Application of Big Data Technology in Internet Financial Risk Control

Jingjing Chen, Guixian Tian, Jia Wang

Abstract—**Big data technique is now a prevalent concentration for specialist and academics due to the development of information technology, cloud computing, and Internet of Things technologies. A fiscal risk controlling model using MSHDS-RS, was innovatively proposed to deal with the current situation of unreasonable design of features in risk controlling technique. This model's innovation is that the model utilizes a normalized sparse approach for optimizing feature fusion after drawing loan customer information sources' hard and soft features, thereby forming integrated features. Then, the feature subset derived from probability sampling is trained as a base classifier, and the results of the base classifier are fused and optimized using evidence reasoning rules. MSHDS-RS's accuracy improvement rate was about 3.0% and 3.6% higher than existing PMB-RS methods', respectively, by observing MSHDS-RS's operating results in different feature sets with soft and integrated feature indicators. Therefore, the proposed optimization fusion method is reliable and feasible. This research contributes to the control of internet financial risks and has certain value in making effective decisions on loan platforms.**

Index Terms—**big data technology, MSHDS-RS, multi-source heterogeneous data structure, risk control, random subspace**

I. Introduction

WITH the progress of social technology, the Internet is imperceptibly affecting all aspects of human life. In the past decade, as online shopping platforms such as Taobao and JD.com gradually penetrates into the public's lives, Internet finance also quietly enters the Chinese market. Various Internet financial services such as P2P, credit, etc., gradually emerge since the development of techniques. With the strong rapidity and convenience, users quantity in the web financial investment is also increasing day by day. Internet finance is a double-edged sword, which can promote the development of China's financial industry, but also brings more risks [1-2].

Manuscript received November 16, 2023; revised May 16, 2024.

This work was supported by the Innovation and Strengthening School Quality Project in 2022 from Guangdong University of Science and Technology. Project name “Research team on collaborative governance of urban agglomeration innovation network in Guangdong Hong Kong Aoda Bay Area”. (GKY-2022CQTD-6)

Jingjing Chen is a lecturer of Guangdong University of Science and Technology, Dongguan, CO 523083 China (e-mail: 1012861876@qq.com).

Guixian Tian is a professor of Guangdong University of Science and Technology, Dongguan, CO 523083 China (Corresponding author to provide e-mail: 16010039@pxu.edu.cn).

Jia Wang is a professor of Zhejiang Wanli University, Ningbo, CO 315100 China. (e-mail: 2004900088@zwu.edu.cn).

Borrowers and lenders have information asymmetry in online credit deals. This will cause financing difficulties for small and medium-sized micro-enterprises, and the financing cost remains high [3]. In addition, lenders will bear greater risks. For example, the credit information of borrowers is incomplete. Therefore, high-quality and bad customers are unable to accurately be identified, and the loan quota and other dangers faced in the repayment circle are unable to be determined. As a result, for enterprises, both borrowers and lenders face difficulties that cannot control business risks, greatly hindering the development of enterprises. Therefore, it is urgent to develop effective technologies for controlling internet financial risks. Big data technology plays an important role in financial risk control. Big data has the characteristics of huge scope, dynamic alteration, et al. The information mainly comes from various channels such as visitors, etc. The web finance industry can enable employees to extract information that has implications for banking performance from various incomplete, fuzzy, and massive data, and help them make further decisions by using big data analysis technology [4]. However, the current big data risk control technology faces many challenges, mainly manifested in the unreasonable design of feature data [5]. Based on the above analysis results, this research innovatively integrates Multi-Source Heterogeneous Data Structure (MSHDS) and Random Subspace (RS). Meanwhile, a financial risk control model is proposed based on big data technology. The experiment aims to enable loan companies to identify potential customers more quickly, increase the pass rate of loan reviews, and bring greater benefits.

II. Related Work

Science and technological advancements increasingly bring the banking sector and Internet technology together. Internet finance offers opportunity to people, but it also comes with many risks and difficulties. The incompatibility between security and ease in Internet finance is the subject of numerous academic discussions. Yan et al. evaluated the risks of Internet insurance firms in 2019 using a Topsis approach as well as four companies' reported data. These findings demonstrated a strong correlation between the asset liquidity and online operation capabilities and the financial risks faced by insurance businesses. Online insurance providers should boost capital and implement an early warning system to offset risks [6]. According to Ou and other academics, although there were numerous risks associated with doing so, the majority of publicly traded corporations in today's society opt to publish financial information via Corporate Internet Report (CIR). They conducted research on the subject, developed a CIR hazard communicating approach using FNN. This model's effectiveness was tested through a study to promote their understanding of this CIR hazard communicating procedure [7]. A banking database approach using big data can be swiftly developed by appropriate specialists, according to a proposal made by Haoru W. and others to deal with the issue that currently developing web trustworthiness was susceptible to a variety of unanticipated hazards. Big data analysis was used to assess the web trustworthiness hazard to lower the web trustworthiness hazard coefficient [8]. From the perspective of the growth of Internet finance, Zhao N and other academics investigated and evaluated the risk transmission mechanisms of third-party payments, P2P platforms, and crowdfunding. Moreover, this paper discussed the preventive measures against the liquidity risk of Internet finance from the aspects of strengthening network security supervision measures, improving pertinent laws and regulations, and social credit system. This research was conducted for the development of Internet finance in China [9]. Feng et al. suggested a radial basis function using an optimizing approach to detect web hazards' occurrence. This approach combined data mining technology and deep learning algorithms, mostly through mining financial data from the Internet to identify potential market hazards. The experimental findings demonstrated the model's good performance, with a minor error of 0.249 and a running time of 2.212 s [10].

