

## Research Article

# Identification of Accounting Fraud Based on Support Vector Machine and Logistic Regression Model

Rongyuan Qin 

*School of Economics and Management, Hanjiang Normal University, Shiyan, Hubei 442500, China*

Correspondence should be addressed to Rongyuan Qin; teacherqin@stu.tjufe.edu.cn

Received 22 February 2021; Revised 10 March 2021; Accepted 30 March 2021; Published 12 April 2021

Academic Editor: Wei Wang

Copyright © 2021 Rongyuan Qin. 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.

The authenticity of the company's accounting information is an important guarantee for the effective operation of the capital market. Accounting fraud is the tampering and distortion of the company's public disclosure information. The continuous outbreak of fraud cases has dealt a heavy blow to the confidence of investors, shaken the credit foundation of the capital market, and hindered the healthy and stable development of the capital market. Therefore, it is of great theoretical and practical significance to carry out the research on the identification and governance of accounting fraud. Traditionally, accounting fraud identification is mostly based on linear thinking to build the fraud identification model. However, more and more studies show that fraud has typical nonlinear characteristics, and the multiobjective of fraud means also determines the limitations of using the linear model for identification. Considering that the traditional identification methods may have the defects of model setting error and insufficient information extraction, this paper constructs the support vector machine and logistic regression model to identify accounting fraud. The support vector machine is used to improve the learning ability and generalization ability of unknown phenomena, and the explanatory power of each variable to the whole model is identified by the logistic regression model. This paper breaks through the linear constraint hypothesis and explores the model setting form which is more suitable for the law of corporate fraud behaviour to extract the fraud identification information more fully and provide more powerful support for investors to effectively identify fraud.

## 1. Introduction

The authenticity of the company's accounting information is an important guarantee for the effective operation of the capital market. Accounting fraud is the tampering and distortion of the company's public disclosure information [1]. The continuous outbreak of fraud cases has dealt a heavy blow to the confidence of investors, shaken the credit foundation of the capital market, and hindered the healthy and stable development of the capital market. Therefore, it is of great theoretical and practical significance to carry out the research on the identification and governance of accounting fraud. For a long time, scholars have carried out rich and fruitful research and exploration [2]. From the early theoretical analysis and questionnaire research to deepen the understanding of the causes of fraud, to the investigation of single fraud identification factor, and then to the construction of the multidimensional index fraud identification

model, the research on fraud identification and governance has achieved fruitful results. However, with the accumulation of research and the deepening of the understanding of the essential characteristics of accounting fraud, as well as the increasing complexity of fraud, the limitations of existing research on fraud identification are becoming increasingly prominent [3]. Due to the complexity and multiobjective of fraud, in order to improve the effect of fraud identification, the dimension of the index system is getting higher and higher, and the multicollinearity of variables is increasing, and it is necessary to consider the elimination of multicollinearity to improve the performance of the model.

This paper reviews the literature on corporate accounting fraud identification and governance at home and abroad, uses the support vector machine to improve the learning ability and promotion ability of unknown phenomena, and uses the logistic regression model to identify

the explanatory power of each variable to the whole model. The reason why the support vector machine has excellent generalization ability is that it is based on the principle of structural risk minimization [4]. By mapping the enter vector to the high-dimensional function space, the finest classification floor can be constructed, which makes up for the defects that cannot be solved by means of the multilayer ahead community [5]. The quadratic programming optimization used by using the aid vector computer can discover the world's most effective solution, which is a hard hassle that the neural community neighborhood minimal trouble cannot overcome; however, the explanatory capacity of the assist vector computer is weak, and the logistic regression mannequin can simply become aware of the explanatory strength of every variable to the entire model, which can assist us look at the impact of every variable on the structured variable [6]. The correlation between each variable and the fraud risk index and the significance of the mean difference of each variable under different levels of fraud risk are tested, and the countermeasures of fraud governance are put forward accordingly.

The rest of this paper is organized as follows: the related work is discussed in Section 2. In Section 3, the support vector machine and logistic regression model are described. In Section 4, based on the support vector machine and logistic regression model, the experiment design and analysis are carried out. Section 5 summarizes the whole paper.

## 2. Related Work

The academic circle has paid close attention to the problem of accounting fraud for a long time and has made a lot of research results in the identification method. People used the logistic regression model to judge most of the accounting fraud and pointed out that accounting data contains effective information to identify accounting fraud. Relevant scholars used the first simulated test data as output variables and built an artificial neural network model to identify accounting fraud based on raw financial data [7]. It finds that this model will greatly improve the ability of independent auditors to discover fraud and suggests that auditors should use this model in the initial stage of audit [8]. Through the analysis of the company's financial indicators by relevant scholars, it is found that the turnover rate of accounts receivable, gross profit rate, asset quality index, sales growth index, and asset liability ratio index can provide useful reference for investors to analyse whether the company carries out accounting fraud [9]. Some scholars introduced five ways to identify accounting fraud from a qualitative perspective: focusing on external financial indicators, focusing on subjects with high audit risk, analysing the relationship between the three major statements, and analysing the structure of cash flow statement and abnormal fluctuation indicators [10]. In addition, we use descriptive statistics, two-population heteroscedasticity test, and binary logistic regression to analyse the financial report fraud of Chinese companies. The empirical research shows that small scale and deteriorating financial situation enterprises are more prone to financial fraud [11]. Based on 29 fraud samples and 29 nonfraud samples, this paper uses univariate and multivariate statistical techniques to

establish a model to identify false financial reports and finds that the correct recognition rate of the model is more than 75% [12]. Using 68 fraud samples and 68 nonfraud samples, four methods were used to establish the recognition model, and the highest recognition rate of the logistic regression model was 77% [13].

From the research status of financial report fraud, most scholars are aware of the significant relationship between financial indicators and the identification of financial statement fraud and pay attention to the identification information provided by abnormal financial indicators [14]. However, most of them are mainly empirical analysis methods, which have higher professional requirements and have not considered the use of many financial indicators for inspection. Some financial indicators may have internal correlation. Therefore, it is very helpful to improve the accuracy of the model if the financial indicators for statistical analysis are tested or processed first.

Among the statistical analysis methods, the logistic regression model has the characteristics of strong explanation and high discrimination accuracy [15]. It is the most widely used model in the application field of fraud identification. Some established logistic regression models, according to KPMG's data, used the logistic regression model for empirical research; some took manufacturing fraud companies as samples, comprehensively used the logistic regression model and principal component analysis method for research; some used the logistic regression model for discrimination and verified the logistic regression model and the effectiveness of regression model in the identification of financial report fraud [16]. However, the logistic regression model also has some problems to be improved, mainly because the model is very sensitive to multicollinearity of variables and the calculation process is complex. Therefore, it is of great significance to effectively display the key indicator variables of economic document fraud identification from many economic indicators. This not only can effectively simplify the human body model and limit the complexity of the calculation, but also can weaken the influence of the multicollinearity of unbiased variables on the human body model and improve the model's recognition ability.

