

## Retraction

# Retracted: A Hybrid Intelligent System for Company Financial Risk Detection Based on Tree-Based Model and Deep Neural Network

### Wireless Communications and Mobile Computing

Received 18 July 2023; Accepted 18 July 2023; Published 19 July 2023

Copyright © 2023 Wireless Communications and Mobile Computing. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] J. Wei, "A Hybrid Intelligent System for Company Financial Risk Detection Based on Tree-Based Model and Deep Neural Network," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 5630319, 10 pages, 2022.

## Research Article

# A Hybrid Intelligent System for Company Financial Risk Detection Based on Tree-Based Model and Deep Neural Network

Jingya Wei 

*The Third Hospital of Wuhan (Tongren Hospital Affiliated to Wuhan University), Wuhan 430000, China*

Correspondence should be addressed to Jingya Wei; [jywei@st.tbtcu.edu.cn](mailto:jywei@st.tbtcu.edu.cn)

Received 3 February 2022; Accepted 31 March 2022; Published 26 May 2022

Academic Editor: Xin Ning

Copyright © 2022 Jingya Wei. 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.

At present, the process of security issuance in China has changed from examination and approval system to approval system. With the increasing stock that enters the capital market through the IPO, establishing an effective financial risk monitoring and control system of listed companies has great practical significance and application value. In the perspective of big data, this paper constructs a new practical financial early warning integration model. A total of 32,283 financial statements of 3,025 listed companies are collected from 2000 to 2017. First, this paper collected and built a financial crisis prediction database: a total of 32,283 financial statements of 3,025 listed companies occurred from year 2000 to 2017 and corresponding financial indicators including 6 primary indicators and 25 secondary indicators are collected. Next, the financial crisis model based on financial indicators is proposed and trained by the classic machine learning method such as random forest and gradient boosting decision tree. And then, through natural language processing and deep learning technology, this paper also builds a financial crisis prediction model based on financial statement text. Finally, an ensemble model is proposed and its performances are evaluated; the results indicate that the proposed ensemble model performed the best and significantly outperforms other benchmark methods for financial crisis prediction and identification, indicating that it can be employed as an intelligent identification system to enhance identification accuracy for early warning and identification of financial crisis of listed companies in China stock marketing.

## 1. Introduction

With the quick improvement of China's financial market as of late, an ever-increasing number of corporate leaders decide to raise assets through IPO to sell corporate protections and utilize a bigger measure of funding to extend the size of endeavors and do advancement, innovative work, venture, and different practices. The financial status of endeavors is impacted by numerous viewpoints and afterward straightforwardly influences the interests of financial backers. During the time spent undertaking activity, regardless of from outer market impact, strategy impact, or from inner factors like endeavor speculation, financing choices and different variables. Any issues in any connection might carry gigantic financial dangers to ventures, making the stock value fall ceaselessly temporarily, monstrous loss of market esteem, and resource misfortune, and even lead to long haul misfortunes of the organization, incapable to

escape the financial emergency, and afterward unique treatment, or constrained delisting, and so on. Along these lines, in the event that a framework can be worked to give early notice to the financial dangers of recorded organizations, from one perspective, it can help financial backers or speculation establishments to track down the emergency of recorded organizations at the earliest opportunity and keep away from huge venture chances. Then again, it can give early admonition to recorded organizations, to track down the venture choice issues behind financial dangers and forestall financial dangers on a very basic level. Accordingly, the investigation of financial danger early admonition framework is of incredible importance.

The exploration objective of this paper is to join state-of-the-art AI calculations, regular language handling innovation, and profound learning hypothesis with financial market, investigate the use of AI and large information apparatuses in the financial field, assemble a new multi-

input financial danger recognizable proof framework dependent on budget summaries and printed information, assess the grouping proficiency of the framework, and investigate its worth in reasonable application.

As for the existence of financial risks in listed companies, we take listed companies' ST standards to measure. ST stands for two consecutive years of losses in the operation of listed companies, special treatment. The company being ST was obviously already at significant financial risk, so we take the ST of the listed companies as the predictive target of the model. The prediction of the model is based on the analysis of financial statements and annual reports of listed companies. As the financial statements and annual reports of listed companies are mostly released in the next year, it can be confirmed whether the next year will be ST. Therefore, we limit the prediction interval to two years; that is, the annual report and financial statements of  $T$  years are used to predict whether the company will become ST in  $T+2$  years. Build the training data as a rule.

As for the financial statement data, we downloaded the income statement, cash flow statement, and balance sheet data of 3747 companies listed in mainland China from CSMAR database. We construct 29 second-level indicators and use Python to build decision tree, random forest, and XGBoost model based on the above financial indicators. K-fold cross-validation was used to obtain the final model score.

