

## Research Article

# A Financial Risk Early Warning of Listed Companies Based on PCA and BP Neural Network

Chen Zhang <sup>1</sup>, Xinmiao Zhou,<sup>1</sup> and Jiaqing Wang <sup>2</sup>

<sup>1</sup>Ningbo University, Ningbo 315211, China

<sup>2</sup>Ningbo University of Finance & Economics, Ningbo 315175, China

Correspondence should be addressed to Chen Zhang; 2101010001@nbu.edu.cn

Received 11 May 2022; Revised 2 October 2022; Accepted 20 October 2022; Published 29 October 2022

Academic Editor: Nicola Biccocchi

Copyright © 2022 Chen zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Financial risk, as one of the most influential and destructive risks in business, will make enterprises unable to escape the fate of bankruptcy if not warned and prevented in time. In the paper, we conducted research on the financial risk early warning of listed companies. A total of 250 companies were randomly selected from the Chinese A-share market from 2019 to 2021. By building the 26 financial indicators of listed companies and constructing the PCA-BP neural network, we compared the financial risk early warning effects among PCA-BPNN, SVM, and Logistic. It is found that the financial data processed by PCA can better adapt to the financial risk early warning model. The PCA-BPNN model improved the prediction accuracy of the financial risk early warning, which has strong generalization ability for the prediction of financial risk. Research findings have certain reference significance for precise judgment on the financial risk of companies.

## 1. Introduction

The good operation of the capital market cannot be done without open, transparent, and true information disclosure. Financial information is an important basis for the business activities and future development of enterprises. The quality of financial information will also have a great influence on the operation and development of enterprises. Therefore, enterprises should pay attention to the quality of financial information and strengthen the management of financial information through effective means. In the complex market economy environment, every decision of the manager can be regarded as a turning point in the operation of the enterprise. At present, due to the downturn in the global economic market and the intensified competition among countries, the market volatility is particularly severe. The uncertainty of the macro environment has reduced the financial growth ability of enterprises [1]. In addition, because of the influence of comprehensive factors such as the epidemic, the business and financial risks faced by enterprises are also increasing simultaneously [2]. Because of this, once a company has financial problems, it may whitewash the

financial data, and the result of accumulating financial risks will cause the company to eventually go bankrupt. According to a report from the China Securities Regulatory Commission [3], financial fraud has become a “hardest hit area” in 20 typical illegal cases in 2021. These companies have inflated huge income through fictitious purchases and sales, compiled false financial book sets, etc., and provided users with wrong financial statements. Wrong market information not only deceived investors but also made financial information users make wrong judgments about the national macroeconomic situation, which seriously affected the normal social and economic order.

Therefore, reducing the financial risk in the operation of enterprises is a crucial basis to ensure their stable development. However, enterprise financial risk management is a continuous, loop, and dynamic process throughout the enterprise decision-making and management of each link. Figure 1 shows the enterprise financial risk management framework, including the source of enterprise financial risk, financial risk early warning, financial risk assessment, financial risk response, supervision, and inspection. The wrong identification of risk will lead to the subsequent

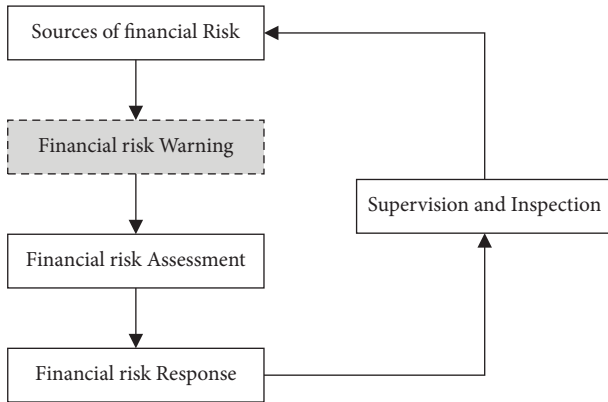


FIGURE 1: Financial risk management framework.

assessment and response plan being wasted, so the early warning and identification of financial risk is the most important link.

The purpose of this paper is to build an effective financial risk early warning system, extract valuable financial information, and use the model proposed in this paper to identify enterprises with potential financial risks. Through empirical analysis and comparison with existing models, the effectiveness of our proposed model is proven.

## 2. Related Work

Financial risk mainly refers to the loss of an enterprise due to various risk factors, or the uncertainty of its operating efficiency [4]. Serious financial risk will have a deep impact on the development direction of an enterprise and the realization of its development goals. From the point of view of the current business situation, a large amount of information and data related to finance will be generated every day, and the generation of this information and data provides the possibility for financial risk prediction. However, in fact, in order to prevent delisting due to continuous losses, enterprises in a financial crisis have the motivation to whitewash their financial data, so the financial statements of most listed companies with potential financial risks are often suspected to be false [5–7]. The more serious the deterioration of the financial condition of listed companies, the stronger the motivation to manipulate financial data. The identification of the quality of these financial data requires the establishment of effective mechanisms. Financial indicators are the core basis of an enterprise’s financial risk early warning indicator system [8], and the authenticity of financial data is a key factor in determining the predictive ability of an enterprise’s financial risk early warning model. Therefore, the construction of an index system and an early warning model should comprehensively consider financial indicators and their data quality [9].

With the in-depth study of machine learning methods, more and more scholars apply related methods in machine learning to financial risk management. Qi et al. used the random forest method to analyze Internet finance and try to find the factors that affect the risk of Internet finance [10]. Liu et al. build a data-driven approach calibrated neural

network to enable artificial neural networks to calibrate financial asset price [11]. Wu and Wu used genetic algorithms, neural network, and PCA to process data and build a risk assessment model for extracting relevant hidden information from financial enterprises [12]. Wang builds a supply chain financial system through blockchain technology, uses fuzzy neural network algorithms for financial data processing and risk assessment [13]. Tavana et al. used ANN and Bayesian network to measure liquidity risk [14]. Du et al. used the BP neural network algorithm to measure Internet credit risk early warning [15].

