International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-15

https://doi.org/10.70917/ijcisim-2026-0369
Article

A corporate bankruptcy early warning framework based on principal component analysis to optimize the construction of financial index system in big data environment

Mingwei Li [1,*]

[1] Business school of zjc, Zhijiang college of zhejiang university of technology, Shaoxing, Zhejiang,312030, China
* Correspondence author: hzlmw314@163.com

Abstract: If we can find and take effective measures when the financial situation of the enterprise is just in crisis, we can make many enterprises avoid suffering greater losses and bankruptcy. This paper uses principal component analysis to construct and standardize the data matrix of enterprise financial early warning indicators, optimizes the indicator system and inputs it into the BP neural network model, and determines the risk of enterprise bankruptcy early warning according to the output of the model. The research results show that the financial early warning indicators selected in this paper can significantly distinguish between enterprises with and without bankruptcy risk, and the average accuracy of the model in early warning of enterprise bankruptcy risk is 92.5%. At the same time, the validity and reliability of the model are found to be high in individual cases. The model in this paper has good predictive ability, has the significance of guiding practice, and the results of the study well support the effectiveness of principal component analysis in the enterprise bankruptcy early warning model.

Keywords: principal component analysis; BP neural network; financial indicators; bankruptcy early warning

1. Introduction

With increased competition in the marketplace and an unstable economic environment, enterprises face an increasing number of risks, including the risk of insolvency. Bankruptcy is a great challenge for business owners, employees and the economic system as a whole, and is an important form of asset restructuring in society, where the bankruptcy of severely loss-making companies frees up resources in those companies and promotes their more efficient use [1-4]. However, within certain limits, bankruptcy also brings many negative effects, such as employees of the company will lose their jobs and creditors of the company will only recover part of their payments. In order to reduce the risk of bankruptcy and protect the interests of enterprises, enterprises need to carry out early warning of bankruptcy risk and take timely measures [5-8].

Bankruptcy early warning is a financial analysis that analyzes the enterprise financial statements and related business information, uses timely financial data and corresponding data-based management methods to inform the enterprise operators and other stakeholders in advance of the dangerous situation facing the enterprise, and analyzes the reasons for the occurrence of financial crises in the enterprise and the hidden problems of the enterprise's financial operation system, as well as to take precautionary measures as early as possible [9-12]. Principal component analysis is a statistical analysis method of the measured multiple indicators, with a small number of potentially independent of each other, the linear combination of principal component indicators to represent the composition of the linear combination can reflect the original multiple measured indicators of the main information of a statistical analysis method, the theory is very mature [13-16]. It can provide enterprise managers with early warning of enterprise's bankruptcy by optimizing the construction of financial indicator system, so as to take effective countermeasures [17-18].

Literature [19] pointed out that effective early warning model is important for enterprises to find out the crisis in order to adjust the marketing strategy, based on a number of listed companies in the new energy industry, the use of principal component analysis and logistic regression model to start the comparison, revealing that the logistic regression model shows a better financial early warning effect. Literature [20] based on the financial indicators of some pharmaceutical listed companies, the use of principal component analysis and factor analysis to sort these companies, to start the financial early warning analysis, indicating the solvency, profitability, operating ability and other important indicators of financial health. Literature [21] constructed an enterprise financial risk analysis and early warning model based on decision tree, which is used to examine the factors affecting the enterprise's finance, and is conducive to providing reference for the enterprise's financial risk early warning and risk control decisions. Literature [22] used a sample of corporate firms that initiated bankruptcy with the aim of identifying the key factors that differentiate between solvent and insolvent firms, indicating that cash mismanagement is the first sign of insolvency, and the research results provide a reference for the company's stakeholders. Literature [23] describes the global pandemic that has brought the economy into difficult times, including financial problems, supply chain disruptions and other aspects of this problem, and researches this issue, emphasizing the importance of having a plan to anticipate and respond to economic crises, verifying the effectiveness of different preventive and response measures. Literature [24] analyzes the challenges and strategies faced by early warning systems for bank insolvency by presenting the main challenges, problems and limitations of early warning systems for bank insolvency and proposing strategies that can be targeted.

This paper proposes a financial indicator system for corporate bankruptcy early warning based on the profitability and solvency of enterprises, constructs the original data matrix of financial early warning indicators based on the principal component analysis method, standardizes the original data, and then confirms the principal component factors to optimize the financial early warning indicator system by using Kaiser's method and the cumulative contribution rate method. Subsequently, the comprehensive financial early warning indicators obtained by principal component analysis are inputted into the BP neural network model, and the best number of hidden layer units of the model is determined to output the prediction results of the enterprise financial crisis, so as to realize the construction of the enterprise bankruptcy early warning model. Finally, we compare and analyze the early warning indicators of enterprises with and without bankruptcy risk to test the significance of the proposed indicators, and verify the accuracy and effectiveness of the model for corporate bankruptcy early warning through case studies.