For the web trustworthiness finance, Wang C. et al. suggested a big data-based method for identifying financial risks that integrates the backpropagation neural network algorithm to evaluate Internet financial risk. The experiment chose financial data from the Internet for examination and verification. The final results demonstrated that the suggested model had a high rate of risk prediction accuracy, which reached 99.98% [11]. Fang W and other researchers developed a deep neural network incorporating the random forest technique and XGBoost algorithm to identify Internet financial crime on a sizable public loan dataset. The greater performance of this model was demonstrated through experiments, which helped different scales of web financial organizations better detect loan fraud [12]. Wen and other researchers built a new LSTM using attention mechanisms and other means considering the lack of a time trend for web hazards. The attention technique introduced allowed the model to focus on key information. The model paid attention to loaning data connection. The model outperformed ARIMA, LSTM, and GRU, according to the verification of available data sets [13]. Researchers such as Wang L analyzed controlling supply chain financing risks to energetically respond to a policy of "Internet plus". The experiment analyzed the supply chain financing and blockchain technology based on theoretical research. At the same time, VAR was introduced to evaluate the supplying chain hazard. This method optimized the supply chain financing risk control system and greatly reduced the risks of supply chain financing parties [14]. Scholars such as Al-Alwan et al. used structured equations to explore the impact of big data technology on the quality of enterprise decision-making. The experiment collected data from three telecommunications service companies to verify. The final result showed that big data had a significant impact on the quality of enterprise decision-making. Therefore, enterprises needed to collect accurate and reliable data and information to timely avoid their own possible financial risks [15-17].

In conclusion, a large number of academics have contributed to the web financial hazard prediction. However, this hazard predicting means can be improved further when used in conjunction with big data-related technology. This paper offers a financial hazard controlling means with MSHDS and RS to control web financial hazard and provide some value in effectively making judgments on lending platforms.

III. Integrated Learning and Risk Control

Firstly, the applicability of ensemble learning algorithm to financial risk classification is analyzed, and RS is proposed as a machine learning method. On this basis, a method based on MSHDS and RS is constructed to better characterize the financial risk characteristics. Finally, the paper proposes to improve the features and results fusions and constructs the final financial hazard controlling means with MSHDS and RS.

Financial risk have a direct bearing on financial markets' volatility, and its most commonly used machine algorithms include ensemble learning-based classification algorithms such as bagging, boosting, and RS. The integrated learning method is the key to realize the financial risk control strategy. This method is obviously superior to the general single classification method in terms of learning ability and learning stability.

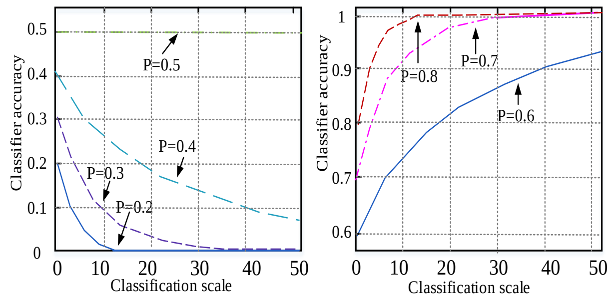


Fig. 1. Comparison of classifying scale and accuracy

The advantage of integrated learning is that it can meet the requirements of learners for accuracy and variety. In Figure 1, the classification size and classifier accuracy are compared. The classification accuracy decreases as data increase, and eventually becomes 0 when the classifying accuracy < 0.5. With the increase of data size, the accuracy is greater than 0.5, and the accuracy of the classifier eventually tends to 1. The main processes of ensemble learning are to create a data subset, train a base classifier, and combine the results. This base classifier's comprehensive output straightly influence this integrated model's quality. There are many kinds of classification methods for ensemble learning, which can be divided into instance and feature ensemble learning methods according to the data aspect, and ensemble learning methods using the same or different learners according to the homogeneity and heterogeneity of classifiers. The training of ensemble learning classifier stacking method is very complicated, and the ensemble model is greatly affected by different classifiers and secondary learning algorithms. The main homogeneous ensemble learning techniques include boosting, bagging, and RS. Bagging generates training sets through random combinations. Bagging mostly uses finite information to constantly sample and create new data that accurately represent the primitive sample. Boosting combines n feeble classifiers into a superior precision classifier by weight voting and turns a weak learning algorithm into a strong learning algorithm. AdaBoost, a common example of Bagging, is modeled as follows:

$$G(y)=\sum_{j=1}^{N}\beta_j g_j(y) \tag{1}$$

In formula (1), β_j denotes weight. g_j refers to the j-th classifier. N refers to the classifiers quantity. y refers to a sample. g_1 is acquired and constantly updated to determine g_j and β_j through direct learning. Assuming that ε_j is the j-th basis classifier's error, the weight updating expression is as follows:

$$\beta_j=1/2\left(\ln(\frac{1-\varepsilon_j}{\varepsilon_j})\right) \tag{2}$$

However, RS is more appropriate for learning issues having higher feature dimensionality. Therefore, the research is conducted based on RS. The specific process of RS is as follows. First, the data samples are randomly sampled with reference to the feature dimension to obtain similar data subsets, and then adjusted according to the subspace proportion parameters. Secondly, the base classifier is trained using data subset. Finally, the different results are synthesized using fusion rules. Among them, the selection of the proportion of RS is the key to affect the learning effect.