In the study field of financial early warning, Huang et al. used the logistic regression method to predict and supervise the systemic risks of China’s financial system, trying to find ways to improve the financial status indicators to measure the degree of fiscal austerity [16]. Since the financial risk of a business is a complex and gradual process, its reasons may be manifold. Businesses face financial risks or difficulties, and there are more and more bankruptcy liquidations. Financial risks have seriously affected businesses and society.

The era of big data has spawned various algorithms for financial data. Based on deep learning algorithms, Cao et al. established a financial early warning model and conducted research from the perspective of building a financial risk early warning mechanism for e-commerce companies [17]. Wei takes manufacturing companies that issue bonds as samples to build a financial early warning model based on decision tree integration, which can effectively improve the correct identification rate of companies in a financial crisis [18]. A big data mining method based on PSO-BPNN for financial risk management of IoT deployment in commercial banks by Zhou and Zhou et al. [19, 20], The method utilizes Apache Spark and Hadoop HDFS technology to build a Nonlinear Parallel Optimization Model. Huang et al. compared the application effects of several common neural network models in the Chinese SME data set to explore enterprise credit risk factors [21]. Shen et al. used an ensemble model based on synthetic minority oversampling technique (SMOTE) and classifier optimization technique for firm’s credit risk assessment [22].

In view of the logical relationship between the financial work in the enterprise, the early-warning model can provide more financial decisions and data support for the managers [23], locate the front-end business problems in advance, and avoid the expansion of the financial crisis in the enterprise and bring greater losses to the company [24, 25]. The existing literature has done some research on enterprise financial risk, but the research on corporate financial risk early warning is not enough. Considering the possibility of manipulation of financial data of listed enterprises and the multicollinearity of data, direct use of these data will inevitably increase noise, which will not only bring dimensional disaster to the operation but also affect the accuracy of financial prediction of enterprises. This paper introduces PCA method into the evaluation of financial data quality, constructs financial indicators and extracts principal components of financial data, introduces BP neural network model, and adopts hybrid PCA-BP neural network model to study the financial early warning of listed companies.

### 3. Model Design and Data Processing

**3.1. Risk Indicator Selection.** The construction of the financial risk early warning system will realize the monitoring of the operation process of the enterprise and the activities of the business process [26]. Through the analysis of early warning indicators, we can get an in-depth understanding of the current business situation of the enterprise. Therefore, in the construction of a financial risk early warning model, it is necessary to select the financial indicators that can reflect the running status of enterprises as explanatory variables. In the preliminary selection of indicators, these indicators can not only comprehensively reflect the operational state of the enterprise but also reflect the financial state of the enterprise. Moreover, all of those selected indicators must be accessible with the corresponding financial data. In addition, much of the corporate financial risk early warning literature often chooses financial ratio indicators as explanatory variables. Referring to the existing research literature [27–29], this paper builds 26 secondary financial indicators from the five aspects of profitability, solvency, development capability, operating ability, and cash flow to establish an enterprise financial risk early warning indicator system.

The analysis is carried out on the basis of the five main indicators. In view of the following reasons:

- (1) **Profitability:** Profitability refers to the ability of an enterprise to obtain profits. Profit is a long-term and stable source of funds for an enterprise, and good profitability can help an enterprise better defend against financial risks. Indicators such as return on equity, return on assets, and operating profit ratio are all representative indicators of profitability.
- (2) **Solvency:** Solvency refers to the ability of an enterprise to repay its debts. Low solvency can easily make an enterprise fall into financial crisis or even go bankrupt. Therefore, the measure of solvency directly reveals the size of the financial risk of the enterprise. Indicators such as current ratio and quick ratio reflect the company's ability to repay debts through the realization of assets.
- (3) **Development Capability:** Development capability refers to the development trend and development potential of an enterprise's future production and operation activities. It is usually evaluated from the perspective of the company's financial status and operating results. The main representative indicators are the net profit margin ratio, net asset growth rate, and so on.
- (4) **Operational ability:** Operational ability refers to the ability of an enterprise to manage and use inventory, accounts receivable, fixed assets, total assets, and other assets for turnover operations to earn profits. The unbalanced financial status of enterprises due to poor operational management will increase the financial risks faced by enterprises.
- (5) **Cash flow:** Cash is the most liquid asset among all the assets of an enterprise, and it is also the basis for an

enterprise to maintain its daily operating and production activities. The shortage of cash flow will directly affect the production and operation of the enterprise and easily lead to a financial crisis.

The selected detailed financial indicators are defined as shown in Table 1.

**3.2. Data Processing.** Due to the different nature of the indicators, normalization is performed before the indicator data is input into the BP neural network model to unify the parameter range for subsequent modeling and analysis. The specific operation formula is as follows:

$$k_{ij}' = \frac{k_{ij} - k_i^{\min}}{k_i^{\max} - k_i^{\min}}. \quad (1)$$

In the formula,  $k_{ij}'$  is the  $j$ th financial profitability indicator of the  $i$  listed company;  $k_{ij}^{\min}$  is the minimum value of the financial company's risk indicator data.  $k_{ij}^{\max}$  is the maximum value of the risk index data of the financial company.