2. Optimization method of financial indicator system based on principal component analysis

2.1. System of financial early warning indicators

The financial status of an enterprise is determined by internal and external factors, therefore, the indicators in this paper introduce corporate governance variables and environmental factors, which comprehensively reflect the profitability, solvency, operating ability and growth ability of an enterprise in order to comprehensively reflect the financial status of listed companies [25]. The enterprise financial early warning indicator system is specifically shown in Table 1. Among the secondary indicators reflecting the profitability of enterprises, four indicators are selected, namely, return on net assets (X11), return on total assets (X12), capital preservation and appreciation rate (X13), and profitability of main business income (X14). These indicators reflect the relationship between an enterprise's operating profit and its assets, equity, income and expenses, thus reflecting its profitability. Among the indicators reflecting the solvency of an enterprise, the gearing ratio indicates the ratio between total liabilities and total assets, reflecting the proportion of liabilities in the total assets of the enterprise and the capital structure of the enterprise. Current ratio refers to the proportion of current assets to current liabilities, reflecting the degree of protection of current assets to current liabilities. Among the indicators reflecting the operating ability of an enterprise, the accounts receivable turnover ratio and the current asset turnover ratio both reflect the asset turnover ability of an enterprise. The higher the value of these two indicators, the more effective the enterprise's operation and the stronger its operating ability. Among the indicators reflecting the growth capacity of the enterprise, the growth rate of total assets is the degree of growth of total assets of the enterprise in the current year compared with the previous year, reflecting the growth capacity of total assets of the enterprise. The indicator of entrepreneurial ability is a non-financial indicator selected for this paper, which indicates the comprehensive ability of the management within the enterprise, taking into account the education and work experience of the managers and other factors for scoring. Generally speaking, the larger the value of this indicator, the stronger the ability of the management of the enterprise, which is conducive to the long-term development of the enterprise, and is not easy to make the enterprise fall into a financial crisis.

Table 1. Enterprise financial warning index system

Primary indicator	Index number	Secondary indicator	Index number
Profitability	X1	Return on equity	X11
		Total asset returns	X12
		Capital preservation and value-added	X13
		Main business income profit margin	X14
Solvency	X2	Asset ratio	X21
		Mobility ratio	X22
		Equity ratio	X23
Operational capacity	X3	Receivable turnover	X31
		Turnover of current assets	X32
Growth ability	X4	Total asset growth rate	X41
		Entrepreneurial ability	X42
		Employee quality	X43

2.2. Indicator significance test methods

The basic idea of the significance test is to use the sample variable information to make a judgment on the overall parameters or overall distribution assumptions that have been made in advance, that is, to determine whether there is a significant difference between the original hypothesis and the overall real situation. In order to determine whether the selected financial indicators can play a role in identifying corporate bankruptcy distress, it is necessary to use the test based on the data of the sample to make a judgment on whether there is a significant difference between the overall mean of the empirical sample. The expression for the test is:

$$\mathrm{t}=\frac{\overline{X}-\mu_0}{S/\sqrt{\mathrm{n}}} \tag{1}$$

where μ_0 is the mean of the sample difference values, S denotes the sample variance, n is the number of paired samples, and the t-statistic obeys a T-distribution with n-1 degrees of freedom.

After standardization and dimensionlessness of the selected indicators using the software, the sample test is carried out to give the value of the t-value corresponding to the probability of companionship based on the results of the test of the T-distribution table. The process of sample testing is the process of screening the early warning variables with significant differences between the financial normal group and the financial crisis group. After the test, it prepares for the next step in the research of corporate bankruptcy early warning model.

2.3. Indicator optimization method based on principal component analysis

2.3.1. Raw Data Matrix Construction

There are n firms, each of which observes m financial early warning variables (indicators), denoted as $y_1, y_2, \cdots, y_m$. $Y_1, Y_2, \cdots, Y_m$ The matrix of raw variables is obtained as:

$$Y=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix}=(Y_1, Y_2, \cdots, Y_m) \tag{2}$$

$$Y_i = \begin{bmatrix} y_{1i} \\ y_{2i} \\ \vdots \\ y_{ni} \end{bmatrix}, i = 1, 2, \cdots, m \tag{3}$$

where Y_i represents the financial ratio of item j for company i.

2.3.2. Standardization and characterization of indicators

Before the principal component analysis [26], each financial indicator is divided into positive, negative lifting and moderate indicators, and the original data are standardized using entropy weighting method [27].

For the positive indicators, the treatment is as follows:

$$x_{ij} = \frac{y_{ij} - \min(y_{1j}, \cdots, y_{mj})}{\max(y_{1j}, \cdots, y_{mj}) - \min(y_{1j}, \cdots, y_{mj})} \tag{4}$$

For negative indicators, the treatment is as follows:

$$x_{ij} = \frac{\max(y_{1j}, \cdots, y_{mj}) - y_{ij}}{\max(y_{1j}, \cdots, y_{mj}) - \min(y_{1j}, \cdots, y_{mj})} \tag{5}$$

For moderate indicators, when $\min(y) \le y_i \le L_{1j}$:

$$x_{ij} = 1 - \frac{L_{1j} - y_{ij}}{\max\left[L_{1j} - \min(y_{1j}, \cdots, y_{mj}), \max(y_{1j}, \cdots, y_{mj}) - L_{2j}\right]} \tag{6}$$

When $L_{1j} \le y_i \le L_{2j}$:

$$x_{ij} = 1 \tag{7}$$

When $L_{2j} \le y_i \le \max(y)$:

$$x_{ij} = 1 - \frac{y_{ij} - L_{2j}}{\max\left[L_{ij} - \min(y_{1j}, \cdots, y_{nj}), \max(y_{1j}, \cdots, y_{nj}) - L_{2j}\right]} \tag{8}$$

Moderate indicators require a predetermined ideal interval $[L_{1j}, L_{2j}]$, such as the current ratio, which is set to (1.5, 2.5).