For text data, this paper mainly uses the text of annual reports of listed companies. We first used crawlers to capture all available annual reports of 15,324 listed companies on the Internet. Python expressions and other tools were used to clean and parse the texts, each annual report corresponds to the corresponding listed company and year, and all chapters of the annual report are analyzed to form a multilayer dictionary structure, and the training set is constructed by pairing with the prediction targets. After that, the text is embedded with pretrained word vector and the numerical form real-value matrix of each text is obtained. Then, the classification model based on deep neural network is constructed.

In the end, we integrate the two models, and based on our experience in the field of machine learning, the algorithm for model integration should not be too complex; therefore, this paper carries out integration by averaging the output results of the two models and calculates various indicators to evaluate the effect of the final model classification and finally analyze the results.

## 2. Literature Review

Capital market started earlier, as early as in the 1930s began to the company financial risk early warning research. Following quite a while of exploration has shaped a moderately full grown hypothetical system and can set up an assortment of compelling expectation models. As a general rule, the examination interaction of financial danger of recorded organizations is from the investigation of few examples to the examination of countless organizations in the full example. The investigation strategies embraced from the early

proportion of financial pointers to assemble auxiliary markers, to the achievement of bringing numerical insights into financial danger examination, fabricating a five-factor Z-score model, and widely applying different AI models in the later stage. With the consistent advancement of PC innovation, measurements, AI, and different fields, the methods for financial investigation are additionally continually enhanced and developed.

In the beginning phases of financial danger research, analysts have taken few firms as tests and concentrated on the contrasts between bankrupt firms and different firms by straightforwardly looking at different markers and the proportions between them. As the examination in this stage is not helped by computers, the ability to analyze samples is greatly limited. The most representative one is Fitzpatrick's univariate determination model in 1932 [1], which analyzes the difference between bankrupt group and nonbankrupt group from the perspective of single financial ratio. It selected 19 enterprises as the research samples and found the ratio of equity to debt of shareholders and the ratio of net profit to shareholder equity; these two indicators can effectively judge whether there is financial risk of listed companies.

During the 1970s, the ceaseless advancement of PC innovation and factual strategies gave novel thoughts and instruments to the investigation of monetary danger. In 1986, Beaver utilized univariate investigation to dissect a bigger number of typical organizations and organizations with monetary danger [2]. Through similar examination, it is observed that the proportion of income to add up to liabilities can adequately caution the organization of monetary emergency. Another achievement was the Z-score model with five factors proposed by Altman [3]. It selected 19 enterprises as the research samples and found the ratio of equity to debt of shareholders and the ratio of net profit to shareholder equity; these two indicators can effectively judge whether there is financial risk of listed companies. In the United States, Brazil, Canada, and other countries, these have been widely used.

Subsequently, in the 1980s, there began to be precedents for introducing machine learning algorithms into financial risk early warning research, which mainly introduced the early popular ANN, SVM, and other models. Because of the innate benefits of AI calculation, the model can productively lead factual examination on a lot of information.

First of all, logistic was used to the financial risk early warning model. Ohlson proposed a logistic advance admonition model and set up the liquidation likelihood appropriation model with in excess of 2,000 bankrupt and nonbankrupt organizations as exploration tests, which was a tremendous jump contrasted with the 19 endeavors concentrated in Fitzpatrick time [4]. He looks at the two contingent likelihood esteems to give an advance admonition of the organization's monetary dangers. Through the exploration, it is found that the forecast capacity of the organization's resources, resource responsibility proportion, productivity, and financing limit is awesome. Because of the low necessity of logistic model on information and the presentation of logit work, it very well may be skilled for arrangement undertakings and produce

strong outcomes, and it is generally utilized and slowly turns into the standard strategy in the field of monetary danger notice research.

Then, at that point, came models dependent on counterfeited neural organizations, which copy the design of neurons in the human cerebrum and have a great deal of boundaries that can demonstrate a ton of information. In 1990, Odom and Sharda creatively presented fake neural organization and built a monetary danger cautioning model dependent on fake neural organization [5]. They took the monetary information of 65 bankrupt undertakings and the matched 65 ordinary endeavors as the preparation set and involved the five trademark factors in the *Z*-score model as the contribution of the neural organization. The expectation exactness of the example came to 80%.

Nonetheless, from the finish of the twentieth century to the start of the 21st century, because the solving algorithm of neural network has not yet appeared, people find that support vector machine can accomplish better execution in an assortment of arrangement undertakings on an enormous number of issues. Support vector machine enjoys an incredible benefit in this period due to its natural development thought and exceptional system of portion capacity to manage nonlinear limit order issue; it is used to study the early warning of bank bankruptcy and confirmed the accuracy of SVM in the advance warning of bank bankruptcy by comparing with other methods [6, 7].

From the perspective of the foreign financial risk early warning research process, the research started from the analysis of a small number of samples and gradually moved to the analysis of a large number of samples. The analysis method is also from the most elementary to calculate the ratio of the original financial indicators to construct new financial indicators for analysis and gradually combined with statistical methods and machine learning methods, from comparative analysis to logistic regression and to artificial neural networks and support vector machines. With the model structure becomes more and more complex and the parameters increase, the research on this problem is also deepened, and better model classification effect can be obtained. Therefore, we have reason to believe that machine learning algorithm has great potential in the field of financial risk warning research.