**3.3. Principal Component Analysis.** Financial indicators include the items in Table 1. While the variable indicators are strictly selected, the independence of each input variable cannot be guaranteed. PCA is used to reduce the dimension of the input variables, the new financial variables after the dimension reduction are used to replace the original multiple variables, which can basically contain all the contents of the original input variables. The process is as follows.

Step 1: Calculate the covariance matrix of the original input financial variable data set, assuming that there are  $n$  historical load data in the original input variable data set  $X$ , and each data has  $p$  influencing factors, that is  $X = [X_1, X_2, \dots, X_p]$ , where  $X_i = [x_{1i}, x_{2i}, \dots, x_{ni}]$ . Then the covariance matrix of the input variable dataset can be expressed as  $C_{p \times p} = \sum i, j = 1 ps_{ij}$ ,  $n$  is a natural number;  $\bar{x}$  is the average of  $x$ , where  $s_{ij} = \sum k = 1n(x_{ki} - \bar{x}_j)/n - 1; i = 1, 2, \dots, p; j = 1, 2, \dots, p$ .

Step 2: Calculate the eigenvalues of the covariance matrix  $\lambda_i$  and the corresponding orthogonalized eigenvector  $a_i$ , the  $i$ th principal component of the original variable  $F_i$  is  $F_i = a_i X$ , In the formula,  $a_i = \lambda / \sum i = 1m\lambda_i$

Step 3: Determine the number of principal components  $m$ , then the variance contribution rate and the cumulative variance contribution rate are respectively

$$\left\{ \eta_i = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k} \eta_{\text{total}} = \sum_{i=1}^m \eta_i. \quad (2) \right.$$

Calculate the principal component load factor  $l$  as  $l(T_i, X_j) = \sqrt{\lambda_i} a_{ij}$ , and then calculate the score of each influencing factor on  $m$  principal components as

TABLE 1: The detail of selected financial indicators.

Types of index	Indicator name	Definition	Calculation	Variable symbol
Label	ST	The company has financial problems and the stock is marked for special treatment (ST)	—	Yes or no
Profitability	Net fixed assets	A metric that evaluates the net value of a company's fixed assets	Total fixed asset purchase price + capital improvements)–(accumulated depreciation + fixed asset liabilities)	X1
	Operating profit	Its total earnings from its core business functions for a given period	Revenue—operating costs—cost of goods sold (COGS)—other day-to-day expenses	X2
	Return on total assets (ROTA)	A ratio that measures a company's earnings before interest and taxes relative to its total net assets	Ebit/average total assets	X3
	Operating profit margin	A performance ratio that reflects the percentage of profit a company produces from its operations	Operating income/net sales revenue	X4
	Cost to sales ratio	A financial ratio that compares a company's expenses generated by sales activity to its revenue	Cost of sales/total revenue	X5
	Return on equity (ROE)	This index reflects the return level of shareholders' equity and is used to measure the efficiency of the company's own capital	Net income/shareholders' equity	X6
	Gross profit margin	A measure of profitability that shows the percentage of revenue that exceeds the cost of goods sold	(Net sales—cost of goods sold)/net sales	X7
	Cost rate	The ratio between the cost of using funds and the amount effectively raised	Total cost/total revenue	X8
Solvency	Current ratio (CR)	A liquidity ratio that measures a company's ability to pay short-term obligations or those due within one year	Current assets/current liabilities	X9
	Quick ratio (QR)	A measure of a company's ability to meet its short-term obligations with its most liquid assets	(Current assets—inventories—prepaid expenses)/current liabilities	X10
	Cash ratio	A liquidity measure that shows a company's ability to cover its short-term obligations using only cash and cash equivalents	Cash and cash equivalents/current liabilities	X11
	Debt-to-assets ratio	Showing how much of a business is owned by creditors compared with how much of the company's assets are owned by shareholders	Total debt/total assets	X12
	Capitalization ratio	A measure of the proportion of debt in a company's capital base, its funds from lenders and stockholders	Total debt/capitalization	X13
	Liquidation value ratio	The ratio of tangible assets to liabilities reflects the company's ability to pay off all debts	(Liquidation value of assets—book value of liabilities)/total liabilities	X14

TABLE 1: Continued.

Types of index	Indicator name	Definition	Calculation	Variable symbol
Development capability	Main business's increasing rate of income	A measure of the company's product life cycle and judge the stage of the company's development	(Current period revenue—prior period revenue)/prior period revenue	X15
	Net profit margin ratio	An index used to evaluate the quality of net income	Net profit/revenue	X16
	Net assets growth rate	A ratio reflects the expansion speed of the capital scale of the enterprise	(Current period net assets—prior period net assets)/prior period net assets	X17
	Total assets growth rate	An index to analyze the capital accumulation ability and development ability of enterprises in that year	Asset growth/total assets	X18
Cash flow	Cash flow from operating activities	Cash flows generated by all transactions and events other than investment activities and financing activities of the enterprise	Funds from operations + changes in working capital	X19
	Current assets	Current assets are all the assets of a company that are expected to be sold or used as a result of standard business operations over the next year	Current assets are all the assets of a company that are expected to be sold or used as a result of standard business operations over the next year	X20
	Current liabilities	Current liabilities are a company's debts or obligations that are due to be paid to creditors within one year	Current liabilities are a company's debts or obligations that are due to be paid to creditors within one year	X21
	Cash flow to sales ratio	Assets that can be realized or utilized by an enterprise within an operating cycle of one year or more	Operating cashflow/net sales	X22
	Return on assets (ROA)	A measure of net profit generated per unit of assets	Net income/average total assets	X23
Operating capacity	Operating cash flow to net profit ratio	A measure of the current profitability of a business	Operating cash flow/net profit	X24
	Operating cash flow ratio	The value of cash flow compared to other item	Operating cash flow/current liabilities	X25
	Closing balance	The total in an account at the end of a reporting period	Net cash flow + opening balance	X26

$$F_i = a_{1i}X_1 + a_{2i}X_2 + \dots + a_{pi}X_p, \quad (3)$$

where  $T_i$  is the principal component matrix;  $a_{ij}$  is the eigenvector corresponding to the eigenvalues of the covariance matrix. Usually, when the cumulative variance contribution rate reaches 90%, the corresponding first  $m$  principal components can contain most of the information in the original load impact data set; therefore, the number of principal components is  $m (m < p)$

**3.4. BP Neural Network.** The BP neural network is trained by the error backpropagation algorithm, which is the most widely used neural network model, The structure diagram of BPNN is shown in Figure 2.

