ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Sustainable corporate financial management: research on intelligent financial internal control system of industrial enterprises based on RPA

Chengbing He[1], *Huanxia Hao*[2], *Yuanyuan Zhou*[1], *Jiaqi Peng*[3] *and Xiongjun Wang*[1,*]

[1] College of Economics and Management, Nanchang Vocational University, Nanchang, Jiangxi, 330500, China
[2] College of Economics and Management, Xinjiang University of Political Science and Law, TumuShuker, Xinjiang Uygur Autonomous Regions, 843900, China
[3] Teaching Service Department, Seentao Technology Co., Ltd., Sanya, Hainan, 572025, China

SUMMARY: *There exists a distinct synergy between enterprise financial sharing and internal control systems. From the perspective of big data intelligence, the present research designs an optimization model for enterprise financial sharing center, combining the application of robotic process automation (RPA), with the aim of making the optimal intelligent internal control system of financial operation and management of industrial enterprises. This financial management optimization model includes three key aspects, namely, (1) enterprise credit risk evaluation through system clustering and factor analysis; (2) employee disciplinary prediction through C4.5 decision tree algorithm; and (3) enterprise financial management risk prediction through BP neural network algorithm. In the empirical study, the researchers find out that the 75 sampled industrial enterprises fall into six groups, while there are three main components in relation to their enterprise credit risk. Moreover, the employee disciplinary prediction model is found effective in recognizing the employee disciplinary behavior, thus helping in enterprise safety management. In addition, compared to the LPM risk prediction model, the new risk prediction model using BP neural network algorithm offers a better fitting effect and accuracy. In summary, the optimization model of enterprise financial management successfully lowers the cost of the financial sharing center operations and management, optimizes the intelligent internal control system of finance, and raises the quality and level of enterprise financial management.*

KEYWORDS: *system clustering; factor analysis; C4.5 decision tree; BP neural network; RPA; financial management*

1 Introduction

With the increasingly strict external regulatory environment, decision support and agile operations and other needs make the enterprise must be driven by data to financial operations, from the rough, single, passive accounting management to refinement, synergistic, proactive foresight type of intelligent management [1-3]. Under this, the traditional financial system in the industry and financial synergy depth, data standards and quality, data value mining, management decision-making support and other aspects of the shortcomings of the information query difficulty, the depth of the application of the data is insufficient, it is urgent to push the original financial system to the “wisdom” of the new height. Intelligent financial system is

*hecbing720823@126.com
https://doi.org/10.65102/is2026438 Published: 30 April, 2026

based on the business model of the new era, through the integration of intelligent networking, big data analysis, cloud computing technology and other emerging technologies, to realize the relevant system of independent collection, convergence, computing, processing, reporting and independent correction of accounting data, support for intelligent decision-making, control of digital staff algorithms and thinking logic of the shared system [4-7]. Some large enterprises have begun to try to utilize the advantages of new technologies such as “big, smart, material, cloud, mobile” to promote the construction of financial systems [8]. But overall, the effect of the construction of smart financial system is not significant.

In this context, robotic process automation (RPA) technology provides a new path for the construction and optimization of intelligent financial systems. RPA is a kind of application technology that can simulate and optimize the interaction process between human beings and computers according to predefined procedures, efficiently execute high-volume, repeatable tasks, and thus significantly improve the level and efficiency of workflow automation [9-11]. After years of information technology layout, many enterprises have formed a situation where various types of information systems coexist, objectively giving rise to the demand for interaction and integration between different information systems, which in turn promotes the popularization and application of RPA technology [12]. RPA technology through the unified enterprise cloud platform technology architecture, the integration of customized development, back-end integration platforms, etc., to centralize the isolated and dispersed financial data and information, and thus promote the construction of intelligent financial system. And then promote the construction and application deployment of intelligent financial system [13-15]. Due to its short deployment cycle and low cost, it can effectively alleviate the cost pressure faced by most enterprises in building smart financial systems [16].

The concept of Robotic Process Automation (RPA) is often referred to as a rules-based digital workforce whose fundamental principle is based on simulating human behavior via software robots to perform business processes that are highly repetitive in nature and with clear rules. Literature [17] carried out a case study on the RPA technology and observed that it could successfully lower the labor cost of an enterprise, increase the efficiency of management, and increase the pace of business-processing, which would unlock the developmental potential of this digital technology and bring about a wide-scale effect on the industrial transformation. Literature [18] indicates that the effectiveness of RPA depends on the regularity of the business process, since robots follow the programed execution logic. The research claims that the one and only means of automating processes and achieving meaningful efficiency improvements is to break the business processes into atomized rules. Literature [19] also states that the RPA technology is a vital instrument in improving business efficiency in case enterprises want to achieve digital transformation and eliminate contradictions due to internal and external changes. Through information-based data management, such as the collection, accumulation, storage, and processing of business processes, enterprises will move towards automation. Literature [20] , to determine the position played by integrating the RPA and ChatGPT technology in enterprise digitalization, analyzed the literature on the topic in the last 20 years and took a new energy service company as a case study. It is found that the joint use of RPA and ChatGPT technology has a positive impact on promoting the digital transformation of an enterprise. Literature [21] observes that most of the modern enterprise business processes have been automated. Replicating the human action, RPA is able to relieve heavy reliance on manual work operations, minimize human errors significantly, and enhance the efficiency of work.

In terms of using RPA technology in enterprise management, the literature [22] reveals the use of RPA-Salesforce in the transformation of enterprise automation. Under this setup, RPA technology creates a link between users and the systems to create human-computer interaction and Salesforce acts as an orchestration hub in evaluating organizational effectiveness. These two are complementary in enhancing business performance, data processing and other performance measures. Literature [23] suggests an automated enterprise development plan based on the combination of AI, machine learning and RPA, and illustrates how these approaches relate to each other and interact. This research confirms the positive changes in the efficiency, accuracy, and other aspects of various industries, including finance, healthcare, and manufacturing. Literature [24] applies an application combining RPA and AI techniques to the automated order processing process of an enterprise, and shows in a case study of German SMEs that the application significantly improves the time-economy gains and organizational performance of the enterprise. Literature [25] uses both AI and RPA solutions in an automated order-processing workflow of a company and, in a case-study of German SMEs, it is shown how the approach leads to substantial time-related financial benefits and better organization performance. The issue of integrating AI technologies, like natural language processing, computer vision, and cognitive computing, with RPA. It introduces a composite cognitive RPA system into enterprise operations and proves that it is effective in enhancing the enterprise capability to process complex documents, context sensitive decision making and managing exceptions.

Studies and relevant literature reviews and theories have proven the applicability of Robotic Process Automation (RPA) in intelligent financial transformation. Literature [26] highlights the possibility of using RPA in enterprise financial management by conducting research and review of literature. Besides cost savings and improvement in operational efficiencies, RPA also supports evidence-based decision making and helps in enterprise digital transformation. Literature [27] emphasizes on the necessity of finance departments using RPA technology to make intelligent decisions and conduct financial analysis of their corporate resources. The study presents an exhaustive rationale about the strengths and future possibilities of RPA in corporate financial systems. Literature [28] explains that RPA technology, due to its effectiveness in processing data and handling transactional processes, is effective in reducing labor costs through high efficiencies in operation. Moreover, the use of artificial intelligence with RPA could lead to improved operations for enterprise financial systems. Literature [29] investigates the application of RPA in enterprise finance digital transformation in different angles. It identifies that automated nature of the financial process at enterprises and the resultant savings in operational costs are other benefits of RPA, besides the benefits of data processing efficiency and data security.

