Research Article

Prediction of Unbalanced Financial Risk Based on GRA-TOPSIS and SMOTE-CNN

Huiru Li ⓘ**, Hongjiu Liu** ⓘ**, and Yanrong Hu** ⓘ

School of Mathematics and Computer Science, Zhejiang A&F University, Zhejiang, Hangzhou 311300, China

Correspondence should be addressed to Hongjiu Liu; joe_hunter@zafu.edu.cn and Yanrong Hu; rosehyr2004@aliyun.com

The financial status of an enterprise is related to its healthy and long-term development, and whether the interests of investors and bank loans can be guaranteed. To improve the prediction accuracy of corporate financial risk, this paper proposes a prediction model for corporate financial risk that integrates GRA-TOPSIS and SMOTE-CNN. First, using GRA-TOPSIS to make a comprehensive evaluation of the financial situation of listed companies. Second, the evaluation results are clustered to obtain the scientific level and interval of financial risk, which lays the foundation for the supervised learning of the convolutional neural network. Then, the SMOTE algorithm is introduced to solve the problem of data imbalance of enterprises at all levels, and the focal loss function is used instead of the cross-entropy loss function to further balance the data. Finally, the listed companies in A shares are randomly selected, and experiments were designed to verify the performance of the model built in this paper. The results show that the prediction accuracy of the financial risk prediction model based on GRA-TOPSIS and SMOTE-CNN is 98.57%, which indicates that the model is feasible and has certain reference value.

1. Introduction

With the deepening of economic reform, China's economy has developed rapidly and become the second largest economy in the world. The resulting social competition for enterprises is becoming more and more intense [1, 2]. Finance also faces many risks and challenges [3]. Listed companies are the representatives of enterprises and the backbone of the national economy. Therefore, to maintain the sustainable and high-quality development of country's economy, it is necessary to ensure the healthy and long-term growth of listed companies. Since corporate financial status is a direct manifestation of achievements of enterprise development and is the focus of all the stakeholders of enterprise including operators, corporate creditors, and investors [4], it is particularly important to accurately evaluate and monitor it.

Financial risk prediction is an early warning mechanism and real-time monitoring method established to prevent enterprises from making mistakes and facing risks [5]. The research started in the 1930s. First Fitzpatrick proposed the univariate financial evaluation model [6], which was simple to operate and has a single indicator, but it was not accurate enough, and then Altman proposed and improved the Z-score model to predict financial risk, but the long-term evaluation ability was weak [7]. Odom and Shardal introduced a neural network for corporate bankruptcy prediction, and the verification found that its accuracy was better than other existing models [8], but technical requirements were high at the time. Later, Shaverdi et al. used fuzzy AHP to determine the index weight and then use fuzzy TOPSIS to determine the financial level of petrochemical enterprises [9]. Deng et al. built a dynamic rating system based on DEA and analytic hierarchy process to determine the financial status of Chinese nuclear power enterprise [10]. Chen constructed a financial performance evaluation system from four aspects: profitability, operating ability, debt payment ability, and development ability and measured the performance level through fuzzy comprehensive evaluation [11].

With the development of enterprises and the increase of financial data of enterprises, traditional statistical methods can no longer accurately predict financial status. Some

scholars began to use machine learning methods for financial risk prediction and the most widely used are BP neural network [12–14], SVM [15, 16], and decision tree [17]. As such, Zhou et al. measured and warned the risks of real estate companies through the implementation of the PSO-SVM model [18]. Feng et al. constructed a corporate finance risk warning model-based BP neural network to predict financial crises and proved the accuracy is at least 2% higher than traditional method [19]. Liao and Liu applied the decision tree method to enterprise financial risk early warning and provided a reference for risk control decision-making [20].

With the wide application of deep learning, scholars have begun to introduce it into the research of the financial field. Chen used convolutional neural network to make financial quantitative investment and obtained investment strategy with higher accuracy and reduced investment risk [21]. Abudureheman et al. built a performance evaluation of enterprise innovation capability based on fuzzy system model and convolutional neural network, which was significant to promote enterprise development [22]. Yin et al. built a convolutional neural network model to supply chain financial risk early warning, but the enterprises were divided into two categories only according to whether they are "ST," and the sample is small [23]. Besides, more and more companies are introducing deep learning into their financial management.

In addition, in the construction of the financial risk evaluation index system, research shows that the quality of financial reports has a certain impact on investment efficiency [24]. Tangible resources and operational performance can promote financial performance [25], and different industries have some different detailed factors that can have a certain impact on financial performance. For example, lean manufacturing has a certain role in promoting the performance of the pharmaceutical industry [26]. Therefore, on the basis of previous research, this paper selects the highly professional NetEase financial website to crawl financial report data. Through correlation analysis, 28 secondary indicators belonging to 5 primary indicators are selected from 69 financial indicators for financial risk prediction and early warning of A-share companies. Regardless of the specific industry, it can better reflect the overall development level of Chinese listed companies.

From the current research, it can be seen that there are still some problems in the research on corporate financial risk early warning: (1) Most early warning research only realizes the measurement and rating of corporate financial situation, and lacks intelligent prediction. When the sample size is large, it is difficult to visually see the financial risk status of the enterprise. (2) When measuring and scoring, only a single model is used to evaluate the financial situation, and the evaluation results are affected due to the different emphasis of information. At the same time, some shallow networks of machine learning are prone to overfitting, which affects the prediction accuracy. (3) When classifying enterprise risk levels, dividing them into two categories only depends on whether the enterprise is processed by "ST," resulting in an extremely unbalanced sample size.

Considering the above problems, this paper proposes a corporate financial risk early warning model based on GRA-TOPSIS and SMOTE-CNN. The GRA-TOPSIS fusion model is used to score the financial situation of the enterprise, and according to the score results, K-means is used for clustering to get the risk level label, and the CNN model is trained to realize intelligent prediction. The integration of supervised learning and unsupervised learning makes up for the lack of intelligent prediction in previous research and the difficulty in obtaining enterprise financial level labels. At the same time, considering that the number of enterprises with financial health and heavy warning is smaller, and the number of enterprises with general finance is larger, it can be regarded as a multiclassification problem of unbalanced data. Therefore, this paper uses the SMOTE algorithm to oversample a small number of samples and uses the focal loss function replaces the traditional multiclass cross-entropy loss function, which further balances the data by assigning different weights to the unbalanced data.

The other parts of the paper are arranged as follows: the second chapter is data and methodology, mainly introduces the source of experimental data and the principles of the methods used in the experiment. The third chapter is results and analysis, mainly through empirical analysis to get the experimental results, according to the experimental results analyze which indicators should be paid more attention in corporate financial risk, and verify the effectiveness and progress of the model proposed in the article through comparison. The fourth chapter is conclusions and prospects, point out the conclusions of this research and the aspects of further research that can be carried out in the future.

2. Data and Methodology

2.1. Data Collection and Index System Construction

2.1.1. Data Collection. This experiment uses the "read_html" function of the pandas module in python to quickly and accurately capture three-year financial data of 4,727 A-share listed companies from the financial website. These companies belong to the industries of information technology, finance, manufacturing, communication, and education. Among them, there are 146 ST enterprises. A total of 14,181 samples were obtained as research objects, and samples with abnormal data and missing data are directly eliminated, leaving 13,190 samples in the end, including 401 "ST" samples. After SMOTE oversampling, the sample size is 23770. The article divides the training, validation, and test according to the ratio of $6:2:2$, and the sample sizes are 14262, 4754, and 4754, respectively.