The data matrix X is obtained after normalization:

$$X = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1m} \\ Y_{21} & Y_{22} & \cdots & Y_{2m} \\ \vdots & \vdots & & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nm} \end{bmatrix} \tag{9}$$

The correlation coefficient matrix R is:

$$r_{ij} = \frac{Cov(X_i, X_j)}{\sqrt{DX_i DX_j}} = EX_i EX_j \tag{10}$$

If no normalization is used to process the relevant raw data, the correlation matrix is calculated directly. The formula is:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \tag{11}$$

$r_{ij}(i,j=1,2,\cdots,n)$ is the correlation coefficient between the original variables x_i and x_j, $r_{ij}=r_{ji}$, then:

$$r_{ij}=\frac{\sum_{k=1}^{n}(x_{ki}-\overline{x}_i)(x_{kj}-\overline{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki}-\overline{x}_i)^2\sum_{k=1}^{n}(x_{kj}-\overline{x}_j)^2}} \tag{12}$$

Calculate the eigenvalues λ and eigenvectors μ of R, since R is a positive definite matrix whose eigenvalues λ_i are all non-negative, ordered by the eigenroots λ_i, in decreasing order, i.e., $0\le\lambda_m\le\cdots\le\lambda_2\le\lambda_1$; the eigenvectors corresponding to λ_i are $\mu_i=(\mu_{1i},\mu_{2i},,\mu_{ni})^T(i=1,2,\cdots,k)$ and $\mu_i^T\mu_i=1$. Since μ is an orthogonal matrix, we have:

$$\mu_i\mu_i^r=\mu_i^r\mu_i=1 \tag{13}$$

Available for launch:

$$R=\mu\begin{bmatrix}\lambda_l & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_k\end{bmatrix}\mu^T \tag{14}$$

from equations (11)-(13):

$$\mu^T X^T X\mu=(n\text{-}1)\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_k\end{bmatrix} \tag{15}$$

Let $Z=X\mu$ then:

$$Z^T Z=(n\text{-}1)\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_k\end{bmatrix} \tag{16}$$

Therefore k principal component is:

$$\begin{Bmatrix}Z_1=X\mu_1\\ Z_2=X\mu_2\\ \cdots\\ Z_k=X\mu_k\end{Bmatrix} \tag{17}$$

Since the mean of $X_j(1\le j\le k)$ is equal to zero, the mean of principal component $z_j(l\le j\le k)$ is also equal to zero.

2.3.3. Principal Component Factor Extraction Methods

There are two commonly used methods for selecting principal component factors, Kaiser's method and cumulative contribution ratio method [28]. Kaiser's method is a test to determine the principal component trade-off by whether the eigenvalue is greater than 1. That is, if the eigenvalue is <1, it is excluded, and if the eigenvalue is >1, that principal component is listed. The cumulative contribution rate method is to make the variance of the extracted principal components to meet the cumulative combined ability to contribute more than a certain percentage, such as 70%. For the first k principal component selected based on the cumulative contribution rate, the sum of the variances of these k principal components must reach a hypothetical percentage of the total variance, indicating that they basically contain the percentage of information that all the indicators have. The eigenvalues of the principal components extracted according to Kaiser's method can be used to some extent as an indicator to evaluate the explanatory strength of the principal components, and if a certain eigenvalue is <1, it indicates that the principal component is not as strong as the original variable directly used in the explanation.

The Kaiser method and the cumulative contribution rate method are used to confirm the principal components, and the cumulative contribution rate is used to confirm the principal component factors:

$$\frac{\sum_{i=1}^{k} r_i}{\sum_{i=1}^{p} r_i} \geq a\%(\text{usual}, 75\% \leq a \leq 95\%) \tag{18}$$

If the eigenvalue of $\lambda_i \geq 1$, it is used as a principal component factor.

2.4. BP network prediction model based on principal component analysis

In this paper, a three-layer BP neural network is utilized to train the network, which provides a series of weights based on a number of samples input to the network, and after the network has been trained, it is possible to classify any new input (company) as either insolvent or non-insolvent, and the results show that the model constructed using the neural network approach has a history of accuracy and greater stability. In this study, the BF artificial neural network analysis tool is used in conjunction with the composite indicators derived from the principal component analysis method used above to solve the technical problem of the inconsistency of the data volume and to improve the accuracy of the early warning.

2.4.1. Overview of BP Neural Networks

Artificial Neural Network (ANN) model [29] is to apply the classification method of neural network to corporate bankruptcy warning. It is a parallel distributed pattern processing system developed by applying mathematical methods from the research results of neuropsychology and cognitive science. It has highly parallel computing capability, distributed storage and processing, self-organization, self-adaptation and self-learning capability and fault tolerance, and is particularly suitable for dealing with imprecise and fuzzy information processing problems that require simultaneous consideration of many factors and conditions. The method makes use of its mapping ability on the one hand and its generalization ability mainly on the other hand in studying the financial situation of enterprises, i.e., after being trained with a certain number of samples with noise, the neural network can extract the feature relationships implied by the samples. And it interpolates and extrapolates the data in the new situation to infer its properties. It is a natural nonlinear modeling process that overcomes the complexity of the traditional analytical process and the difficulty of choosing the appropriate functional form.

2.4.2. Early warning models of corporate insolvency

This part does not discuss the mathematical derivation process of the learning rule, but only gives the learning process and steps for the following design and analysis of BP networks as a reference. This paper still takes a three-layer BP network as an example to introduce the learning process and steps of BP network.