### 3. Methodology

*3.1. Forecasting Model Based on Financial Statement.* In the examination of fiscal reports of recorded organizations, various pointers have diverse significance. During the time spent breaking down explanations, monetary investigators will likewise examine different information in a specific request, continually change their judgment in this cycle, and afterward get the last end. This investigation technique is like choice tree model in rationale, so we choose to utilize choice tree model to dissect report information.

Decision tree (DT) is an essential AI calculation, which is predominantly used to tackle grouping and relapse issues; it is used in monetary field generally [8].

Random forest (RF) calculation is an execution of troupe learning in choice tree. Group learning is a calculation that joins different students.

Gradient boosting decision tree (boosting decision tree), proposed by Friedman, is similar to random forest algorithm and builds strong learners by integrating weak learners; it is also used in financial market commonly [9].

Figure 1 shows the simplified process of random forest algorithm based on decision tree.

*3.2. Prediction Model with Financial Statement Text.* In terms of financial statement text, this paper decides to try a model based on DNN. The fundamental reason for this piece of the model is to discover a few qualities or explicit examples through the examination and mining of the yearly reports of recorded organizations with monetary dangers and ordinary organizations. Contrasted and the immediate counting and insights of words, DNN can partially understand the semantic meaning of the text, considering the word order relationship of these words and some permutation and combination phenomena on the basis of the occurrence frequency of words. The center of this piece of the model is the profound neural organization model addressed by convolutional neural organization and cyclic neural organization. These models will be momentarily presented in this segment. What is more, during the time spent text numeralization, the one-hot encoding of words will prompt meager grid comparing to the text, which is not helpful for model preparing; hence, common methods in natural language processing are adopted here, and the word vector based on mass text pretraining is used to complete the text embedding in the first step. As there are few companies with financial risks in the training process, it is easy to determine all companies as normal companies when the model is optimized directly. To solve this problem, focal loss is adopted in our experiment to replace the default loss function to optimize the network [10]. The models and techniques used above are described next.

*3.2.1. The Word Vector.* In the process of natural language processing, the first problem we need to face is how to numeralize the text and then process the text by computer. Probably, the least complex thought is to fabricate a word reference so that each word has an extraordinary number that compares to it. Then, at that point, you can construct a vector that has similar length as the word reference and address the word by setting the number to 1 and the rest to 0. But there are many problems with this approach. First of all, assuming another term shows up, the length of the vector portrayal of the first term should be changed, which is not helpful for the support and extension of the model. Second, such vector representations completely discard information about the words themselves. Profits and profit margins, for example, have an inner vector product of zero, whereas in practice, they are highly correlated. To tackle the above issues, Mikolov et al. proposed the word vector model in 2013 [11]. Unlike traditional dictionary method to construct the numerical form of words, vector can train a family of fixed dimension real-value vector. Thus, fixed-length

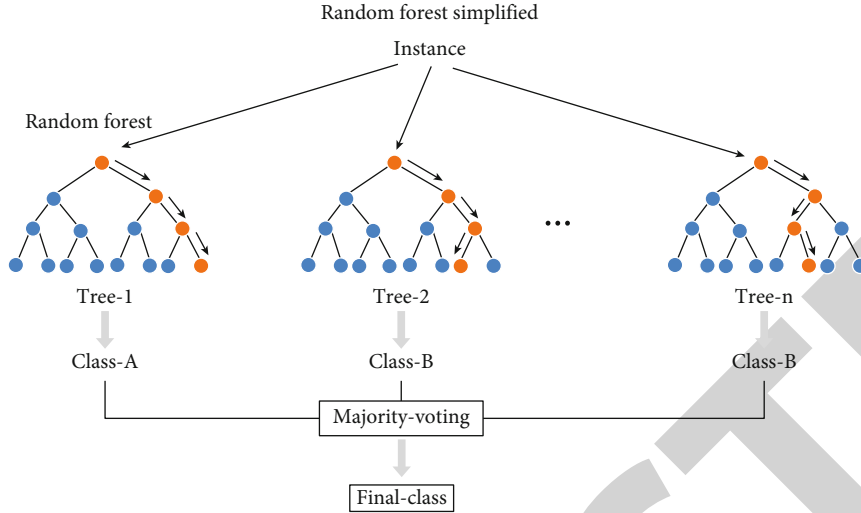


FIGURE 1: The simplified process of random forest algorithm.

vectors can be obtained, and similitude between vectors can be determined, mostly safeguarding the data of the actual word.

Word vector training mostly depends on the NN language method. It can be trained in an unsupervised way by using a large amount of text. Essence can be understood as the analysis of the cooccurrence relation of words; that is, words in the same semantic environment should also have the substitutability. The development cycle of word vector is to foresee the current words dependent on the setting in the text, or vice versa to forecast the possible words based on the current words in the context to achieve the above analysis process. Here, we use the extensive CBOW model as an example to explain.