2.1.2. Construction of Index System. This paper establishes a sound evaluation index system for prediction of corporate financial risk. First of all, 69 financial indicators are used as evaluation indicators, combined with the correlation coefficient method, and using the IBM SPSS Statistics 22 software for correlation analysis. Finally, 28 evaluation indicators in this study are selected from the five aspects of main financial indicators, solvency, growth, profitability, and operating, as shown in Table 1.

2.2. Methods Based on GRA-TOPSIS and SMOTE-CNN

2.2.1. GRA-TOPSIS Model.

The GRA-TOPSIS model is a combination model of gray correlation analysis and TOPSIS. The GRA model mainly judges the correlation of the sequence based on the similarity of the curve change trend between the corresponding points of the sequence. While the TOPSIS model is an approach to the ideal solution, which is sorted by the distance between the evaluation target and the best and worst sequences. The combined algorithm steps are as follows:

(1) Determine the multi-attribute evaluation matrix. Assuming that n factors are influencing corporate financial, and there are m companies, the evaluation matrix is

$$A = \left(a_{ij}\right)_{m \times n}. \tag{1}$$

Standardize the index using the maximum and minimum normalization methods:

Positive index:

$$a_{ij} = \frac{a_{ij} - \min a_j}{\max a_j - \min a_j}. \tag{2}$$

Negative index:

$$a_{ij} = 1 - \frac{a_{ij} - \min a_j}{\max a_j - \min a_j}. \tag{3}$$

(2) The entropy method determines the index weight, the weight of the j-th index is w_j, and the weight matrix is

$$W = \left[w_1, w_2, \cdots, w_n\right]. \tag{4}$$

(3) The TOPSIS method determines the Euclidean distance between the evaluation object and the positive and negative ideal solutions. Calculate the evaluation matrix: $A = w_i * (a_{ij})_{m \times n}$

The positive and negative ideal solution of a_{ij} is

$$\begin{aligned}
R_i^+ &= \left\{\left(\max_j a_{ij} | i \in I\right) | j = 1, 2, \cdots n\right\} \\
&= \left(a_1^+, a_2^+, \cdots a_n^+\right)(i = 1, 2, \cdots, m), \\
R_i^- &= \left\{\left(\min_j a_{ij} | i \in I\right) | j = 1, 2, n\right\} \\
&= \left(a_1^-, a_2^-, \cdots a_n^-\right)(i = 1, 2, \cdots, m),
\end{aligned} \tag{5}$$

d_i^+, d_i^- represent the euclidean distance between a_{ij} and its positive and negative ideal solutions:

$$\begin{aligned}
d_i^+ &= \sqrt{\sum_j^n \left(a_j^+ - a_{ij}\right)^2}(i = 1, 2, \cdots, m), \\
d_i^- &= \sqrt{\sum_j^n \left(a_j^- - a_{ij}\right)^2}(i = 1, 2, \cdots, m).
\end{aligned} \tag{6}$$

Then the closeness between the ith enterprise and the ideal enterprise is

$$G_i = \frac{d_i^-}{d_i^+ + d_i^-}(i = 1, 2, \cdots, m). \tag{7}$$

(4) The GRA method determines the degree of relevance. The correlation coefficient matrix between each comparison sequence and the best reference sequence and the worst reference sequence is as follows:

$$\begin{aligned}
R^+ = r_{ij}^+ &= \frac{\min\min\left|a_{ij}^+ - a_{ij}\right| + \lambda\max\max\left|a_{ij}^+ - a_{ij}\right|}{\left|a_{ij}^+ - a_{ij}\right| + \lambda\max\max\left|a_{ij}^+ - a_{ij}\right|}, \\
R^- = r_{ij}^- &= \frac{\min\min\left|a_{ij}^- - a_{ij}\right| + \lambda\max\max\left|a_{ij}^- - a_{ij}\right|}{\left|a_{ij}^- - a_{ij}\right| + \lambda\max\max\left|a_{ij}^- - a_{ij}\right|}.
\end{aligned} \tag{8}$$

The formula for calculating the gray correlation degree is as follows:

$$\begin{aligned}
r_i^+ &= \frac{1}{n}\sum_{j=1}^n r_{ij}^+ (i = 1, 2, \cdots, m), \\
r_i^- &= \frac{1}{n}\sum_{j=1}^n r_{ij}^- (i = 1, 2, \cdots, m).
\end{aligned} \tag{9}$$

(5) Integrated Euclidean distance and gray correlation degree for comprehensive evaluation. Build a more reasonable comprehensive evaluation model through weighted processing, and the calculation formula is

$$\begin{aligned}
S_i^+ &= \alpha d_i^- + \beta r_i^+ (i = 1, 2, \cdots, m), \\
S_i^- &= \alpha d_i^+ + \beta r_i^- (i = 1, 2, \cdots, m).
\end{aligned} \tag{10}$$

Finally, calculate the comprehensive evaluation score of each enterprise, the calculation formula:

$$C_i = \frac{S_i^+}{S_i^+ + S_i^-}(i = 1, 2, \cdots m). \tag{11}$$

2.2.2. Clustering Algorithm.

After the comprehensive evaluation, the classification of financial risk is achieved by clustering the unlabeled data. Clustering is a typical unsupervised learning algorithm, which refers to dividing samples into multiple clusters according to a certain standard. Commonly used clustering methods are K-means clustering, hierarchical clustering, SOM clustering, and FCM clustering. Because K-means clustering has the characteristics of high efficiency, high accuracy, and strong interpretability, this paper chooses K-means clustering algorithm to generate corporate financial risk grade labels. The algorithm steps are as follows:

TABLE 1: Evaluation indicator system of enterprise performance.

Primary index	Secondary index	Variable	Nature of indicator
Main financial indicators	Basic earnings per share	X_1	Positive
	Revenue from main business	X_2	Positive
	Net non-operating income and expenditure	X_3	Positive
	Net increase in cash and cash equivalents	X_4	Positive
	Return on equity weighting	X_5	Positive
Solvency	Current ratio	X_6	Positive
	Cash ratio	X_7	Positive
	Interest payment multiple	X_8	Positive
	Asset-liability ratio	X_9	Negative
	Ratio of shareholders' equity to fixed assets	X_{10}	Positive
	Net fixed assets ratio	X_{11}	Positive
	Equity ratio	X_{12}	Negative
	Proportion of fixed assets	X_{13}	Positive
Growth	Growth rate of main business revenue	X_{14}	Positive
	The growth rate of total assets	X_{15}	Positive
Profitability	Total assets margin	X_{16}	Positive
	Main operating profit margin	X_{17}	Positive
	Cost profit margin	X_{18}	Positive
	Operating profit margin	X_{19}	Positive
	Return on assets	X_{20}	Positive
	Proportion of three expenses	X_{21}	Negative
	Non-main business proportion	X_{22}	Positive
Operating	Accounts receivable turnover rate	X_{23}	Positive
	Inventory turnover rate	X_{24}	Positive
	Total asset turnover rate	X_{25}	Positive
	Ratio of net operating cash flow to sales revenue	X_{26}	Positive
	Return on operating cash flow of assets	X_{27}	Positive
	Cash flow ratio	X_{28}	Positive

(1) Initialize the cluster centers: according to experience, select K from the sample set as the initial clustering centers, and determine the maximum number of iterations.