A single hidden layer BP neural network consists of an input layer, a hidden layer and an output layer. Let the training data set be  $D = \{(\mathbf{x}_n, \mathbf{d}_n), 1 \leq n \leq N\}$ .

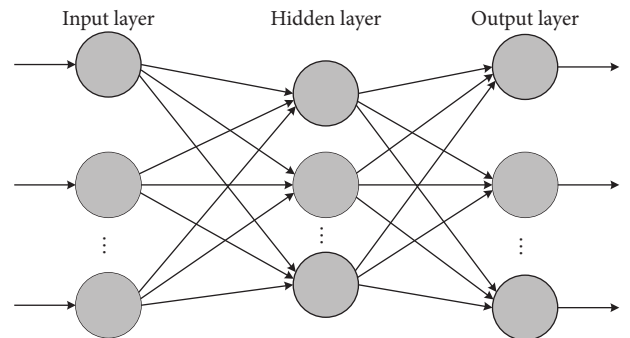


FIGURE 2: BP neural network structure.

The input layer has  $m + 1$  neuron nodes, and the input vector is a  $m + 1$  dimension vector considering the bias, that is,  $\mathbf{x}_n = (x_{n0}, x_{n1}, x_{n2}, \dots, x_{nm})^T$ , where  $x_{n0} = 1, n = 1, 2, \dots, N$ . The hidden layer contains  $L$  nodes,  $h_i$  Represents the weight

between the neurons  $l$  and  $x_{ni}$  connecting the hidden layer,  $= 0, 1, 2, \dots, m$ .  $\varphi(g)$  is the activation function of the hidden layer neurons.

$$z_{nl} = \varphi \left( \sum_{i=0}^m h_{li} x_{ni} \right), l = 1, 2, \dots, L, \quad (4)$$

is the output of the hidden layer. The output layer contains  $K$  nodes, and the input vector of the output layer is  $z_n = (1, z_{n1}, z_{n2}, \dots, z_{nL})^T$ ,  $w_{kj}$  represents the weight between the neurons  $k$  and  $z_{nj}$  connecting the output layer,  $i = 0, 1, 2, \dots, m$ .  $\varphi(g)$  represents the activation function of the output layer,

$$y_{nk} = f \left( \sum_{l=0}^L w_{kl} z_{nl} \right), k = 1, 2, \dots, K. \quad (5)$$

The error backpropagation algorithm used to train the above model consists of two processes of forward propagation of the input signal and backpropagation of the error signal. Let the training data set at step  $t$  be

$$D_t = \{(x_n(t), d_n(t)), 1 \leq n \leq N\}, \quad (6)$$

where  $x_n(t) = [1, x_{n1}(t), x_{n2}(t), \dots, x_{nm}(t)]^T$

- (1) Signal forward propagation process. The input signal is first input to the input layer, then passes through the hidden layer, and finally reaches the output layer. Calculate the input and output signals of the hidden layer and the output layer respectively, set the activation function  $\varphi(g)$  of the neurons in the hidden layer, and the activation function  $f(g)$  of the output layer, then:

The input vector of the hidden layer:  $\mathbf{x}_n(t) = [1, x_{n1}(t), x_{n2}(t), \dots, x_{nm}(t)]^T, n = 1, 2, \dots, N$

The input signal of the hidden layer neuron  $l$ :

$$\alpha_{nl}(t) = \sum_{i=0}^m h_{li}(t) x_{ni}(t), l = 1, 2, \dots, L. \quad (7)$$

The output signal of the hidden layer neuron  $l$ :

$$z_{nl}(t) = \varphi(\alpha_{nl}(t)), l = 1, 2, \dots, L. \quad (8)$$

The input vector of the output layer is  $z_n(t) = [1, z_{n1}(t), z_{n2}(t), \dots, z_{nL}(t)]^T, n = 1, 2, \dots, N$

The input signal of the output layer neuron  $k$ :

$$\beta_{nk}(t) = \sum_{j=0}^L w_{kj}(t) z_{nj}(t), k = 1, 2, \dots, K. \quad (9)$$

The output signal of the output layer neuron  $k$ :

$$y_{nk}(t) = f[\beta_{nk}(t)], k = 1, 2, \dots, K. \quad (10)$$

- (2) Error backpropagation process. From the output layer, the error of each layer is calculated in reverse layer by layer, and the weight of each layer is updated according to the gradient descent method, so that the

actual output of the network is as close to the expected output as possible.

First let  $d_n(t) = [d_{n1}(t), d_{n2}(t), \dots, d_{nK}(t)]^T$  represents the expected output of  $K$  neurons in the output layer, and

$$\mathbf{y}_n(t) = [y_{n1}(t), y_{n2}(t), \dots, y_{nK}(t)]^T, \quad (11)$$

represents the actual output of  $K$  neurons in the output layer, then the  $n$  difference for each training sample is:

$$E_n(t) = \frac{1}{2} \sum_{k=1}^K d_{nk} [(t) - y_{nk}(t)]^2, n = 1, 2, \dots, N. \quad (12)$$

We can summarize the training process of the error backpropagation algorithm as follows.