Before the introduction, the symbols and their meanings are first explained. network input vector $P_k = (a_1, a_2, \cdots a_n)$, network target vector $T_k = (y_1, y_2, \cdots y_q)$, input vector of intermediate layer units $S_k = (s_1, s_2, \cdots s_p)$. Output vector $B_k = (b_1, b_2, \cdots b_p)$, input vector of output layer units $L_k = (l_1, l_2, \cdots l_q)$, output vector $C_k = (c_1, c_2, \cdots c_q)$. connection right from input layer to intermediate layer $w_{ij}, I = 1, 2, \cdots, n, j = 1, 2, \cdots, p$, connection right from intermediate layer to output layer $v_{ji}, j = 1, 2, \cdots, p, j = 1, 2, \cdots, p$, output threshold of each unit in intermediate layer $\theta_j, j = 1, 2, \cdots, p$, output threshold of each unit in output layer $\gamma_j, j = 1, 2, \cdots, p$. Parameters $k = 1, 2, \cdots, m$.

(1) Initialization. Assign random values in the interval (-1, 1) to each connection weight w_{ij}, v_{ji}, threshold θ_j & γ_t.

(2) Randomly select a set of input and target samples $P_k = (a_1^k, a_2^k, \cdots, a_n^k)$, $T_k = (s_1^k, s_2^k, \cdots, s_p^k)$ to provide to the network.

(3) Calculate the inputs s_j of each unit in the middle layer using input samples $P_k = (a_1^k, a_2^k, \cdots, a_n^k)$, connection weights w_i, and thresholds θ_j, and then calculate the outputs b_j of each unit in the middle layer using s_j via a transfer function. Eq:

$$s_j = \sum_{i=1}^{n} w_{ij} a_i - \theta_j \quad j = 1, 2, \cdots, p \tag{19}$$

$$b_j = f(s_j) \quad j = 1, 2, \cdots, p \tag{20}$$

(4) Calculate the output L_t of each unit of the output layer using the output b_j, connection weights v_{ji} and threshold γ_t of the intermediate layer, and then calculate the response C_t of each unit of the output layer using the pass-through function.

(5) Calculate the generalized error d_t^k for each unit of the output layer using the network target vector $T_k = (y_1, y_2, \cdots y_q)$, the actual output of the network C_t:

$$d_t^k = (y_t^k - C_t) \cdot C_t \cdot (1 - C_t) \quad t = 1, 2, \cdots, q \tag{21}$$

(6) Calculate the generalized error e_j^k of each cell of the middle layer using the connection right v_{jt}, the generalized error d_t of the output layer and the output b_j of the middle layer, i.e:

$$e_j^k = \left[\sum_{t=1}^{q} d_t \cdot v_{jt} \right] \cdot b_j \cdot (1 - b_j) \tag{22}$$

(7) The generalized error d_t^k of each unit of the output layer and the output b_j of each unit of the intermediate layer are used to correct the connection right v_{jt} and the threshold γ_t, calculated as:

$$v_{jt}(N+1) = v_{jt}(N) + \alpha \cdot d_t^k \cdot b_j \tag{23}$$

$$\gamma_t(N+1) = \gamma_t(N) + \alpha \cdot d_t^k \tag{24}$$

$$t = 1, 2, ..., q \quad j = 1, 2, ..., p \quad 0 < \alpha < 1 \tag{25}$$

(8) The generalized error e_j^k of each cell in the middle layer and the input $P_k = (a_1, a_2, \cdots a_n)$ of each cell in the input layer are used to correct the connection right w_{ij} and the threshold θ_j, Eq:

$$w_{ij}(N+1) = w_{ij}(N) + \beta \cdot e_j^k \cdot a_i^k \tag{26}$$

$$\theta_j(N+1) = \theta_j(N) + \beta \cdot e_j^k \tag{27}$$

$$i = 1, 2, \cdots, n \quad j = 1, 2, \cdots, p \quad 0 < \beta < 1 \tag{28}$$

(9) Randomly select the next learning sample vector to be provided to the network and return to step (3) until m training samples are trained.

(10) Re-select a set of input and target samples randomly from the m learning samples and return to step (3) until the network global error E is less than a pre-determined a very small value, i.e., the network converges. If the number of learning times is greater than a pre-set value, the network fails to converge.

(11) End of learning.

It can be seen that in the above learning steps, (7) ~ (8) steps for the network error “back propagation process”, (9) ~ (10) steps are used to complete the training and convergence process.

The BP neural network model used in this paper, the input layer of variables for the use of principal component analysis method to obtain a comprehensive index, because the output results are 0 and 1, so the output layer needs only 1 neuron, 1 represents the bankruptcy crisis of the company, 0 on behalf of the non-bankruptcy crisis of the company.

And the choice of the number of neurons in the hidden layer is a very complex problem, which often needs to be determined based on the experience of the designer and many experiments. The number of hidden layer units is directly related to the requirements of the problem and the number of input/output units. An increase in the number of hidden layer units helps to fit the training samples, but at the same time reduces the degree of freedom of the model and weakens its generalization ability, i.e., over-training problems may occur. Therefore, there must exist an optimal number of hidden layer units, and the following formula can be used as a reference formula when choosing the optimal number of hidden layer units:

$$n_1 = \log_2 n \tag{29}$$

Where, n is the number of input units.

Combined with the above equation, the finalized number of hidden layer units is 3. Therefore, a network structure of 8*3*1 will be used, while the neuron function is a Sigmoid eigenfunction:

$$f(x) = 1/(1+e^{-x}) \tag{30}$$

The matlab 6.5 software was used, the training function was the Traingdm function, the transfer function used the logarithmic S-shaped transfer function logsig, the maximum number of iterations was set to 100,000, the target error value was set to 0.01, and the other parameters used the system default values.

