

Retraction

Retracted: Identification and Early Warning of Financial Fraud Risk Based on Bidirectional Long-Short Term Memory Model

Mathematical Problems in Engineering

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

Copyright © 2023 Mathematical Problems in Engineering. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] X. Liu and M. Fan, "Identification and Early Warning of Financial Fraud Risk Based on Bidirectional Long-Short Term Memory Model," *Mathematical Problems in Engineering*, vol. 2022, Article ID 2342312, 8 pages, 2022.

Research Article

Identification and Early Warning of Financial Fraud Risk Based on Bidirectional Long-Short Term Memory Model

Xiaoli Liu¹ and Min Fan ²

¹School of Business, Southwest Jiaotong University Hope College, Chengdu, Sichuan 610400, China

²School of Civil Engineering, Southwest Jiaotong University, Chengdu, Sichuan 610031, China

Correspondence should be addressed to Min Fan; fanmin@swjtu.edu.cn

Received 20 April 2022; Revised 14 June 2022; Accepted 15 June 2022; Published 5 July 2022

Academic Editor: Zaoli Yang

Copyright © 2022 Xiaoli Liu and Min Fan. 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 modern market economy, corporate financial frauds emerge one after another, which have a huge impact on the stock market and triggered an unprecedented credit crisis. Therefore, it is particularly important to identify financial frauds. Improving the efficiency, accuracy, and coverage of fraud identification in financial reports through digital and intelligent means is one of the important links to improve credit risk control of securities companies and also for securities companies to accurately price related financial products of target companies. Traditional recognition methods based on artificial rules can cover relatively limited indicators; rule parameters are set randomly. Besides, it is difficult to make rules based on high-dimension indicators and to dig the hidden deep relationships between indicators. This paper summarizes the relevant indicators of financial fraud identification from the perspective of financial and nonfinancial characteristics. Then, the identification and early warning of financial fraud risk based on bidirectional long-short term memory model are proposed. This method uses the idea of ensemble learning, weights the probability of financial key indicators, and uses the optimal transfer probability to solve the financial fraud risk results. The results show that the industry-specific modeling can significantly improve the accuracy of the financial fraud identification model and it also can effectively help the government regulatory departments, investors, and audit departments to correctly identify the financial fraud of listed companies.

1. Introduction

1.1. Introduction Part. Financial fraud refers to the deliberate creation of false accounting information by the parties involved in accounting activities in order to evade taxes, reap high dividends, and withdraw secret reserve funds [1, 2]. The American Institute of Certified Public Accountants defined financial fraud as deliberate error, omission, or disclosure in financial statements to deceive users of financial statements in consideration of fraud in the Audit of Financial statements. Since the establishment of China's securities market, corporate financial fraud cases emerge one after another. After entering the twenty-first century, although the relevant regulatory authorities have strengthened supervision, there are still many companies' financial fraud information that has not been found by the securities and futures commission (CSRC), which not only leads to the distortion of accounting

information in the whole society and forms the hotbed of insider trading, but also weakens market resources. The configuration of the function endangers the healthy development of the whole social economy. In order to actively create the market economic order of fair competition and promote sustainable, stable, and rapid development of economy, it is urgent to accurately identify the company's financial frauds [3, 4].

Financial fraud in listed companies is of great harm. Firstly, for enterprises, the high performance achieved through financial fraud is doomed to be a flash in the pan. If listed company financial fraud one day is exposed, it is hard to escape the punishment of regulatory departments. At the same time, company's investors and creditors may decide to withdraw, worsening the company's financial situation [5, 6]. In addition, financial fraud scandals will also affect the brand image of listed companies and reduce consumers'

brand trust, resulting in a decline in sales performance. Secondly, investors will make wrong decisions by using false financial information. When financial fraud is disclosed, the stock price of the enterprise suspected of fraud often falls sharply or even goes bankrupt, which will seriously damage the interests of investors [7, 8]. Finally, the financial fraud scandal will deal a heavy blow to investors and the public's confidence in the whole capital market, accounting circles, and auditing departments, which will have a huge impact on the capital market and form an unprecedented crisis of confidence. It can be seen that financial fraud is a serious disease in China's capital market, which must be prevented and managed [9, 10]. Therefore, how to build an efficient financial fraud identification model of listed companies and research on this issue is of great significance to all parties involved in the capital market.

In general, domestic financial fraud identification research mostly comes from the characteristics of financial fraud and case analysis of listed companies in China, selecting characteristics with significant correlation indicators, based on supervised learning and other methods to build models; the selection of samples pays more attention to a certain industry or a certain type of financial fraud [11, 12]. In machine learning model construction, there is room for further research. Academic research abroad on the identification of financial statement fraud is earlier than domestic, with a relatively mature research system. To the problem of financial fraud identification, the current research achievements can be divided into supervised and unsupervised learning algorithms, such as logistic regression model, the decision tree model, support vector machine (SVM) model, and naive Bayesian and neural network model [13, 14].