Step 1: Initialize the network weight to a small random number,  $t = 0$ .

Step 2: Enter the training data for step  $t$ :

- (a) Forward propagation of the signal, calculating the output of each neuron layer by layer
- (b) Error backpropagation, adjusting the weights of the output layer and the hidden layer

Step 3: Repeat step 2 until the error of the training data set is less than the preset threshold or reaches the maximum number of iterations.

**3.5. Evaluation Indicators.** In order to evaluate the prediction accuracy of the prediction model proposed in this paper, the root mean squared error (RMSE) and the confusion matrix are selected as evaluation indicators. The calculation formula is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_f - y_r)^2}, \quad (13)$$

where  $y_f$  is the predicted value;  $y_r$  is the real value

Confusion Matrix is a two-way matrix where one axis shows the distribution of the actual classes and the other axis shows the predicted classes. The schematic diagram of confusion matrix is shown in Table 2.

The accuracy of the model is calculated based on the diagonal elements of the classification matrix. Because they represent the correct classification made by the classification model, the following formula shows the calculation of model accuracy.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}. \quad (14)$$

The flow chart of the financial early warning mechanism can be obtained as shown in Figure 3. It is mainly divided into three stages.

*Phase 1:* Randomly select the financial data of China A-share listed companies marked with and without

TABLE 2: Confusion matrix.

		Prediction category	
		Yes	No
Actual Category	Yes	True positive (TP)	False negative (FN)
	No	False positive (FP)	True negative (TN)

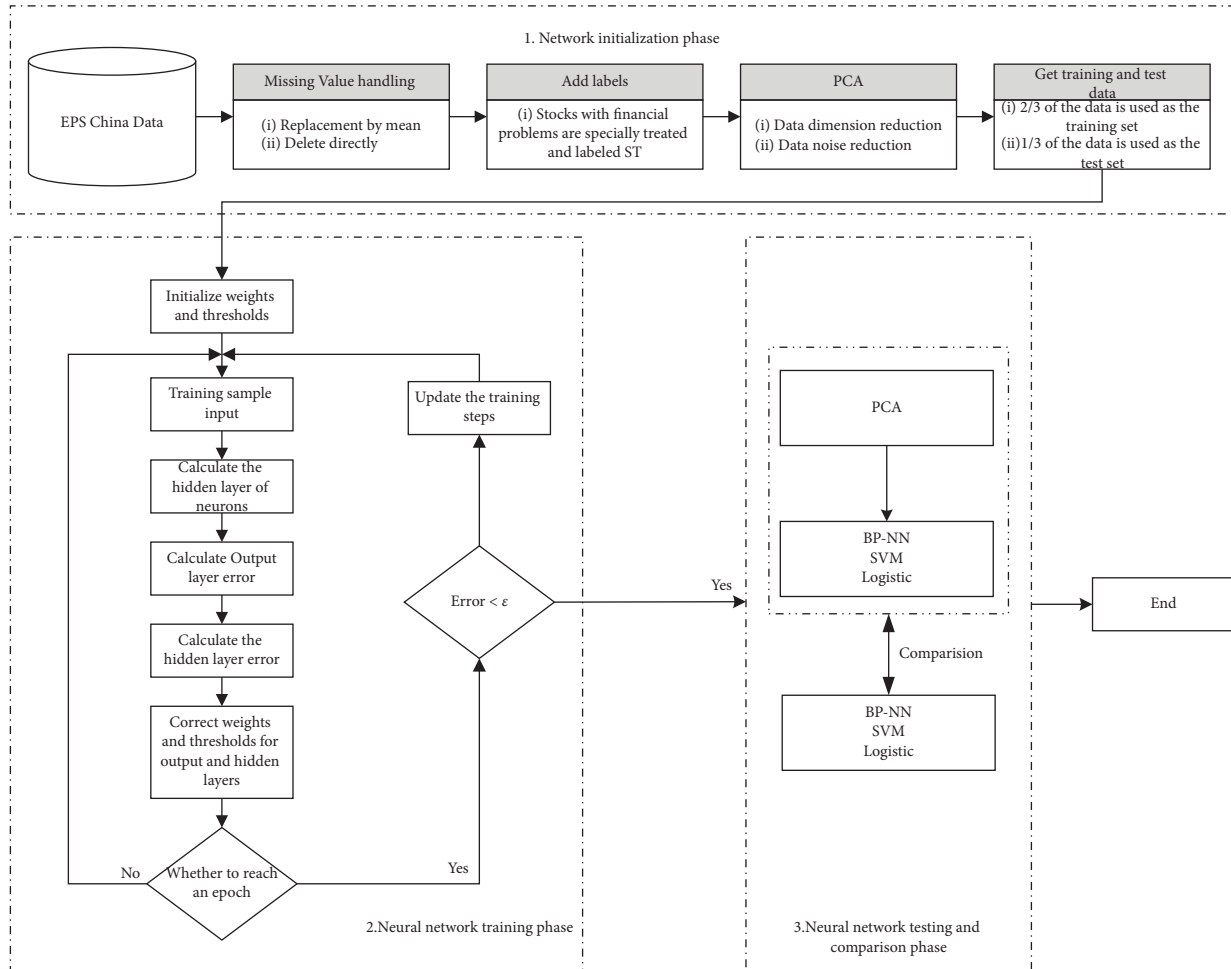


FIGURE 3: Flow chart of modeling financial risk early warning mechanism.