3. Validation of the accuracy of early warning models for corporate insolvency

3.1. Analysis of significance tests for early warning indicators

3.1.1. Research sample

In the study of this paper, ST (specially treated) companies in listed companies are defined as enterprises with financial bankruptcy risk. 40 ST companies in the stock markets of Shanghai and Shenzhen from 2020 to the end of 2023 are selected, and another 40 non-ST companies corresponding to them totaling 80 enterprises (excluding B shares) are selected as the research sample. Among the 40 ST companies, 6 are declared as ST in 2020, 14 are declared as ST in 2021, and the other 20 are declared as ST in 2023. Given the research needs of this paper, the financial data selected are for the three years prior to the declaration of special treatment for these companies. The other 40 non-ST companies are selected according to the year in which the financial ratio data of their corresponding ST companies were selected. Now these 40 ST companies are randomly divided into two groups and another 40 non-ST companies are paired up separately and correspondingly. This makes the estimation sample group 40 (20 ST companies and 20 non-ST companies) and the test sample group also 40 (20 ST companies and 20 non-ST companies).

3.1.2. Results of the financial early warning indicator test

In order to analyze whether the indicators in the optimized financial early warning indicator system constructed in this paper can effectively differentiate between ST companies and non-ST companies, the following 80 sample enterprises respectively carry out univariate T-test on the above financial early warning indicators, and the test results of each indicator under the first-level indicator of profitability are shown in Table 2. According to the T-test results, taking return on net assets (net profit), total return on assets and capital preservation and appreciation ratio as examples, it can be found that when the number of years before being declared ST is 1, the two-tailed T-test probability of significance of ST companies and non-ST companies are 0.015, 0.048 and 0.003 respectively, which are all less than 0.05. Therefore, it can be concluded that the use of return on net assets, total return on assets and capital preservation and appreciation ratio can effectively distinguish between ST companies and non-ST companies in the distance from the ST companies. Capital Preservation and Appreciation Ratio in the first 1 year from the special treatment can significantly differentiate between ST companies and non-ST companies.

Table 2. Financial warning index test results (Profitability)

Financial warning indicators	Number of years before declaration of ST	Company type	Mean	T	P
Return on equity (X11)	1	ST Non-ST	-0.049 0.024	2.008	0.015
	2	ST Non-ST	-0.015 0.04	4.577	0.007
	3	ST Non-ST	-0.043 0.059	5.295	0.033
Total asset returns (X12)	1	ST Non-ST	-0.088 0.083	2.042	0.048
	2	ST Non-ST	-0.022 0.088	3.352	0.003
	3	ST Non-ST	-0.051 0.074	4.518	0.006
Capital preservation and value-added (X13)	1	ST Non-ST	0.004 0.01	2.750	0.003
	2	ST Non-ST	0.01 0.02	3.996	0.000
	3	ST Non-ST	0.008 0.012	3.856	0.025
Main business income profit margin (X14)	1	ST Non-ST	0.018 0.074	5.794	0.015
	2	ST Non-ST	0.03 0.063	3.193	0.031
	3	ST Non-ST	0.021 0.069	2.082	0.024

The test results of solvency indicators are shown in Table 3, and the test results of operating ability are shown in Table 4. In the significant difference test of the second level indicators under solvency, gearing ratio, current ratio and shareholders' equity ratio, there is a significant difference between the mean values of ST companies and non-ST companies in each of the indicators (P<0.05). In terms of the mean value of accounts receivable turnover indicator under operating capacity, there is a difference between the mean value of accounts receivable turnover of ST and non-ST companies in the first 1~3 years before ST companies were declared special treatment. Also the probability of significance of two-tailed t-test for the first 1~3 years is 0.016, 0.035 and 0.045, which are less than 0.05, so there is a significant difference between the accounts receivable turnover ratio of ST companies and non-ST companies.

Table 3. Financial warning index test results (Solvency)

Financial warning indicators	Number of years before declaration of ST	Company type	Mean	T	P
Asset ratio (X21)	1	ST Non-ST	0.887 0.362	5.342	0.025
	2	ST Non-ST	0.745 0.342	2.860	0.022
	3	ST Non-ST	0.703 0.364	2.804	0.044
Mobility ratio (X22)	1	ST Non-ST	0.694 1.584	5.931	0.049
	2	ST Non-ST	0.789 2.361	3.270	0.039
	3	ST Non-ST	0.885 2.654	5.157	0.042
Equity ratio (X23)	1	ST Non-ST	0.165 0.425	3.572	0.005
	2	ST Non-ST	0.189 0.562	4.814	0.031
	3	ST Non-ST	0.194 0.215	5.014	0.042

Table 4. Financial warning index test results (Operational capacity)

Financial warning indicators	Number of years before declaration of ST	Company type	Mean	T	P
Receivable turnover (X31)	1	ST	4.95	3.684	0.016
		Non-ST	12.65		
	2	ST	5.24	3.617	0.035
		Non-ST	15.45		
	3	ST	5.98	4.659	0.045
		Non-ST	14.56		
Turnover of current assets (X32)	1	ST	0.34	5.310	0.043
		Non-ST	0.59		
	2	ST	0.36	5.262	0.018
		Non-ST	0.67		
	3	ST	0.39	2.475	0.038
		Non-ST	0.68		

Table 5 shows the results of the T-test of the growth ability of the first-level indicators in the early warning indicator system of enterprise financial bankruptcy, among the three second-level indicators, there is no significant difference between the average total asset growth rate of ST companies (0.112) and non-ST companies (0.174) when the number of years before ST is declared to be 3 (T=0.526, P=0.085>0.05), but the total asset growth rate of ST companies in the first year and the first two years of ST is much lower than that of non-ST companies, and there is a significant difference (P<0.05). This indicates that the financial early warning indicators selected in this paper can effectively differentiate between ST and non-ST companies, and can be analyzed as input indicators for the corporate bankruptcy risk model based on principal component analysis.