1.2. Background and Related Work. At home and abroad, the research on financial fraud identification of listed companies began from the research on the characteristics of financial indicators, which is also a main content of financial fraud research [15, 16]. Research on identifying fraud through financial indicators is mature and comprehensive, involving all aspects of financial indicators, including solvency, profitability, operating capacity and cash flow capacity, and other indicators. These studies are based on the assumption that the problems of accounting fraud of listed companies through the study of the fraud of financial data, so it can be unusual by the listed company's financial indicators and historical data or other normal company, comparing the data to identify the existence of financial fraud. However, with the increasing concealment of financial report fraud, the need to identify financial fraud can no longer be satisfied simply by the anomaly of financial information. More and more researchers begin to focus on nonfinancial characteristics. Many scholars have studied the characteristics of board of directors, board of supervisors, and audit committee of listed companies from the perspective of internal control, as well as external nonfinancial characteristics such as accounting firm characteristics, shareholders characteristics, transaction characteristics, and industry characteristics [17, 18].

The identification and early warning of financial fraud risk based on neural network are mainly simulated distribution processing and the establishment of calculation model. The neural network-based financial fraud risk identification and early warning implicitly decentralized storage of financial indicators. Then it uses autonomous learning to modify the internal weight, get the correct identification results, and finally can get the neural network-based financial fraud risk identification as well as the early warning automatic identification results [19, 20]. Compared with traditional methods, the identification method based on statistical learning uses the probability of occurrence of risk indicators as the basis. However, the conditional random field model requires strong domain knowledge and self-defined feature templates [21, 22]. In order to avoid such feature engineering, the recognition methods based on neural network emerged and have attracted more and more attention in academic circles. In the 1990s, some scholars have applied neural network to study risk identification, but the effect is not outstanding. In recent years, with the rapid development of computer technology and deep learning, methods based on neural networks have also achieved good results in risk identification. Some scholars try to introduce deep learning methods to risk identification problems and have made some progress. At present, most neural network-based recognition is on the basis of indicators; that is, the financial risk problem is regarded as a labeling problem of index sequence, and each index in the sequence is assigned a label, which is transformed into a classification problem. The recognition method based on neural network takes the index embedding vector as input, trains the correctly recognized data to automatically learn features and train models, and then identifies the test data with the trained model. The recognition method based on neural network avoids the traditional feature engineering and has a good effect on the problem of unregistered recognition [23, 24]. Compared with other traditional recognition methods, the overall recognition effect of bidirectional long-short term memory model has obvious advantages. Bidirectional long-short term memory (LSTM) neural network model combines the advantages of bidirectional cyclic neural network model and short-long term memory neural network model. Bidirectional cyclic neural network model solves the problem that the traditional cyclic neural network model can only input sequence information in one direction but can not use future information [25].

2. Bidirectional Long and Short Memory Neural Network Model

Neural networks can interact with objects in the real world by mimicking the biological nervous system. Neural network is composed of neuron model. As a machine learning technology that simulates human brain neural network, neural network is composed of a large number of neurons and the interactions between neurons. Generally speaking, neural network can be divided into three layers: input layer, hidden layer, and output layer. Each layer is composed of several neurons, and neurons of adjacent two layers are

connected to each other to achieve the purpose of signal transmission and processing. Figure 1 shows the M-P neuron model. The neuron receives input signals x_i from n other neurons and transmits these input signals through weighted connections. The threshold value θ of the neuron is compared with the total input value (1) received.

$$\sum_{i=1}^n w_i x_i. \quad (1)$$

The output of the generated neuron is then processed by the “activation function” f :

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right). \quad (2)$$

In practical applications, the Sigmoid function is commonly used in neural network:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}. \quad (3)$$

The characteristic of recurrent neural network (RNN) is that it can use contextual information, but the contextual information is very limited. As a result, the influence of the hidden layer on the output layer decreases. However, with the birth of LSTM network model, it overcomes this problem because of its unique gate structure. LSTM controls cell behavior through input gate, forget gate, and output gate. The operation on the gate is based on the Sigmoid network layer and the element-level dot product. The Sigmoid output values between 0 and 1 are used to indicate the degree of information passing, where 1 indicates that all information can pass, and 0 indicates that no information can pass. The process of the algorithm is detailed as follows:

- (1) Cells can forget information through the Sigmoid layer of forgetting gate, where h_{t-1} represents the output of the previous layer and x_t is the input of the current layer.