The full name of CBOW model is continuous word bag model [12], which predicts the possibility of the current word in a given context. For example, if we take the word within 5 distances from the word as the context, the probability formula of the center word can be obtained as follows:

$$P(w_t | w_{t-k}, w_{t-(k-1)}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+k}) = P(W_t | \text{context}). \quad (1)$$

CBOW model develops a solitary layer neural organization to finish this forecast task. The first assertions are organized by the above window rules to build the preparation set. The info is ten words, and the result is a solitary word situated in the middle. After completing the training process, the required word vectors can be obtained. One of the advantages of this approach is the efficient use of large volumes of unmarked corpus. As we all know, a lot of supervised learning requires a lot of manpower to label data. Unsupervised learning is adopted here, which eliminates the process of labeling, so that large-scale corpus can be utilized, such as a large number of announcements, news, and reports in the financial field. In this way, a large number of data can be used effectively, and the parameters in the model can be greatly reduced, the difficulty of training can be

reduced, and the generalization performance can be improved. During our training, we found that the word vector contained more than 90 million parameters, while the parameter synthesis of other parts of the model was only in the order of hundreds of thousands. Therefore, separating the training process of these two parts could greatly alleviate the phenomenon of overfitting of the model.

**3.2.2. Recurrent Neural Network.** The course of monetary text demonstrating is really a course of monetary text jargon arrangement displaying. Text data, unlike other panel data set, tends to be strongly contextual. To handle these kinds of problem in the field of deep learning, recurrent neural network is widely used for analysis [13].

**3.2.3. Convolutional Neural Network.** In the past, many kinds of recurrent neural network architectures have been widely used in normal language handling and accomplished great outcomes. With the continuous advancement of profound learning hypothesis and innovation, individuals step by step find that convolutional neural organization additionally has a solid capacity in the field of text handling [14, 15]. Therefore, in the analysis of financial texts, we will also try to build a model based on CNN.

## 4. Preparation for the Experiment

**4.1. Predictive Variable Processing.** There are numerous ways of characterizing the monetary danger of recorded organizations. In order to obtain uniform rules to complete the marking of all listed companies, we adopt the company to become ST or \*ST as the norm of organizations with monetary danger. Since the budget summaries and yearly reports of recorded organizations in t year will be declared toward the start of t+1, it implies that the organization will know whether it becomes ST or \*ST in t+1. In this manner, the info information and result information of the forecast model are isolated by two years, and the budget summaries and yearly reports of the recorded organization in t+2 are

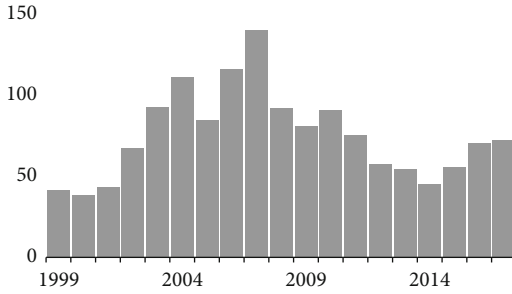


FIGURE 2: Statistical information of ST company.

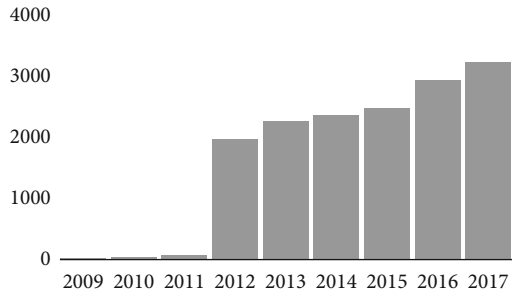


FIGURE 3: Statistical information of annual report data.

utilized to foresee whether the organization will become ST in  $t+2$ . Subsequently, the handling of expectation factors is finished. Thus, the process of prediction indicators is collected successfully. Figure 2 shows the statistical information of ST company.

In Figure 2, it tends to be seen that the quantity of recorded organizations with monetary dangers expanded essentially during the monetary emergency around 2007; there were 115 in 2006 and 139 in 2007. In subsequent years, the data gradually showed a steady decline. From 2012 to 2015, the number of companies in financial crisis is less, and it shows good economic environment during this period. In 2017, a total of 72 listed companies were treated by ST or +ST.

**4.2. Financial Statement Text Processing.** Figure 3 shows the statistical information of financial statements data, including 15,838 financial statements of 3465 corporations.