ST in the past three years from 2019 to 2021; all of those financial data are from the EPS China Data, involving a total of 250 companies. When there are missing values in the same company and the same financial variable, replace them with the average value of the past three years. If there is no data in the past three years, delete the variable and other financial indicators of the company. Companies whose shares are marked with ST are identified as having potential financial problems. After standardizing the data, PCA dimensionality reduction is performed to obtain the principal component financial data with a variance contribution rate of 90%, and 2/3 of the data is randomly selected as training data, and the remaining 1/3 is used as test data.

*Phase 2:* Input the training data into the BP neural network for modeling and analysis. The parameters of BP-NN are obtained by selecting different numbers of hidden layers. The optimal number of neurons in the hidden layer of the BP model is selected by the method of minimizing the error. The BP-NN model should repeatedly train the dataset to obtain the optimal parameters in case the obtained parameters are not globally optimal.

*Phase 3:* Observe the performance of the financial data with and without PCA dimension reduction in the SVM, Logistic, and BP neural network models respectively, and record the respective prediction value on confusion matrices. Calculate and compare the correct classification accuracy of predicted corporate financial warnings and finally draw conclusions.

TABLE 3: Principal components and cumulative variance contribution.

Components	Standard deviation	Proportion of variance	Cumulative proportion (%)
Comp.1	0.1911	0.1685	16.85
Comp.2	0.1832	0.1549	32.34
Comp.3	0.1654	0.1263	44.97
Comp.4	0.1419	0.0929	54.26
Comp.5	0.1393	0.0895	63.21
Comp.6	0.1133	0.0592	69.14
Comp.7	0.0991	0.0453	73.67
Comp.8	0.0919	0.039	77.57
Comp.9	0.0849	0.0333	80.91
Comp.10	0.0748	0.0258	83.48
Comp.11	0.0718	0.0238	85.86
Comp.12	0.0695	0.0223	88.09
Comp.13	0.0656	0.0199	90.08
Comp.14	0.0623	0.0179	91.87
Comp.15	0.052	0.0125	93.12
Comp.16	0.0513	0.0122	94.34
Comp.17	0.0494	0.0113	95.46
Comp.18	0.0465	0.01	96.46
Comp.19	0.0442	0.009	97.37
Comp.20	0.0427	0.0084	98.21
Comp.21	0.0397	0.0073	98.93
Comp.22	0.0316	0.0046	99.40
Comp.23	0.0286	0.0038	99.77
Comp.24	0.0194	0.0017	99.95
Comp.25	0.0083	0.0003	99.98
Comp.26	0.0069	0.0002	100.00

#### 4. Empirical and Result Analysis

Data processing is carried out in the manner of stage one. The company is assigned the label of “Yes” according to the listed company’s being marked with ST as the dependent variable, and the normal company without potential financial risk is assigned the label of “No.” In order to improve the quality of data and increase the prediction accuracy of the financial risk early warning model, it is necessary to use PCA to extract principal components from data. After normalization the financial data, the information of each principal component and its cumulative variance contribution rate are get as shown in Table 3.

It can be seen from Table 3 that the cumulative variance contribution rate of the first 13 principal components reaches 90%, indicating that the first 13 principal components basically contain all the information of the original data, and the last 13 principal components can be ignored as noise, so we selected The first 13 principal components replace the original input data for neural network training. Before PCA processing, there were 26 input variables, and after PCA processing, there are only 13. In the case of ensuring the maximum retention of information, the dimension of the data is reduced, thereby improving the efficiency of the algorithm. The 13 principal components were analyzed by the BP neural network. By setting different numbers of neurons, different measurement results are obtained, as shown in Table 4.

According to Table 4 and Figure 4, it can be concluded that when the number of hidden neurons is 15, the training

error and test error are relatively low, and the neural network fits well. This is selected as the final fitting result. Through the confusion matrix of SVM, Logistic and BP neural networks, respectively. The prediction results are shown in Table 5.

According to the results in Table 5, we can calculate the training and testing accuracy of various models with and without PCA through formula (13), and the results are shown in Table 6. From the results in Tables 5 and 6, we can clearly see that without PCA preprocessing, SVM can have very good fitting results on the training data, and its accuracy is the highest, which represents very serious overfitting, and followed by the neural network model. However, for the test data, the SVM model prediction results are not as good as logistic, but the most accurate prediction rate is the neural network model. This indicates that the prediction ability of SVM for financial risk is not good after the serious overfitting phenomenon. After the data is processed by PCA, the SVM model does not have any substantial improvement in the performance of the financial risk warning. The prediction of the training data is as high as 99.7%, and the prediction of the test data is still low. It is further shown that SVM is easy to cause overfitting. The performance of the logistic model is not as good as that of the SVM, but in general, the PCA-BPNN performs the best. Not only did the accuracy of training data achieve 95.1%, but the test accuracy was also as high as 86.6%. It shows that its generalization ability is good, and PCA-BPNN can accurately capture financial risk problems of listed companies.



TABLE 4: Measurement errors for different numbers of hidden layers neurons.