(2) The Euclidean distance between each sample and the clustering center is calculated, and the samples to be clustered are judged to belong to the same class according to the distance, with the following formula:

$$d_i = \sqrt{\sum_{j=1}^{n} \left(x_j - p_{ij}\right)^2}. \tag{12}$$

(3) Calculate the average of each category of samples as the new cluster center, the formula is as follows:

$$K_i = \frac{1}{|S_i|} \sum_{x_i \in S_i} x_i. \tag{13}$$

(4) Repeat Steps 2 and 3 until the clustered data points do not change or the number of iterations is reached.

2.2.3. SMOTE Algorithm. Among the five risk classes obtained by clustering, the enterprises with excellent and poor finance are few, so the SMOTE algorithm is used to balance the samples. SMOTE is an improved scheme based on the random oversampling algorithm [27], and its basic idea is to analyze the minority class samples and add new samples to the dataset by artificially synthesizing them according to the minority class samples, which can effectively avoid the overfitting problem. The algorithm steps are as follows:

(1) Calculate the Euclidean distance from each minority sample x to the minority sample set to obtain k nearest neighbors.

(2) Determine the oversampling magnification N, randomly select a sample x_i from k nearest neighbors, take a random number between 0 and 1, and synthesize a new sample for each sample x according to formula (14).

$$x_{new} = x + a * \left(x_i - x\right). \tag{14}$$

(3) Repeat Step 2 to get a new data set.

2.2.4. Convolutional Neural Network. After generating labels by cluster analysis, convolutional neural networks are used to achieve corporate financial risk classification prediction. Convolutional neural network is a feed-forward neural network, one of the representative algorithms of deep learning, whose most important feature is the sharing of weights, which can greatly improve the time required for learning and reduce the amount of data
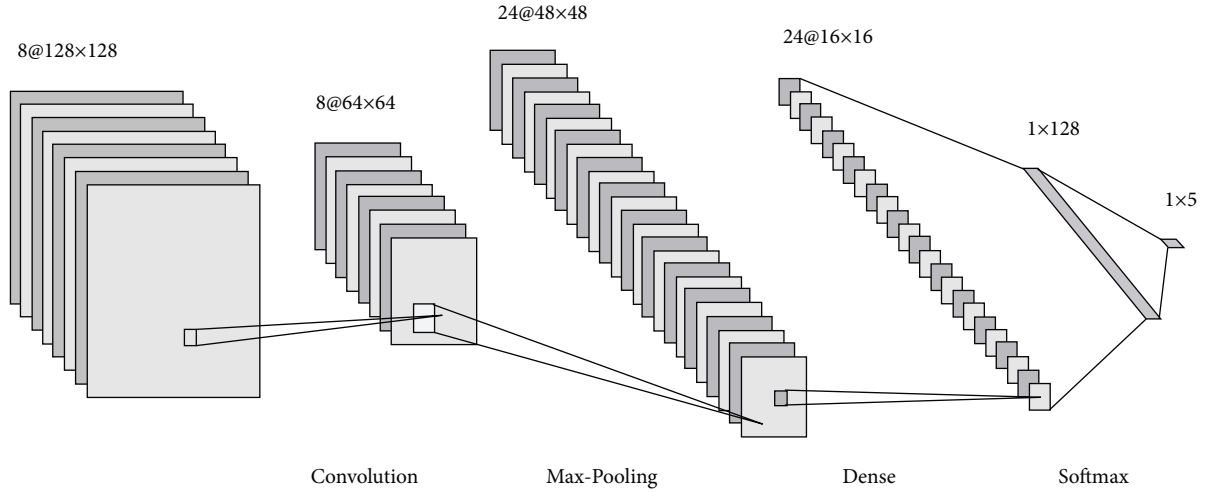
Figure 1: CNN structure chart.

needed to train the model by reducing the parameters [28]. The traditional convolutional neural network is generally composed of convolutional layer, pooling layer, and fully connected layer. The input data are extracted through the convolutional layer, and the output result is passed to the pooling layer to further extract and filter information, and finally through the fully connected layer output, the structure is shown in Figure 1. Convolutional neural networks are generally used in computer vision, natural language processing, and other fields. In the past two years, studies have introduced them into corporate bankruptcy prediction [29], indicating that it is feasible for convolutional neural networks to process financial statement information.

The data enters the convolution layer through the input layer, and the convolution operation formula is as follows:

$$X_i = f\left(X_{i-1} \otimes W_i + b_i\right). \tag{15}$$

Among them, X_i is the output of the ith layer of convolution, X_{i-1} is the output of the previous layer, W_i is the weight parameter of the ith layer of convolution, and b_i is the offset. The formula of the shape value output by convolution is as follows:

$$\text{outputshape} = \frac{m - n + 2 * u}{v} + 1, \tag{16}$$

where m is the output of the upper layer, n is the number of convolution kernels, u is the number of edge padding, and v is the step size.

The convolutional data enter the pooling layer. In this experiment, max-pooling is selected, and the shape of the output data is the same as the calculation formula of the convolutional layer. The last is the fully connected layer, which acts as a classifier in the convolutional neural network. The commonly used activation functions are Softmax and sigmoid. The activation function used by the fully connected layer in the author's experiment is Softmax:

$$\delta(x)_i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}, \tag{17}$$

$$i = 1, 2, \cdots, n.$$

To further reduce the impact of data imbalance on the classification accuracy, the focal loss function is used instead of the cross-entropy loss function during training. The formula is as follows:

$$L_{fl} = \begin{cases} -\alpha\left(1 - y'\right)^{\gamma} \log y', & y = 1, \\ -(1 - \alpha)\left(1 - y'\right)^{\gamma} \log\left(1 - y'\right), & y = 0. \end{cases} \tag{18}$$

2.2.5. Model Evaluation Indicators. In essence, prediction of financial risk can be regarded as an unbalanced multiclassification problem. For multiclassification problems, microaverage and macroaverage are generally used to evaluate the performance of the model. This experiment uses macro averaging to evaluate model performance. It is to calculate the precision and recall of each category separately and then calculate the average as the macroprecision and the macro recall. The macro_F1 is the harmonic average of the macroprecision and the macro recall. The calculation formula is as follows:

$$\text{macro}_P = \frac{1}{n} \sum_{i=1}^{n} P_i,$$

$$\text{macro}_R = \frac{1}{n} \sum_{i=1}^{n} R_i,$$

$$\text{macro}_{F1} = \frac{2 * \text{macro}_P * \text{macro}_R}{\text{macro}_P + \text{macro}_R},$$