As for practice in the financial industry, the literature [30] makes use of Robotic Process Automation (RPA) technology to solve the problem of low accuracy and flexibility in financial data processing. With the analysis of historical financial data using data mining, this literature is able to rationally identify and optimize key financial processes. As shown in the research, using the RPA technology can achieve more than 97.36% automation of financial operations and above 90% accuracy of financial operations' execution, thus greatly improving the efficiency of financial operations. In literature [31] , the authors apply RPA technology to the configuration of enterprise accounting system and evaluate the reliability of this technology by means of expert questionnaires and interviews. Further, the feasibility of its application is proved by the case implementation from the perspective of privacy, security, and sustainability of the system. In literature [32] , the authors design an enterprise financial statement filing robot with the help of RPA to solve the problems arising during financial analysis due to large volume of repetitive and routine tasks. Through performing tasks with high efficiency and accuracy, the robot is helpful in promoting RPA among enterprises. Literature [33] uses RPA technology to optimize the financial management process, involving data collection, building of a cloud procurement platform, and cost analysis. In doing so, it promotes the building of artificial intelligence society and gives guidance for the automation of enterprise financial shared services. Literature [34] studies the optimization of banking operations using RPA in a financial sharing framework. It makes the audit process better than before, making the auditing procedure more systematic and standardized and offering accurate feedback on the result, which enables it to solve many financial issues faced by banks. Literature [35] combines big data analysis with RPA and uses this combination in analyzing and giving risk warning of enterprise financial data. According to the results, there is a 32.78% reduction of task errors when comparing it with traditional methods.

This paper combines techniques such as system clustering, factor analysis, C4.5 decision tree, back propagation (BP) neural network, and robotics process automation in order to create an optimization model regarding the operation and management of financial sharing service centers under a big data-intelligence environment. This paper provides support for improving the intelligent financial internal control systems of industry enterprises. More specifically, system clustering and factor analysis methods are adopted to build an enterprise credit risk evaluation model; C4.5 decision tree method is used to build an employee disciplinary prediction model; and finally, BP neural network is used to build an enterprise financial management risk prediction model. The following empirical verification will be conducted in order to verify the effectiveness of the models.

2 Optimization of the design of intelligent financial internal control system for industrial enterprises based on RPA

The essence of internal control, as part of enterprise risk management, is the ability and influence to make sound decisions and the trust that develops over time through the flow of information, the chain of interests and the ability to mobilize resources. Financial shared service centers achieve the goal of internal control by being a trusted decision-making body. In view of this, this study achieves the optimal design of an intelligent financial internal control system for an industrial enterprise by optimizing the operation and management of the enterprise's financial shared center model (FSSC).

The study selects the operation and management of FSSC of Z industrial enterprise as the object, and designs the operation and management optimization framework of enterprise FSSC based on big data intelligence as shown in Fig. 1.

The present research is concerned with identifying and analyzing the limitations associated with Enterprise Z’s FSSC operation and management framework. The research aims at proposing a modified operating framework for the enterprise financial sharing center (FSSC) based on the systematic clustering algorithm, factor analysis, the C4.5 decision tree algorithm, the BP neural network, and the RPA financial robot. On a macro level, the research is designed to be implemented via the utilization of business modules involved in the scope and platform of FSSC operation and management. In other words, data mining tools will be utilized to gather information from each system module involved in the FSSC operation and management framework. Hence, it will become possible to solve the issues identified in relation to the operation and management of Enterprise Z’s FSSC. On this basis, the study provides recommendations related to implementing the FSSC operation and management framework.

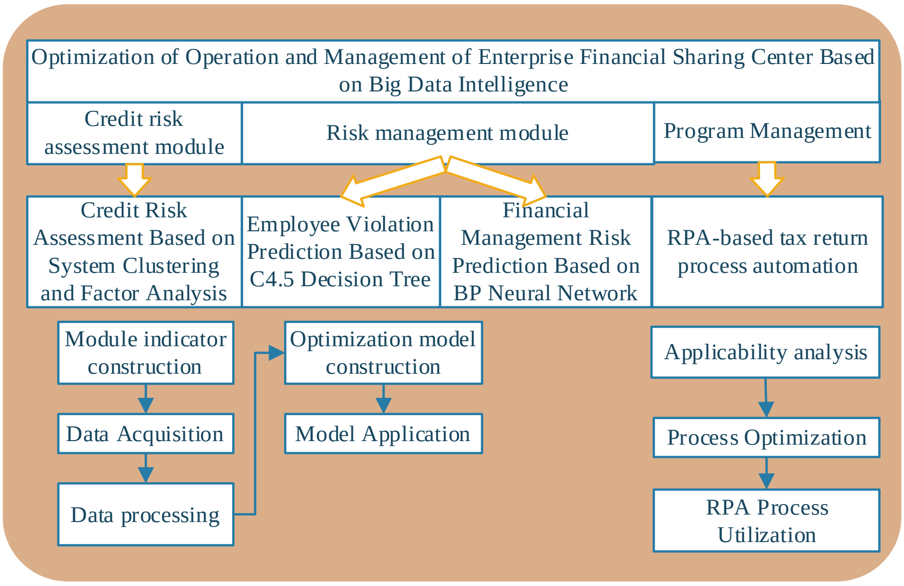

Figure 1: Overall framework for optimizing FSSC operation management based on RPA

The operation management optimization implementation process of Enterprise Z's financial sharing center under big data intelligence is shown in Figure 2, which mainly includes two parts: the algorithm-based optimization process and the optimization process based on the RPA financial robot. The data collection, data depth processing, and model construction of each module of FSSC operation management together constitute the algorithm-based FSSC operation management optimization process. Among them, the in-depth data processing of FSSC operation management includes three elements: index data conversion, cleaning and order of magnitude unification. The RPA-based FSSC operation management optimization process includes applicability analysis, process optimization, and process application. In short, according to the performance management optimization, risk management optimization, and process management optimization needs of Enterprise Z's FSSC operation management, system algorithms, factor analysis, C4.5 decision tree algorithm, BP neural network, and RPA technology are used to implement the optimization, respectively.

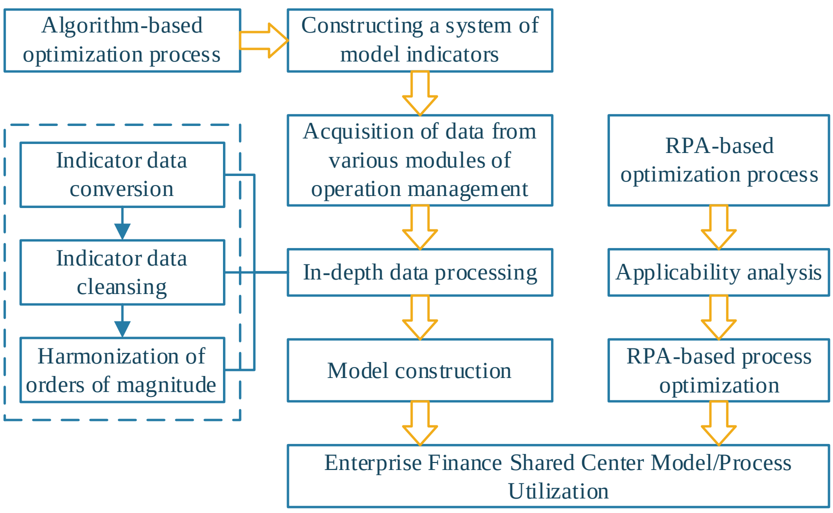

Figure 2: Implementation process of FSSC operation management optimization

3 Evaluation of corporate credit risk based on cluster analysis and factor analysis methodology

3.1 Systematic cluster analysis

Many clustering algorithms have been devised according to different clustering approaches. In particular, K-means is relatively easy to implement and usually results in good clustering effects, which explains why it is frequently applied in practice. However, in K-means clustering, it is essential to determine beforehand the number of clusters. If it is not known how the data samples are classified, predicting the number of clusters (value of K) in advance becomes a hard task. The method of systematic clustering serves as a solution to this problem.