Table 5. Financial warning index test results (Growth ability)

Financial warning indicators	Number of years before declaration of ST	Company type	Mean	T	P
Total asset growth rate (X41)	1	ST	-0.154	4.071	0.044
		Non-ST	0.162		
	2	ST	0.015	5.761	0.006
		Non-ST	0.185		
	3	ST	0.112	0.526	0.085
		Non-ST	0.174		
Entrepreneurial ability (X42)	1	ST	1.393	3.206	0.045
		Non-ST	4.347		
	2	ST	1.814	3.127	0.021
		Non-ST	4.675		
	3	ST	1.281	2.075	0.003
		Non-ST	3.523		
Employee quality (X43)	1	ST	1.000	5.367	0.039
		Non-ST	4.097		
	2	ST	1.206	2.622	0.004
		Non-ST	3.638		
	3	ST	1.292	3.020	0.019
		Non-ST	3.640		

3.2. Results of the optimization analysis of financial early warning indicators

The eigenvalues and contribution rates of each principal component were obtained using principal component analysis as shown in Table 6. According to the principle of selecting principal components with eigenvalue greater than 1 and cumulative contribution rate around 70%, the 5th eigenvalue in the table is 0.998, which is extremely close to 1, so that it has little impact on the calculation of principal component coefficients. Therefore, this paper selected 5 principal components, using 5 principal components to replace the original 12 financial early warning indicators.

Table 6. The main component characteristics value and contribution rate

Principal component	Initial value			After extraction		
	Eigenvalue	Contribution rate (%)	Cumulative contribution (%)	Eigenvalue	Contribution rate (%)	Cumulative contribution (%)
1	3.695	27.47%	27.47%	3.695	27.47%	27.47%
2	2.365	17.58%	45.06%	2.365	17.58%	45.06%
3	1.926	14.32%	59.38%	1.926	14.32%	59.38%
4	1.653	12.29%	71.67%	1.653	12.29%	71.67%
5	0.998	7.42%	79.09%	0.998	7.42%	79.09%
6	0.748	5.56%	84.65%	——	——	——
7	0.653	4.86%	89.50%	——	——	——
8	0.175	1.30%	90.80%	——	——	——
9	0.165	1.23%	92.03%	——	——	——
10	0.098	0.73%	92.76%	——	——	——
11	0.068	0.51%	93.26%	——	——	——
12	0.052	0.39%	93.65%	——	——	——

3.3. Bankruptcy Early Warning Model Test Analysis

In this paper, according to the optimized and validated financial warning indicators and combined with the training samples as a training set, the corresponding target output results are shown in Table 7, the node target output values of ST companies and non-ST companies in the bankruptcy risk early warning model should be 1 and 0. It is found by the model node actual output value that one of the 20 ST companies (No. 7, with the node actual output value of 0.5294) cannot be judgment, the actual output value of the node of the ST company numbered 20 is 0.0654, indicating that it is misjudged as a non-ST company. One of the non-ST companies is misjudged, and the actual node output value is 0.8591. After calculation, the accuracy rate of bankruptcy risk early warning for ST and non-ST companies is 90% and 95% respectively, which indicates that the financial early warning index system and early warning model proposed in this paper are able to accurately predict corporate bankruptcy risk and make early warning.

Table 7. Financial bankruptcy model output

Number	Node target output	Node actual output	Number	Node target output	Node actual output
1	1	0.9764	1	0	0.0444
2	1	0.9914	2	0	0.0429
3	1	0.9781	3	0	0.0336
4	1	0.9741	4	0	0.0383
5	1	0.9838	5	0	0.0244
6	1	0.9945	6	0	0.0488
7	1	**0.5294**	7	0	0.0232
8	1	0.9743	8	0	0.0379
9	1	0.9876	9	0	0.0174
10	1	0.9894	10	0	0.0071
11	1	0.9904	11	0	0.0296
12	1	0.9932	12	0	0.0342
13	1	0.9806	13	0	0.0390
14	1	0.9775	14	0	**0.8591**
15	1	0.9746	15	0	0.0135
16	1	0.9892	16	0	0.0428
17	1	0.9738	17	0	0.0388
18	1	0.9972	18	0	0.0096
19	1	0.9921	19	0	0.0437
20	1	**0.0654**	20	0	0.0454

4. Results of the analysis of empirical applications by enterprises

4.1. Case background

The predecessor of XX Electronics Co., Ltd. was Electronics Co., Ltd. established in 1987. The company was founded by the former Electronics Co., Ltd. and other six enterprises in May 1997, with its net assets, cash discount 69.5 million shares of legal persons and state-owned legal persons 4.8 million shares, after the issue of July 1998, the total share capital of the listing of 125 million shares. XX Electronics Co., Ltd. has a wide range of business scope, including the sound and image of the electronic products and other mechanical and electronic product development, manufacturing, real estate development and operation, wholesale and retail construction materials, department stores, hardware and electricity, packaging materials, etc. , real estate development and operation, wholesale and retail construction materials, department stores, hardware and electric appliances, packaging materials, etc.. The most ideal cycle for observing the financial situation of a company is 5 years, so the financial data from 2018 to 2022 will be selected for analysis in this paper. XX Company was classified as a special treatment enterprise in May 2018. After 3 consecutive years of losses, the enterprise successfully completed the transformation of its main business into the direction of mobile communication products at the beginning of 2021, and realized the reversal of losses and gained abundant profits at one stroke through the adjustment of industrial structure. After the main business of the company entered the field of mobile communication products, with the successful positioning in the market of medium and high-end cell phones and the core technology centered on application development as the focus of development, the company's cell phones stood out among many brands in terms of style, function, signal, environmental protection and personalization.