The essence of screening the variables in the model is to select the best model in the optional model set. The early variable selection methods mainly include the optimal subset method and forward or backward stepwise regression model, but there are some defects in the application [17]. The optimal subset method is difficult to solve when the variable dimension is large, while the forward or backward stepwise regression model is more sensitive to the change of the variable structure, and the stability of the model needs to be improved. In order to overcome the shortcomings of these traditional methods, the variable selection method based on the penalty function has gradually attracted the attention of most researchers. The lasso method proposed by Toshigami has become one of the most used methods for variable selection [18]. It transforms the penalty function into the form of the absolute value and compresses the coefficients of some variables to 0 by compressing the regression coefficients in order to achieve the effect of variable screening. In this paper, the support vector machine is used to improve the learning ability and generalization ability of unknown phenomena,

and the explanatory power of each variable to the whole model is identified by the logistic regression model. By mapping the input vector to the high-dimensional feature space, the support vector machine can construct the optimal classification surface, which can make up for the defects of the multilayer feedforward network. The quadratic programming optimization used by the support vector machine can find the global optimal solution, which is a difficult problem that the neural network local minimum problem cannot overcome, but the explanatory ability of the support vector machine is weak, and the logistic regression model can just identify the explanatory power of each variable to the whole model, which can help us observe the influence of each variable on the dependent variable.

### 3. Variable Selection and Model Construction

This paper constructs the support vector machine and logistic regression model to identify accounting fraud. The support vector machine is used to improve the learning ability and generalization ability of unknown phenomena, and the explanatory power of each variable to the whole model is identified by the logistic regression model. The details are shown in Figure 1.

**3.1. Variable Selection.** Accounting fraud refers to the intentional omission or falsification of financial accounting information in the process of external reporting. According to the types of fraud, accounting fraud is mainly divided into financial statement fraud and illegal disclosure of accounting

information. The latter usually involves the relevant laws and regulations and the rules and regulations of the exchange, so it is not a simple accounting problem [19, 20]. This study only focuses on financial statement fraud; from the balance sheet as the core and the income statement as the core, financial statement fraud can be divided into false net assets and false profit. The common manipulation items are accounts receivable, accounts payable, inventory, depreciation, prepaid expenses, and other accruals; the false profit is often achieved by means of recognizing or fabricating revenue in advance, understating costs and expenses, and concealing losses, and the report items involved include operating income, operating profit, net profit, and period expenses. In addition, previous studies have shown that financial health is an important factor affecting corporate accounting fraud, and Altman's Z-score provides a comprehensive observation to measure corporate financial health.

**3.2. Support Vector Machine.** The main idea of using the support vector machine for classification prediction is to establish an optimal classification hyperplane as the decision surface, to maximize the interval between various types, as shown in Figure 2.

SVM is based on the optimal classification hyperplane in the case of linear separability. Suppose training set T:

$$T = \{(x_i, y_i) | i = 1, 2, \dots, n\}, x_i \in R^n, y_i = \{-1, 1\}. \quad (1)$$

To find such an optimal classification hyperplane, the Lagrange function can be used:

$$L(w, b, \xi, \alpha, \beta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n s_i \xi_i - \sum_{i=1}^n \alpha_i [y_i (w \cdot \varphi(x_i) + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i \xi_i. \quad (2)$$

When the training set is linearly nonseparable, in order to construct the optimal hyperplane in the case of linearly nonseparable data, we need to introduce a nonnegative relaxation variable  $\xi_i \geq 0 (i = 1, 2, \dots, n)$ ; for the nonlinear classification problem, SVM maps the sample space to the Hilbert space by introducing a kernel function  $\kappa(x_i, x_j)$ . In this way, the nonlinear classification problem can be transformed into a linear classification problem.

At this point, by introducing the kernel function  $\kappa(x_i, x_j)$ , the optimization function becomes

$$Q(a) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \alpha_i \alpha_j \kappa(x_i, x_j). \quad (3)$$

And, the corresponding decision function becomes

$$f(x) = \text{sgn} \left[ \sum_{i=1}^n \alpha_i^* y_i \kappa(x_i, x) + b^* \right]. \quad (4)$$

Different kernel functions can be used to construct different types of nonlinear decision surface learning machines in the input space.

**3.3. The Model of Logistic Regression.** Logistic regression model is a discrete choice model based on maximum likelihood estimation (MLE). Unlike the least squares (OLS) method commonly used in linear regression, which takes the least square deviation between the observed value of the variable and the estimated value of the model as the optimization condition, the maximum likelihood estimation (MLE) method takes the maximum probability to reproduce the observed value of the sample in random selection as the optimization condition of parameter estimation [21]. Suppose the probability of accounting fraud is  $p (0 < p < 1)$ , and the probability of not implementing accounting fraud is  $1 - p$ . There are  $Q$  indicators to judge whether accounting fraud occurs, which are  $x_1, x_2, \dots, x_q$ . Generally, the logistic regression model was set as follows:

$$p_i = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}. \quad (5)$$

Take the natural logarithm on both sides of the equation:

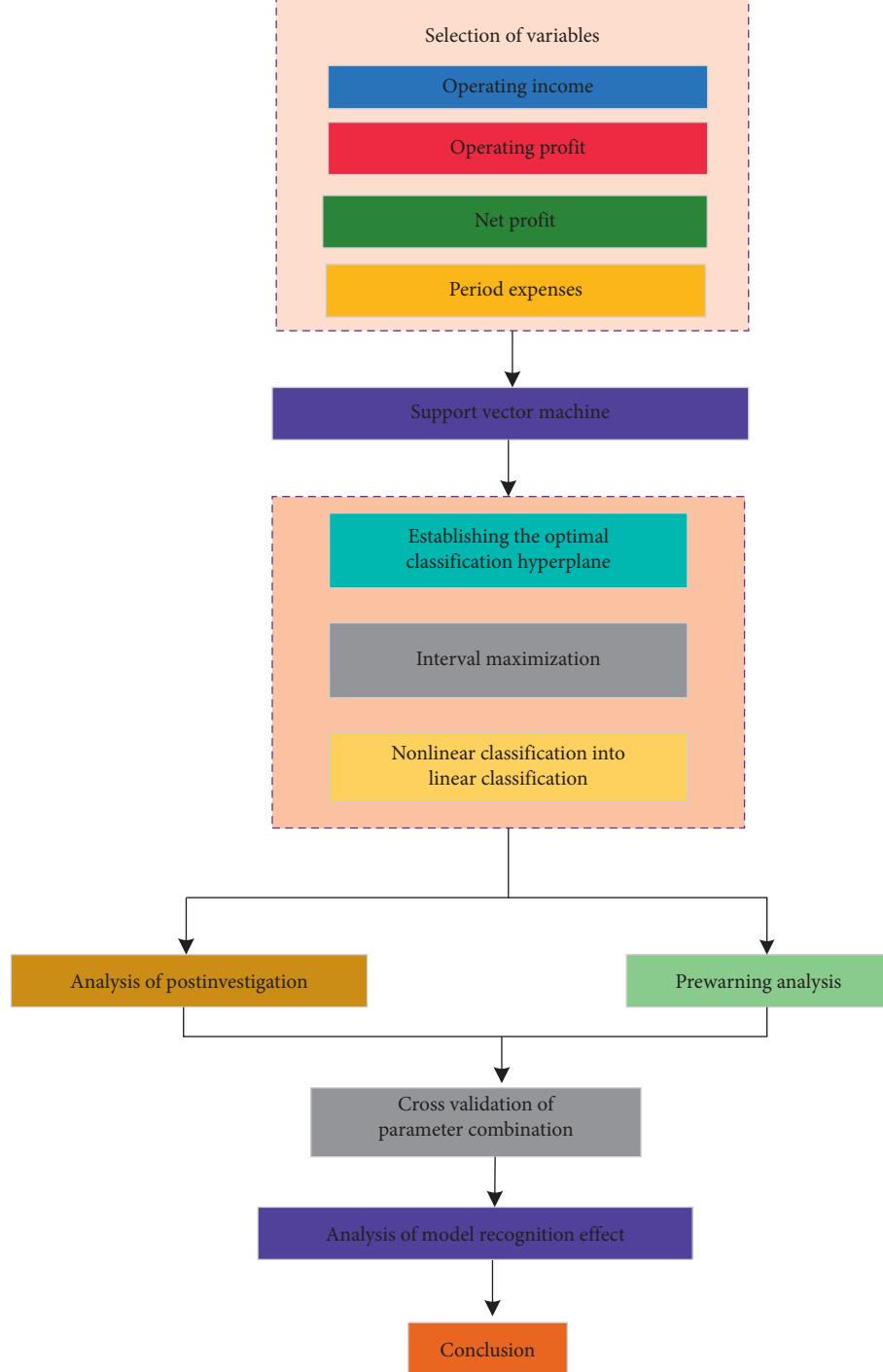


FIGURE 1: Framework of the overall study.

$$\ln\left(\frac{p}{1-p}\right) = \alpha_0 + \sum_{i=1}^n \alpha_i F_i + \sum_{i=1}^n \beta_i F_i^2 + \sum_{i=1, j=2, i < j}^n \gamma_{ij} F_i F_j. \quad (6)$$

In order to effectively identify accounting fraud, the model usually contains multidimensional variables.

However, the increase of variable dimensions easily leads to high correlation between multiple explanatory variables. When there are many explanatory variables in the model, principal component analysis uses principal component transformation to extract unrelated principal components instead of the original variables [22]. On the basis of retaining the information of the original variables to the

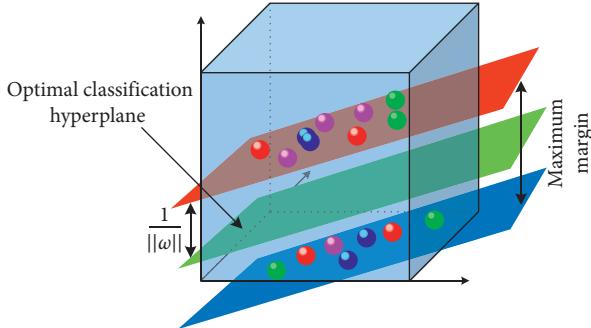


FIGURE 2: Schematic diagram of the optimal classification hyperplane.

greatest extent, it can eliminate the correlation between variables and achieve the purpose of model dimension reduction.

**3.4. The Framework of Fraud Analysis.** In this paper, the accounting information fraud in enterprises except for the embezzlement of physical assets is generally called accounting fraud, which is further divided into illegal disclosure of accounting information and financial report fraud. And, the definition, the so-called accounting information disclosure, refers to the enterprise in order to achieve a certain purpose in violation of the relevant laws and regulations of the disclosure behaviour, which is limited to the relevant information disclosure link. Fraud in the financial report refers to the manipulation of substantial activities by an enterprise in order to achieve a certain purpose [23]. The manipulation is limited before the disclosure of information, and the disclosed information is true after being distorted. In essence, it is a systematic planning activity for the whole enterprise accounting information. If we say that financial statement fraud is a true reflection of false economic business, accounting information disclosure is a false reflection of real economic business. This division can provide convenience for the study of the causes of fraud and then put forward antifraud countermeasures. Based on this, this paper proposes an analysis framework, as shown in Figure 3.

From the perspective of research methods, domestic and foreign scholars tend to empirical analysis and data mining on the causes of financial fraud. The main methods involved are probability analysis, building the linear probability model, regression model, normal distribution model, and artificial neural network model to build the identification model. The regression model uses the maximum likelihood estimation method to calculate the probability of fraud, and it does not have strict assumptions and does not need to obey the normal distribution, so it is used by most scholars to establish the fraud model [24]. Next, this paper uses the logistic method to build the financial report fraud prewarning model and postinvestigation model. After the recognition efficiency is obtained, the recognition rate of the two models is compared.

According to the theory of fraud, through the deep analysis of the greed factor, opportunity factor, need factor,

and violence factor, we can accurately identify the possibility of fraud in the financial report. Therefore, the prewarning model established in this paper is based on the fraud theory and constructs the prewarning model of fraud behaviour with the setting of the relevant substitution variables of four factors. The process of accounting fraud identification model is shown in Figure 4.

## 4. Experimental Design and Result Analysis

**4.1. Regression Analysis of Prewarning and Postinvestigation Model.** It can be seen from Figure 5 that the chi square value of the prewarning model is 75.718, and the significance level is 0.000. Therefore, the independent variable test of the pre-warning model is significant. The goodness of fit of the model is 0.435, which shows that the model has good explanatory power. From the perspective of indicators, if the estimated coefficient of managers' risk preference, chairman and general manager's concurrent appointment, and management's shareholding ratio are positive, it indicates that managers' risk preference, chairman and general manager's concurrent appointment, and management's shareholding ratio are positively correlated with financial reporting fraud; if the estimated coefficient of directors' personnel replacement frequency, related party transaction frequency, and their influence is positive [25], it indicates that the risk preference of managers, chairman and general manager's concurrent appointment, and management's shareholding ratio are positively correlated with financial reporting fraud. If the estimated coefficient of the size of the regulatory council and the number of shareholders' meetings of the company is negative, then it is inversely proportional to the occurrence of fraud; if the estimated coefficient of the audit opinion type is negative, then it is inversely proportional to the occurrence of fraud; if the estimated coefficient of the current ratio and asset liability ratio is negative, then it is inversely proportional to the occurrence of fraud. The debt paying ability of the company is inversely proportional to the fraud.