The essence of systematic clustering, sometimes referred to as hierarchical clustering, consists in forming groups of the most similar objects taking into account their proximity relationship, then dividing these groups into subgroups (or combining them into bigger ones) until all cases or variables are grouped into one cluster. The process of systematic clustering can be shown schematically in Figure 3.

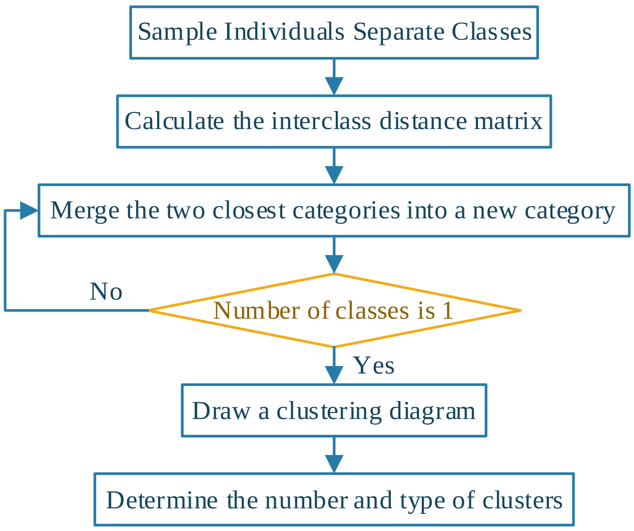

Figure 3: System clustering process

3.1.1 Data pre-processing

The degree of variability in the measurements and variables affects the comparability of the data and therefore requires data standardization. Data standardization methods are generally of the following types:

(1) Standard deviation standardization:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{1}$$

Among them:

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{2}$$

$$s_j = \left[\frac{1}{n} \sum_{i=1}^{n} \left(x_{ij} - \overline{x_j} \right)^2 \right]^{\frac{1}{2}} \tag{3}$$

x_{ij} is denoted as the raw data of the i th sample in the j th indicator; x'_{ij} the standardized data of the i th sample in the j th indicator.

(2) Extreme difference standardization:

$$x''_{ij} = \frac{x'_{ij} - \min_{1 \le i \le n} \left\{ x'_{ij} \right\}}{\max\limits_{1 \le i \le n} \left\{ x'_{ij} \right\} - \min\limits_{1 \le i \le n} \left\{ x'_{ij} \right\}}, (j = 1, 2, \ldots, n) \tag{4}$$

where x''_{ij} is the data after normalization of the j th indicator of the i th sample. After the polar transformation process, all the indicator data $x'_{ij} \in [0,1]$.

3.1.2 Calculating inter-sample distances

In cluster analysis of practical problems, the proximity between any two sample points is usually measured by measuring the distance between them. If d_{ij} is defined as the distance between the sample x_i and the sample x_j, there are several methods of calculating the distance between samples as follows:

(1) Absolute value distance:

$$d_{ij} = \sum^{m} \left| x_{ik} - x_{jk} \right| \tag{5}$$

(2) Euclidean distance:

$$d_{ij} = \left[\sum_{k=1}^{m} \left(x_{ik} - x_{jk} \right)^2 \right]^{\frac{1}{2}} \tag{6}$$

(3) Minkowski distance:

$$d_{ij} = \left[\sum_{k=1}^{m} \left(x_{ik} - x_{jk} \right)^q \right]^{\frac{1}{q}} \tag{7}$$

(4) Chebyshev distance:

$$d_{ij} = \max_{1 \le k \le m} \left| x_{ik} - x_{jk} \right| \tag{8}$$

(5) Marginal distance:

$$d_{ij} = \left[\left(x_i - x_j \right)' s^{-1} \left(x_i - x_j \right) \right]^{\frac{1}{2}} \tag{9}$$

where S is the covariance.

(6) Lang's distance:

$$d_{ij} = \sum_{k=1}^{m} \frac{\left| x_{ik} - x_{jk} \right|}{x_{ik} + x_{jk}} \tag{10}$$

3.1.3 Calculating class-to-class spacing

In the longest-distance strategy, the inter-class distance is defined as the largest distance between any two samples chosen from different classes. The shortest-distance strategy measures the inter-class distance through the smallest distance found between any two samples from the various groups. In average intergroup linkage technique, the inter-class distance is measured by computing the average distance between samples in the various classes. The average intragroup linkage computes the inter-class distance by treating all people in the combined group as one and finding the average distance among all people in the combination group. Both the centroid and median use the averages or medians of the variables to find the class distance. In Ward's minimum variance, when n samples are classified into n clusters, the total sum of squared deviations becomes larger when the number of clusters decreases. This means that at the start, only a very small portion of the total sum of the squares deviates is assigned to the two classes, then the process repeats itself until all the samples are clustered into one.

3.1.4 Methods for Determining Sample Spacing and Class Spacing

Before conducting cluster analysis, the similarity among samples must first be measured by calculating the distance between them. In this paper, squared Euclidean distance or Euclidean distance is selected as the method for measuring sample spacing. Under the premise that the sample distance is the squared Euclidean distance method, the samples are clustered by using seven distance calculation methods respectively, and the aggregation coefficients obtained from the seven methods are compared and studied so as to select the best distance calculation method.

The aggregation coefficient is the sum of the distortion degree of all classes in the sample. The distortion degree of each class is the sum of the squares of the distances between the center of the class and its internal members, assuming that the number of classes in n samples is $K(K \le n-1)$, and defining C_{k} to be the k th class, and U_K to be the location of the center of gravity of the class, the distortion degree of the k th class is:

$$\sum \left| x_i - u_k \right|^2 \tag{11}$$

The coefficient of aggregation is:

$$J = \sum_{k=1}^{K} \sum_{i \in c_k} \left| x_i - u_k \right|^2 \tag{12}$$

3.2 Factor analysis

The objective of factor analysis lies in identifying certain latent factors which cannot be observed directly, but which can be observed indirectly by a series of measurable factors. Factor analysis involves four basic processes: (1) analyzing the correlation between the initial factors, (2) establishing latent factors, (3) understanding the factor structure using rotation, and (4) calculating the factor score.

First, the latent demand is explored by determining whether there is a good correlation between the initial variables. In this study, the covariance formula is used to achieve this:

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] \tag{13}$$

If there are n-dimensional primitive variables in the experiment, one can construct a covariance matrix consisting of $\frac{n!}{(n-2)!*2}$ covariance matrices:

$$Cov = \begin{pmatrix} \mathrm{cov}(x_1,x_1) & \mathrm{cov}(x_1,x_2) & \dots & \mathrm{cov}(x_1,x_n) \\ \mathrm{cov}(x_2,x_1) & \mathrm{cov}(x_2,x_2) & \dots & \mathrm{cov}(x_2,x_n) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \mathrm{cov}(x_m,x_1) & \mathrm{cov}(x_m,x_2) & \dots & \mathrm{cov}(x_m,x_n) \end{pmatrix} \tag{14}$$

Next is the construction of factor variables. Constructing factor variables is to summarize the original variables into a few common factors and solve the factor loadings based on the correlation matrices of the variables, and the method of solving the factor loadings is factor analysis based on the principal component model. This method uses the principle of rotating coordinates: first the original variables are transformed through linear relationships, and then another set of non-corresponding variables are transformed next to get the second transformation of these variables called principal components. In factor analysis, the determination of the common factor is based on the characteristic root $\lambda_1(\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0)$ found by the matrix of correlation coefficients R, and the characteristic root λ is the value of the variance contribution of the common factor. Factor's variance contribution value, calculating the characteristic root can calculate the variance contribution rate R_1 and cumulative contribution rate R_2 of each common factor F:

$$R_1 = \frac{\lambda_i}{\sum_{k=1}^{n} \lambda_k} (i = 1,2,\dots,n) \tag{15}$$

$$R_2 = \frac{\sum_{k=1}^{i} \lambda_k}{\sum_{k=1}^{n} \lambda_k} (i = 1,2,\dots,n) \tag{16}$$

When the cumulative contribution of the factor R_2 is greater than 80%, it means that the previous principal component $m(m <= n)$ explains most of the information of all variables, where m is the final number of factors.