4.2. Early warning analysis of bankruptcy of case companies

4.2.1. Model early warning results

This paper selects the data of XX company from 2018 to 2022 to be brought into the enterprise bankruptcy early warning model to discriminate, and the output of the model for early warning analysis is shown in Table 8.The output values of the bankruptcy early warning model of XX Electronics Company Limited from 2018 to 2020 (0.9697, 0.9474, and 0.9509) are all close to 1, which indicates that the enterprise has a poorer financial condition during these three years , there is a risk of bankruptcy, which should be worth to let the enterprise related personnel pay more attention to. From 2021, the output value of bankruptcy warning of XX Electronics Co., Ltd. is 0.0307, which is close to 0, indicating that the company's financial situation is more favorable and there is no bankruptcy risk.

Table 8. XX company bankruptcy warning results

Year	2018	2019	2020	2021	2022
X1	0.9399	0.9598	0.9059	0.0565	0.0386
X2	0.9856	0.9653	0.9642	0.0054	0.0477
X3	0.9851	0.9534	0.9382	0.0259	0.0081
X4	0.9680	0.9109	0.9953	0.0348	0.0476
Y	0.9697	0.9474	0.9509	0.0307	0.0355

4.2.2. Cash flow early warning results

The basic condition for the survival of the enterprise is to have sufficient cash flow, so this paper will analyze the financial situation of XX Electronic Enterprise from 2018 to 2022, and then compare the output of the above model to explore the effectiveness of the bankruptcy early warning model.XX Electronic Enterprise in the period from 2018 to 2022, the net cash flow from each economic activity is shown in Table 9.The net cash flow from operating and investing activities is negative, and the net cash from financing activities is positive, indicating that the company mainly relies on borrowing to maintain production and operation relative to the previous year, and there is insufficient cash. year has negative net cash from operating and investing activities and positive net cash from financing activities, indicating that it mainly relies on borrowing to maintain its production and operation, and the company's financial position continues to deteriorate relative to the previous year, and it is short of cash.In 2020, although the company's net cash from operating activities (524,635,100 yuan) rises to a positive value, its net cash from financing activities (-485,163,000 yuan) relative to 2018 (195.6842 million yuan) declined substantially, and it can be concluded that at this time the cash generated from the company's operating activities is mainly used to repay debts, and the company's outlook is still not optimistic.In 2021, the company's net cash from operating activities (125.4863 million yuan) is positive, with a relatively large increase compared to 2020, as it successfully completes the transformation of its main business in the direction of mobile communication products, and through the adjustment of industrial structure to turn around losses and gain substantial profits in one fell swoop. Net cash from investing and financing activities) is negative, indicating that the company's operating situation is good and the cash is more abundant, on the one hand, repaying previous debts, on the other hand, continue to invest. Finally, combined with the output results of the early warning model, we find that the model of this paper has high validity and reliability in the empirical analysis of XX Electronics Co.

Table 9. Detailed analysis of cash clearance (10,000yuan)

Year	2018	2019	2020	2021	2022
Net operating cash	-2648.54	-1542.69	52463.51	65263.25	99856.54
Investment activity	-3265.34	-3954.68	-3015.26	-2956.36	-2145.85
Financing activities	19568.42	3956.54	-4851.63	-2154.63	-1558.95
Net cash gain	-7845.62	-9856.35	-9975.95	12548.63	49563.85

5. Conclusion

In this paper, after designing the enterprise financial early warning indicators including profitability and solvency, we construct the enterprise financial and bankruptcy risk early warning model based on principal component analysis and BP neural network model. We analyze whether the indicators in the optimized financial early warning indicator system can effectively distinguish between ST (special treatment) companies and non-ST companies, and find that the indicators of profitability, solvency, operating ability and growth ability can significantly distinguish between ST companies and non-ST companies. For the 40 enterprises in the test set, it is found that the accuracy of the model in early warning of bankruptcy risk for ST and non-ST companies is 90% and 95%, respectively. In addition, the application of case examples found that the model of this paper has high validity and reliability in the financial and bankruptcy early warning of XX Electronics Co.

To summarize, this study proposes a set of practical financial crisis early warning methods, which are not only conducive to the control of financial crises by listed companies, but also serve as a bankruptcy early warning.