As can be seen from Figure 6, the chi square value of the ex-postinvestigation model is 23.249, and the significance level is 0.010. Therefore, the independent variable test of the ex-postinvestigation model is significant. The goodness of fit of the model is 0.152, which shows that the ex-postinvestigation model has good explanatory power. From the index point of view, the company's abnormal financial indicators, that is, the worse the company's profitability, solvency, and operation ability, will lead to the implementation of financial reporting fraud.

**4.2. Cross-Validation Analysis of Parameter Combination.** According to the theory of fraud, through the deep analysis of the greed factor, opportunity factor, need factor, and exposure factor, we can accurately identify the possibility of fraud in the financial report. Therefore, the prewarning model established in this paper is based on the fraud theory and constructs the prewarning model of fraud behaviour with the setting of the relevant substitution variables of four factors. In this model, the value range of  $C$  is  $[2^{-10}, 2^{10}]$ , the value range of  $\gamma$  is  $[2^{-10}, 2^{10}]$ , and the single change step is  $2^{-1}$ , with a total of  $21 * 21$  ( $C, \gamma$ )

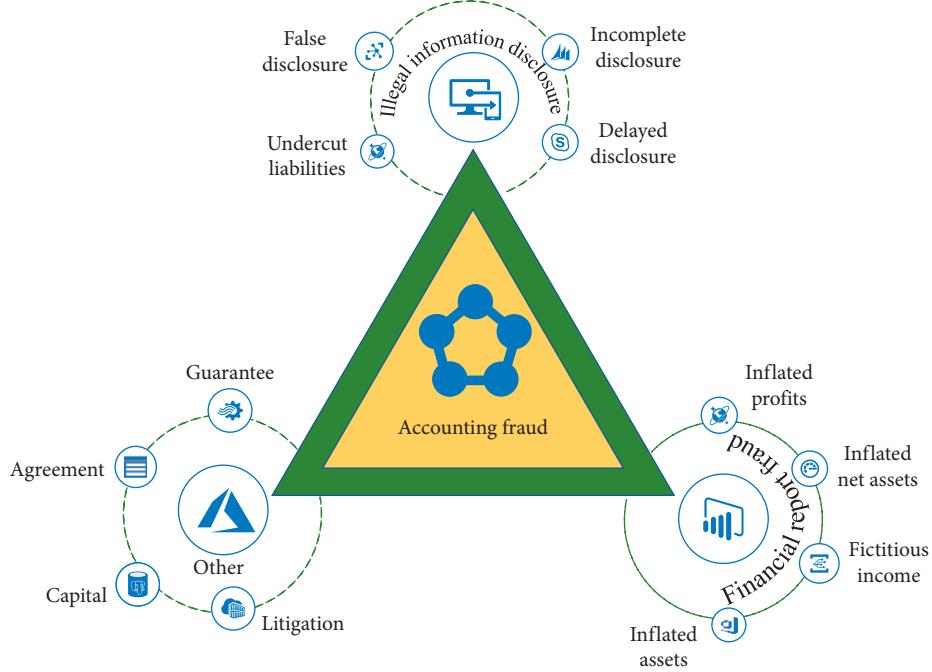


FIGURE 3: Analysis framework of accounting fraud.

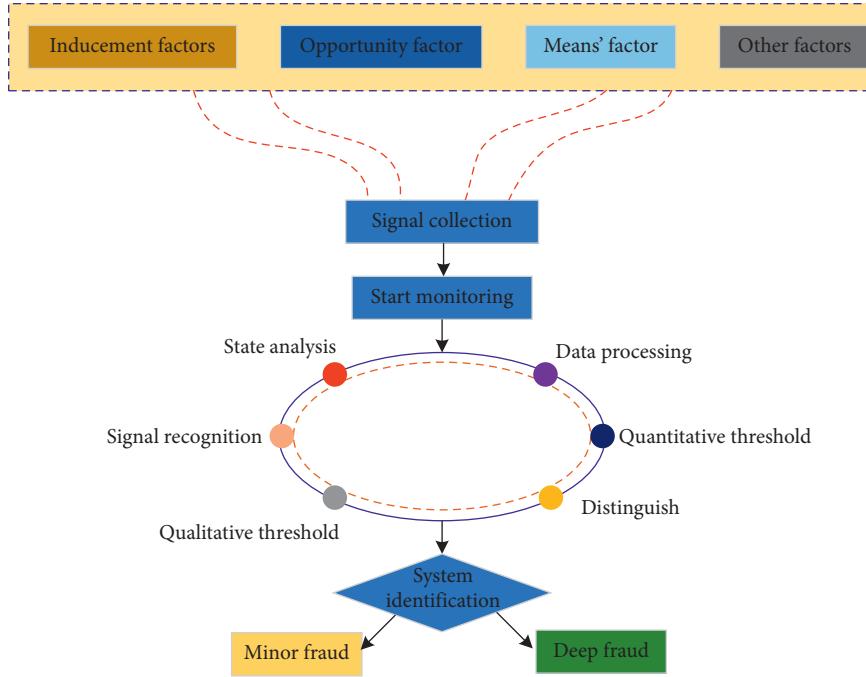


FIGURE 4: The process of the accounting fraud identification model.

parameter combinations. The details of cross-validation rate parameter combinations are shown in Figure 7.

The establishment of the prewarning model mainly takes four fraud factors as the starting point to identify the possibility of fraud. Afterwards, the reconnaissance model is used to detect the abnormal data of financial statements after the occurrence of fraud, in order to identify the possibility of financial fraud. In this model, the value range of  $C$  is  $[2^{-10}, 2^{10}]$ , the value range of  $\gamma$

is  $[2^{-10}, 2^{10}]$ , and the single change step is  $2^{-1}$ , with a total of  $21 * 21$  ( $C, \gamma$ ) parameter combinations. The cross-validation details are shown in Figure 8.

