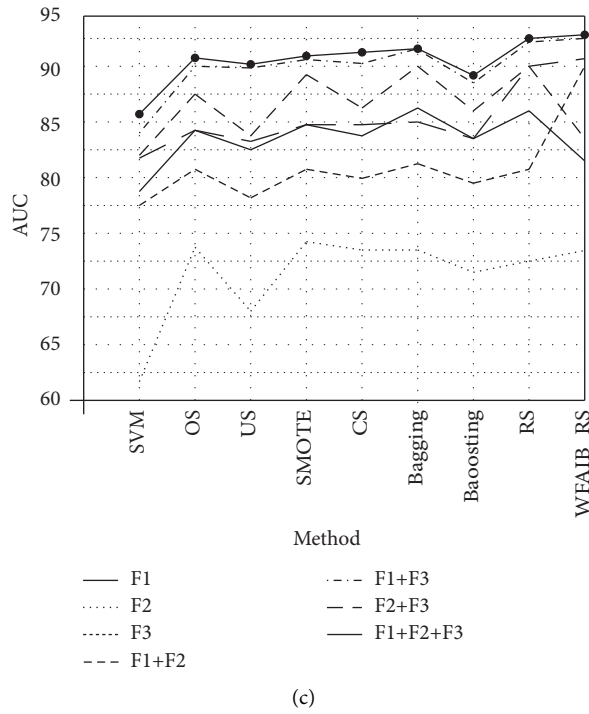


FIGURE 2: Feature comparison results under $T-3$, $T-4$, and $T-5$ time panels: (a) $T-3$ time panels, (b) $T-4$ time panels, and (c) $T-5$ time panels.

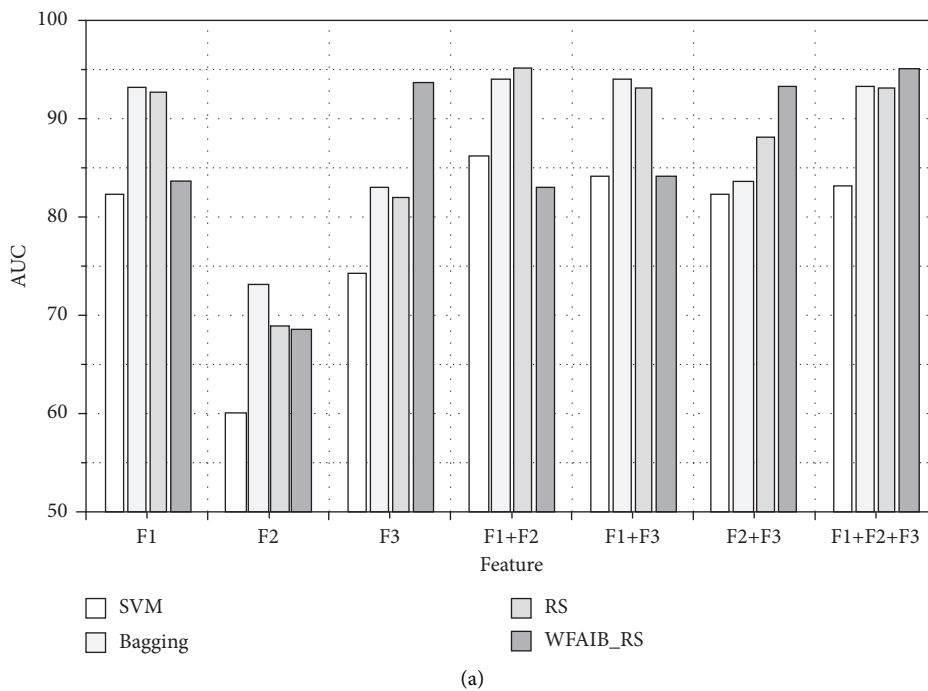


FIGURE 3: Continued.