$$f_t = \sigma[W_f \cdot (h_{t-1}, x_t) + b_f]. \quad (4)$$

- (2) When storing information in cells, the information to be updated is first obtained through Sigmoid of the input gate, and then a new vector C_t is created through \tanh function. Finally, f_t was multiplied by the old cell state to realize the information to be forgotten and then added to the product of $i_t C_t$ to obtain cell renewal.

$$\begin{aligned} i_t &= \sigma[W_i \cdot (h_{t-1}, x_t) + b_i], \\ \tilde{C}_t &= \tanh[W_C \cdot (h_{t-1}, x_t) + b_C], \\ C_t &= f_t C_{t-1} + i_t \tilde{C}_t. \end{aligned} \quad (5)$$

- (3) By determining the output information through the Sigmoid layer of the output gate, the cell state is then processed using the \tanh function, and the product of the last two parts is the output value.

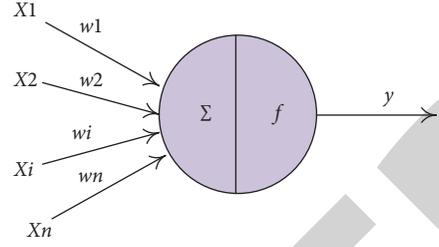


FIGURE 1: M-P neuron model.

$$\begin{aligned} O_t &= \sigma[W_O(h, x) + b_O], \\ h_t &= O_t \tanh(C_t), \end{aligned} \quad (6)$$

where O_t is the output.

When propagating back, the model shares the weight parameter matrix at the same layer, and the scaling trend of the gradient is also affected by the eigenvalues of the weight parameter matrix at each time step. If the gradient approaches 0 or infinity, gradient disappearance and gradient explosion will occur in RNN model. The problem of gradient disappearance can be alleviated by optimization algorithm (RMSprop method or Adam method). The problem of gradient explosion can be improved by sacrificing part of training efficiency and appropriately lowering the learning rate parameter.

Some people proposed a bidirectional RNN model, which solved the problem that traditional RNN could only input information one-way but could not utilize future information. The key to LSTM is how to control the long-term state gate. The gate control unit of LSTM includes three components: forgetting gate, input gate, and output gate. We first introduce the concept of a gate here. Assuming W is the weight of the gate and b is the offset term, then $g(x)$ is the Sigmoid function, and then the range of $g(x)$ is a vector of real numbers between 0 and 1. In short, a gate multiplies its output by the element vector in order to control information. When the gate output is 0, any vector after the gate processing is 0 vector, which is equivalent to prohibiting access. When the output is 1, any vector multiplied by it is still itself.

Among the three units of LSTM, forgetting gate F can control the amount of information that can be transmitted to the unit state at t moment and at $t - 1$. The input gate can control the amount of input information at time t be stored in the unit state at time: the output gate determines how much information of the unit state can be transmitted to the output of LSTM. The bidirectional LSTM neural network model is formed by combining the advantages of bidirectional RNN and LSTM; see Figure 2. Bi-LSTM neural network consists of two parts: (1) forward LSTM; (2) backward LSTM. For the choice of neural network, compared with BP neural network, CNN, and other neural network models, RNN model has the advantages of timing and directional circulation, while LSTM model solves the gradient explosion problem existing in model, so LSTM model is gradually applied to financial fraud risk analysis. The calculation formula of Bi-LSTM neural network is as follows:

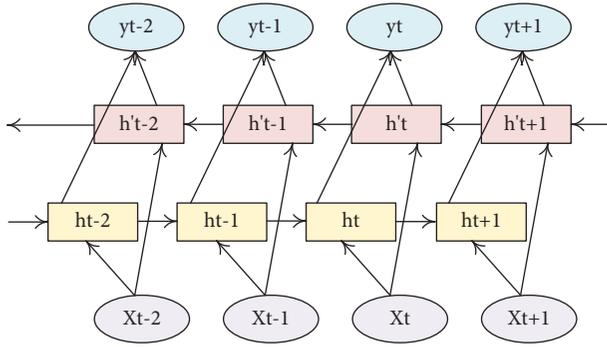


FIGURE 2: Bi-LSTM model.

$$\begin{aligned}
 h_t &= O_t \tanh[f_t \cdot C_{t-1} + i_t \tilde{C}_t], \\
 h'_t &= O'_t \tanh[f'_t \cdot C'_{t-1} + i'_t \tilde{C}'_t], \\
 y_t &= [h_t, h'_t].
 \end{aligned} \quad (7)$$

3. Fraud Identification Feature Extraction of Financial Reports

The financial index of an enterprise's annual report is a comprehensive reflection of its financial status and operating conditions. Some empirical studies have found that financial fraud will make the financial structure of enterprises abnormal, so that they are significantly different from similar companies in some financial indicators. Previous studies have also confirmed that some nonfinancial indicators (such as senior personnel structure, audit opinions, etc.) are also significantly associated with financial fraud. Based on the relevant literature of financial fraud identification at home and abroad [26], this paper selects indicators in accordance with the principles of validity, authenticity, and comprehensiveness and uses them to preliminarily build an indicator system of financial fraud identification.