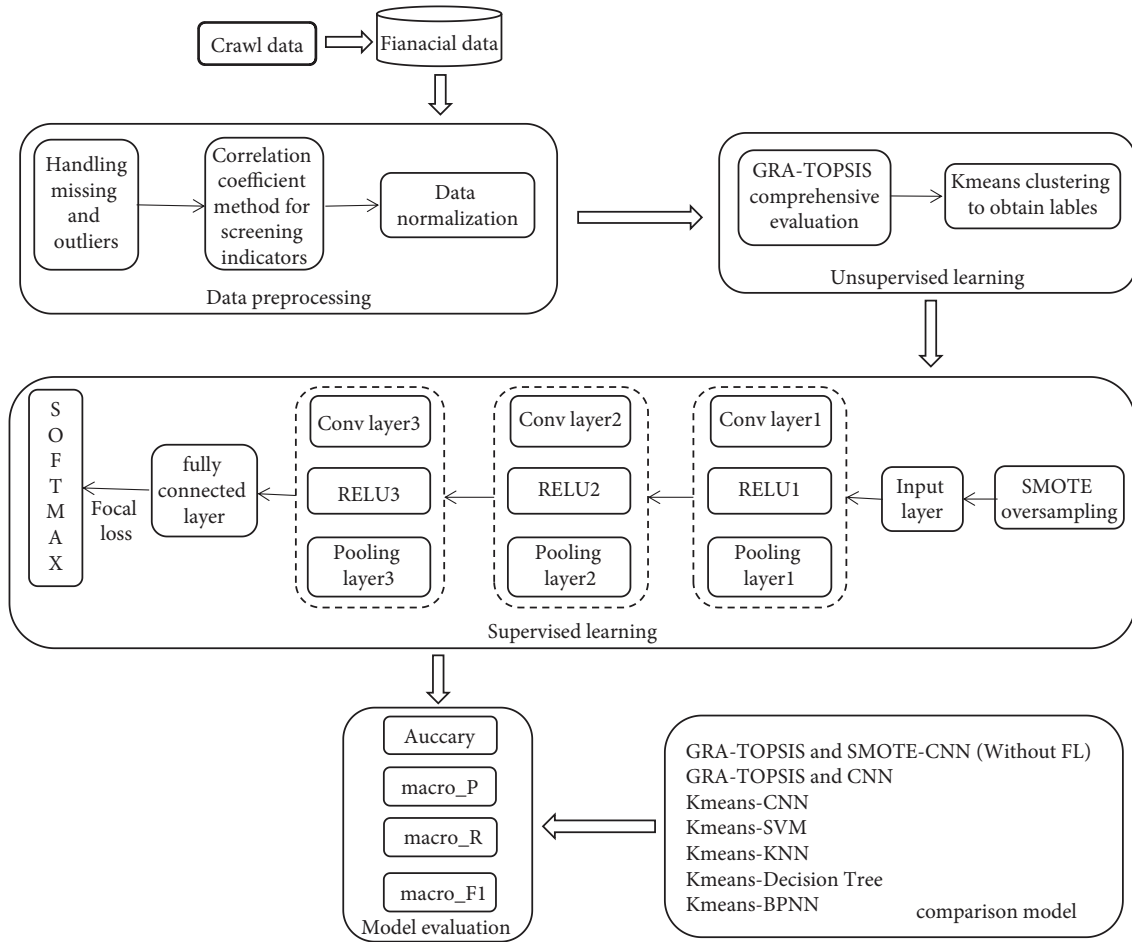$$\text{auccary} = \frac{TN + TP}{TN + FN + TP + FP}. \tag{19}$$

FIGURE 2: Financial risk prediction model flowchart.

2.2.6. Financial Risk Prediction Model. The financial risk prediction model based on GRA-TOPSIS and SMOTE-CNN is shown in Figure 2. The model is mainly divided into four parts: data preprocessing, unsupervised learning, supervised learning, and model performance evaluation. The preprocessing module mainly deletes the missing values and outliers of the data crawled on the financial website and then removes the correlation between the indicators through the Person correlation analysis to determine the evaluation index system. The unsupervised learning module focuses on comprehensive evaluation of financial status and clustering based on the evaluation results to generate class labels. The supervised learning module mainly uses convolutional neural networks to classify and predict financial risk levels based on the samples processed by SMOTE oversampling and the labels generated by clustering. The model performance evaluation module is mainly to observe the accuracy, macroprecision, macrorecall, macro_F1, and other indicators by comparing with other models to prove the advancement and effectiveness of the model proposed in this paper.

3. Results and Discussion

3.1. GRA-TOPSIS Comprehensive Evaluation and Regression Analysis. Calculate the Euclidean distance d_i^+ and d_i^- between the evaluation object and its positive and negative ideal solutions by formulas (5)~(8). Calculate the correlation degree r_i^+, r_i^- between each comparison sequence and the best and worst reference sequence by formulas (11)~(13). Then, the weighted closeness S_i^+ and S_i^- of the coupled TOPSIS and GRA are calculated by formulas (14)~(16), and the comprehensive evaluation score C_i of the coupled model is further calculated. The six companies with the highest scores and the six companies with the lowest scores in the single evaluation model and the GRA-TOPSIS coupling model are shown in Table 2. The evaluation results of the coupling model and standardized indicator data are brought into the multiple linear regression model to further analyze the relationship between the indicators and enterprise financial risk. The coefficients of the indicators calculated by IBM SPSS Statistics 22 software are shown in Table 3.

It can be seen from Table 2 that among the top 6 companies, the coupled model is exactly the same as that of single TOPSIS, and it is the same as that of single GRA with three companies, and in the last six companies, the coupled model and single TOPSIS and single GRA are exactly the same. Calculate the range and variable coefficient of the comprehensive evaluation of corporate financial risk under single GRA, single TOPSIS, and GRA-TOPSIS models. The range is 0.1134, 0.0002, and 0.1233, and the variable coefficient is 0.0081, 0.0113, and 0.0111. Larger range and variable coefficient indicate a higher level of dispersion and

TABLE 2: Corporate financial risk scores and rankings.

Rank	GRA		TOPSIS		GRA-TOPSIS	
	Code	Score	Code	Score ($*10^{-5}$)	Code	Score
1	300896(2020)	0.5966	300896(2020)	8.33	300896(2020)	0.5082
2	301047(2021)	0.5897	301047(2021)	8.30	301047(2021)	0.5059
3	688298(2021)	0.5861	301080(2021)	8.17	301080(2021)	0.4986
4	688068(2021)	0.5839	603879(2021)	8.11	603879(2021)	0.4950
5	605336(2019)	0.5818	688068(2021)	8.10	688388(2019)	0.4948
6	688008(2020)	0.5806	688388(2019)	8.10	688068(2021)	0.4944
...	...	...	...	...	...	...
13185	002157(2021)	0.5327	002157(2021)	7.23	002157(2021)	0.4422
13186	000587(2021)	0.5312	000587(2021)	7.22	000587(2021)	0.4419
13187	002670(2021)	0.5252	002670(2021)	7.07	002670(2021)	0.4327
13188	688176(2021)	0.4994	688176(2021)	6.58	688176(2021)	0.4039
13189	000820(2019)	0.4927	000820(2019)	6.55	000820(2019)	0.4016
13190	000820(2020)	0.4832	000820(2020)	6.26	000820(2020)	0.3849

TABLE 3: Regression model index coefficient table.

Index	Factor b	Index	Factor b	Index	Factor b	Index	Factor b	Index	Factor b
X_1	0.023	X_7	0.023	X_{13}	0.023	X_{19}	0.026	X_{25}	0.023
X_2	0.024	X_8	0.024	X_{14}	0.024	X_{20}	0.025	X_{26}	0.024
X_3	0.024	X_9	−0.025	X_{15}	0.024	X_{21}	−0.024	X_{27}	0.026
X_4	0.023	X_{10}	0.026	X_{16}	0.022	X_{22}	0.024	X_{28}	0.021
X_5	0.025	X_{11}	0.025	X_{17}	0.027	X_{23}	0.025	ε	0.231
X_6	0.026	X_{12}	−0.024	X_{18}	0.027	X_{24}	0.024		

discrimination of the composite evaluation scores, so the coupled model is better than a single composite evaluation model for distinguishing the financial risk of each firm.