IV. FINANCIAL HAZARD CONTROLLING MEANS USING MSHDS AND RS

A. Multi-source Heterogeneous Data Preparation

The core of financial risk control lies in the comprehensive analysis of the multifaceted risk characteristics of individual lending behavior, with the data preparation stage being particularly crucial. Data collection is validated, cleaned, and processed into feature engineering to construct refined risk descriptors. The integration of hard and soft information provides a comprehensive foundation for risk assessment and optimizes the construction of risk control models. This section discusses in detail the organization and extraction of multi-source heterogeneous data, and explains the process of risk feature analysis and its importance in the financial field.

Risk business losses and personal default rates can BE lowered by controlling financial. The key point to creating a risk control model is learning how to undertake thorough and multi-faceted analysis of personal lending risks [18-19]. The prerequisite for creating a hazard controlling means is to obtain and extract data that describe the customer risk's characteristics. Data collection, verification, cleansing, feature screening, and feature derivation are all parts of data preparation [20]. In the preparation stage, the data are first gathered, examined, and cleaned before feature extraction, integration, and screening, which are carried out using the feature derivation approach. The data features are chosen for training. Soft and hard information are acquired to effectively control businesses with P2P financial risks. Figure 2 shows the findings of the information gathering and research.

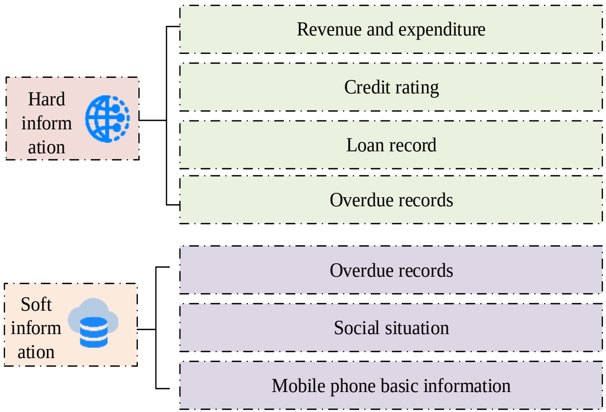


Fig. 2. Composition of information sources

The following five forms of data are specifically included in information source data. There are seven basic characteristics of loan customers: the historical borrowing platforms, the historical borrowings, location, etc. There are 13 fields that constitute mobile phone's applying properties: the necessity of real-name, calls' length and volume with friends, parents, etc. during the loan term. Real name, etc. are only a few of the payment consumption aspects. In all, there are eight fields. Five factors make up the features of online platform consumption data: total online purchases, average price per purchase, largest single purchase amount, total orders, and average price of each transaction. The number of fans, followers, and friends are three categories that make up social media characteristics.

Both Soft and hard information is used in feature extraction. The public basic information platform is mostly used to gather hard feature information. The term "soft feature information" refers to the 12 emotional characteristics as well as some text features that are obtained when word vectors are used to transform qualitative and quantized data into learnable domains. Various soft and hard information sources have different effects on financial risks. Therefore, handling multi-source various features is crucial before using various information sources for risk assessment. Second, these features of different apps are classified into different groups using these qualitative and quantized data. First, the information from various sources is divided into various feature groups. Moreover, the emotional characteristics are split into three groups: positive, neutral, and negative emotions according to the emotional qualities. The terms "positive emotion group" stands for active or strong qualities. "Neutral emotional group" stands for unidentified qualities. Finally, a multivariate heterogeneous feature set $B=\left\{B^{(1)},B^{(2)},B^{(3)}\cdots B^{(i)},\cdots B^{(s)}\right\}$ is acquired. s stands for information origins quantity. $1,2$ stand for quantized as well as qualitative properties. A property space which is separated into J groups is $X=(x_1^{(1)},\cdots x_{P_1}^{(1)},\cdots x_1^{(2)},\cdots x_{P_2}^{(2)},\cdots x_1^{(j)},\cdots x_{P_j}^{(j)})$.

B. Construction of Financial Hazard Controlling Means by Integrating MSHDS and RS

Risk control models in the financial field are usually based on a single data source, which leads to model instability and errors. This section introduces a hazard controlling means using MSHDS and RS. This model improves the diversity and accuracy of classifiers by adaptively fusing features and adopting evidence reasoning rules to address the challenges of risk control.

The established models are unstable and erroneous since most of the current borrower default hazard controlling means are on the foundation of a unitary origin. Among these three ensemble learning techniques previously described, it is determined after thorough investigation that RS is fitter for financial hazard control's reality [21-23]. The risk control model on the foundation of MSHDS and RS is proposed in this study simultaneously considers classifier diversity and accuracy [24-26]. The traditional RS method contains three procedures: gathering feature sub-aggregate, building base classifiers, and conflating outcomes. Modifications are proposed to features and outcomes fusions' two stages, namely MSHDS-RS. The specific process is shown in Figure 2. The first stage is the adaptively feature fusing, which mainly improves these extracted features' quality and comprehensively considers various risk indicators. This stage reduces the complexity of the control model, thereby achieving high stability and accuracy [27-29]. A feature weighting method is first introduced to clarify the significance of characteristics while determining key features. Additionally, the research presents a Sparse Group Lasso (SGL) model for enhancing these results considering that customer risk indicators exhibit multi-source heterogeneity. Support Vector Machine (SVM) can learn categorization in the second phase. Kernel functions can decrease the complexity to accomplish high-latitude linear segmentation and streamline. SVM's uppermost component is straightly decided with a kernel function utilized.