This study is based on the factor loading matrix to explain the factor variables:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \tag{17}$$

The correlation between the constructed factors and the original variables can be obtained by performing the analysis. Where $a_{ij} = \sqrt{\lambda_i} l_{ij} (i, j = 1, 2, \ldots, n)$.

To make the original factor variables easier to interpret, factor rotation is introduced to describe the variables more clearly. The central idea of orthogonal rotation is to maximize the sum of the variance of the squared loadings for each factor while keeping the orthogonal properties of the original factors unchanged.

The score of each factor is obtained through a linear combination of the original variables. In other words, the original variables are combined on the basis of the eigenvalues of the factors. The corresponding linear combination can be expressed as follows:

$$\begin{cases} F_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ F_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ F_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{cases} \tag{18}$$

where F represents the factor variables and x represents the original variables.

The study used Barlett's estimation method, which is calculated based on the principle of weighted least squares regression. Let (x, y) be the pair of variables that need to be fitted together to satisfy the relationship $y = f(x, \omega)$, where $x = \left[x_1, x_2, \ldots, x_n\right]^T$ and $\omega = \left[\omega_1, \omega_1, \ldots \omega_n\right]^T$ is the parameter to be determined, and the least squares method aims to solve the objective function model:

$$L(y, f(x, \omega)) = \sum_{i=1}^{m} \left[y_i - f\left(x_i, \omega_i\right)\right]^2 (i = 1, 2, \ldots, n) \tag{19}$$

where $\omega = \left[\omega_1, \omega_1, \ldots \omega_n\right]^T$ is the minimum value.

Lastly, the weight assigned to each factor is based on the percentage of variation caused by each individual factor. The sum total of these factors forms the index function for evaluating the sample on the basis of the newly formed factors. Following the rules and procedures of factor analysis and the methodology adopted in SPSS statistical software, the sequence of steps involved in conducting factor analysis using the correlation matrix is shown in Figure 4 below.

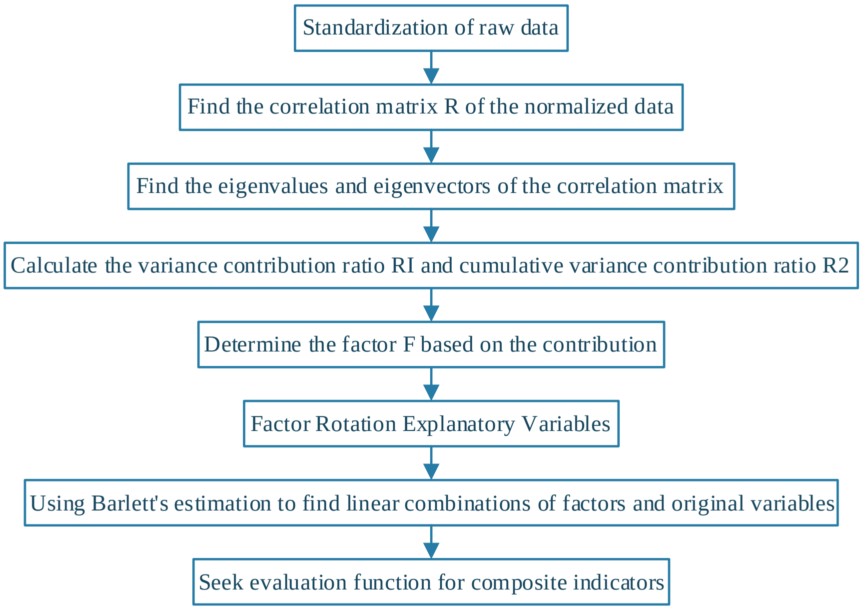

Figure 4: Factor analysis process

3.3 Empirical analysis of credit risk evaluation of industrial enterprises

This section empirically studies the credit risk situation of seventy samples of industrial enterprises based on the credit risk evaluation model of industrial small enterprises that were built up by systematic cluster analysis and factor analysis.

3.3.1 Cluster analysis of sample firms

Before conducting the cluster analysis, the indicators of the samples need to be normalized so as to avoid the influence caused by the difference in data size and outlier value. The process of cluster analysis is completed with the aid of SPSS 28.0. According to the clustering results and combining with practical situations, the seventy-five samples can be divided into six major categories.

3.3.2 Discriminant analysis

Discriminant analysis is utilized to verify whether the classification results are reasonable, whether they are affected by several factors such as the variables used and the distance measurement method, and whether the classification results are unstable.

The typical discriminant function eigenvalues are shown in Table 1. It can be seen that discriminant functions 1 to 5 explain 67.88%, 16.05%, 9.48%, 4.99% and 1.60% of the data, respectively. The five discriminant functions explained 100% and the typical correlation coefficients were also high.

Table 1: Characteristic values of typical discriminant functions

Function	Characteristic value	Percentage of variance	Cumulative percentage	Typical correlation
1	12.438a	67.88	67.88	0.971
2	2.941a	16.05	83.93	0.868
3	1.736a	9.48	93.41	0.804
4	0.915a	4.99	98.40	0.695
5	0.293a	1.60	100.00	0.479

The discriminant function significance tests are shown in Table 2. It can be seen that the five discriminant functions directly have significant differences and discriminatory power, all of which are statistically significant. In general, the classification results are better, the special characteristics of each category are obvious, and the classification is more successful.

Table 2: Significance test of discriminant function

Function test	Wilkes λ value	Chi-square	df	Sig.
1-5	0.004	361.283	37	0.000
2-5	0.043	205.428	25	0.000
3-5	0.162	121.639	16	0.000
4-5	0.425	58.472	9	0.000
5	0.793	16.854	4	0.001

The classification result judgment is shown in Table 3. The typical discriminant function plot is shown in Figure 5. It can be concluded that the accuracy of the classification results obtained by applying cluster analysis in the previous section is as high as 97.33%, which means that all 75 original grouped cases have been correctly classified. And it can also be seen from Figure 5 that the sample enterprises are all centered around 6 groups, indicating that intuitively the grouping discriminant results are better and the cluster analysis results are reasonable.

Table 3: Judgment of classification results

Classification results									
		Average connection of each group	Predict the information of the group members						Total
			1	2	3	4	5	6	
Original	Count	1	16	0	1	0	0	1	18
		2	0	4	0	0	0	0	4
		3	0	0	43	0	0	0	43
		4	0	0	0	3	0	0	3
		5	0	0	0	0	4	0	4
		6	0	0	0	0	0	3	3
		1	88.88%	0	5.56	0	0	5.56	100.00%
		2	0	100.00%	0	0	0	0	100.00%
		3	0	0	100.00%	0	0	0	100.00%
		4	0	0	0	100.00%	0	0	100.00%
		5	0	0	0	0	100.00%	0	100.00%
		6	0	0	0	0	0	100.00%	100.00%

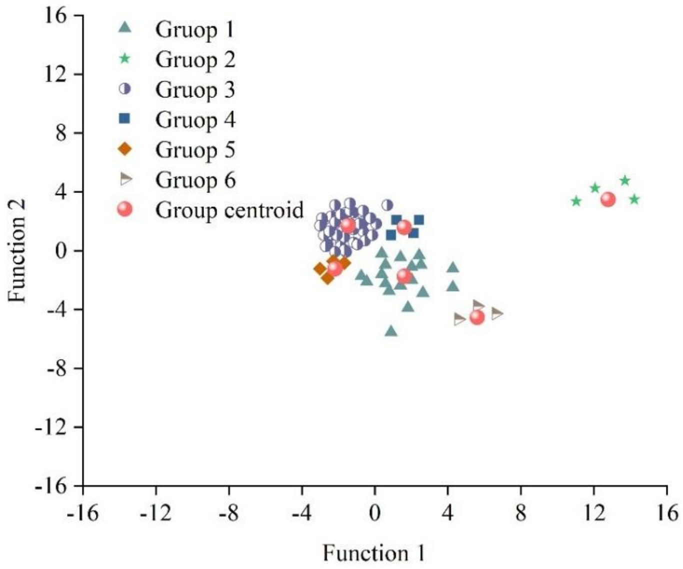

Figure 5: Diagram of a typical discriminant function

3.3.3 Factor analysis of credit risk indicators

Three factors that have significant impacts on the normal operation of industrial enterprises include profitability, solvency, and growth capacity. Seven credit risk indicators are selected from the three factors mentioned above, including gearing ratio, equity ratio, return on total assets, return on net assets, gross profit margin, current ratio, and quick ratio. These seven indicators are then analyzed by the method of factor analysis, and the final scores are determined.