Selection criteria of fraud samples: (1) the sample only includes listed companies that committed fraud from 2008 to 2018. (2) The sample only includes the listed companies with fraud in their annual report and excludes the listed companies with fraud in their interim report. (3) The sample is with their first fraud from 2008 to 2018. For listed companies with financial fraud in different years, this paper chooses the accounting year in which financial fraud occurred for the first time. The listed companies selected in this paper that committed financial fraud in 2009 did not commit financial fraud in 2008.

The financial fraud samples selected in this paper cover 19 industries: agriculture, forestry, animal husbandry, and fishery; 3 extractive industries; textile, clothing, petroleum, chemical, and plastic.

3.1. Selection of Nonfinancial Indicators. In the method of financial fraud identification, a large amount of literature based on enterprise fraud triangulation is to build an indicator system. According to the theory, financial fraud is composed of pressure, opportunity, and self-rationalization.

In today's fiercely competitive market economy, in order to pursue high profits, accounting firms omit necessary procedures to reduce costs and take poor means such as relationship building and kickbacks to compete for market share. This objectively provides an opportunity for listed companies to commit fraud. Information asymmetry is also the cause of financial fraud. As long as one party does not have sufficient knowledge of the other party, it is likely to make a wrong decision. Information demanders include investors, listed companies, government departments, and securities intermediaries. The understanding among these information demanders is not balanced, so investors may make wrong decisions. The separation of ownership and management rights of modern listed companies results in information asymmetry between investors and managers. The pressure factor is the motivation of corporate fraud, which is related to the operating conditions of enterprises. Therefore, the selection of nonfinancial indicators is mainly constructed from pressure factors and opportunity factors.

3.2. Selection of Financial Indicators. In order to reflect the company's financial situation as comprehensively, concretely, and scientifically as possible, 207 financial indicators are selected from the main financial analysis indicators table and financial notes table in Berg Financial database. The company's main financial analysis index table is selected by the three financial statements, the company's balance sheet, cash flow statement, and the company's income statement calculation. The selected financial notes include financial notes-accounts payable table and financial notes-prepaid accounts table, a total of 15. Opportunity factor refers to the opportunity to escape punishment for corporate fraud, which is mainly related to internal control of enterprises, internal punishment measures of enterprises, audit system, etc.

Financial fraud is a continuous process, so it is necessary to calculate the mean and variance of the above financial indicators in the last three years as the new features and standardize all financial indicators according to the Shenwan secondary industry category of the company and finally obtain 629 variables. Indicators with a missing rate of more than 80% were removed, leaving 454 indicators.

The data used in this article are all from Berg Financial database. Among them, the fraud samples were selected from a-share listed companies punished by CSRC for false information disclosure in their annual reports from 2008 to 2018. Due to the incompleteness of quarterly statements and semiannual statements, the financial index data are selected from the data in the annual financial statements, and the fraudulent behaviors of the same company in different years are regarded as different sample data. This paper finally selected a total of 106 samples of financial fraud, including 61 companies.

Before empirical analysis, the integration, cleaning, and feature screening of original data are essential. Since the financial indicator data of listed companies in Berg Financial database are divided into multiple tables according to different categories, the data in these tables need to be splicing

and merging according to company code, deadline date, and industry code. In data identification, companies with fraudulent behaviors were recorded as positive samples and marked as 1, while companies without fraudulent behaviors were recorded as negative samples and marked as 0.

Because the listed company annual report information is incomplete, the database form information is missing. There are a lot of missing values in the original data set obtained. In the experiment in this paper, we believe that features with a high completion miss rate make it easy to cause data distortion, so the features with a miss rate of more than 80% are removed, and the median is used to fill in the remaining features.

In order to make the features of different dimensions additive and comparable, data need to be processed in dimensionless manner. As listed companies in the same industry are relatively consistent in the scale change trend of financial indicators, which has great reference significance for the identification of fraudulent companies, each sample data is standardized according to its industry. Opportunity factor refers to the opportunity to escape punishment for corporate fraud, which is mainly related to internal control of enterprises, internal punishment measures of enterprises, audit system, etc.