The financial statements collected by us gradually increased from 2012 to a total of 3243 in 2017. Since the principle text of the yearly report does not straightforwardly show the organization and year, we broke down the title of the yearly report and compare every yearly report to the relating organization and year. From that point forward, we dissected each segment of the yearly report through text handling and separated the title of each segment. Since the configuration and title of a few yearly reports are marginally not the same as the standard circumstance, ordinary articulation and different methods are utilized here, and a few sections that cannot be parsed with less recurrence are disposed of. At long last, the accompanying parts in the yearly report text are joined into the model: essential profile of the organization, corporate administration, significant things, organi-

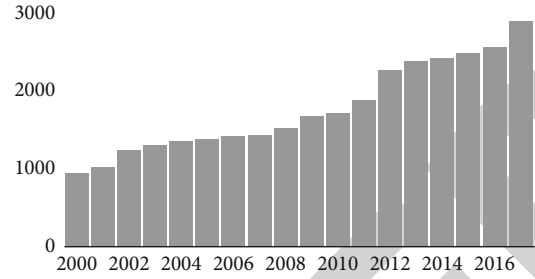


FIGURE 4: Statistical information of financial statement data.

zation profile and the vitally monetary markers, working circumstance conversation and examination, board report, and the executive conversation and investigation. According to the general text processing process, we first perform word segmentation for all texts.

We use Jieba to complete the task. For the text message, the total length is limited to 200 words. For the text exceeding the length, due to the language habit of Chinese, the more important text will be placed at the later position of the sentence. We truncate them from back to front. For paragraphs with less than 200 words in length, we repeatedly fill them and then truncate them backwards to ensure that each listed company has an annual report every year, which contains nine paragraphs with a total length of 200 words for each paragraph. After that, we use the pretrained 256-dimensional word vector to embed it, so that each paragraph with 200 words in length will be turned into a  $200 \times 256$  real-value matrix, which will serve as the input of various subsequent deep neural networks. After that, we can pair the annual report data of each company with the data of the company's  $T+2$  years by ST to form our training data.

**4.3. Financial Index Construction.** As for the financial indicators of listed companies, we consistently download the budget summaries of all recorded organizations in mainland China after 2000 from the CSMAR information base. Figure 4 shows the details.

Since the fiscal report information comes from the data set, the information can be destined to be moderately finished. From 2000 to 2017, the greater part of the budget summaries of recorded organizations has been effectively obtained, including an aggregate of 32,283 fiscal reports of 3,025 recorded organizations. Based on perusing countless writing [16–19], this paper constructed six first-level indicators according to the above financial statement data, namely, solvency index, profitability index, operating ability index, development ability index, cash flow index, and risk level index. Under this, we figured out 25 auxiliary markers as indicated by the current writing on monetary danger and monetary extortion. Our AI model will learn dependent on these second-level pointers and create new markers through the screening and further mix of these pointers through the model. These markers will be presented independently beneath.

**4.3.1. Indicators of Solvency.** Dissolvability is a broadly involved file in monetary examination, which mirrors the

TABLE 1: Introduction of listed companies' solvency indicators.

The index name	Short for index	Design formulas
Asset-liability ratio	DAR	$\frac{\text{Total liabilities}}{\text{Total assets}}$
Cash flow debt ratio	CCR	$\frac{\text{Annual net operating cash flow}}{\text{Current liabilities at year-end}}$
Quick ratio	ATR	Quick ratio = $\frac{\text{quick assets}}{\text{current liabilities}}$ Quick assets = current assets – inventory

TABLE 2: Profitability indicators of listed companies.

The index name	Short for index	Design formulas
Whether to report small book profit	SP	$SP = 1$ $\text{ROE} = \frac{\text{after-tax profits}}{\text{owner's equity (between 0\% and 1.5\%)}}$
Profit margin	GPM	$\frac{\text{Main business income} * \text{main business cost}}{\text{Main business income}} * 100\%$
Return on total assets	ROA	$\frac{\text{Net profit}}{\text{Average total assets}} * 100\%$ Average total assets = $\frac{\text{total assets at beginning of year} + \text{year-end total assets}}{2}$

capacity of endeavors to reimburse obligations. Obviously, these days, endeavors by and large take on the method of obligation activity to get more accessible assets for big business the executives and activity. Yet, the obligation activity additionally brings hazards; in the event that the undertaking would not control its obligation be able to even out sensibly, it will prompt the inability to reimburse the obligation and different circumstances, which may straightforwardly prompt the insolvency of the venture. In this manner, dissolvability is a significant proportion of monetary danger. In this paper, indexes in Table 1 are used to measure the solvency of listed companies.

4.3.2. *Profitability Index.* Productivity list can mirror the capacity of endeavors to acquire benefits through creation and activity, and a venture with solid productivity can more readily oppose monetary dangers. Benefit is the motivation behind big business activity, so it is additionally a significant record to investigate the monetary status of ventures. To measure the profitability of enterprises, we adopt the following indicators in Table 2.

4.3.3. *Operating Capacity Indicators.* Working limit mirrors the working productivity of a venture and furthermore mirrors the proficiency of creating gains by utilizing different resources from the side, just as the capacity to manage stowed away monetary dangers. The development of monetary emergency is brought about by the helpless administration of ventures. Table 3 employs some classic indicators in this paper.