(1) Extraction of principal components

The extraction method chosen in the factor analysis module of SPSS 28.0 is the principal component analysis. The KMO is 0.726 and the approximate chi-square statistic of the Bartlett sphericity test is 538.615 which is large enough. The significance level of $p = 0.000 < 0.01$ indicates that there are significant correlations between the variables and that the sample data can be subjected to factor analysis.

Cumulative proportion of total variance explained is given in Table 4. These findings suggest that reduction of the indicator system to three principal components would be more efficient when principal component analysis is applied, and the cumulative explanatory power is equal to 88.114% which is higher than 85%.

Table 4: Total variance explanation cumulative

	Initial eigenvalue			Extract the sum of squared loads			Rotating load sum of squares		
Component	Total	Percentage of variance	Cumulative /%	Total	Percentage of variance /%	Cumulative /%	Total	Percentage of variance /%	Cumulative /%
1	3.392	48.457	48.457	3.392	48.457	48.457	2.987	42.671	42.671
2	1.655	23.643	72.100	1.655	23.643	72.100	1.836	26.229	68.9
3	1.121	16.014	88.114	1.121	16.014	88.114	1.345	19.214	88.114
4	0.516	7.371	95.485						
5	0.202	2.886	98.371						
6	0.083	1.186	99.557						
7	0.031	0.443	100.000						

Table 5 contains the rotated component matrix, and the variables may be clustered into three fundamental components.

The first principal component, F1, consists primarily of the gearing ratio, equity ratio, current ratio, and quick ratio. The gearing ratio and equity ratio are negatively correlated with F1 and they are referred to as debt-related indicators that indicate a company capability to settle its long-term debts. Conversely, the current ratio and quick ratio are positively related to F1 and can be categorized as indicators of liquidity change which indicates the ability of a company to pay off its short-term debt. This is why F1 is primarily the indicator of the ability of an enterprise to pay off its debts, and the higher the F1 value, the better the solvency will be.

Second principal component, F2, is mostly based on return on net assets and return on total assets. Of these, return on net assets indicates the earnings level of the enterprise whereas return on total assets indicates the efficiency of the operation of assets. Hence, F2 mainly denotes the profitability of the enterprise and the higher the value of F2, the more profitable the company is.

The third principal component, F3, consists mainly of gross profit margin and equity ratio. Gross profit margin is an indicator of the degree of sustained competitive advantage of the enterprise, which can also be horizontally compared across industries. Consequently, F3 is primarily related to the competitive advantage of the enterprise and its tendency toward sustainable development. The bigger value of F3 indicates how competitive advantage of the enterprise is more noticeable and how prospects of the future development are promising

Table 5: The component matrix after rotation

Index	Component		
	1	2	3
Asset-liability ratio	-0.871		
Property rights ratio	-0.758		-0.502
Return on net assets		0.982	
Return on total assets		0.915	
Current ratio	0.931		
Quick ratio	0.926		
Gross profit margin			0.807

(2) Credit risk evaluation scores

The factor analysis model was run through SPSS28.0 statistical software to derive the three public factor scores for each enterprise and calculate the composite factor score F. The factor scores for the six categories of enterprises are shown in Figure 6.

Overall, the first category belongs to high-risk enterprises. The second, third and fourth categories belong to medium-high risk enterprises, which trigger risks from different perspectives. Among them, the second category is weak overall strength, the third category is excessive debt pressure, and the fourth category is limited profitability. The fifth and sixth categories are enterprises with a good level of development and healthy corporate finances.

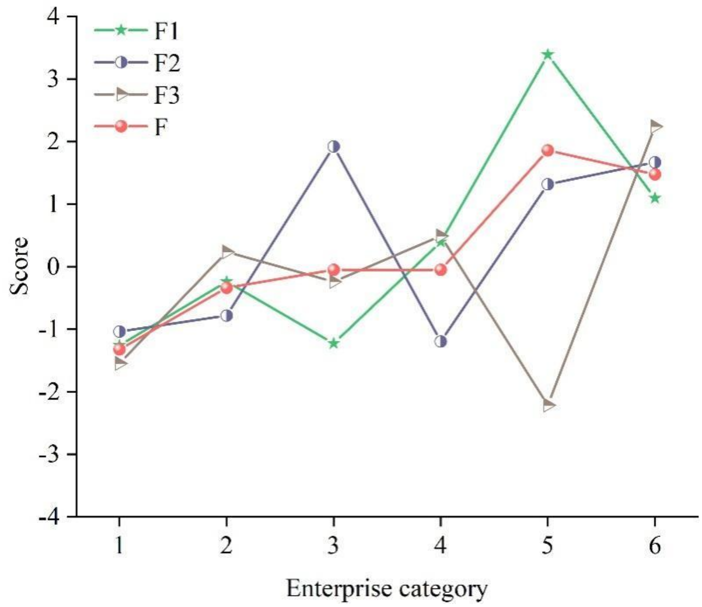

Figure 6: Enterprise factor scores

4 C4.5 Decision Tree Based Prediction of Employee Discipline in Enterprises

This chapter takes Z industrial enterprise as the research object and develops a prediction model for employee disciplinary behavior based on the C4.5 decision tree algorithm.

4.1 Decision Tree Classification Algorithm

A decision tree to analyse and process an available dataset has become one of the typical solutions in the field of data mining. Many different algorithms can be used to construct decision trees, the most popular of them being the ID3 algorithm, the C4.5 algorithm, and the CART algorithm.

4.1.1 ID3 algorithm

The ID3 algorithm is a classification algorithm based on the information entropy. Its fundamental concept is to classify all samples in one of the various categories depending on the values that are given to the attributes. The main objective of the algorithm is to find the right classification attribute among the list of conditional attributes. Information gain is the splitting condition in the ID3 algorithm, and the conditional attribute with highest information gain is commonly used as the splitting attribute of the current node such that the information entropy needed to derive the resulting subset after partitioning is minimum.

The concept of information entropy describes the uncertainty of a source in physics whereas in mathematics it refers to the relation between information redundancy and probability. The concept of information gain is defined as the difference between two entropy values. Particular entropy value is an entropy of some particular class in the sample data set and the second entropy value is calculated when the entropy of that same class measured in the data set. The exact definition is stated below:

Information gain is also the decrease in the quantity of information needed to be classified after the classification process, in comparison to the initial quantity of information, and it is obtained by summing up the initial entropy with the weighted entropy of every subsegmented subset. The associated equation is given below:

Assuming that P_i is the probability that the sample takes the attribute value C_i and specifying that $\sum p_i = 1$, entropy is defined as follows:

$$H\left(p_1, p_2, \cdots, p_s\right) = \sum_{i=1}^{s} -\left(p_i \log p_i\right) \tag{20}$$

The entropy determines the positional status of the attribute in terms of its sorting in the dataset. The $H = 0$ indicates the best classification setting for the attribute.