4. Empirical Analysis

This section elaborates the functions of the fraud identification system and the realization process of each function module. The structural design of the whole system is closely related to the actual application requirements of the system. In view of the practical application of financial fraud identification system requirements, such as fraud identification model of the evaluation results, fraud data identification result output, etc., the financial fraud identification system adopting the operation is convenient, intuitive, simple, and more humanized client/server structure model, in order to shorten the system response time.

The working mechanism of client/server structure is as follows: the client is responsible for providing the system working interface, sending the data that needs to be trained and identified to the application server through the network, receiving the identification result of the application server and outputting it through the terminal device. The application server runs and provides the computed results and data to the client. Therefore, most of the core algorithms of the financial fraud identification system run on the server side. This structure mode can give full play to the processing capacity of the client PC, and a lot of work can be processed by the client and then submitted to the server. It is characterized by strong interactivity, safe access mode, low network traffic, fast response speed, and large amount of data processing. Self-rationalization factor is closely related to the moral concept and code of conduct of the perpetrators of corporate fraud, which is difficult to quantify here.

After the completion of the model construction, the following attempts to build an intelligent financial fraud identification and monitoring system to improve the automatic monitoring ability. The overall logical architecture

of the fusion system shows that the system obtains the basic data required by the model through a unified data bus. The basic data can be divided into two types: one is the standard financial data of listed companies. The other is the alternative data related to financial fraud of listed companies, whose main characteristics are mainly text data, which need to be further processed and mined. In order to ensure the system identification effect and adaptability to new financial fraud means, the basic data processing link of the system will be adjusted regularly. The analysis engine includes the supervised learning financial index fraud mining model and outlier detection model. Supervised learning financial index fraud mining model is based on the sample data of listed companies that have experienced financial fraud and the sample data of other normal listed companies, using supervised machine learning modeling classification and analysis prediction. The outlier detection model is based on the classification and statistics of all the sample data of listed companies and uses the method of outlier detection to mine the financial anomalies of listed companies.

The financial fraud risk comprehensive scoring framework of the financial fraud fusion system is as follows: the financial fraud risk value of listed companies is obtained by the fusion of "supervised learning score" and "outlier detection model score." "Outlier detection model score" is divided into "single index anomaly detection score," "multidimensional anomaly detection score," and "multi-index consistency detection score," and the score of each subitem integrates the results of different constant detection models. In the "score of supervised learning model," the score is evaluated based on the supervised learning model obtained from the above analysis, and the XGBoost model is used to score after sample imbalance treatment. The higher the risk of financial fraud, the higher the score of listed companies.

The training data set and validation data set are, respectively, 25% of the total data sample, and the test data set is 50% of the total data sample, and the data set is discretized. Experimental comparison methods include bidirectional long-short term memory model method, a classical static combination method, and dynamic combination method. Specifically, $C4.5$, S_D_Tree , $A(C4.5)$, $A(S_D_Tree)$, proposed method, and $D(S_D_Tree)$ are studied in the following. The accuracy rates of nonfraud and fraud prediction of these six methods are compared, respectively. The comparison of experimental results is shown in Figure 3, which also can be seen in Figure 4. Meanwhile, methods D and S_D are compared in terms of time cost. The experimental results are shown in Figure 5.

It can be seen from Figure 4 that the classification accuracy of the method based on bidirectional long-short term memory model and that of method $C4.5$ are very similar. The classification accuracy of method AdaBoost ($C4.5$) and method AdaBoost (S_D_Tree) is higher than that of method $C4.5$ and method $D(S_D_Tree)$. Methods The classification accuracy of $A(C4.5)$ and $D(S_D_Tree)$ was higher than that of $A(S_D_Tree)$, $C4.5$, and AdaBoost ($C4.5$ and S_D_Tree).

As can be seen from Figure 5, the running time of the method based on bidirectional long and short memory

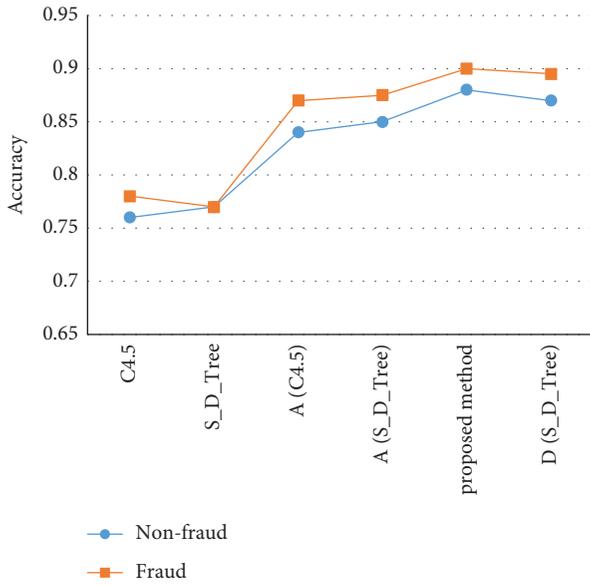


FIGURE 3: The accuracy of fraud and nonfraud condition.