4.3.4. *Development Capability Indicators.* The improvement development capacity list is primarily used to gauge the

capacity of endeavors to persistently extend the size of activity. An organization with a superior pattern will normally be better ready to endure monetary dangers. The development pace of fundamental business is determined by the following equation: (current year's deals – last year's business)/last year's absolute deals. This list can quantify the improvement capability of big business items and afterward mirror the advancement limit of undertakings.

4.3.5. *Cash Flow Indicators.* The cash flow indicators are introduced in Table 4 for details.

4.3.6. *Risk-Level Indicators.* In order to measure the risk level of an enterprise, we adopt the financial leverage index of the company, calculated by the following formula: EBIT/(EBIT – I), where EBIT is Ebit before changes, where I is interest. The level of financial leverage directly affects the company's ability to withstand external risks.

## 5. Empirical Results and Analysis

In this part of the model, performance indicators are generated by comparing the anticipated consequences of tests and the real outcomes. Among them, the most fundamental model precision addresses the extent of a similar tag in the forecast outcome as the genuine circumstance, which mirrors the general order aftereffect of the model. The exactness rate and review rate focus harder on the grouping capacity of the model for the issue tests to be identified.

Finally, F1 index integrated accuracy rate and recall rate and evaluated the model from both aspects. Next, these indicators are introduced, respectively.

TABLE 3: Operating capacity indicators of listed companies.

The index name	Short for index	Design formulas
Whether to turn losses into profits	LP	Net profit was less than 0 in the previous year and more than 0 in the current year
Whether to avoid falling profits	Dec	The ratio of current net profit minus last year's net profit and the total assets at the beginning of last year is within the range [0,0.005]
Whether to avoid losses	AL	The net profit is greater than 0 in the previous year and the ratio of the current year to the total assets at the beginning of the period is within the range [0,0.005]
Accounts receivable turnover	SALAR	$\frac{\text{Net income from sales on credit}}{[\text{Initial accounts receivable} + \text{closing accounts receivable}]/2}$
Inventory turnover	SALINV	$\frac{\text{Operating income}}{[\text{Initial inventory} + \text{the ending inventory}]/2}$
Total asset turnover	SALTA	$\frac{\text{Sales revenue}}{\text{Total assets}}$
Financing ability	FINANCE <sub>t</sub>	$\frac{\text{Operating cash flow}_t - \text{average capital expenditure}_{t-1-t-3}}{\text{Current assets}_m}$
Total accruals	TATA	$\frac{\text{Accrued items}}{\text{Total assets}}$ Accrued items = (Δcurrent assets – Δmonetary fund) – (Δcurrent liabilities – long-term liabilities due within one year – Δpayable taxes) – depreciation expense
Administrative rate	MR	Administrative rate = $\frac{\text{administrative expenses}}{\text{main business income}} * 100\%$
Sales expense ratio	SR	Ratio of marketing expense = $\frac{\text{operating expense}}{\text{main business income}} * 100\%$
Inventory is accounted for	SR	Inventory and camping business costs

TABLE 4: Cash flow indicators of listed companies.

The index name	Short for index	Design formulas
Relationship between operating profit and operating cash flow	CATA	$\frac{\text{Operating profit} \cdot \text{operating cash flow}}{\text{Total assets}}$
Free cash flow	PREEC	Ebit + depreciation • income tax • capital expenditure • net increase in working capital
Proportion of other receivables	OTHREC	$\frac{\text{Other receivables}}{\text{Total assets}}$

5.1. *Accuracy.* The meaning of model exactness is as per the following, given  $D$  set with test size  $m$ . Each sample in the training set is  $\chi_i$ , and the corresponding label is  $y_i$ . The accuracy of model is  $f$ .

$$\text{acc}(f; D) = \frac{1}{m} \sum_{i=1}^m I(f(\chi_i) = y_i). \quad (2)$$

5.2. *Precision Rate and Recall Rate.* In our application scenario, we also focus on how much of the problematic samples really exists in financial risk and for all the financial risk of the company how many model successfully was identified. Therefore, precision and recall often attract more attention. Table 5 gives a confusion matrix in our experiment.

With the obfuscation matrix, we can define accuracy  $P$  and recall rate  $R$  as

$$P = \frac{TP}{TP + FP}, \quad (3)$$

$$R = \frac{TP}{TP + FN}. \quad (4)$$

Precision rate and review rate are disconnected factors. As a rule, the higher the precision rate, the lower the review rate, and the higher the review rate, the lower the exactness rate. For instance, in outrageous cases, arranging all organizations as monetarily dangerous would yield a 100 percent review rate; however, exactness would be extremely low. If by some stroke of good luck the most dependable



TABLE 5: Confusion matrix.

Real situation	Predicted results	
	Positive example	Counter-examples
Positive example	True positive	True negative
Counter-examples	False positive	False negative

TABLE 6: The result of financial indicators forecasting model.