The information gain, which is the difference between the amount of information originally required and the new information needs after categorization, is determined by calculating the sum of the value of the original entropy and the weighted entropy in each segmented out dataset with the following formula:

$$G(D,S) = \mathrm{H}(D) - \sum P\left(D_i\right)\mathrm{H}\left(D_i\right) \tag{21}$$

where D denotes a dataset, D_i denotes a sample belonging to the i th class, and S denotes the number of class attributes of the data.

The core idea of ID3 algorithm is as follows:

For a dataset T, let T be a training set of samples with known class attributes. Let :

$C_i (i = 1, 2, \ldots\ldots, n)$: be a class labeling attribute having n distinct values.

TC_i : the set of samples belonging to class C_i in training set T.

$|T|$: is the number of samples in the training set T.

$|TC_i|$: is the number of samples in the training set TC_i.

The desired information, i.e., information entropy, required for sample classification is denoted as:

$$Info(T) = -\sum_{i=1}^{n} p_i \log_2\left(p_i\right) \tag{22}$$

p_i: is the non-zero probability that any sample in T belongs to the class C_i and is obtained by estimating with $\frac{|TC_i|}{|T|}$.

$Info(T)$: is the average information entropy required to identify the class labeling of the samples in the dataset T.

Then the samples in the training set are divided by attribute A with m distinct values, which are $\{a_1, a_2, \cdots, a_m\}$, and assuming that A is a discrete attribute, the values are directly and one-to-one corresponding to m outputs tested on the attribute A, and the attribute A divides the training set T into m subsets $\{T_1, T_2, \cdots, T_m\}$, where T_j contains tuples in T with a_j values on attribute A.

Then the information entropy of the tuple classification of T by A is:

$$Info_A(T) = \sum_{j=1}^{m} \frac{|T_j|}{|T|} Info\left(T_j\right) \tag{23}$$

$\frac{|T|}{|T|}$: denotes the number of samples out of the total number of samples when attribute A is the j th value.

$Info_A(T)$: indicates the expected value information needed to classify the samples of the data set T by attribute A. The smaller the value, the higher the purity of the category to which each attribute value belongs.

According to the definition of information gain, the specific calculation formula is:

$$Gain(A) = Info(T) - Infoo_A(T) \tag{24}$$

Equation (24) is the reduction in the expected value of the information demand resulting from the condition of the known value of attribute A. A conditional attribute with the largest information gain is chosen as the split attribute of the node of the moment.

The algorithmic flow of the ID3 algorithm to create a decision tree is shown in Figure 7.

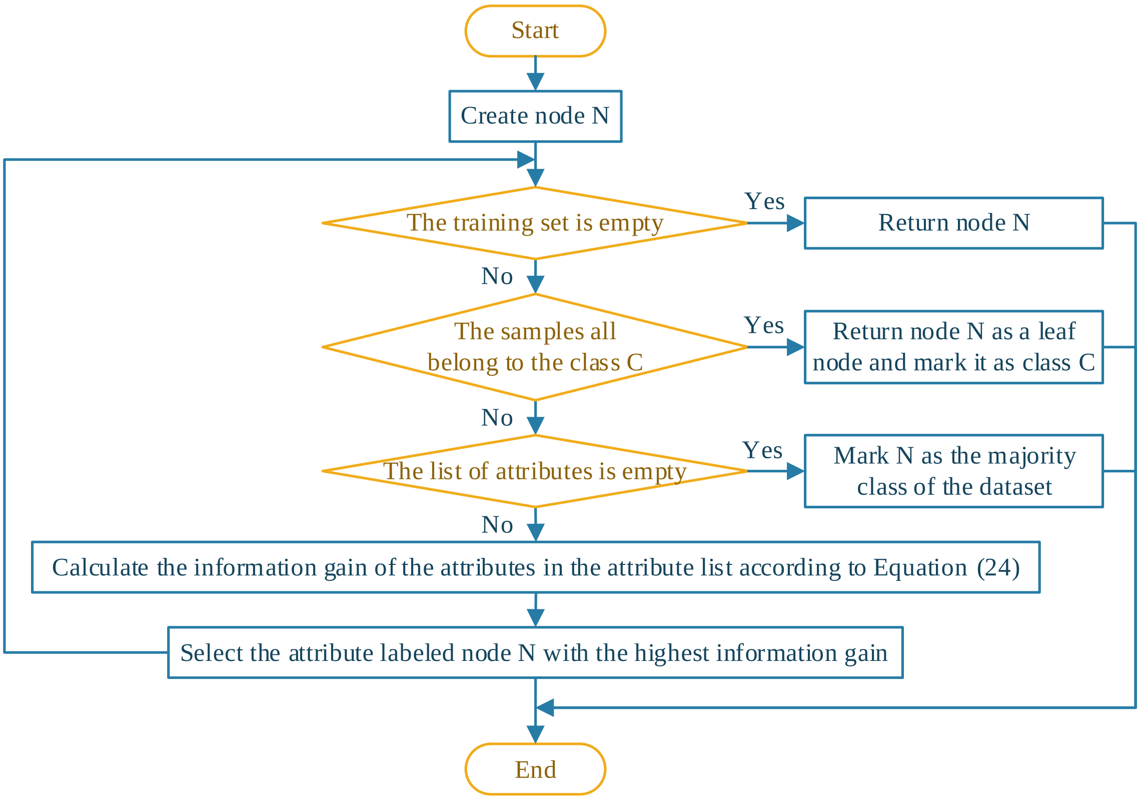

Figure 7: Flowchart of creating a decision tree using the ID3 algorithm

4.1.2 C4.5 algorithm

The C4.5 algorithm is one of the classical algorithms that generate decision trees and it may be viewed as an extension and enhancement of the ID3 algorithm. The step-wise mechanisms of C4.5 and ID3 are generally alike in terms of decision-tree design, but the key distinction between them is how they handle continuous features and what measure of goodness they use to select attributes. The ID3 algorithm does not process continuous attributes directly, and the C4.5 algorithm can first discretize continuous attributes and then perform attribute-selection computations. Information gain is used as an attribute evaluation metric by ID3 and C4.5 utilises the gain ratio.

Gain ratio is an information gain normalized. To put it differently, C4.5 presents the gain-ratio formula to substitute the information-gain criterion in ID3.

In the training set T, Eq. (25) expresses the splitting information of the attribute A:

$$SplitInfo_A(T) = -\sum_{j=1}^{m} \frac{|T_j|}{|T|} \log_2 \frac{|T_j|}{|T|} \tag{25}$$

$SplitInfo_A(T)$: denotes the information generated by dividing the training set T into m partitions of j outputs corresponding to the attribute A.

Different from information gain, the gain ratio is used to evaluate the information obtained from the same partition in a normalized way. The gain ratio of an attribute is calculated by the following formula:

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo_A(T)} \tag{26}$$

In the construction of a decision tree with C4.5, the attribute that has the maximum gain ratio is used for splitting at each step. With recursive calculations, the gain ratios of the attributes are decreased continuously. In the latter stages of the process, the attribute having relatively higher gain ratios is selected.

By combining Eq. (25)–(26) with Eq. (24), the specific computational procedure of the C4.5 algorithm can be obtained.

$$GainRatio(A) = \frac{Info(T) - Info_A(T)}{SplitInfo_A(T)} \tag{27}$$

The process by which C4.5 generates a decision tree can be summarized as follows:

Step1: Create a node N.

Step2: IF the training dataset is empty, THEN return a single node N as an empty leaf node.

Step3: IF all the samples in the training set belong to the same class C, THEN return node N as a leaf node and label the node as class C.