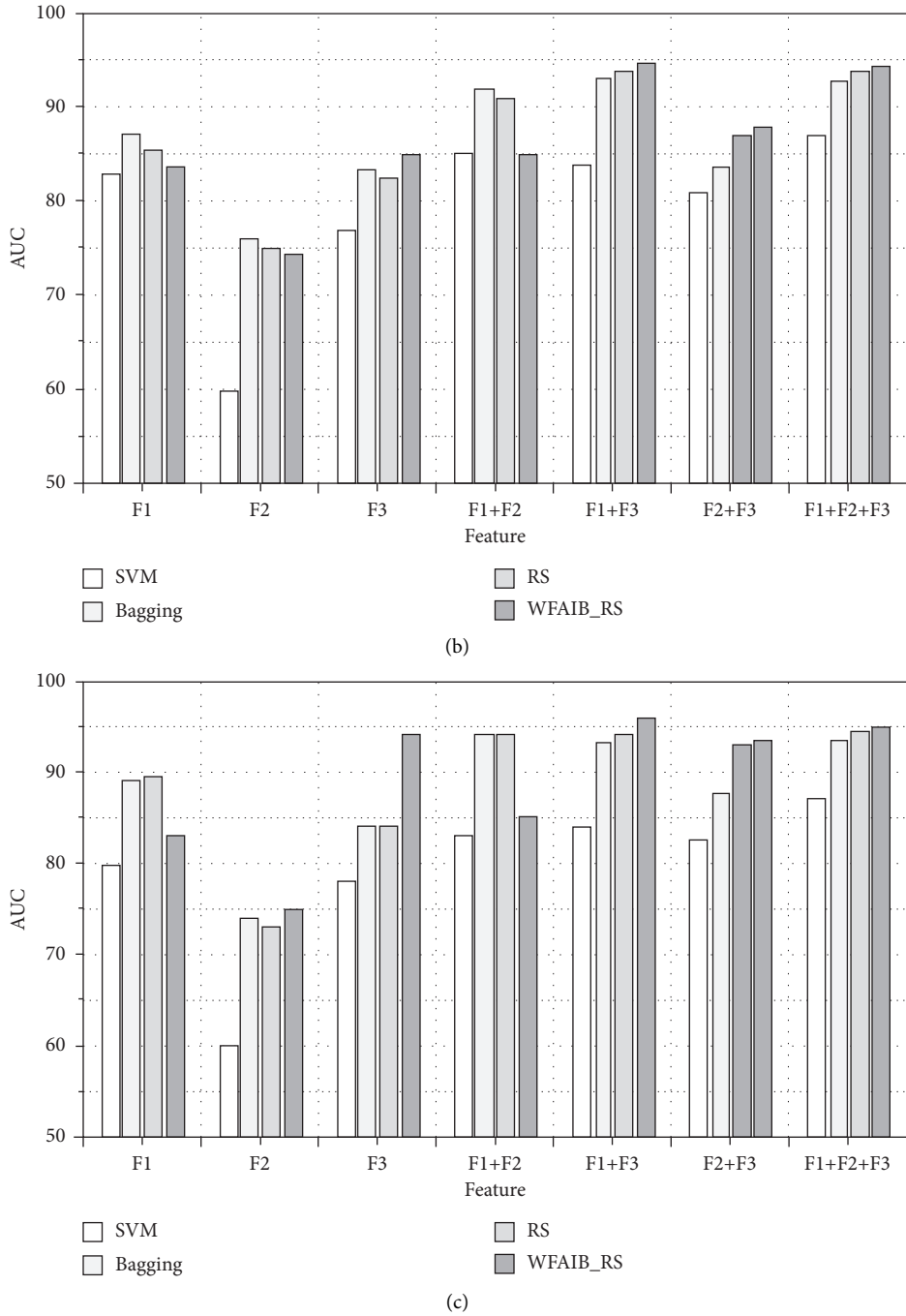


FIGURE 3: AUC prediction results of $T-3$, $T-4$, and $T-5$: (a) AUC of $T-3$, (b) AUC of $T-4$, and (c) AUC of $T-5$.

of the other eight methods. The possible reason why these methods improve the sum of $F2$ is that compared with financial features and text features, $F2$ has smaller feature dimensions and contains less prediction information, while SVM and bagging methods can effectively supplement samples, thus playing a good role in strengthening $F2$. Therefore, it shows that the improved random subspace method is effective in financial risk prediction and can predict financial risk more accurately.

5. Conclusion

In this paper, a new financial risk prediction method, the improved stochastic subspace method, is constructed. Firstly, multisource heterogeneous features are extracted based on multisource data sets, including quantitative financial features and emotional features and text features based on qualitative text information. Secondly, the advantages of the stochastic subspace method are fully

absorbed in the model construction, and it is taken as the model foundation. At the same time, a new adaptive fusion method considering the relationship between features is obtained by introducing the regularized sparse model to integrate adaptive and weighted feature fusion strategies into the stochastic subspace method.