Method	Accuracy	Precision	Recall	F1	AUC
DT	0.8834	0.3323	0.7162	0.454	0.8059
RF	0.8837	0.3461	0.8078	0.4846	0.9066
GBDT	0.9083	0.36	0.8227	0.508	0.9247

TABLE 7: The result of financial statement forecasting model.

Method	Accuracy	Precision	Recall	F1	AUC
CNN	0.9001	0.3184	0.6898	0.4357	0.8012
RNN	0.91	0.2779	0.6631	0.3916	0.7922

TABLE 8: The result of hybrid forecasting model.

Method	Accuracy	Precision	Recall	F1	AUC
CNN+RF	0.9452	0.2675	0.6369	0.3768	0.7952
CNN+DT	0.9404	0.4722	0.5651	0.5145	0.9175
CNN+GBDT	0.9535	0.3261	0.6977	0.4444	0.8291
RNN+RF	0.9429	0.2866	0.8036	0.4225	0.9584
RNN+DT	0.9244	0.4561	0.6064	0.5206	0.8930
RNN+GBDT	0.9679	0.3860	0.3929	0.3894	0.9363

organizations were chosen to further develop precision, the review rate would be low.

In many cases, we can rank the sample according to the model's predictions, putting the companies most likely to be at financial risk first, representing the companies that the model thinks are "most likely" to be at risk. At the bottom of the list are companies that the model says are "least likely" to be at risk. In this order, the samples are taken as the companies with risks one by one for prediction, and the corresponding predicted value of the company is taken as the segmentation point of the model to mark the samples, so as to calculate the current accuracy rate and recall rate. The graph was drawn with accuracy as the vertical axis and recall rate as the horizontal axis. The accuracy-recall curve can be obtained, also known as "P-R graph." This figure can intuitively compare multiple models. If the P-R curve corresponding to model A completely "covers" the curve corresponding to model B, it can be said that the performance of model A is better than that of model B.

**5.3. F1-Score.** Generally speaking, it is necessary to take into account both accuracy and recall rate in model evaluation; in any case, it is hard to contrast various models just and P-R bend, since it is hard to think about the presentation of two models when bends meet. At this time, a comprehensive

index *F1*-score is adopted, which is defined as follows:

$$F1 = \frac{2}{1/p + 1/R} = \frac{2 \times P \times R}{P + R}. \quad (5)$$

It can be seen from the nature of harmonic mean that the mean value is sensitive to small values; that is, at the point when a little worth shows up in the exactness and review pace of the model, the *F1*-score of the model will be a lot more modest, albeit the other worth is huge. This property just meets our requirement for model evaluation; that is, the model should give the consideration to accuracy rate and recall rate, and not lose one or the other. Therefore, *F1*-score will be taken as one of the main evaluation indexes in the evaluation process of the model.

According to Table 6, we can find that GBDT model has a certain improvement compared with RF and DT model. In terms of model accuracy, GBDT model is better, and the model also achieves the best *F1*-score. DT model has almost the same accuracy as RF model, but RF model performs better in other evaluation indicators than DT model. Generally speaking, GBDT is an ideal model.

The model dependent on the text of yearly report of recorded organizations is chiefly developed by profound neural organization. Here, we first use Jieba in Python to perform word division for text. After word division, we really want to execute word vector-based embedding handling. We attempt a text classifier dependent on profound neural organization. During the time spent addressing the enhancement boundaries of the model, because of the difficult issue of test unevenness, we embraced focal loss as the loss work. During the time spent preparing, we observed that the improvement of loss capacity could empower the model to successfully become familiar with the qualities of information and try not to decide all organizations as organizations without monetary danger. Table 7 shows the result of financial statement forecasting model as follows.

In this paper, two sorts of profound neural organization structures are attempted altogether, CNN and RNN individually. It tends to be seen from the outcomes in the table that the impact of the model dependent on RNN design is altogether better compared to that of the CNN model dependent on convolutional neural organization.

More importantly, based on the constructed classification model based on financial statements, we utilize the monetary danger cautioning model dependent on the text of the yearly report to create a helper order result and afterward integrate the result consequence of the past model to yield the last forecast outcome. Table 8 shows the result of hybrid forecasting model.

The test results show that our proposed group model played out awesome and significantly outperforms other benchmark methods for financial crisis prediction and identification, showing that it very well may be applied as a wise recognizable proof framework to further develop classification exactness and proficiency for early warning and identification of financial crisis of listed companies in China stock marketing.