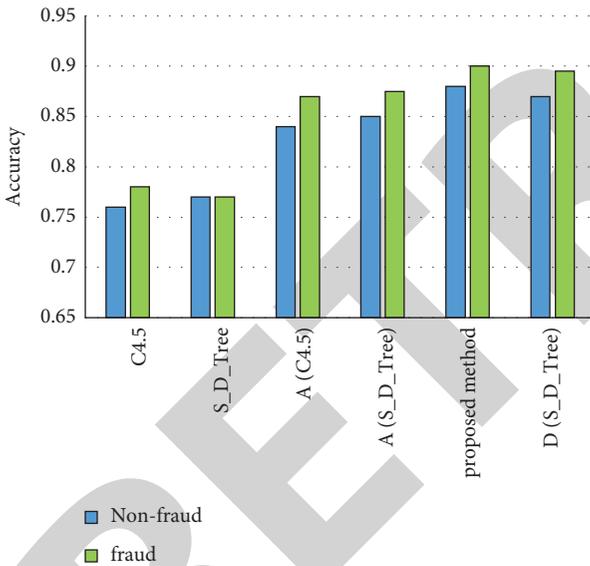


FIGURE 4: Column of the accuracy comparison.

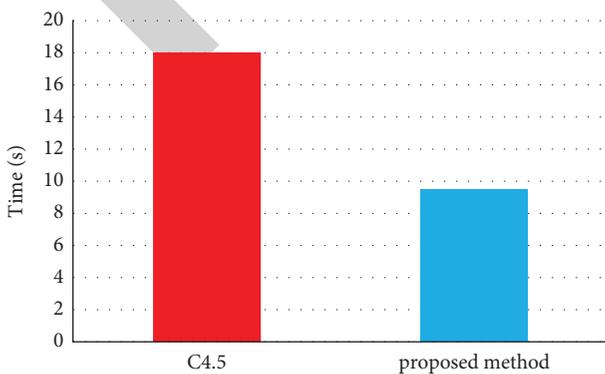


FIGURE 5: Time consumed of both C4.5 and proposed method.

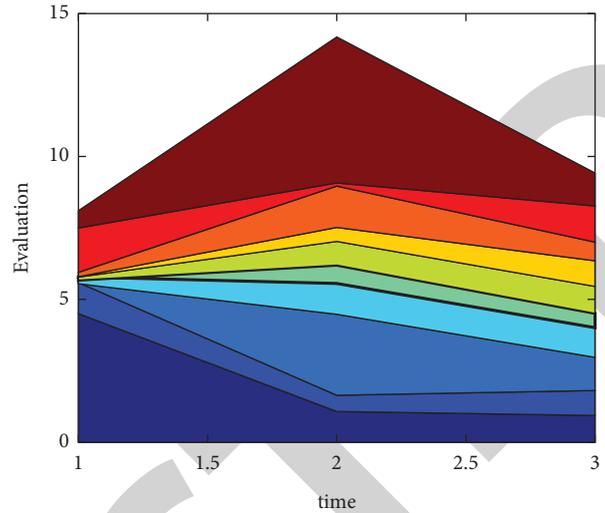


FIGURE 6: The evaluation by the proposed model.

model is significantly less than that of DCC-CD (C4.5). This is because the time complexity of the method based on bidirectional long and short memory model is lower than that of algorithm C4.5 when selecting test attributes.

In addition, it can be seen from Figure 6 that manufacturing and nonmanufacturing using own training model to rate their misclassification rate are the lowest. Besides, financial fraud identification model has industry adaptability, using the model of manufacturing and nonmanufacturing model for manufacturing and nonmanufacturing sample grade crossing. In addition, its classification error rate is significantly higher than their own training model of forecast error rates. Therefore, the model recognition has the characteristic of industry. The industry nature of model identification may be caused by the difference in production and operation between listed manufacturing companies and nonlisted manufacturing companies, which is inevitably reflected in the difference in financial indicators.