In the 3rd phase, these fused findings lack some hazard controlling messages as these existing fusion approaches' base classifiers are ambiguous. This research uses an Evidence Inference (ER) rule to combine the data while also taking the overall effect into account, as follows:

$$E_i=\{(\theta,p_{\theta,i})\forall\theta\subseteq\Theta,\sum_{\theta\subseteq\Theta}p_{\theta,i}=1\} \quad (3)$$

In formula (4), E_i denotes evidence. Θ denotes reciprocally exclusive and complete assumptions. $(\theta,p_{\theta,i})$ denotes a testimony element. θ denotes proposition. $p_{\theta,i}$ denotes support. Integrating evidence's reliability and weight, the supporting degree $m_{\theta,i}$ is listed as:

$$m_i=\{(\theta,m_{\theta,i})\forall\theta\subseteq\Theta,(p(\Theta),m_{p(\Theta),i})\}$$
$$m_{\theta,i}=\begin{cases}0, & \theta=\varnothing\\ c_{Rw,i}m_{\theta,i}\ , & \theta\subseteq\Theta,\theta\neq\varnothing\\ c_{Rw,i}(1-R_i), & \theta=p(\Theta)\end{cases} \quad (4)$$

Where, $c_{Rw,i}=\frac{1}{1+w_i-R_i}$. Ignoring the local ignorance, the proposition's total support is listed as follows:

$$m_{\theta,E(N_c)}=[(1-R_i)m_{\theta,E(i-1)}+m_{p(\Theta),E(i-1)}m_{\theta,i}],\forall\theta\subseteq\Theta \quad (6)$$

The total classifiers' supporting possibility for θ derived from normalization is:

$$p_\theta=\frac{m_{\theta,E(L)}}{1-m_{p(\Theta),E(L)}},\forall\theta\subseteq\Theta \quad (5)$$

The hazard controlling outcome is a lopsided classifying scene. Then, the Area under Curve (AUC) is utilized to assess the eventual outcome's quality. The reliability $R=(R_1,R_2,\cdots R_M)^T\in R_+^M$ is expressed as AUC acquired from a base classifier test. In addition, these risky samples' reliability is utilized as the recall base classifier's weight $w^{(2)}=(w_1^{(2)},w_2^{(2)},\cdots w_M^{(2)})\in R_+^M$.

V. RESULTS ANALYSIS OF MSHDS-RS MODEL FOR FINANCIAL HAZARD CONTROLLING MEANS

The study selected 6000 borrowers having a one-year loaning circle from P2P platforms as an experimental samples, with 120 days overdue as the default limit period. The experiment obtained 1000 risk and 5000 normal samples. A ten-fold cross validation method was used in this study. The comparison methods used include SVM, Bagging, Boosting, and PBM_ RS and RS algorithms. The experiment also supported TSAIB-RS's upgraded algorithm at the same time. Figure 3 shows the evaluation and comparison findings of these eight classification methods' precision and recall rates in a conjoined dataset. The classification accuracy value of the risk-free sample consumers was higher than that of the risk-distributed sample customers by observing accuracy. A hazard-free group's clients owned a lower recall than these clients from the hazard-category. The sample distribution's imbalance could cause this circumstance. However, the misclassification was the result of inadequate classifier training. For dangerous customers, the MSHDS-RS technique outperformed the other seven classification methods for precision and recall. MSHDS-RS provided superior recall and precision for dangerous consumers than other classifying means and had benefits.

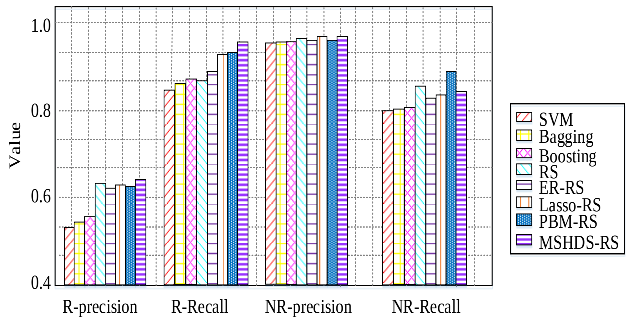


Fig. 3. Evaluation results of eight classification methods

Figure 4 shows the outcomes of the MSHDS-RS model based on conventional basic, soft, and multi-source heterogeneous fusion features. In the SVM method, the AUC which had integrated features increased by about 5.2% compared to the AUC based on hard indicators, and increased by about 5.0% compared to the AUC which had soft features. In this Bagging method, these integrated features' AUC value was 0.9234, which was increased by about 2.4% compared to the Soft Index (SI) of 0.9016, and by about 2.2% compared to the Hard Index (HI) of 0.9034. In this Boosting method, these integrated features' AUC value was 0.9141, which was increased by about 1.4% compared to SI of 0.9013, and was increased by about 1.2% compared to HI of 0.9028. In this RS method, these integrated features' AUC value was 0.9294, which was increased by about 2.1% compared to SI of 0.9103, and by about 2.9% compared to HI of 0.9024. In this ER-RS method, these integrated features' AUC value was 0.9324, which was increased by about 1.6% compared to SI of 0.9181, and by about 3.0% compared to HI of 0.9044. In this Lasso-RS method, these integrated features' AUC value was 0.9385, which was increased by about 1.8% compared to SI of 0.9213, and by about 2.9% compared to HI of 0.9086