References

1. Durst, S., & Zieba, M. (2020). Knowledge risks inherent in business sustainability. Journal of cleaner production, 251, 119670.
2. Dvorsky, J., Belas, J., Gavurova, B., & Brabenec, T. (2021). Business risk management in the context of small and medium-sized enterprises. Economic Research-Ekonomska Istraživanja, 34(1), 1690-1708.
3. Bărbuță-Mișu, N., & Madaleno, M. (2020). Assessment of bankruptcy risk of large companies: European countries evolution analysis. Journal of risk and financial management, 13(3), 58.
4. Macey, J., & Salovaara, J. (2019). Bankruptcy as bailout: coal company insolvency and the erosion of federal law. Stan. L. Rev., 71, 879.
5. Voda, A. D., Dobrotă, G., Țîrcă, D. M., Dumitrașcu, D. D., & Dobrotă, D. (2021). Corporate bankruptcy and insolvency prediction model. Technological and Economic Development of Economy, 27(5), 1039-1056.
6. Horváthová, J., & Mokrišová, M. (2018). Risk of bankruptcy, its determinants and models. Risks, 6(4), 117.
7. Karas, M., & Reznakova, M. (2017). Predicting the bankruptcy of construction companies: a CART-based model. Engineering Economics, 28(2), 145-154.
8. Zhang, W. (2017). Machine learning approaches to predicting company bankruptcy. Journal of Financial Risk Management, 6(04), 364.
9. Klopotan, I., Zoroja, J., & Meško, M. (2018). Early warning system in business, finance, and economics: Bibliometric and topic analysis. International Journal of Engineering Business Management, 10, 1847979018797013.
10. Ying, W., Mayburov, I. A., & Leontyeva, Y. V. (2024). Assessing the Bankruptcy Risks of China's Emerging Port Industries: Modeling and Early Warning. Journal of Applied Economic Research, 23(3), 776-800.

11. Zhang, C., Zhou, X., & Wang, J. (2022). A Financial Risk Early Warning of Listed Companies Based on PCA and BP Neural Network. Mobile Information Systems, 2022(1), 8320329.
12. Zavalko, N. A., Kozhina, V. O., Kovaleva, O. P., Kolupaev, R. V., & Lebedeva, O. Y. (2018). System approach to diagnostics and early prevention of a financial crisis at an enterprise. Journal of Applied Economic Sciences, 13(1).
13. Kurita, T. (2021). Principal component analysis (PCA). In Computer vision: a reference guide (pp. 1013-1016). Cham: Springer International Publishing.
14. Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., ... & Laishram, M. (2017). Multivariate statistical data analysis-principal component analysis (PCA). International Journal of Livestock Research, 7(5), 60-78.
15. Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. Nature Reviews Methods Primers, 2(1), 100.
16. Kherif, F., & Latypova, A. (2020). Principal component analysis. In Machine learning (pp. 209-225). Academic Press.
17. Kitowski, J., Kowal-Pawul, A., & Lichota, W. (2022). Identifying symptoms of bankruptcy risk based on bankruptcy prediction models—A case study of Poland. Sustainability, 14(3), 1416.
18. Yuxia, S., Mei, S., Congyuan, Y., & Zhiya, L. (2024). An empirical study of micro‐and small‐enterprise bankruptcy protection under the COVID‐19 pandemic: New evidence from China. International Insolvency Review, 33(2), 231-262.
19. He, X. (2021, April). New Energy Enterprises Financial Risk Early Warning Model Based on the Comparison of Principal Component and Logistic Regression Model Analysis. In Proceedings of the 6th International Conference on Information and Education Innovations (pp. 125-134).
20. Yang, L., Yang, M., & Xu, Y. (2023, February). An Early Warning Analysis of Pharmaceutical Listed Companies: Based on PCA and Factor Analysis Method. In International Conference on Business and Policy Studies (pp. 1749-1759). Singapore: Springer Nature Singapore.
21. Liao, S., & Liu, Z. (2022). Enterprise financial influencing factors and early warning based on decision tree model. Scientific Programming, 2022(1), 6260809.
22. Kušter, D., Vuković, B., Milutinović, S., Peštović, K., Tica, T., & Jakšić, D. (2023). Early insolvency prediction as a key for sustainable business growth. Sustainability, 15(21), 15304.
23. Ling, B. (2025, February). Economic Crisis Early Warning Signs and Response Strategies for Corporate Bankruptcy Prevention. In International Workshop on Navigating the Digital Business Frontier for Sustainable Financial Innovation (ICDEBA 2024) (pp. 482-493). Atlantis Press.
24. Keshavarz Hadadha, H., Alipour, H., & Kheradyar, S. (2024). Analysis of Challenges and Strategies to Develop More Effective Early Warning System of Bank Bankruptcy. Environmental Energy and Economic Research, 8(1).
25. Nguyen Thi Mai Anh,Nguyen Thanh An,Nguyen Thi Minh Ngoc & Vu Ngoc Xuan. (2024). The Influence of Financial Indicators on Vietnamese Enterprise's Sustainability Reports Disclosing Process. Journal of Risk and Financial Management(4),146-.
26. Lingbin Shen,Liping Tian,Hongbing Yao,Dongpeng Tian,Yifan Ge,Zhongmou Sun & Yuzhu Liu. (2024). Rapid detection and recognition of phosphors using laser-induced breakdown spectroscopy and principal component analysis method—back propagation neural network algorithm. Laser Physics(5).
27. Sun Runfang,Yan Qixiang,Zhang Chuan,Qiao Minjie & Ren Junnan. (2025). Comprehensive Evaluation of Grouting Effectiveness Combining Qualitative On-Site Tests and Improved Fuzzy Integration with Entropy Weight Method: Case Study of a Mountain Tunnel. Journal of Construction Engineering and Management(4).

28. Lili Li,Liangliang Zhang,Gang Liu,Qing Li & Xing An. (2019). Sensor fault localization with accumulated residual contribution rate for bridge SHM. Cluster Computing: The Journal of Networks, Software Tools and Applications(4s),7821-7829.
29. Isaac Oyeyemi Olayode,Alessandro Severino,Frimpong Justice Alex & Elmira Jamei. (2025). Traffic flow modelling of vehicles on a six lane freeway: Comparative analysis of improved group method of data handling and artificial neural network model. Results in Engineering104094-104094.