Step4: IF the attribute list in the training set is empty, THEN return node N as a leaf node and mark the node as a class with many samples in the dataset.

Step5: FOREACH AttributeList in the attribute list.

Step6: IF the attribute is continuous, THEN discretize the attribute.

Step7: Calculate the information gain rate of the attribute in the AttributeList according to Equation (27).

Step8: Select the attribute A that has the highest information gain rate in the attributeAttributeList and mark node N as attribute A.

Step9: Delete attribute A in the AttributeList.

Step10: FOREACH attribute value a of attribute A. Branch from node N with a condition $A = a$ to get the subtree.

Step11: Loop Step3-10 in a recursive manner to get a preliminary decision tree.

Step12: Prune the decision tree using a larger training dataset.

When the decision tree construction is performed, it will be judged whether the conditions for stopping the tree construction are met according to the information in the dataset, otherwise the iteration will continue. In general, the main conditions for ending are: the attribute list is empty, all the samples in the dataset have been categorized, and all the remaining samples belong to the same class. If one of these conditions is met, tree building is terminated and the initial decision tree is obtained. Then the post pruning strategy is applied to prune and simplify the decision tree.The flow of the C4.5 algorithm is shown in Figure 8.

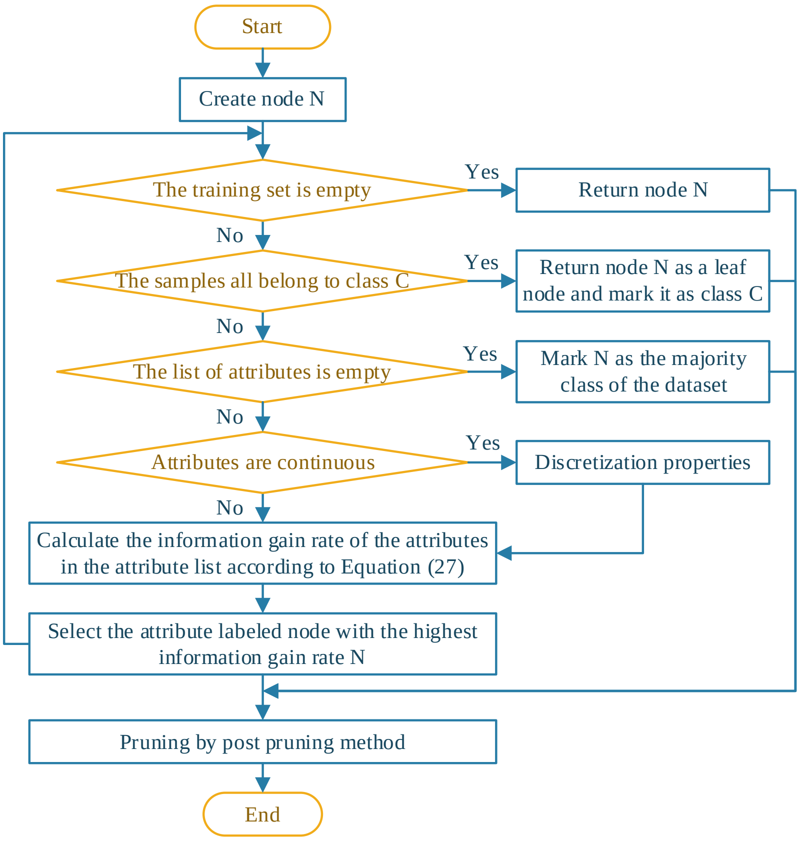

Figure 8: Flowchart of the C4.5 algorithm

4.2 Employee Discipline Predictor Construction

In this research, basic information about employees in the FSSC of Enterprise Z is obtained, and C4.5 decision tree analysis is performed to predict whether employees are likely to commit indiscipline or not, which helps reduce operational risks faced by the FSSC of Enterprise Z. Accuracy in the choice of indicators in evaluating indiscipline among employees is important since it determines the accuracy of the assessment results. Therefore, in order to conduct an accurate assessment of the employees' indiscipline situation in the enterprise and to develop a predictive tool, the choice of indicators needs to be carried out comprehensively. Based on the previous research and with reference to the prediction indicators of default risk in the financial industry, this paper establishes a system of indicators based on the FSSC employee number, employee age, education level, number of errors, years of employment, monthly income, debt assumption, average number of training sessions, and indiscipline by analyzing the basic status of FSSC employees in Enterprise Z.

Among them, specific data on employee number, employee age, education level, years of service, monthly income and average number of training can be collected from the personnel office of FSSC of enterprise Z. Regarding the number of errors and disciplinary cases, the relevant data are collected from the existing Financial Shared Service Center of Enterprise Z FSSC. Data on debt obligations are collected through actual visits and surveys, with approximations or averages used to approximate the extent to which debts can be concealed.

4.3 Initial processing of employee data

Based on the collected employee data, the existing numbering sequence of the employees is ranked. Since the four levels of educational level into specialist, undergraduate, postgraduate as well as PhD cannot be recognized in the decision tree algorithm, they are replaced with Arabic numerals 1 to 4 respectively. At the same time, the disciplinary cases were replaced with 1 and 0 respectively, with 1 representing a disciplinary offense and 0 representing no disciplinary offense. After screening, about 300 valid data were compiled. After screening, a few more representative data were selected as shown in Table 6.

Table 6: Processed employee prediction data

ID	Age	Educational attainment level	Number of errors	Years of service	Monthly income	Bear the debt	Average number of trainings	Violation situation
1	50	2	10	25	13000	300000	25	0
2	44	2	5	18	8500	480000	14	0
3	30	2	12	6	5000	0	8	1
4	27	1	3	4	4000	0	3	0
5	25	1	2	2	3000	0	1	0
6	42	2	9	20	8000	360000	5	1
7	35	3	6	13	6500	0	12	0
8	41	2	7	13	7000	720000	12	1
9	38	3	4	15	8000	540000	12	1
10	39	3	6	14	8800	100000	13	0
…	…	…	…	…	…	…	…	…

4.4 Use of Employee Discipline Prediction Models

In this paper, the data in Table 6 are used to start the C4.5 decision tree method. First of all, the information entropy of the target “disciplinary cases” is calculated, and the organized data is shown in Table 7.

Table 7: The classification of "disciplinary violations" factors

ID	Age	Educational attainment level	Number of errors	Years of service	Monthly income	Bear the debt	Average number of trainings	Violation situation
1	50	2	10	25	13000	300000	25	0
2	44	2	5	18	8500	480000	14	0
4	27	1	3	4	4000	0	3	0
5	25	1	2	2	3000	0	1	0
7	35	3	6	13	6500	0	12	0
10	39	3	6	14	8800	100000	13	0
3	30	2	12	6	5000	0	8	1
6	42	2	9	20	8000	360000	5	1
8	41	2	7	13	7000	720000	12	1
9	38	3	4	15	8000	540000	12	1
…	…	…	…	…	…	…	…	…

From this, then, we learn:

$$Ent(\text{Violation situation}) = -\frac{4}{10}\log_2\frac{4}{10} - \frac{6}{10}\log_2\frac{6}{10} = 1.04465 \tag{28}$$

Next, we select a factor and run the C4.5 decision tree algorithm on it. The first choice is the “education level” factor, and its calculation process is shown in Table 8.