Through the grid search method based on cross validation, the two key parameters  $c$  and  $\gamma$  of the RBF kernel function are determined to be (8, 0.25). Figure 9 is a three-dimensional graph of the cross-validation rate changing with parameters. It shows that the cross-validation accuracy

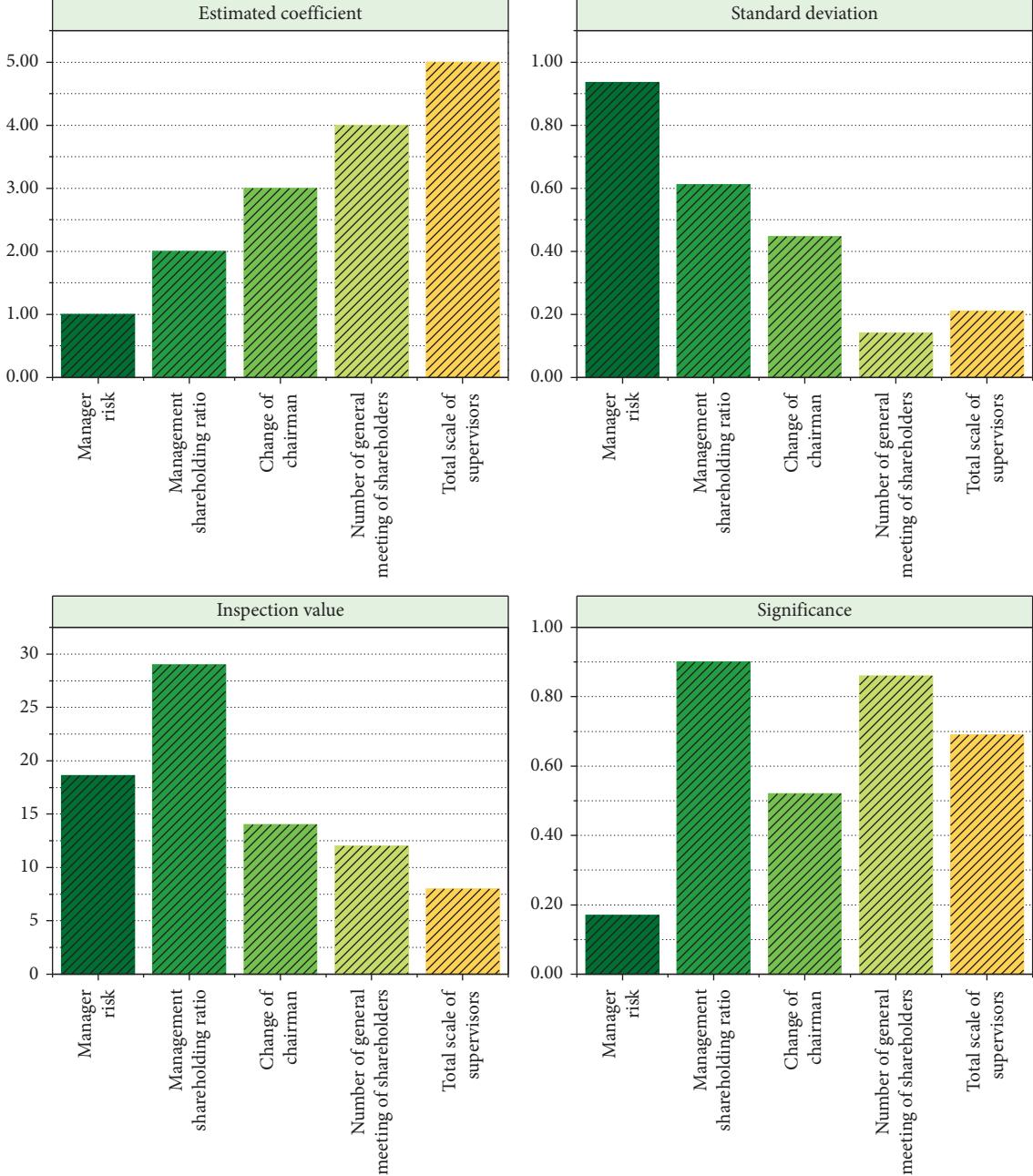


FIGURE 5: Analysis on the regression model of prewarning.

rate is between 30% and 79.17% and fluctuates with the change of  $(C, \gamma)$  parameter combination. It shows that the cross-validation accuracy rate cannot reject the influence of  $C$  and  $\gamma$ . The results also show that the accuracy of cross validation is stable, and the support vector machine has good anti-microvariation characteristics. The recognition accuracy of cross validation is 79.17% at the best, and 79.17% at the parameter combination (8, 0.25).

Through the grid search method based on cross validation, the two key parameters  $c$  and  $\gamma$  of the RBF kernel function are determined as (4, 2). Figure 10 is a three-dimensional graph of the cross-validation rate changing with parameters. It shows that the accuracy of cross validation is floating in the range of

30%–65.10% due to the change of  $(C, \gamma)$  parameter combination. It shows that the accuracy of cross validation cannot reject the influence of  $C$  and  $\gamma$ . The results also show that the accuracy of cross validation is stable, and the support vector machine has good anti-microvariation characteristics. The recognition accuracy of cross validation is 65.10% at the best, and only the combination of parameters (4, 2) is 65.10%.

**4.3. Comparative Analysis of the Model Recognition Effect.** Under the fixed distribution and fixed number of samples, by comparing the experimental results of Step 1, Step 2, and Step 3, it is concluded that the accuracy of the experimental results of Step 1 is the highest, which determines that, in this