From the regression coefficients in Table 3, we can see that in the evaluation of enterprise risk, the asset–liability ratio X_9, the equity ratio X_{12}, and the proportion of three expenses X_{21} have a negative effect on corporate performance, and other indicators have a positive effect, which is consistent with the indicator system proposed in this article. And the coefficients of the indicators do not differ much, among which the more important ones are main operating profit margin X_{17}, cost profit margin X_{18}, current ratio X_6, ratio of shareholders' equity to fixed assets X_{10}, operating profit margin X_{19}, and return on operating cash flow of assets X_{27}, with regression coefficients of 0.027, 0.027, 0.026, 0.026, and 0.026, respectively. The less significant effects are cash flow ratio X_{28}, total assets margin X_{16}, total assets turnover rate X_{25}, cash ratio X_7, and proportion of fixed assets X_{13} with regression coefficients of 0.021, 0.022, 0.023, 0.023, and 0.023, respectively.

3.2. Unlabeled Data Clustering Based on K-Means Clustering.
The authors chose to cluster them into five levels, "AAA," "AA," "A," "B," "C," which represent financial health, financial good, financial general, financial light warning, and financial heavy warning. Substitutes the data into formulas (17)~(18) for clustering. For visual display, randomly select 800 samples to draw the clustering results, as shown in Figure 3, and the specific clustering results are shown in Table 4.

As can be seen intuitively in Figure 3, K-means clusters the samples into five levels, with five colors in the figure indicating five levels, respectively. where the first level "AAA" has a higher score in the composite measure of the GRA-TOPSIS model, the financial situation is better but the least number. The three categories in the middle are more concentrated and more numerous, that is, the number of companies with intermediate financial status in the composite measure is higher. Level "C" measure scores the lowest and has a low number. The sample companies are clustered into five levels, and companies are no longer simply divided into two categories according to whether they are "ST," which makes the prediction results more accurate, and can achieve the purpose of early warning.

From the clustering results in Table 4, it can be seen that the average score difference of each category after clustering according to the GRA-TOPSIS model score is 0.0087, 0.0045, 0.0043, and 0.0056, respectively. That is, the two categories of financial health and heavy warnings are significantly different from the average scores of the middle three categories. It also shows that the samples of healthy and heavy warnings have a large degree of dispersion from the middle three types of samples, and the financial status of most companies is at a medium level. This can also be seen from the final number of clusters in each category, there are 298 samples of financial level "AAA," 3019 samples of level "AA," 4754 samples of level "A," 3877 samples of level "B," and 1242 samples of level "C."

To provide targeted reference opinions for various levels of enterprises, it is necessary to understand the indicators that play an important role in different categories, so that
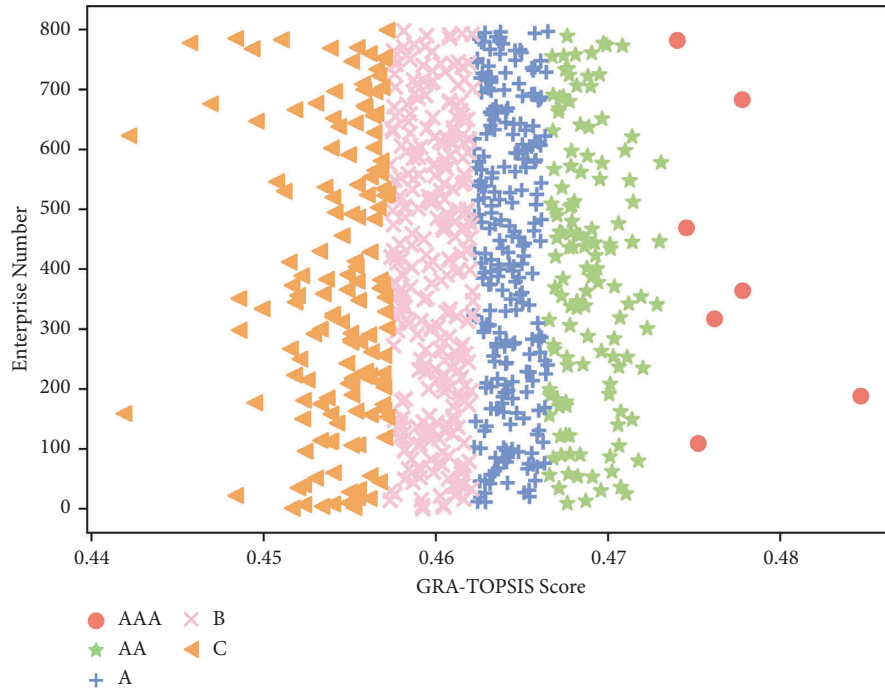
FIGURE 3: 800 samples clustering results.

TABLE 4: K-means clustering results.

Financial risk level	Number of samples	GRA-TOPSIS score range	average score
AAA	298	0.4732~0.5082	0.4775
AA	3019	0.4665~0.4731	0.4688
A	4754	0.4622~0.4665	0.4643
B	3877	0.4572~0.4622	0.4600
C	1242	0.3849~0.4572	0.4544

enterprises can centrally monitor and adjust in time. Therefore, the entropy method is further used to analyze the weights of various financial indicators of various levels of enterprises, as shown in Figure 4. The results show that (1) among the enterprises whose finances are healthy, the indicators with greater weight are inventory turnover rate X_{24}, accounts receivable turnover rate X_{23}, ratio of shareholders' equity to fixed assets X10, revenue from main business X_2, growth rate of main business revenue X_{14}, and the cumulative total weight is 0.6132; (2) among the enterprises with good finance, the more weighted indicators are the ratio of shareholders' equity to fixed assets X_{10}, revenue from main business X_2, accounts receivable turnover rate X_{23}, and inventory turnover rate X_{24}, the cumulative total weight is 0.7186; (3) among the enterprises with a general finance, the indicators with greater weight are inventory turnover rate X_{24}, revenue from main business X_2, ratio of shareholders' equity to fixed assets X_{10}, the cumulative total weight is 0.5524; (4) among enterprises with financial light warning, the indicators with greater weight are accounts receivable turnover rate X_{23}, ratio of shareholders' equity to fixed assets X_{10}, and revenue from main business X_2, the cumulative total weight is 0.6573; (5) among the enterprises with

financial heavy warning, the indicators with greater weight are inventory turnover rate X_{24}, revenue from main business X_2, ratio of shareholders' equity to fixed assets X_{10}, accounts receivable turnover rate X_{23}, and the cumulative total weight is 0.5504, which has a greater impact on the finance of the fifth category of enterprises.