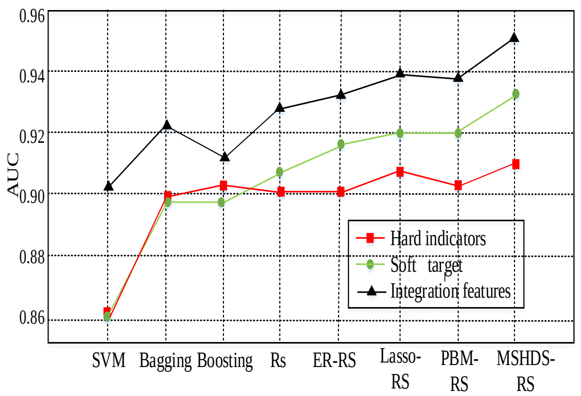


Fig. 4. Experimental results based on different feature sets

In Figure 5, the comparative outcomes for the eight approaches are displayed. The best risk control level was attained in integrated, soft, and hard features sets. Overall, this MSHDS-RS approach's hazard controlling outcome was the most steady and the effect was the most evident. These results demonstrated this two-stage adaptive fusion method's merits and demonstrated the robustness and adaptability of the suggested strategy for reducing financial risk. MSHDS-RS's impact on integrated feature financial hazard control was noticeably superior compared to pure hard indicators and soft indicators. MSHDS-RS provided the improved financial hazard control with an increase in features. MSHDS-RS was more feasible than the PMB-RS now in use by comparing the two methods. The hard feature's accuracy of MSHDS-RS was roughly 1.5% higher than the PMB-RS method currently in use. MSHDS-RS's accuracy improving rate was between 3.0% and 3.6% when compared to the existing PMB-RS method for the soft and integrated features' accuracy.

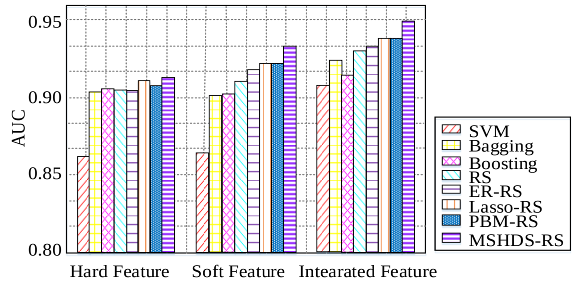
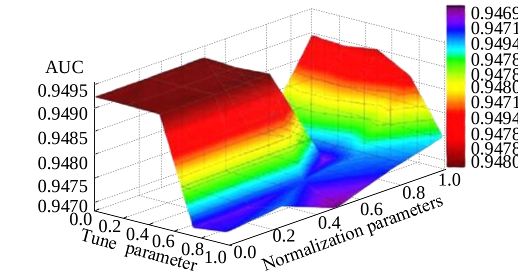


Fig. 5. Results of different methods comparison

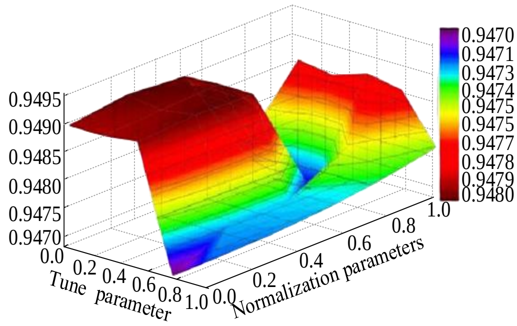
MSHDS-RS's performance tuning depends on the precise configuration of key parameters, including the normalization parameter λ, subspace ratio α, and the setting of adjustment parameters. The subspace ratios shown in Figure 6 (1)~(3) are 0.3, 0.5, and 0.7, correspondingly. With the established aggregated characteristics, an analysis was conducted on the normalized and differenced parameters. Sensitivity analysis reveals that specific parameter combinations can significantly optimize model performance, such as when α =0.3, λ =0.01, and subspace ratio is 0.5, this model's AUC reaches the upmost outcome of 0.9492. The increasing α leads to an increase in sparsity within this group, which affects the controlling outcome and shows a trend of decreasing AUC values with the increase of α. λ exhibits a V-shaped relationship, indicating that appropriate selection of normalization parameters is crucial for improving model performance. MSHDS-RS demonstrates excellent financial hazard controlling capabilities within an effective parameter range, verifying the feasibility and rationality of the model design.

This study selected the financial risk control dataset of e-commerce financial risk dataset for supplementary experiments to further verify the reliability of the MSHDS-RS risk assessment model. This dataset comes from innovative business data of e-commerce enterprises in the past 5 years, containing 28476 pieces of data, including 321 risk records. There are 21 characteristics of innovative business risks. Except for time, quantity, and type, the fields in the dataset are all insensitive. Therefore, the original purpose of these fields cannot be determined. This study selected 80% of the samples and 20% of the samples as test samples.

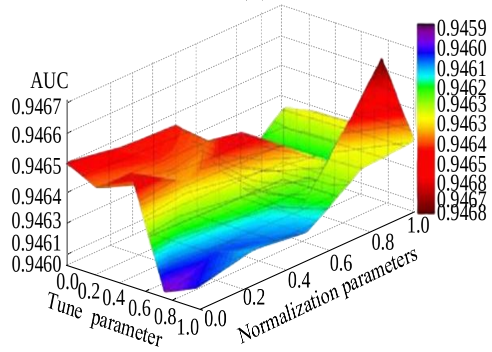
Figures 7 (a) and 7 (b) show the AUC and F1 score prediction curves of MSHDS-RS for each layer in the e-commerce innovation business risk dataset. In Figure 7 (a), the AUC value reached the optimal value of 0.928 at the third layer in the training set. The AUC value also reached the optimal value of 0.888 at layer 3 in the test set. In Figure 7 (b), the F1 value reached the optimal value of 0.894 at layer 3 in the training set. The F1 value reached the optimal value of 0.843 at layer 3 in the test set. In summary, both of the above two indexes achieved the maximum value in layer 3 and the optimal value in the test set. Therefore, the proposed model had accurate evaluation function. Therefore, the constructed model effectively evaluated the financial risk situation of the e-commerce industry.