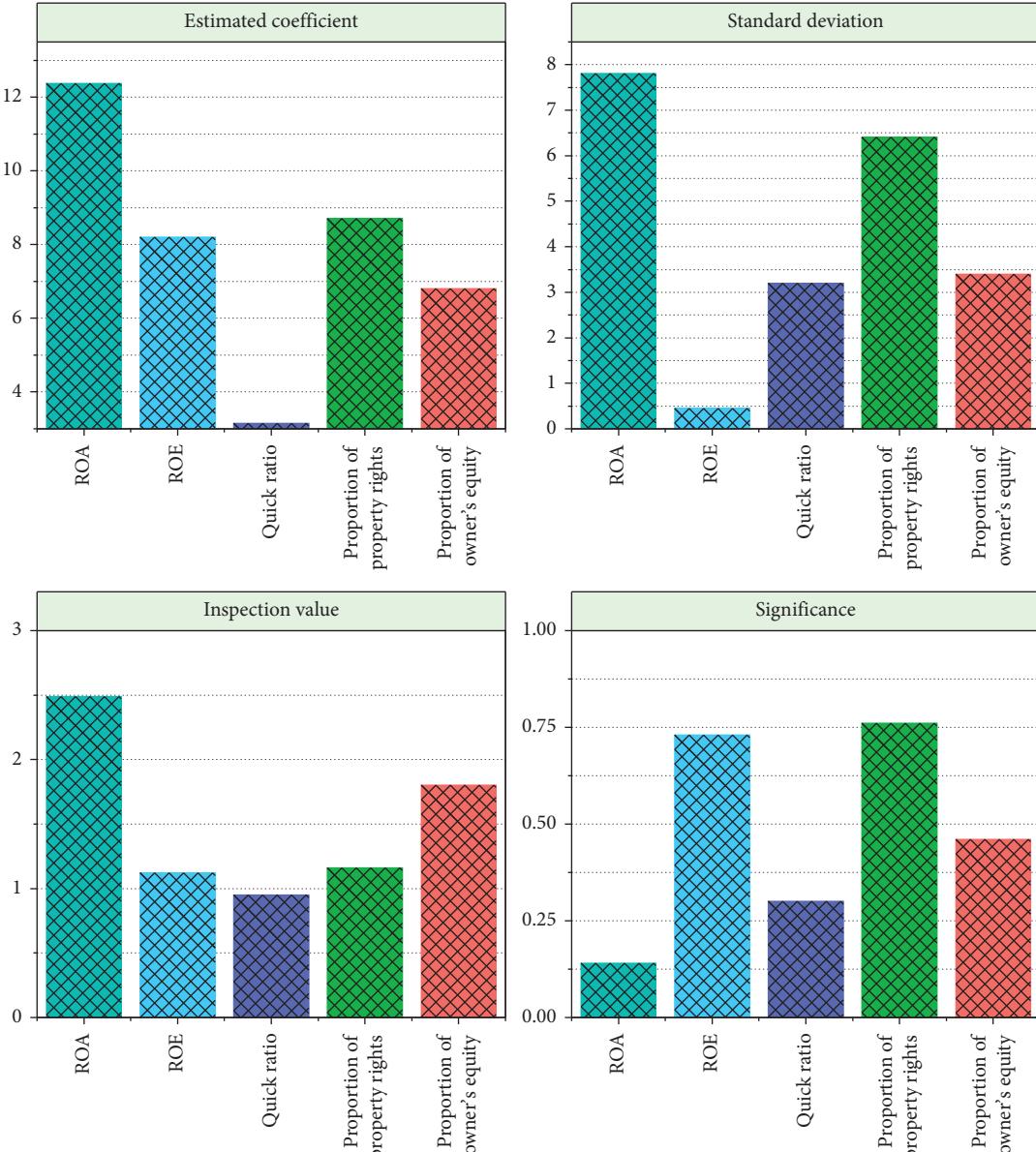


FIGURE 6: Regression analysis on the model of investigation after the event.

experiment, the accounting fraud identification model is directly established by logistic regression. The fourth step is to know the influence of data distribution and training sample size on the prediction accuracy under the condition of a certain accounting fraud identification model. From the two groups of experimental data, 50, 100, 150, 200, 250, and 300 samples of fraud and nonfraud samples were selected, respectively, for pairing, and then, logistic regression was used directly. It can be seen that, in the data with the same distribution, the sample size has a significant impact on the prediction accuracy. The smaller the sample is, the higher the accuracy is, and the larger the sample is, the lower the accuracy is. However, when the sample is large enough, the accuracy will remain at a certain level, which is the real fraud identification level of the model. In this experiment, when the sample size is 200, it is basically maintained between 71%

and 74%. Therefore, when using the logistic model to build the accounting fraud identification model, it is better to have a larger sample size, to avoid excessively high prediction accuracy, which may be seriously inconsistent with the reality; in different distribution of data, the sample size is the same, but the prediction accuracy is different. The accuracy rate of recognition results of the accounting fraud identification model is shown in Figure 11.

When the accounting fraud identification model and the financial indicators that should be included in the model are determined, the possibility of accounting fraud of a company can be calculated. However, why these financial indicators are used to identify accounting fraud and what are the differences and commonalities between the fraud samples and the nonfraud samples can be understood through descriptive statistics of the original data. The financial

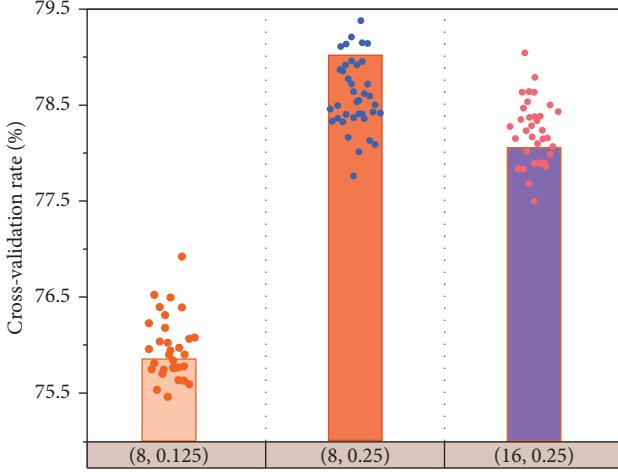


FIGURE 7: Parameter combination and cross-validation accuracy of prewarning.

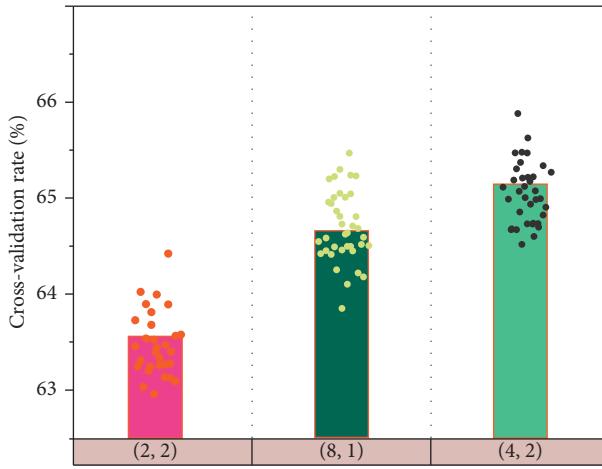


FIGURE 8: Parameter combination and cross-validation accuracy of postinvestigation.

indicators included in the model are divided into fraud group and nonfraud group for descriptive statistics, and the results are shown in Figures 12 and 13. From the experimental results of Step 5, we can see that the variance of financial indicators  $x_1$ ,  $x_3$ ,  $x_5$ , and  $x_{11}$  in the fraud samples is larger than that of the same financial indicators  $A_1$ ,  $A_3$ ,  $A_5$ , and  $A_{11}$  in the nonfraud samples; but the variance of  $x_2$  in the fraud samples is smaller than that of the same financial indicators  $A_2$  in the nonfraud samples, which proves that most of the financial indicators included in the accounting fraud identification model are smaller than those in the fraud samples in the nonfraud samples, and the volatility is large.

Under the support vector machine method, with the increase of the number of training samples, the training accuracy is also improved, and the corresponding classifier promotion ability is also on the rise, and the test accuracy is more than 94. It can be seen that the machine learning method of the support vector machine is more reliable and effective in processing financial data. In terms of the overall recognition rate, the overall recognition rate of the SVM algorithm is higher than

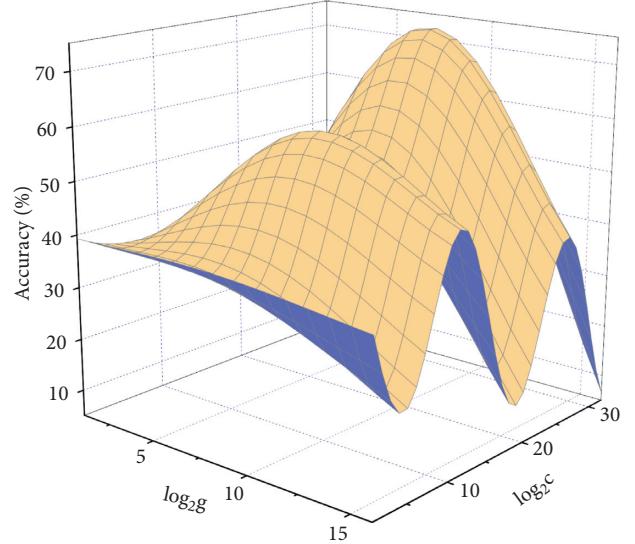


FIGURE 9: The change chart of the accuracy of cross validation in advance warning.