Numbers	Training error	Test error	Total error
1	4.9468	3.0361	7.983
2	5.9912	3.5811	9.5723
3	4.4522	3.8146	8.2668
4	4.4522	4.5153	8.9675
5	4.2873	3.8925	8.1798
6	1.9787	3.0361	5.0149
7	3.4078	3.7368	7.1446
8	3.7926	4.5153	8.3079
9	2.0337	4.671	6.7047
10	3.188	3.4254	6.6134
11	3.3529	3.8925	7.2454
12	1.3741	3.5032	4.8774
13	1.8138	3.8925	5.7063
14	2.6383	1.7127	4.351
15	0.8794	2.9583	<b>3.8377</b>
16	2.2536	3.3475	5.6011
17	1.4291	3.2697	4.6988
18	1.539	3.3475	4.8866
19	1.9787	4.0482	6.0269
20	2.1986	3.1918	5.3904

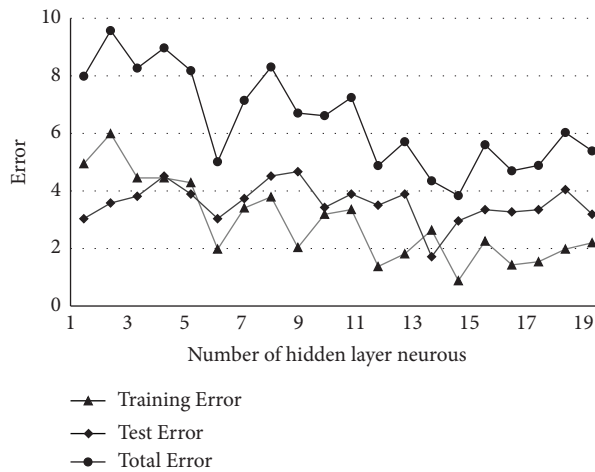


FIGURE 4: Measurement error for different number of hidden layers neurons.

TABLE 5: Cross-comparison of prediction results of different models.

		—				PCA			
		Training		Test		Training		Test	
		No	Yes	No	Yes	No	Yes	No	Yes
SVM	No	164 (0.495)	0 (0)	45 (0.273)	39 (0.236)	167 (0.505)	0 (0)	37 (0.224)	44 (0.267)
	Yes	1 (0.003)	166 (0.502)	16 (0.097)	65 (0.394)	1 (0.003)	163 (0.492)	11 (0.067)	73 (0.442)
Logistic	No	147 (0.444)	17 (0.051)	69 (0.418)	15 (0.091)	151 (0.456)	16 (0.048)	65 (0.394)	16 (0.097)
	Yes	70 (0.211)	97 (0.293)	30 (0.182)	51 (0.309)	96 (0.29)	68 (0.205)	46 (0.279)	38 (0.23)
BP-NN	No	156 (0.471)	8 (0.024)	61 (0.37)	23 (0.139)	162 (0.489)	7 (0.021)	71 (0.43)	8 (0.048)
	Yes	12 (0.036)	155 (0.468)	17 (0.103)	64 (0.388)	9 (0.027)	153 (0.462)	14 (0.085)	72 (0.436)

Note. The numbers in parentheses are the proportion of the measurement results to the total.

TABLE 6: Prediction accuracy of different models.

	—		PCA	
	Training	Test	Training	Test
SVM	<b>99.7%</b>	66.7%	<b>99.7%</b>	66.8%
Logistic	73.7%	72.7%	66.1%	62.4%
BP-NN	93.9%	<b>75.8%</b>	95.1%	<b>86.6%</b>

## 5. Conclusion

Financial data manipulation exists in listed companies, and the distortion of financial information reduces the prediction accuracy of the financial risk early warning model. PCA is a data dimensionality reduction method, which can filter out excess noise data. In this paper, PCA is introduced into financial information data and a PCA-BPNN model is constructed for the study of financial risk early warning of listed companies. Taking 250 Chinese A-share listed companies from 2019 to 2021 as research samples, this paper empirically analyzes the impact of the PCA-BPNN model on corporate financial risk. The fitting effect and prediction accuracy are compared with SVM and logistic, respectively. The conclusions of this paper are as follows: First, the preprocessing of financial data with the PCA method can significantly improve the prediction accuracy of BP-NN financial early warning model. Second, the PCA-BPNN model has significant financial risk generalization ability. In the process of identifying enterprises with financial problems, the prediction accuracy of the PCA-BPNN model is significantly higher than that of the SVM and logistic models. This study provides new ideas for financial risk early warning, and relevant research findings have certain reference significance for regulators, financial institutions, and market investors; that is, when predicting corporate financial risks, PCA can be combined with BP-NN. The model PCA-BPNN is used in financial risk early warning to make more accurate judgments. Though our model has much better performance than other models, there is still room for improvement. First of all, the establishment of indicators is a dynamic process. It is worth thinking about how to more accurately describe risk indicators. Secondly, the selection of principal components by the PCA is often judged based on personal experience. How to reduce the dimensionality of the data and retain the original characteristics as much as possible is also a research direction.

## Data Availability

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

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This research was supported by the National Social Science Fund General Project of China (21BJY002), the Humanities

and Social Sciences Research Project of the Ministry of Education of China (20YJA790097), and Philosophy and Social Sciences Major Tender Project in Zhejiang Province of China (20XXJC03ZD).