In judging the quality of a model, besides the total misclassification rate, misrecognition rate and rejection rate are also of importance. In real life, the cost of identifying a fraudulent company as a nonfraudulent company is far greater than the cost of identifying a nonfraudulent company as a fraudulent company. Therefore, in the case of a certain classification error rate, the smaller the rejection rate, the better. Figure 7 shows that the identification error rate of the neural network model established for the manufacturing industry is 28.82%, in which the rejection rate is 26.47%, much higher than the error rate 0. The false judgment rate of the model for fraudulent companies is 11.3 times that for nonfraudulent companies. In this paper, the paired samples are selected in accordance with the ratio of 1:1, and the fraudulent samples and nonfraudulent samples in all industries account for half and half, so the probability of a fraudulent company being misjudged as a nonfraudulent company is 2. For example, the probability of fraudulent samples being misjudged as nonfraudulent samples in the manufacturing industry is $2 * 26.47\% = 52.94\%$, which

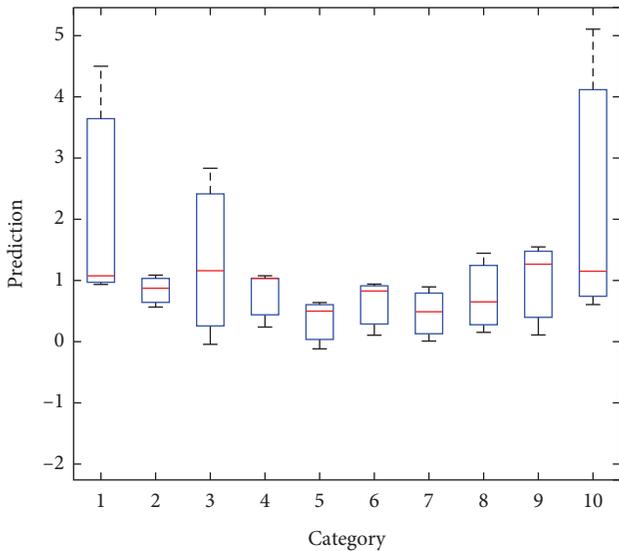


FIGURE 7: Comparison of industry.

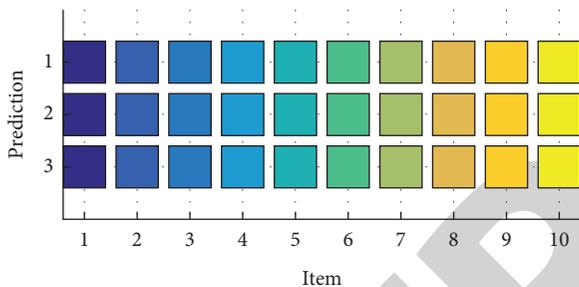


FIGURE 8: Prediction.

means that more than half of fraudulent companies have not been identified. Such a high proportion of misjudgment brings very serious losses to the users of the company's financial statements in real life. However, compared with the nonmanufacturing industry, the number of unrecognized fraudulent companies only accounted for $2 * 7.94\% = 15.88\%$. The prediction is also shown in Figure 8.

5. Conclusion

This paper constructs a financial fraud identification model, which mainly includes two key parts:

- (1) Feature selection is based on bidirectional long and short memory neural network algorithm and dynamic combination classification is based on clustering division. The validity of the fraud identification model is verified by using corporate financial data.
- (2) At the same time, the object-oriented design method is adopted to design and implement the financial fraud identification system.

However, this paper only selects some samples of the financial indicators of listed companies and it is not

complete. The insufficiency of the data in the index database of listed companies in China is not released by the denominator or the indicators. All these missing indices in this paper with the average instead may have influence on the result of the reality.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] H. Wu, Y. Chang, J. Li, and X. Zhu, "Financial fraud risk analysis based on audit information knowledge graph," *Procedia Computer Science*, vol. 199, no. 1, pp. 780–787, 2022.
- [2] Wu Xiuguo and S. Shengyong, "An analysis on financial statement fraud detection for Chinese listed companies using deep learning," *IEEE Access*, vol. 10, pp. 22516–22532, 2022.
- [3] K. Khan, J. Narumoto, and S. Watanabe, "Who is next? A study on victims of financial fraud in Japan," *Frontiers in Psychology*, vol. 12, no. 1, p. 649565, 2021.
- [4] T. Xiong, Z. Ma, Z. Li, and J. Dai, "The analysis of influence mechanism for internet financial fraud identification and user behavior based on machine learning approaches," *International Journal of System Assurance Engineering and Management*, vol. 1, no. 2, pp. 1–12, 2021.
- [5] B. Ma, P. A. Zhong, Z. Yu, J. Zhang, and F. Zhu, "Bargaining model of synergistic revenue allocation for the joint operations of a multi-stakeholder cascade reservoir system," *Water Resources Management*, vol. 32, no. 14, p. 4642, 2018.
- [6] H. Solomon and S. Soltes, "Is "not guilty" the same as "innocent"? Evidence from SEC financial fraud investigations," *Journal of Empirical Legal Studies*, vol. 18, no. 2, pp. 287–327, 2021.
- [7] Al-H. Khaled Gubran and M. Pritheega, "Financial fraud detection applying data mining techniques: a comprehensive review from 2009 to 2019," *Computer Science Review*, vol. 40, pp. 1–9, 2021.
- [8] L. Xu, "The financial fraud of listed companies —a case study of lukin coffee," *Journal of Social Science and Humanities*, vol. 3, no. 3, pp. 23–35, 2021.
- [9] D. Albareda, L. Goodman, and J. Goodman, "Contestation in multi-stakeholder initiatives: enhancing the democratic quality of transnational governance," *Business Ethics Quarterly*, vol. 30, no. 2, pp. 169–199, 2020.
- [10] H. Xia and H. Ma, "A novel structure-based feature extraction approach for financial fraud detection," *Journal of Physics: Conference Series*, vol. 1865, no. 4, pp. 1–9, 2021.
- [11] M. Wilson, A. Van Citters, I. Khayal et al., "Designing an electronic point-of-care dashboard to support serious illness clinical visits: a multi-stakeholder coproduction project (TH341B)," *Journal of Pain and Symptom Management*, vol. 59, no. 2, pp. 430–431, 2020.
- [12] L. Borge and S. Bröring, "What affects technology transfer in emerging knowledge areas? A multi-stakeholder concept mapping study in the bioeconomy," *The Journal of Technology Transfer*, vol. 45, no. 1, pp. 430–460, 2020.