## 6. Conclusion

The examination objective of this paper is to join state-of-the-art AI calculations, regular language handling innovation, profound learning hypothesis, and monetary market and investigate the utilization of AI and enormous information apparatuses in the monetary field. In the investigation in this paper, for an aggregate of 3465 recorded organizations, we have taken on in the model dependent on fiscal reports of the six classes a total of 25 indicators, and we adopted the annual report of listed companies in terms of text data as a basis analysis while receiving numeric type and text input and output of enterprise financial risk probability model. To be specific, this paper first builds a financial risk prediction method with collected corporation financial statements, assembles auxiliary markers dependent on the current writing on monetary danger investigation, and afterward utilizes different tree models to construct models and assesses the characterization results. Secondly, this paper continues to construct a monetary danger expectation model dependent on the text of the yearly report of recorded organizations. It utilizes a few normal profound neural organization models for regular language handling errands to construct the model and uses the model with the best arrangement impact as the result to fabricate the mix model. Finally, this paper integrates the output results of the two models through model integration to assemble a total monetary danger examination arrangement of recorded organizations, dissects the commitment of each model, and obtains the final model classification evaluation index. The results show that through the combination of text information and financial statement information, its performance in the experiment is better than the single financial warning model.

In addition, a class of tree structure model dependent on choice tree and profound neural organization have been generally utilized in different fields, for example, picture acknowledgment and regular language handling. In this paper, we demonstrate through training that these models can likewise assume a part in the monetary field. What's more according to the application perspective, this sort of model has more boundaries, can manage high layered information, and can be led on countless informational indexes. The adaptability of the fit itself allows the model to be used to perceive and mine the complexity of the monetary business sector, which is more flexible than traditional time series models. RF model can obtain the best overall score among financial statement based models and might be more reasonable for down to earth application situations in the monetary field.

Moreover, in the model dependent on the text of the yearly report of recorded organizations, holding more text for the text of the yearly report can guarantee the model to get more compelling data, yet with the expansion of text arrangement, it additionally carries challenges to the model learning. This is additionally one of the issues we want to settle for our examination.

## Data Availability

The supporting data are available from the corresponding author.

## Conflicts of Interest

The author declares no conflicts of interest.

## References

- [1] C. Title, "A comparison of the ratios of successful industrial enterprises with those of failed companies," *Análise Molecular Do Gene Wwox*, vol. 12, pp. 598–605, 1932.
- [2] W. H. Beaver, "Market prices, financial ratios, and the prediction of failure," *Journal of Accounting Research*, Autumn, vol. 6, no. 2, pp. 179–192, 1968.
- [3] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.
- [4] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–131, 1980.
- [5] M. D. Odom and R. Sharda, "A neural network model for bankruptcy prediction," in *1990 IJCNN International Joint Conference on neural networks*, pp. 163–168, San Diego, CA, USA, June 1990.
- [6] D. Niu, S. Hou, Y. Zhang, and X. Sun, "Financial risk early-warning based on RS-SVM hybrid model," *Proceedings of SPIE-The International Society for Optical Engineering*, vol. 7490, article 74902I, 2009.
- [7] Y. Z. Ma, "Computer simulation evaluation of financial risk based on cuckoo search and SVM algorithm," *Journal of Physics Conference Series*, vol. 1533, no. 3, article 032045, 2020.
- [8] A. A. Soltan and M. M. Mohammadi, "A hybrid model using decision tree and neural network for credit scoring problem," *Management Science Letters*, vol. 2, no. 5, pp. 1683–1688, 2012.
- [9] S. Deng, C. Wang, M. Wang, and Z. Sun, "A gradient boosting decision tree approach for insider trading identification: an empirical model evaluation of China stock market," *Applied Soft Computing*, vol. 83, p. 105652, 2019.
- [10] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 99, pp. 2999–3007, 2017.
- [11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, *Efficient Estimation of Word Representations in Vector Space*, 2013, <http://arxiv.org/abs/1301.3781>.
- [12] Q. Wang, J. Xu, H. Chen, and B. He, "Two improved continuous bag-of-word models," in *International Joint Conference on Neural Networks*, pp. 2851–2856, Anchorage, AK, USA, May 2017.
- [13] P. Tino, C. Schittenkopf, and G. Dorffner, "Financial volatility trading using recurrent neural networks," *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 865–874, 2001.
- [14] N. N. Dereli and M. Saralar, "Convolutional neural networks for financial text regression," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 331–337, Florence, Italy, July 2019.

- [15] J. Dong, "Financial investor sentiment analysis based on FPGA and convolutional neural network," *Microprocessors and Microsystems*, vol. 103418, p. 103418, 2020.
- [16] S. Deng, C. Wang, J. Li et al., "Identification of insider trading using extreme gradient boosting and multi-objective optimization," *Information*, vol. 10, no. 12, p. 367, 2019.
- [17] X. Xi, X. Gao, J. Zhou, H. Zheng, J. Ding, and J. Si, "Uncovering the impacts of structural similarity of financial indicators on stock returns at different quantile levels," *International Review of Financial Analysis*, vol. 76, p. 101787, 2021.
- [18] M. Adnan, A. T. Pasha, R. Mehmood, M. Sadiq, M. Waris, and R. T. Naveed, "Impact of the financial indicators on bank performance and a comparison of small and large banks-evidence from the Bahrain," 2021.
- [19] Z. Bao and C. Wang, "A multi-agent knowledge integration process for enterprise management innovation from the perspective of neural network," *Information Processing & Management*, vol. 59, no. 2, article 102873, 2022.