## References

- [1] K. Van Barneveld, M. Quinlan, P. Kriesler et al., "The COVID-19 pandemic: lessons on building more equal and sustainable societies," *Economic and Labour Relations Review*, vol. 31, no. 2, pp. 133–157, 2020.
- [2] X. Liu, Y. Liu, and Y. Yan, "China macroeconomic report 2020: China's macroeconomy is on the rebound under the impact of COVID-19," *Economic and Political Studies*, vol. 8, no. 4, pp. 395–435, 2020.
- [3] *InThe Securities Regulatory Commission Inspected 20 Typical Illegal Cases in 2021* China securities regulatory commission, Beijing China, 2021, <http://www.csrc.gov.cn/csrc/c100028/c2265190/content.shtml>.
- [4] A. R. Razali and I. M. Tahir, "Review of the literature on enterprise risk management[J]," *Business Management Dynamics*, vol. 1, no. 5, p. 8, 2011.
- [5] W. Weng, X. Zhu, N. Wang, and B Peng, "Analysis on the Whitewashing Motivation of Financial Statement of Listed Real Estate Companies in China," in *Proceedings of the 2018 International Conference on Education Science and Social Development (ESSD 2018)*, pp. 193–195, Atlantis Press, Paris, July 2018.
- [6] C. Liu, K. Su, and M. Zhang, "Water disclosure and financial reporting quality for social changes: empirical evidence from China," *Technological Forecasting and Social Change*, vol. 166, Article ID 120571, 2021.
- [7] E. Okoye and E. N. Ndah, "Forensic accounting and fraud prevention in manufacturing companies in Nigeria[J]," *International Journal of Innovative Finance and Economics Research*, vol. 7, no. 1, pp. 107–116, 2019.
- [8] H. Shang, D. Lu, and Q. Zhou, "Early warning of enterprise finance risk of big data mining in internet of things based on fuzzy association rules," *Neural Computing & Applications*, vol. 33, no. 9, pp. 3901–3909, 2021.
- [9] B. Lev, "The deteriorating usefulness of financial report information and how to reverse it," *Accounting and Business Research*, vol. 48, no. 5, pp. 465–493, 2018.
- [10] S. Qi, K. Jin, B. Li, and Y Qian, "The exploration of internet finance by using neural network," *Journal of Computational and Applied Mathematics*, vol. 369, Article ID 112630, 2020.
- [11] S. Liu, A. Borovykh, L. A. Grzelak, and C. W Oosterlee, "A neural network-based framework for financial model calibration," *Journal of Mathematics in Industry*, vol. 9, no. 1, pp. 1–28, 2019.
- [12] Z. Wu and R. Wu, "Using machine learning approach to evaluate the excessive financialization risks of trading enterprises," *Computational Economics*, vol. 59, no. 4, pp. 1607–1625, 2021.
- [13] Y. Wang, "Research on supply chain financial risk assessment based on blockchain and fuzzy neural networks," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 2021–8, 2021.
- [14] M. Tavana, A. R. Abtahi, D. Di Caprio, and M Poortarigh, "An Artificial Neural Network and Bayesian Network model for liquidity risk assessment in banking," *Neurocomputing*, vol. 275, pp. 2525–2554, 2018.
- [15] G. Du, Z. Liu, and H. 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, 2021.
- [16] A. Huang, L. Qiu, and Z. Li, “Applying deep learning method in TVP-VAR model under systematic financial risk monitoring and early warning,” *Journal of Computational and Applied Mathematics*, vol. 382, Article ID 113065, 2021.
- [17] Y. Cao, Y. Shao, and H. Zhang, “Study on early warning of E-commerce enterprise financial risk based on deep learning algorithm,” *Electronic Commerce Research*, vol. 22, pp. 21–36, 2021.
- [18] X. 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, 2021.
- [19] H. Zhou, G. Sun, S. Fu, J. Liu, X. Zhou, and J. Zhou, “A big data mining approach of PSO-based BP neural network for financial risk management with IoT,” *IEEE Access*, vol. 7, pp. 154035–154043, 2019.
- [20] P. Prasetyawan, I. Ahmad, R. I. Borman, and A. P. Yogi, “Classification of the Period Undergraduate Study Using Back-Propagation Neural Network,” in *Proceedings of the 2018 International Conference on Applied Engineering (ICAE)*, pp. 1–5, IEEE, Batam, Indonesia, October 2018.
- [21] X. Huang, X. Liu, and Y. Ren, “Enterprise credit risk evaluation based on neural network algorithm,” *Cognitive Systems Research*, vol. 52, pp. 317–324, 2018.
- [22] F. Shen, X. Zhao, Z. Li, K. Li, and Z. Meng, “A novel ensemble classification model based on neural networks and a classifier optimisation technique for imbalanced credit risk evaluation,” *Physica A: Statistical Mechanics and Its Applications*, vol. 526, Article ID 121073, 2019.
- [23] K. Valaskova, T. Kliestik, L. Svabova, and P. Adamko, “Financial risk measurement and prediction modelling for sustainable development of business entities using regression analysis,” *Sustainability*, vol. 10, no. 7, p. 2144, 2018.
- [24] S. Ren, “Optimization of enterprise financial management and decision-making systems based on big data,” *Journal of Mathematics*, vol. 2022, pp. 1–11, Article ID 1708506, 2022.
- [25] A. Cheaitou, R. Larbi, and B. Al Housani, “Decision making framework for tender evaluation and contractor selection in public organizations with risk considerations,” *Socio-Economic Planning Sciences*, vol. 68, Article ID 100620, 2019.
- [26] I. Klopotan, J. Zoroja, and M. Meško, “Early warning system in business, finance, and economics: bibliometric and topic analysis[],” *International Journal of Engineering Business Management*, vol. 10, Article ID 184797901879701, 2018.
- [27] Z. Zhang, Y. Xiao, Z. Fu, K. Zhong, and H. Niu, “A study on early warnings of financial crisis of Chinese listed companies based on DEA-SVM model,” *Mathematics*, vol. 10, no. 12, p. 2142, 2022.
- [28] S. K. Onsongo, S. M. A. Muathe, and L. W. Mwangi, “Financial risk and financial performance: evidence and insights from commercial and services listed companies in Nairobi Securities exchange, Kenya,” *International Journal of Financial Studies*, vol. 8, no. 3, p. 51, 2020.
- [29] V. Verbraak-Kolevska, “I,” *nternal Audit Effectiveness: Factors Influencing Management to Listen to the Internal Auditor’s Risk Warnings: A Multi-Method Study from a Behavioural Decision Making perspective*, Erasmus University Rotterdam, Rotterdam, Netherlands, 2018.