3.3. 1DCNN Classification Prediction. The goal of this experiment is to classify the financial status of the enterprise to achieve the purpose of intelligent forecasting the risk level. The essence is a multiclassification problem. The commonly used loss function is the categorical_crossentropy. Due to the imbalance of the classification samples, the article chooses focal loss to replace the traditional cross-entropy loss function. Focal loss was proposed by He Mingkai in 2017 to improve the effect of dense target detection [30] and has been often used in the field of target detection and natural language processing in the past two years. The optimizer uses the Adam optimizer with faster convergence speed, and the activation function uses Relu and Softmax, batch_size is set as 64 and max epoch is set as 10000. Since there are only 298 "AAA" and 4757 "A" samples in the
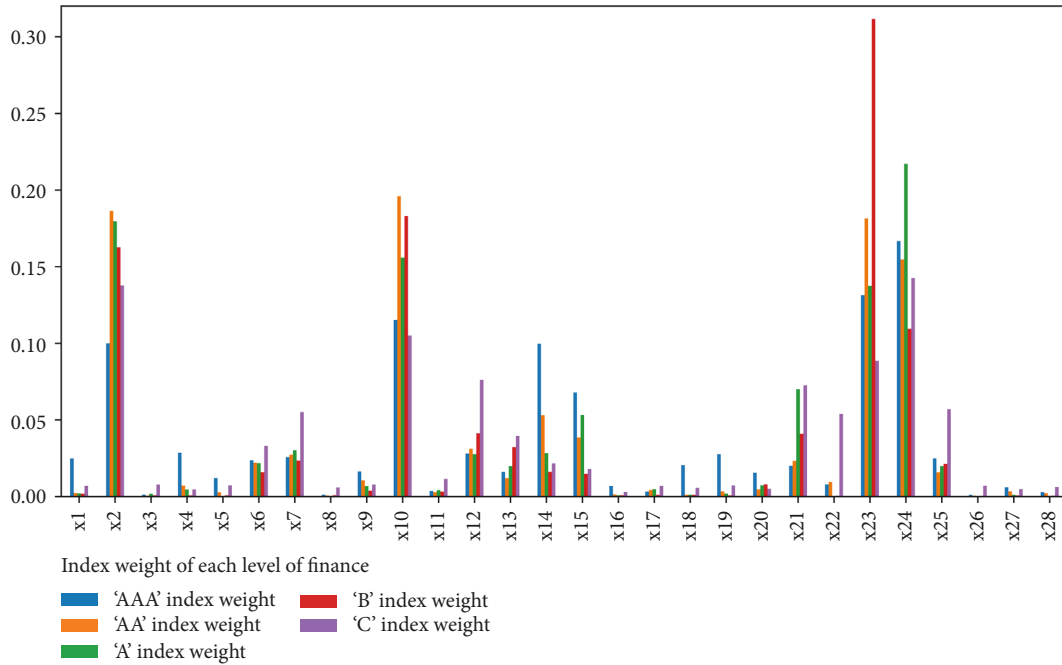
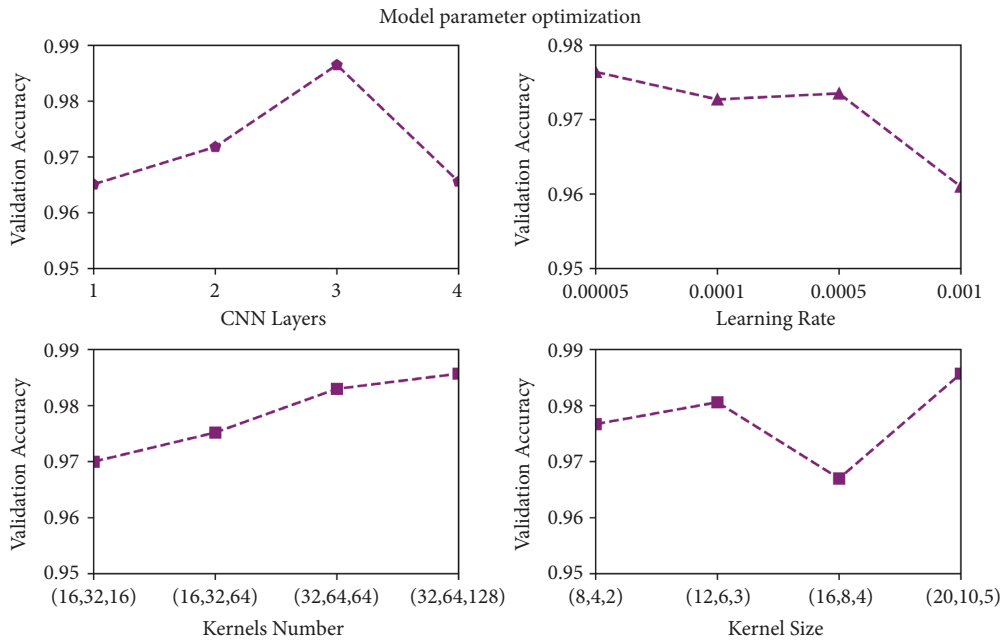FIGURE 4: Comparison chart of the weighting of indicators for each level of financial status.

FIGURE 5: Model parameter optimization.

sample, the ratio is close to 1 : 16, which is a typical sample imbalance multiclassification problem. Therefore, before the data are brought into the neural network for training, they are first oversampled by the SMOTE algorithm. Besides, the model prevents overfitting by adding an Early Stopping mechanism. The patience parameter is set to 50, that is, when the loss function of the verification set does not decrease significantly during 50 iterations, the training is stopped. In addition, L2 regularization is added to control the complexity of the model.

The model hyperparameter setting has a large impact on the model accuracy, and this experiment obtains the optimal parameters of the model by adjusting the parameters to observe the accuracy of the validation set. The more important parameters in the convolutional neural network are the learning rate, the number of hidden layers, the number of convolution kernels, and the size of the convolution kernel. Focal loss has parameters a and r. The experiment adjusts the model parameters through the controlled variable method, that is, keeps other parameters unchanged, adjusts one

TABLE 5: Model parameter table.

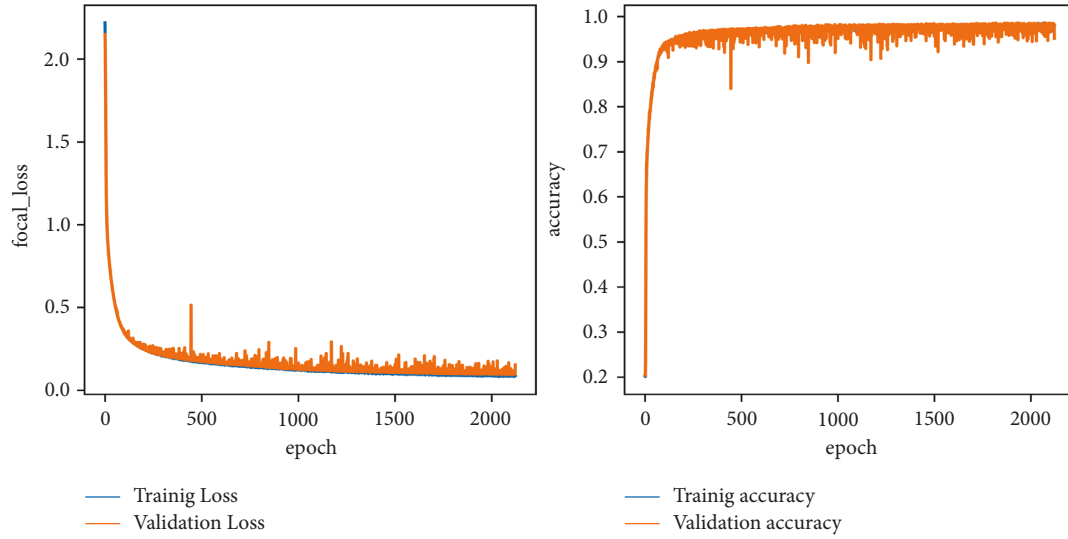| lr	Number of CNN	Number of kernels	Length of kernels	a	r
0.00004	3	32-64-128	12-6-3	[3, 2, 1, 2, 3]	0.2