In order to verify the effectiveness of the proposed method, an experiment is carried out on the real data set of Chinese listed companies, and different characteristics and different methods are compared in the experiment. Finally, through the analysis of the experimental results, the effectiveness and stability of the proposed method in the financial risk prediction of listed companies are fully demonstrated. The future work will focus on the reanalysis of the prediction effect of data sets under different models, such as the discussion of prediction accuracy and time.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

References

- [1] D. Marcek and L. Falat, "Volatility forecasting in financial risk management with statistical models and ARCH-RBF neural networks," *Journal of Risk Analysis and Crisis Response*, vol. 4, no. 2, p. 77, 2014.
- [2] H. Sun and D. L. Zhi, "Estimate of capital stock of provinces in China and the typical fact from 1978 to 2008," *Journal of Financial Economics*, vol. 25, pp. 107–115, 2010.
- [3] H. Akkermans and K. E. van Oorschot, "Pilot error? Managerial decision biases as explanation for disruptions in aircraft development," *Project Management Journal*, vol. 47, no. 2, pp. 79–102, 2016.
- [4] A. Huang and F. Wu, "Two-stage adaptive integration of multi-source heterogeneous data based on an improved random subspace and prediction of default risk of micro-credit," *Neural Computing & Applications*, vol. 33, no. 4, pp. 1–11, 2020.
- [5] Z. Jia, Y. Shi, C. Yan, and M. Duygun, "Bankruptcy prediction with financial systemic risk," *The European Journal of Finance*, vol. 26, 2020.
- [6] G. Vicente, A. I. Marqués, and S. J. Salvador, "Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction," *Information Fusion*, vol. 47, pp. 88–101, 2019.
- [7] Y. Li, "Research on Financial Risk Prediction and Prevention Countermeasures Based on Big Data," in *Proceedings of the 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* Qiqihar, China, IEEE, April, 2019.
- [8] Y. Ying, W. Jun, W. Gang, and C. Yu-Wang, "Research and development project risk assessment using a belief rule-based system with random subspaces," *Knowledge-Based Systems*, vol. 178, pp. 51–60, 2019.
- [9] Y. Zhu, L. Zhou, C. Xie, G.-J. Wang, and T. V. Nguyen, "Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach," *International Journal of Production Economics*, vol. 211, no. MAY, pp. 22–33, 2019.
- [10] A. Mosavi, A. Shirzadi, B. Choubin et al., "Towards an ensemble machine learning model of random subspace based functional tree classifier for snow avalanche susceptibility mapping," *IEEE Access*, vol. 8, Article ID 145968, 2020.
- [11] N. García-Pedrajas and D. Ortiz-Boyer, "Boosting random subspace method," *Neural Networks*, vol. 21, no. 9, pp. 1344–1362, 2008.
- [12] G. Chen and S. Li, "Research on location fusion of spatial geological disaster based on fuzzy SVM," *Computer Communications*, vol. 153, pp. 538–544, 2020.
- [13] U. Bhosle and J. Deshmukh, "Mammogram classification using AdaBoost with RBFSVM and Hybrid KNN-RBFSVM as base estimator by adaptively adjusting γ and C value," *International Journal on Information Technology*, vol. 11, no. 4, pp. 719–726, 2019.
- [14] G. Chen, X. Xie, and S. Li, "Research on complex classification algorithm of breast cancer chip based on SVM-RFE gene feature screening," *Complexity*, vol. 2020, pp. 1–12, 2020.
- [15] X. Ning, K. Gong, W. Li, and L. Zhang, "JWSAA: joint weak saliency and attention aware for person re-identification," *Neurocomputing*, vol. 453, pp. 801–811, 2021.
- [16] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, Article ID 102053, 2021.
- [17] G. Wang, W. Li, and L. Zhang, "Encoder- X,: Solving Unknown Coefficients Automatically in Polynomial Fitting by Using an Autoencoder," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [18] Z.-G. Liu, Q. Pan, J. Dezert, and A. Martin, "Combination of classifiers with optimal weight based on evidential reasoning," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1217–1230, 2018.
- [19] Z.-f. Ma, H.-p. Tian, Z.-c. Liu, and Z.-w. Zhang, "A new incomplete pattern belief classification method with multiple estimations based on KNN," *Applied Soft Computing*, vol. 90, no. 4, Article ID 106175, 2020.
- [20] M. Kottas, O. Kuss, and A. Zapf, "A modified Wald interval for the area under the ROC curve (AUC) in diagnostic case-control studies," *BMC Medical Research Methodology*, vol. 14, no. 1, p. 26, 2014.
- [21] W. Chen, X. Yan, Z. Zhao, H. Hong, D. T. Bui, and B. Pradhan, "Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China)," *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 1, pp. 247–266, 2019.