(1)

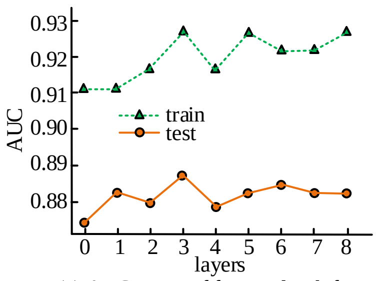


(2)

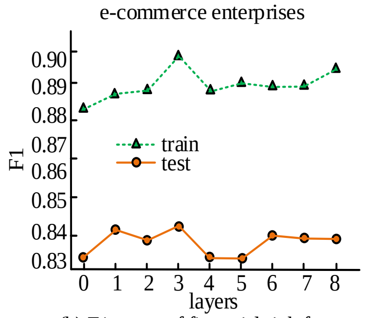


(3)

Fig. 6. Sensitivity analysis results of three parameters



(a) AUC curve of financial risk for e-commerce enterprises



(b) F1 curve of financial risk for e-commerce enterprises

Fig. 7. AUC and F1 score prediction curves of MSHDS-RS in e-commerce financial risk

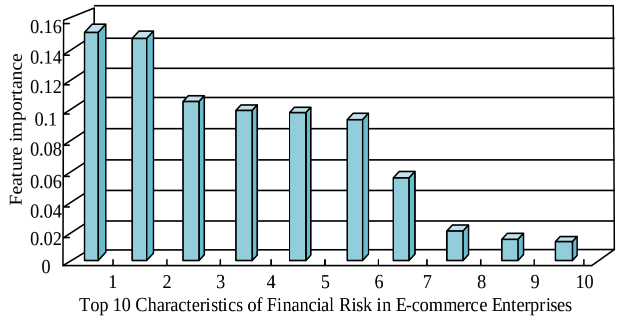
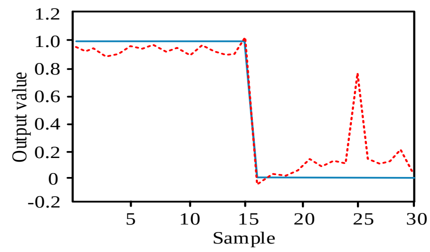
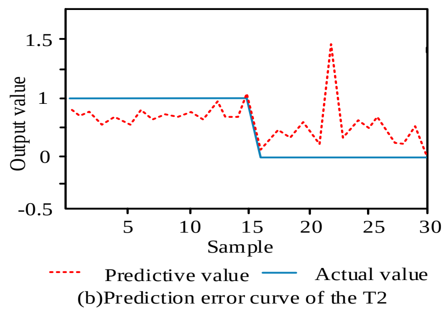


Fig. 8. AUC and F1 score prediction curves of MSHDS-RS in e-commerce financial risk

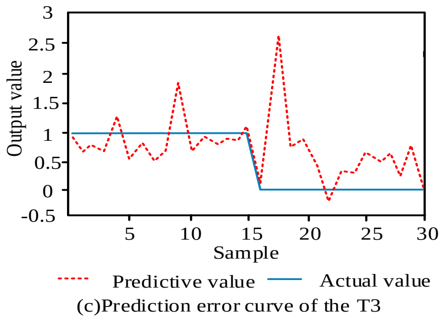
Figure 8 shows the 10 main characteristics of business innovation risks in e-commerce enterprises, of which 4 are newly generated, namely: features 1, 8, 9, and 10. The newly generated features also made a significant contribution to classification. This result indicated that the proposed model helped to explore the potential relationships of feature columns and met the financial risk assessment needs of integrated e-commerce enterprises.



(a)Prediction error curve of the T1



(b)Prediction error curve of the T2



(c)Prediction error curve of the T3

Fig. 9. Prediction error curve of the MSHDS-RS model

This study divided the e-commerce financial risk dataset into three time periods: T1, T2, and T3 to verify the actual predictive performance of the MSHDS-RS model. 30 samples were tested from T1, T2, and T3 using the model. The final simulation results are shown in Figure 7. In year T1, the prediction accuracy of this model for crisis samples was 94.49%, the prediction accuracy for normal samples was 90.59%, and the average accuracy was 92.54%. In T2, the accuracy of the model in predicting crisis samples was 90.19%, the accuracy in predicting normal samples was 88.29%, and the average accuracy was 89.24%. In T3, the accuracy of the model for crisis samples was 87.19%, for normal samples it was 85.79%, and the average accuracy was 86.49%. The average prediction accuracy of the MSHDS-RS model for crisis samples was 93.64%, the average prediction accuracy for normal samples was 90.19%, and the overall average prediction accuracy was 92.11%.

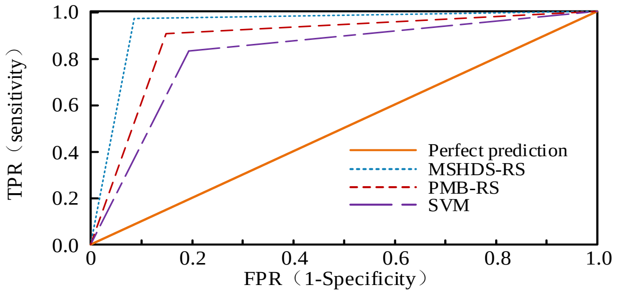


Fig. 10. Comparison results of ROC curves of model

ROC curves were plotted based on each predicted result using the X-axis as the False Positive Rate (FPR) and the Y-axis as the True Positive Rate (TPR). The experimental results are shown in Figure 10. The MSHDS-RS prediction model had the highest AUC of 0.911. The second AUC of the PMB-RS prediction model was 0.814. The AUC of the SVM prediction model was the lowest, at 0.701. In summary, the MSHDS-RS prediction model achieved better predictive performance.