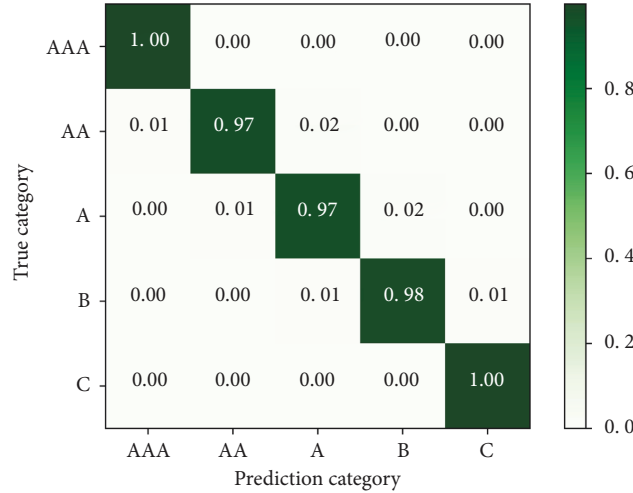FIGURE 6: Training set and validation set loss function and accuracy curve during training.

FIGURE 7: 1DCNN model test set confusion matrix.

parameter successively, observes the accuracy of the validation during model training, and makes it optimal. The accuracy of the validation set corresponding to different parameters is shown in Figure 5, and the optimal parameters of the final model are shown in Table 5.

During the training process of the model, the loss and accuracy change curves of the training set and the validation set are shown in Figure 6. Use the test set to verify the performance of the model and get the confusion matrix as shown in Figure 7. It can be seen that the recognition accuracy of samples with levels "AAA" and "C" is 100%, the

recognition accuracy of samples with levels "AA" and "B" is 98%, the level is "A," that is, the recognition accuracy rate of general financial samples is low, which is 96%. The performance of the test shows that the model proposed in this paper is feasible and effective.

3.4. Multimodel Performance Comparison. To further verify the classification effect of the fusion model proposed in this paper, it is compared with GRA-TOPSIS and SMOTE-CNN (without Focal Loss), GRA-TOPSIS and CNN, Kmeans-
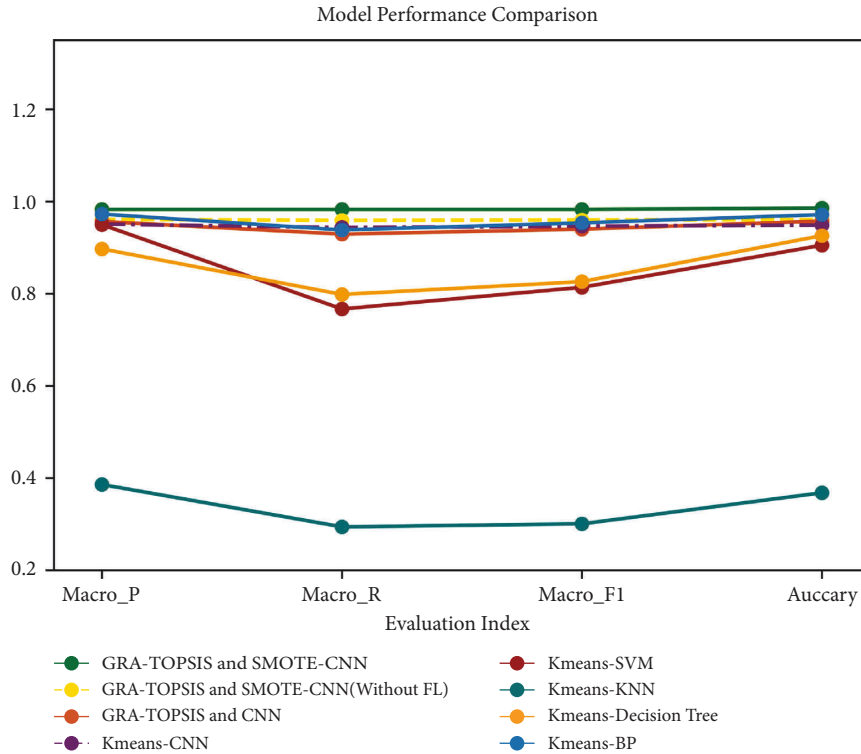
FIGURE 8: Models performance comprehensive comparison.

TABLE 6: Models classification results comparison.

Model name	Macro_Precision	Macro_Recall	Macro_F1	Accuracy
GRA-TOPSIS and SMOTE-CNN	0.9830	0.9830	0.9830	0.9857
GRA-TOPSIS and SMOTE-CNN (without FL)	0.9612	0.9590	0.9599	0.9602
GRA-TOPSIS and CNN	0.9568	0.9263	0.9401	0.9576
Kmeans-CNN	0.9510	0.9437	0.9466	0.9490
Kmeans-SVM	0.9498	0.7669	0.8139	0.9052
Kmeans-KNN	0.3860	0.2943	0.3008	0.3681
Kmeans-decision tree	0.8972	0.7984	0.8262	0.9261
Kmeans-BPNN	0.9729	0.9383	0.9537	0.9717

CNN, Kmeans-SVM, Kmeans-KNN, Kmeans-Decision Tree, and Kmeans-BPNN. The comparison of the evaluation indicators in the validation set in each model is shown in Figure 8, and the specific values of the evaluation indicators of each model are shown in Table 6.

It can be seen from Figure 8 that the model in this paper performs better than other comparison models in the four evaluation indicators of macroprecision, macrorecall, macro_F1, and accuracy. And from Table 6, we can see that the values of each evaluation index of the model in this paper are 0.9830, 0.9830, 0.9830, and 0.9857, which are, respectively, 0.0218, 0.0270, 0.0231, and 0.0255 higher than the model without focal loss, compared with the model that has not been processed by the SMOTE algorithm increase by 0.0262, 0.0567, 0.0429, and 0.0281. In addition, compared with SVM, KNN, decision tree and BPNN commonly used in current research, it can also be seen that the model constructed in this paper performs better in each evaluation index.

The reasons for the higher accuracy of the model in this paper are firstly, the comprehensive evaluation using the GRA-TOPSIS fusion model and then the clustering process, which is more reasonable compared with the direct clustering results of the indicators; second, the SMOTE algorithm and the focal loss function are introduced to balance the data considering the corporate financial performance data as an unbalanced sample, which has some influence on the classification accuracy.

4. Conclusion

To address the two problems that the research of corporate financial risk warning only achieves financial status measurement and rating, intelligent prediction is lacking, and the corporate financial data sample is extremely unbalanced. This paper randomly selects a total of 13190 samples of three years of financial data of 4727 listed

companies in A-share as the research object, uses the GRA-TOPSIS model to make a comprehensive evaluation of the enterprises, and realizes the combination of unsupervised learning and supervised learning through K-means clustering and convolutional neural network and then achieve intelligent prediction of risk level. The clustering results are processed by SMOTE, and the focal loss function is introduced to solve the data imbalance problem of each category and improve the model prediction accuracy. The specific research conclusions and prospects are as follows:

(1) This paper constructs a GRA and TOPSIS fusion model to make a comprehensive evaluation of the financial status of enterprises and measures the final closeness in terms of similarity and Euclidean distance, which is more scientific and reasonable than a single evaluation method, among which 300896(2020) Imeik has the best finance and 000820(2020) *ST energy saving has the worst finance.