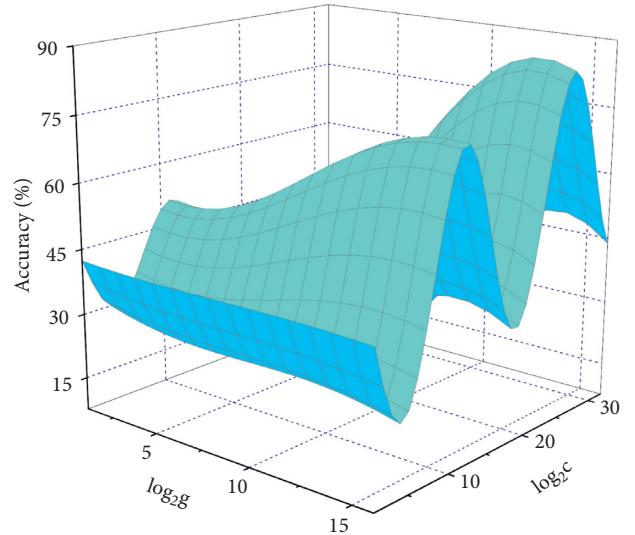


FIGURE 10: The change chart of the accuracy of cross validation in postinvestigation.

that of the logistic regression model, which shows the good performance of the classifier. The logistic regression model has a high explanatory power in the influence of various variables on the probability of fraud. Second, in terms of fraud pressure, growth rate, bankruptcy risk, asset liability ratio, and loss or not, the number of senior executives' shareholding has significant positive correlation with the possibility of financial fraud, while net cash flow generated from operating activities/total liabilities, growth rate of operating income, and annual compensation of senior executives have significant negative correlation with the probability of fraud. Thirdly, in terms of fraud opportunities, the more the board meetings, the higher the probability of fraud; the lower the control degree of the largest shareholder, the attendance rate of shareholders' meeting, and the attendance rate of independent directors, the higher the probability of fraud.

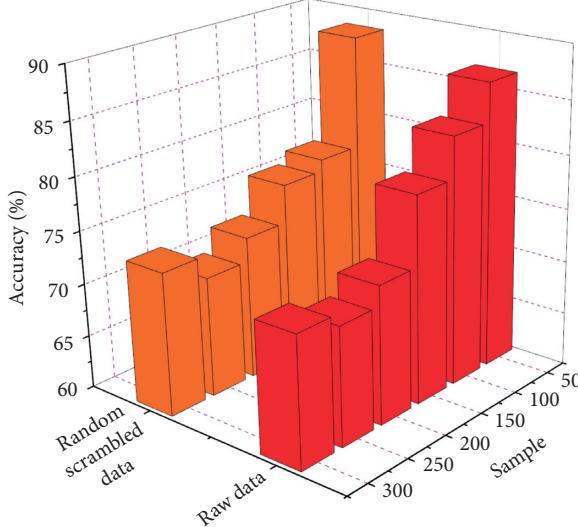


FIGURE 11: Accuracy rate of recognition results of the accounting fraud identification model.

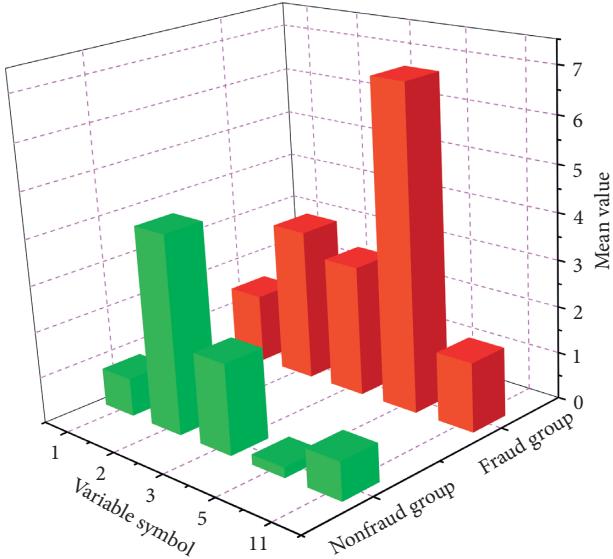


FIGURE 12: Mean value statistics of different groups.

Fourth, in terms of fraud excuse, the more times we have obtained nonstandard audit opinions, the higher the possibility of fraud. Therefore, to effectively prevent and detect fraud, we should not only strengthen the construction and operation process of internal control but also strengthen the external supervision of shareholders, independent directors, and independent auditors. First, ease the internal and external pressure faced by enterprises and management. By formulating a scientific development strategy, designing a stable organizational structure, and creating a reasonable incentive mechanism, a healthy corporate culture can be established, and the internal control environment can be improved. Second, internal control and risk management evaluation should be carried out to prevent fraudsters from being provided with any opportunities for fraud. Third, strengthen internal and external supervision to eliminate fraud excuse. The empirical results of this paper show that the problem of the internal supervision function is still

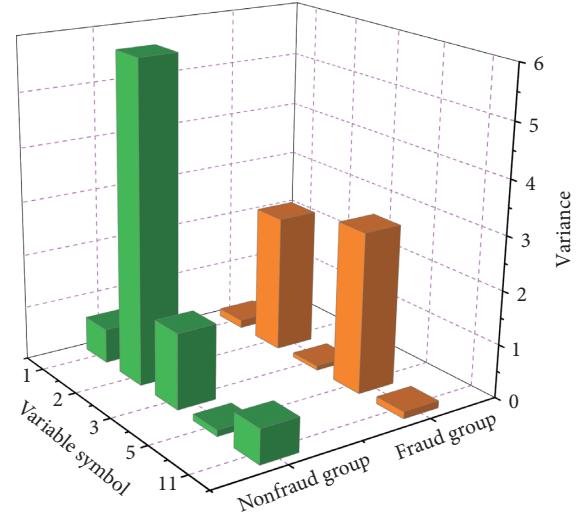


FIGURE 13: Variance statistics of different groups.

serious in China. While improving the efficiency and effect of internal supervision, we should strengthen the supervision of external investors and auditors.