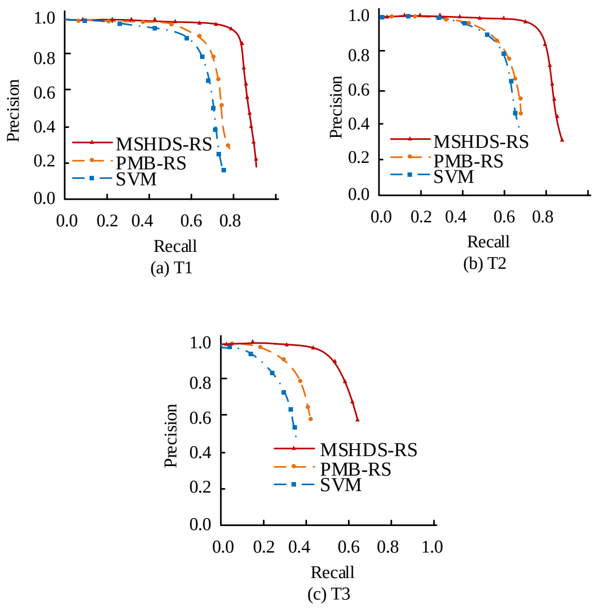


Fig. 11. Comparison results of PR curves of the model

Figure 11 shows the Precision-Recall (PR) curve results of three algorithms on different year datasets. The PR curves of all three models decreased, and the area under the PR curve of the MSHDS-RS model was higher than that of the other three models. The results indicated that the risk prediction performance of the MSHDS-RS model was superior to the other two models.

TABLE I
PERFORMANCE COMPARISON RESULTS OF THREE MODELS

Model	ROC	Accuracy	Running time
MSHDS-RS	0.911	96.23%	251s
PMB-SVM	0.814	88.14%	301s
SVM	0.701	83.22%	358s

Table III shows the summary results of ROC, accuracy, and running time of the three models. The ROC, accuracy, and running time of the MSHDS-RS model were 0.911, 96.23%, and 251s, respectively, which were superior to the comparison models selected by the research. The MSHDS-RS model integrated three stages: adaptive fusion feature, SVM classification, and result fusion based on ER rules. These three methods reduced this model's complexity, enhanced the accuracy of classification, and enhanced the risk control information, respectively. Therefore, the MSHDS-RS model had better performance in running time, classification accuracy, and ROC.

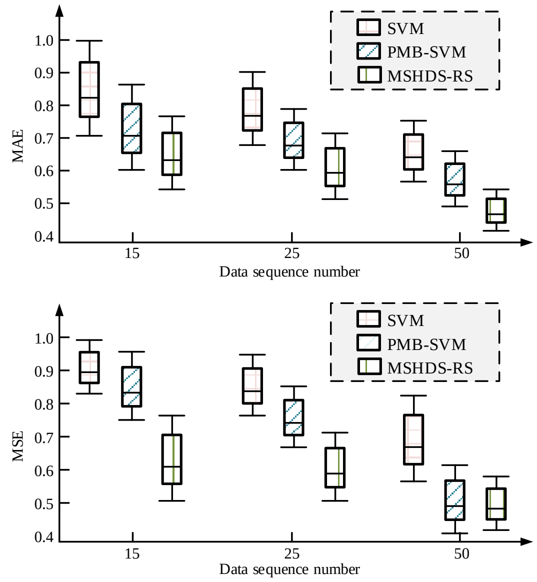


Fig. 12. Comparison results of absolute error (MAE) and squared error (MSE) of the model

Figure 12 shows the comparison results of absolute error (MAE) and squared error (MSE) of three models. As shown in the figure, the MSHDS-RS model shows significant performance advantages in comparison. On both MAE and MSE evaluation metrics, the results of this model are superior to SVM and PMB-SVM, indicating that both MAE and MSE have smaller variations. This result demonstrates that MSHDS-RS has better predictive and generalization performance when processing this dataset.

VI. CONCLUSION

While the rapid development of the Internet brings convenience to people's lives, it also brings some financial security risks. This paper proposed a financial hazard controlling means using MSHDS and RS, namely MSHDS-RS, to deal with the current financial hazard. This paper utilized a normalized means for optimizing property syncretism to form aggregated features using the original financial hazard controlling means dividing feature subsets, constructing base classifiers, and synthesizing results. Then, MSHDS-RS was applied in different feature sets. The experimental results showed that the proposed optimized fusion method was reliable, feasible, and practical. For hard feature indicators, MSHDS-RS's accuracy was improved by about 1.5% compared to PMB-RS. This new approach's accuracy improving rate was about 3.0% and 3.6% higher than other two approaches, respectively, in soft and integrated features. Due to the limited conditions of current experimental research, there is no research on the behavior of customers after loans, which is also the next research direction.

REFERENCES

[1] Q. Yang, Y. Wang and Y. Ren, "Research on financial risk management model of internet supply chain based on data science," *Cogn Syst Res*, 2019, vol. 56, pp. 50-55.

[2] X. Lyu and J. Zhao, "Compressed sensing and its applications in risk assessment for internet supply chain finance under big data," *IEEE Access*, 2019, vol. 7, pp. 53182-53187.

[3] Y. Liu, "Development and risk of internet finance based on big data," in *The International Conference on Cyber Security Intelligence and Analytics*, Shanghai, 2021, vol. 1342, pp. 518-525.

[4] L. Wang and Y. Wang, "Supply chain financial service management system based on block chain IoT data sharing and edge computing," *Alex Eng J*, 2022, vol. 61, no. 1, pp. 147-158.