(2) Each indicator of enterprise financial data contributes differently to financial status. Based on correlation analysis screening to construct the indicator system, regression model is used to further analyze the indicators that have significant impact, and it is found that the most important indicator is the profitability of main business, and the least important indicator is the cash flow ratio.

(3) The classification accuracy of the GRA-TOPSIS and SMOTE-CNN model proposed in this paper reaches 98.57%, with an average improvement of 4.075% compared to other models except KNN.

(4) The model built in this article can be applied to enterprise financial risk warning, providing a reference for related comprehensive evaluation and other issues. In addition, the financial status evaluation index system of this article only selects financial indicators, and nonfinancial indicators can be added for deeper research in the future.

Data Availability

The experimental data in this paper are available from the NetEase Finance website (https://money.163.com).

Disclosure

Hongjiu Liu and Yanrong Hu are the joint first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this paper.

Acknowledgments

References

[1] C. M. Huang, W. Yang, and R. Q. Zeng, "Analysis on the efficiency of risk management in the Chinese listed companies," *Mathematics*, vol. 8, no. 10, p. 1831, 2020.

[2] X. H. Huang, J. H. Sun, and X. Y. Zhao, "Credit risk assessment of supply chain financing with a grey correlation model: an empirical study on China's home appliance industry," *Complexity*, vol. 2021, Article ID 9981019, 12 pages, 2021.

[3] Y. Cao, Q. F. Wang, W. Cheng, S. Nojavan, and K. Jermsittiparsert, "Risk-constrained optimal operation of fuel cell/photovoltaic/battery/grid hybrid energy system using downside risk constraints method," *International Journal of Hydrogen Energy*, vol. 45, no. 27, Article ID 14108, 2020.

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

[5] A. Q. Wang and H. X. Yu, "The construction and empirical analysis of the company's financial early warning model based on data mining algorithms," *Journal of Mathematics*, vol. 2022, pp. 1–9, Article ID 3808895, 2022.

[6] P. J. Fitzpatrick, "A comparison of the ratios of successful industrial enterprises with those of failed companies," *Certified Public Accountant*, vol. 1, no. 1, pp. 598–605, 1932.

[7] E. I. J. T. J. O. F. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[8] M. D. Odom and R. Sharda, "A neural network model for bankruptcy prediction," *International Joint Conference on Neural Networks*, vol. 2, pp. 163–168, 2012.

[9] M. Shaverdi, I. Ramezani, R. Tahmasebi, and A. A. A. Rostamy, "Combining fuzzy AHP and fuzzy TOPSIS with financial ratios to design a novel performance evaluation model," *International Journal of Fuzzy Systems*, vol. 18, no. 2, pp. 248–262, 2016.

[10] Y. L. Deng, S. L. Zou, and D. M. You, "Financial performance evaluation of nuclear power-related enterprises from the perspective of sustainability," *Environmental Science and Pollution Research*, vol. 27, no. 10, Article ID 11349, 2020.

[11] Z. Y. Chen, "Research on accounting intelligence system modeling of financial performance evaluation," *Security and Communication Networks*, vol. 2021, no. 1, Article ID 5550382, 9 page, 2021.

[12] G. S. Du, Z. X. Liu, and H. F. Lu, "Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment," *Journal of Computational and Applied Mathematics*, vol. 386, Article ID 113260, 2020.

[13] H. J. Liao, H. B. Yue, Y. B. Lin, D. Li, and L. Zhang, "Enterprise financing risk analysis and internal accounting management based on BP neural network model," *Mathematical Problems in Engineering*, vol. 2022, Article ID 8627185, 13 pages, 2022.

[14] J. C. Lyu, "Construction of enterprise financial early warning model based on logistic regression and BP neural network," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2614226, 7 pages, 2022.

[15] B. Y. Liu, "Influence of enterprise cloud computing database application and enterprise debt financing based on SVM

algorithm: internal control quality analysis," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2021.

[16] G. Yao, X. J. Hu, and G. X. Wang, "A novel ensemble feature selection method by integrating multiple ranking information combined with an SVM ensemble model for enterprise credit risk prediction in the supply chain," *Expert Systems with Applications*, vol. 200, Article ID 117002, 2022.

[17] X. F. Wei, "A method of enterprise financial risk analysis and early warning based on decision tree model," *Security and Communication Networks*, vol. 2021, Article ID 6950711, 9 pages, 2021.

[18] W. W. Zhou, M. Y. Chen, Z. L. Yang, and X. B. Song, "Real estate risk measurement and early warning based on PSO-SVM," *Socio-Economic Planning Sciences*, vol. 77, Article ID 101001, 2020.

[19] Q. Feng, H. Chen, and R. H. Jiang, "Analysis of early warning of corporate financial risk via deep learning artificial neural network," *Microprocessors and Microsystems*, vol. 87, Article ID 104387, 2021.

[20] S. Y. Liao and Z. S. Liu, "Enterprise financial influencing factors and early warning based on decision tree model," *Scientific Programming*, vol. 2022, Article ID 6260809, 8 pages, 2022.

[21] C. C. Chen, P. Zhang, Y. Liu, and J. Liu, "Financial quantitative investment using convolutional neural network and deep learning technology," *Neurocomputing*, vol. 390, pp. 384–390, 2020.

[22] A. Abudureheman, A. Nilupaer, and Y. He, "Performance evaluation of enterprises' innovation capacity based on fuzzy system model and convolutional neural network," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 2, pp. 1563–1571, 2020.

[23] L. L. Yin, Y. W. Qin, Y. Hou, and Z. J. Ren, "A convolutional neural network-based model for supply chain financial risk early warning," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–16, Article ID 7825597, 2022.

[24] P. Wajeetongratana, T. Sriyakul, and K. J. I. J. O. I. Jermsittiparsert, "Creativity, and Change, 334 Financial effectiveness, investment efficiency, and quality of financial reporting," *Evidence From ASEAN States*, vol. 8, no. 8, pp. 334–350, 2019.

[25] S. Wongwilai, S. Putnuan, A. Banyongpisut et al., "The influence of tangible resources and operational performance to promote financial performance of electronic industry," *Uncertain Supply Chain Management*, vol. 10, no. 2, pp. 315–324, 2022.

[26] C. Pamornmast, T. Sriyakul, and K. J. S. R. I. P. Termsittiparsert, "Can lean manufacturing and 4.0 industry enhance the financial performance of pharmaceutical industries of Thailand?" *Mediating Role of Waste Reduction Behavior*, vol. 10, no. 2, pp. 318–327, 2019.

[27] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[28] N. Shukla, *Machine Learning with TensorFlow*, Manning Publications, Greenwich, England, 2018.

[29] T. Hosaka, "Bankruptcy prediction using imaged financial ratios and convolutional neural networks," *Expert Systems with Applications*, vol. 117, pp. 287–299, 2019.

[30] T. Y. Lin, P. Goyal, R. Girshick, K. M. He, and P. Dollar, "Focal loss for dense object detection," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.