## 5. Conclusion

In this study, with the increase of the number of training samples, the training accuracy is also improved, and the corresponding classifier promotion ability is also on the rise, and the test accuracy is more than 94%. The machine learning method of the support vector machine is used to process the financial data, and its predictive ability shows that the judgment of unknown data is more reliable and effective. In terms of the overall recognition rate, the overall recognition rate of the SVM algorithm is higher than that of the logistic regression model, which shows the good performance of the classifier. The logistic regression model has a high explanatory power in the influence of various variables on the probability of fraud. Second, in terms

of fraud pressure, growth rate, bankruptcy risk, asset liability ratio, and loss or not, the number of senior executives' shareholding has significant positive correlation with the possibility of financial fraud, while net cash flow generated from operating activities/total liabilities, growth rate of operating income, and annual compensation of senior executives has significant negative correlation with the probability of fraud. Relieve the inside and exterior stress confronted via companies and management. Through scientific improvement strategy methods, safe organization chart, and realistic incentive mechanism, a healthy company life style can be established to improve the internal manipulation environment and enhance the efficiency of internal manipulation and hazard management assessment.

## Data Availability

The 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 the publication of this paper.

## References

- [1] Z. Wang, Z. Zou, and C. Guedes Soares, "Identification of ship manoeuvring motion based on nu-support vector machine," *Ocean Engineering*, vol. 183, no. 11, pp. 270–281, 2019.
- [2] Q. Zheng, X. Tian, M. Yang et al., "The email author identification system based on support vector machine (svm) and analytic hierarchy process," *International Journal of Computer Science*, vol. 46, no. 2, pp. 178–191, 2019.
- [3] X. Z. Qiu, G. D. Bai, B. H. Huang, and W. Xu, "Identification of thermal process based on relevance vector machine and firefly algorithm," *Applied Mechanics and Materials*, vol. 742, no. 2, pp. 344–348, 2015.
- [4] L. Huang, Z. Wu, W. Huang et al., "Identification of fusarium head blight in winter wheat ears based on Fisher's linear discriminant analysis and a support vector machine[J]," *Applied Sciences*, vol. 9, no. 18, pp. 38–44, 2019.
- [5] D. Iswati, M. Nindito, and A. Zakaria, "The effect of internal financial indicators on the tendency of accounting fraud," *Jurnal Dinamika Akuntansi*, vol. 9, no. 2, pp. 123–131, 2017.
- [6] I. D. Pamungkas, I. Ghozali, T. Achmad et al., "Corporate governance mechanisms in preventing accounting fraud: a study of fraud pentagon model," *Journal of Applied Economic Sciences*, vol. 13, no. 2, pp. 549–560, 2018.
- [7] X. Wang, M. Liu, S. Chen et al., "Identification of continuous rotary motor based on improved particle swarm optimizing support vector machine model," *International Journal of Innovative Computing Information and Control*, vol. 14, no. 6, pp. 2189–2202, 2018.
- [8] Z. Taha, R. M. Musa, A. P. Majeed et al., "The identification of high potential archers based on fitness and motor ability variables: a support vector machine approach," *Human Movement Science*, vol. 5, no. 7, pp. 184–186, 2017.
- [9] C. Rahmad, F. Rahutomo, M. A. Gustalika et al., "Identification of sugarcane maturity scale based on RGB, gabor feature extraction and support vector machine," *IOP Conference Series Materials Science and Engineering*, vol. 43, no. 4, pp. 176–180, 2018.
- [10] A. Wang and J. Sun, "Automatic identification of sedimentary facies based on a support vector machine in the aryskum graben, Kazakhstan," *Applied Sciences*, vol. 9, no. 21, pp. 44–49, 2019.
- [11] Y. Sun, N. Yang, Z. Zhang et al., "Structural damage identification study based on kernel component analysis and support vector machine," *Journal of Basic Science and Engineering*, vol. 26, no. 4, pp. 888–900, 2018.
- [12] W. W. Li, J. Peng, M. C. Dong et al., "Automatic identification and extraction of clouds from astronomical images based on support vector machine," *Chinese Astronomy and Astrophysics*, vol. 43, no. 1, pp. 128–142, 2019.
- [13] S. P. Das and S. Padhy, "A novel hybrid model using teaching-learning-based optimization and a support vector machine for commodity futures index forecasting," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 1, pp. 97–111, 2018.
- [14] J. Hou, "Online teaching quality evaluation model based on support vector machine and decision tree," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2193–2203, 2021.
- [15] I. Ridwan, M. Bisri, F. H. Yusran, L. Hakim, and S. Kadir, "Identification of characteristics of land cover in mangkauk catchment area using support vector machine (SVM) and artificial neural network (ANN)," *American Journal of Applied Sciences*, vol. 14, no. 7, pp. 726–736, 2017.
- [16] S. Rabeea, J. Oláh, J. Popp et al., "An investigation of the influence of the worldwide governance and competitiveness on accounting fraud cases: a cross-country perspective," *Sustainability*, vol. 10, no. 3, pp. 588–590, 2018.
- [17] K. Larsen, P. Jørgen, B. Esben et al., "Interpreting parameters in the logistic regression model with random effects," *Biometrics*, vol. 56, no. 3, pp. 909–914, 2015.
- [18] V. Nagarajah and P. Wijkkoen, "Stochastic restricted maximum likelihood estimator in logistic regression model," *Open Journal of Statistics*, vol. 5, no. 7, pp. 1–6, 2015.
- [19] K. Suriya and S. Niwattisaiwong, "Adaptive logit model: enhancement of predictability in logistic regression," *International Journal of Intelligent Technologies and Applied Statistics*, vol. 12, no. 2, pp. 169–181, 2019.
- [20] A. Mustafa, M. Cools, I. Saadi, and J. Teller, "Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM)," *Land Use Policy*, vol. 69, no. 8, pp. 529–540, 2017.
- [21] M. Marbac, P. Tubert-Bitter, and M. Sedki, "Bayesian model selection in logistic regression for the detection of adverse drug reactions," *Biometrical Journal*, vol. 58, no. 6, pp. 1376–1389, 2016.
- [22] H. Hemasinghe, R. S. S. Rangali, N. L. Deshapriya, and L. Samarakoon, "Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka)," *Procedia Engineering*, vol. 212, pp. 1046–1053, 2018.
- [23] R. K. Poudel and S. P. Khanal, "Binary logistic regression model for assessing factors associated with nutritional status of children under five years among chepang community in siddi, chitwan, Nepal," *Nepalese Journal of Statistics*, vol. 4, no. 2, pp. 43–56, 2020.
- [24] S. K. Shah, "Multinomial logistic regression model to identify factors associated with food insecurity in rural households in Nepal," *Nepalese Journal of Statistics*, vol. 4, no. 1, pp. 17–32, 2020.
- [25] I. A. I. Ahmed and W. Cheng, "The performance of robust methods in logistic regression model," *Open Journal of Statistics*, vol. 10, no. 01, pp. 127–138, 2020.