[5] K.D. Dutta and M. Saha, "Nexus of governance, macroprudential policy and financial risk cross-country evidence," *Econ Chang Restruct*, 2021, vol. 54, no. 4, pp. 1253-1298.

[6] X. Yan and Y. Wu, "Financial risk assessment based on entropy weight topsis method: Take the internet insurance industry as an example," *J Simul*, 2020, vol. 8, no. 6, pp. 7-15.

[7] L. Ou and L. Chen, "Predicting risk propagation of corporate internet reporting based on fuzzy neural network," *Ingénierie des Systèmes d'Information*, 2020, vol. 25, no. 4, pp.481-488.

[8] W. Haoru, Y. Zhixuan and Yujia, "Risk assessment of internet credit based on big data analysis," *E3S Web of Conferences, EDP Sciences*, 2020, vol. 214, pp. 01012-01017.

[9] N. Zhao and F. Yao, "The transmission and preventive measures of internet financial risk," in *International Conference on Business Intelligence and Information Technology*, Harbin, 2021, vol. 107, pp. 623-631.

[10] R. Feng and X. Qu, "Analyzing the Internet financial market risk management using data mining and deep learning methods," *J Enterp Inf Manag*, 2021, vol. 35, no. 4/5, pp. 1129-1147.

[11] C. Wang and S. Liu, "Innovative risk early warning model based on internet of things under big data technology," *IEEE Access*, 2021, vol. 9, pp. 100606-100614.

[12] W. Fang, X. Li and P. Zhou, "Deep learning anti-fraud model for internet loan where we are going," *IEEE Access*, 2021, vol. 9, pp. 9777-9784.

[13] C. Wen, J. Yang and L. Gan, "Big data driven Internet of Things for credit evaluation and early warning in finance," *Future Gener Comp Sy*, 2021, vol. 124, pp. 295-307.

[14] L. Wang and Y. Wang, "Supply chain financial service management system based on block chain IoT data sharing and edge computing," *Alex Eng J*, 2022, vol. 61, no. 1, pp. 147-158.

[15] M. Al-Alwan, S. Al-Nawafah and H. Al-Shorman, "The effect of big data on decision quality Evidence from telecommunication industry," *Int J Data Netw Sci*, 2022, vol. 6, no. 3, pp. 693-702.

[16] T. M. Choi, H. K. Chan and X. Yue, "Recent development in big data analytics for business operations and risk management," *IEEE T Cybernetics*, 2016, vol. 47, no. 1, pp. 81-92.

[17] D. Yang, P. Chen and F. Shi, "Internet finance: Its uncertain legal foundations and the role of big data in its development," *Emerg Mark Financ Tr*, 2018, vol. 54, no. 4, pp. 721-732.

[18] K. S. Namahoot and T. Laohavichien, "Assessing the intentions to use internet banking: The role of perceived risk and trust as mediating factors," *Int J Bank Mark*, 2018, vol. 36, no. 2, pp. 256-276.

[19] N. Metawa and S. Metawa, "Internet financial risk early warning based on big data analysis," *Am J Bus Oper Res*, 2021, vol. 3, no. 1, pp. 48-60.

[20] H. Shang, D. Lu and Q. Zhou, "Early warning of enterprise finance risk of big data mining in internet of things based on fuzzy association rules," *Neural Comput Appl*, 2021, vol. 33, no. 9, pp. 3901-3909.

[21] Q. Ding, "Risk early warning management and intelligent real-time system of financial enterprises based on fuzzy theory," *Int J Fuzzy Log Inte*, 2021, vol. 40, no. 4, pp. 6017-6027.

[22] S. Hongjin, "Analysis of risk factors in financial supply chain based on machine learning and IoT technology," *J Intell Fuzzy Syst*, 2021, vol. 40, no. 4, pp. 6421-6431.

[23] A. Akinola, "The nexus between bank size and financial performance, pp. Does internal control adequacy matter?" *J Account Tax*, 2022, vol. 14, no. 1, pp. 13-20.

[24] Y. E. Park, "Developing a COVID-19 crisis management strategy using news media and social media in big data analytics," *Soc Sci Comput Rev*, 2022, vol. 40, no. 6, pp. 1358-1375.

[25] G. Jakka, N. Yathiraju and M. F. Ansari, "Artificial intelligence in terms of spotting malware and delivering cyber risk management," *J Posit Sch Psychol*, 2022, vol. 6, no. 3, pp. 6156–6165.

[26] S. Sahoo, "Big data analytics in manufacturing, pp. A bibliometric analysis of research in the field of business management," *Int J Prod Res*, 2022, vol. 60, no. 22, pp. 6793-6821.

[27] J. Liu, Y. Jiang and S. Gan, "Can digital finance promote corporate green innovation?" *Environ Sci Pollut R*, 2022, vol. 29, no. 24, pp. 35828-35840.

[28] Ayoub Mniai and Khalid Jebari. "Credit card fraud detection by improved SVDD," Proceedings of The World Congress on Engineering 2022, 6-8 July, 2022, London, UK, pp6-8.

[29] Narong Wichapa, and Suphan Sodsoon. A Relative Closeness Coefficient Model Based on the Distance of Virtual DMUs Cross-Efficiency Method for Ranking Thai Economic Development. Engineering Letters, vol. 31, no.1, pp122-135, 2023

[30] Jingxian Guo, and Wenjie Gong, "The Impact of Carbon Emissions Trading Scheme on Corporate Financial Performance: Based on the Moderating Effect of Carbon Prices," IAENG International Journal of Applied Mathematics, vol. 53, no.4, pp1162-1169, 2023