- [13] H. Patel, S. Parikh, A. Patel, and A. Parikh, "An application of ensemble random forest classifier for detecting financial statement manipulation of Indian listed companies," *Advances in Intelligent Systems and Computing*, vol. 740, pp. 349–360, 2019.
- [14] M. Du, "Corporate governance: five-factor theory-based financial fraud identification," *Journal of Chinese Governance*, vol. 6, no. 1, pp. 1–19, 2021.
- [15] M. King, "Financial fraud investigative interviewing - corporate investigators' beliefs and practices: a qualitative inquiry," *Journal of Financial Crime*, vol. 28, no. 2, pp. 345–358, 2020.
- [16] A. Albattah, A. Rassam, and A. Rassam Murad, "A financial fraud detection model based on LSTM deep learning technique," *Journal of Applied Security Research*, vol. 15, no. 4, pp. 498–516, 2020.
- [17] J. J. Nzau, B. M. Denemadjbe, E. F. Dumas, and M. A. Rodriguez, "Catalysing change for reproductive health in Chad through a multi-stakeholder coalition," *Sexual and Reproductive Health Matters*, vol. 27, no. 1, Article ID 1626185, 2019.
- [18] V. Moseley, M. Vetter, and V. Thorsten, "Multistakeholder advice at the European medicines agency: is it still needed?" *Clinical Pharmacology & Therapeutics*, vol. 105, no. 4, pp. 819–821, 2019.
- [19] C. Gong, X. Wang, and Y. He, "Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network," *Energy*, vol. 232, pp. 1–9, 2021.
- [20] R. Saia and S. Carta, "Evaluating the benefits of using proactive transformed-domain-based techniques in fraud detection tasks," *Future Generation Computer Systems*, vol. 93, pp. 18–32, 2019.
- [21] V. Balandina, A. Chernyshov, and A. Chernyshov, "Application of long-short memory neural networks in semantic search engines development," *Procedia Computer Science*, vol. 169, no. C, pp. 388–392, 2020.
- [22] L. Leyns, M. De, and W. Willems, "Using concept mapping to identify policy options and interventions towards people-centred health care services: a multi stakeholders perspective," *International Journal for Equity in Health*, vol. 17, no. 1, p. 177, 2018.
- [23] N. H. Chan and S. Ling, "Correction: residual empirical processes for long and short memory time series," *The Annals of Statistics*, vol. 38, no. 6, p. 3839, 2020.
- [24] A. Corsaro, L. G. Sisti, P. Parente et al., "Evolution of a multi-stakeholder initiative of health promotion in 20 suburban Roman districts," *European Journal of Public Health*, vol. 28, no. 4, pp. 1–9, 2018.
- [25] T. Pourhabibi, K.-L. Ong, B. H. Kam, and Y. L. Boo, "Fraud detection: a systematic literature review of graph-based anomaly detection approaches," *Decision Support Systems*, vol. 133, Article ID 113303, 2020.
- [26] Z. Huangfu, H. Hu, N. Xie, Yu-Qi Zhu, H. Chen, and Y. Wang, "The heterogeneous influence of economic growth on environmental pollution: evidence from municipal data of China," *Petroleum Science*, vol. 17, no. 4, pp. 1180–1193, 2020.