Table 8: Classification of the "educational attainment" factor

ID	Age	Educational attainment level	Number of errors	Years of service	Monthly income	Bear the debt	Average number of trainings	Violation situation
1	50	2	10	25	13000	300000	25	0
2	44	2	5	18	8500	480000	14	0
3	30	2	12	6	5000	0	8	1
6	42	2	9	20	8000	360000	5	1
8	41	2	7	13	7000	720000	12	1
4	27	1	3	4	4000	0	3	0
5	25	1	2	2	3000	0	1	0
7	35	3	6	13	6500	0	12	0
10	39	3	6	14	8800	100000	13	0
9	38	3	4	15	8000	540000	12	1
…	…	…	…	…	…	…	…	…

Next,we calculate the information entropy for each of the three classifications:

Because there are a total of 5 undergraduate students in these 10 sample data extracted, 3 of them have disciplinary cases, therefore:

$$Ent(2) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97095 \tag{29}$$

The 10 sample data centers drawn had a total of 2 specialists and each of them did not have a disciplinary case, so the calculations are shown below:

$$Ent(1) = -\frac{2}{2}\log_2\frac{2}{2} - \frac{2}{2}\log_2\frac{2}{2} = 0 \tag{30}$$

The 10 sample data centers drawn had a total of only three graduate students, one of whom had a disciplinary case, so their information entropy was calculated as shown below:

$$Ent(3) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.91830 \tag{31}$$

After calculating the information entropy of the three categories, we calculate the total information entropy of the “education level” factor:

$$\begin{aligned} Ent(\text{Educational attainment level}) &= \frac{5}{10}*0.97095 - \frac{2}{10}*0 + \frac{3}{10}*0.91830 \\ &= 0.76097 \end{aligned} \tag{32}$$

After calculating the information entropy of "education level", the information gain of "education level" is calculated as follows:

$$Gain(\text{Educational attainment level}) = 1.04465 - 0.76097 = 0.28368 \tag{33}$$

Next, the split information for "level of education" is calculated:

$$Spl(\text{Educational attainment level}) = -\frac{5}{10}\log_2\frac{5}{10} - \frac{2}{10}\log_2\frac{2}{10} - \frac{3}{10}\log_2\frac{3}{10} \\ = 1.48548 \tag{34}$$

Next, the information gain rate for "education level" is calculated:

$$Gain\_Ratio(\text{Educational attainment level}) = 0.28368/1.48548 = 0.19097 \tag{35}$$

Having computed the information gain of the education level variable, we proceed with calculating the information gain of the remaining variables using the same method. We choose the greatest one among the sizes of information gain of these factors as the root node and solve all the next nodes to create a full decision tree. During actual implementation, the management is able to forecast whether an employee is disciplinary or not using the available decision tree of the organization and the given circumstances of the employee.

In addition to the above decision tree model for predicting employee disciplinary behavior, in order to solve the problem of internal employee discipline, the characteristics of FSSC's information system and process setup can also be utilized, and internal control can be considered when setting up the information system, so that the employee can be traced and traces can be found at every important step. In this way, the traces left in the system can be used to find the responsible person in case of a security problem with the information. At the same time, employees are made aware of the existence of such a method so that they can be reminded not to violate discipline. In addition, it is important to strengthen the ideology and morality of employees to eliminate the root causes of internal employee discipline risks. If an employee ever violates discipline, he or she should be held strictly accountable to demonstrate discipline.

5 Enterprise financial management risk prediction based on BP neural network

In this chapter, a BP neural network is employed to construct an enterprise financial management risk prediction model, with the aim of providing decision support for enterprise financial risk management.

5.1 Financial risk prediction model based on BP neural network

In the process of integrating risk prediction indicators various indicators tend to be measured on different scales. In order to overcome the issue that outcomes cannot be directly compared due to the different scales, the multilayer feedforward neural network trained with the error back-propagation algorithm, i.e. the BP neural network is used to conduct dimensionless process on every evaluation indicator. Then learning is applied to change the connections between the neurons of each layer.

Set the input neuron as $m(i = 1, 2, \ldots, x)$ and the output neuron as $n(j = 1, 2, \ldots, y)$. The specific number of hidden layer neurons is obtained based on the relationship between the hidden layer neurons and the input neurons, and the range of values of the y th neuron $\left(m_y, n_y\right)$ of the input neuron is known, then $m_y = \left(m_1^y, m_2^y, m_3^y, \ldots, m_x^y\right), n_y = \left(n_1^y, n_2^y\, n_3^y, \ldots, n_x^y\right)$.

The weighted sum for the j th input cell n_j is:

$$S_{nf} = \sum_{k-1}^{q} R_x c_k \tag{36}$$

In Eq. (36), c_k denotes the k th hidden layer unit and R_x is the neuron weights, then the actual output of the unit is as follows:

$$n_f = f\left(S_{nf}\right) = \frac{1}{B^{-S_{nf}}} \tag{37}$$

In Eq. (37), $f(\cdot)$ denotes the sigmoid function, B is the neuron output layer output value, and the weighted input sum of the k th hidden layer unit is as follows:

$$S_{ck} = \sum_{k=1}^{q} R_{jk} c_k \tag{38}$$

In Eq. (38), R_{jk} denotes the connection weights of the j th output unit and the k th hidden layer.

Due to the fact that the number of samples chosen to conduct the experiment is fairly small, random filtration can cause losing information. Consequently, working capital asset ratio, debt-equity ratio, long-term debt-equity ratio, accounts receivable turnover ratio, return on net assets, loss ratio, main business growth rate, and total asset growth rate are used as the predictive variables on enterprise financial management risk. The input layer has 8 units in this model, but the output layer has anywhere from 0 to 1 unit. A 0 is used to signify that there is no financial risk in the enterprise, and a 1 is used to signify that there is financial risk. In simpler terms, the BP neuron network has 8 neurons in the input layer, and the output layer generates a value between 0 and 1, with outcomes near 0 indicating less financial risk and outcomes near 1 indicating more financial risk.

5.2 Model empirical tests

The research object of this paper is the sample data of financial information risk management of 75 listed industrial enterprises in the A-share market in 2021-2023. The comparative research approach has been used to compare the predictive performance of the LPM risk prediction model and the BP-neural-network-based financial information management risk prediction model.

Table 9: Financial information risk management index system

Dimension	Index	Coding
Short-term debt-paying ability	Working capital asset ratio	A_1
Long-term debt-paying ability	Debt-to-equity ratio	A_2
	Long-term debt-to-equity ratio	A_3
Operational capability	Accounts receivable turnover ratio	A_4
Profitability	Return on net assets	A_5
	Loss ratio	A_6
Growth ability	Growth rate of main business	A_7
	Total asset expansion rate	A_8

Based on the stepwise regression method to select the data with F-value less than 0.1 from the sample data of the indicator system in the last three years, the results of the variables obtained from the stepwise regression in each year are shown in Table 10.

Table 10: The results of stepwise regression variables for each year

Year		2021	2022	2023
Nodal increment		0.364	0.635	0.203
Dimension	Index			
Short-term debt-paying ability	A_1	0.538*	-	0.654*
Long-term debt-paying ability	A_2	-0.282*	-	-
	A_3	-	-	-0.176
Operational capability	A_4	-1.803*	-1.502*	-
Profitability	A_5	-0.014	0.006*	0.035*
	A_6	-	-	-0.206*
Growth ability	A_7	-	-	-0.044
	A_8	-0.275	-	-

In order to prevent multicollinearity in the course of predicting the risk in management of financial information, BP neural network prediction model is initially checked on multicollinearity and the findings are presented in Table 11. It can be observed in the table that there is no significant multicollinearity between any of the eight chosen evaluation indicators.

Table 11: Multicollinearity detection

Index	TOF	VIF
A_1	0.381	2.628
A_2	0.938	1.047
A_3	0.864	1.163
A_4	0.252	3.954
A_5	0.873	1.137
A_6	0.495	2.063
A_7	0.643	1.834
A_8	0.979	1.572

The reported regression results of the LPM risk prediction model and the BP-neural-network-based financial information management risk prediction model applied to the chosen 75 listed industrial enterprises are given in Tables 12 and 13, respectively. By comparing the results of prediction of the two models it can be seen that regression results of the BP neural network model are closer to the maximum likelihood value indicating that this model has a better fit and is more accurate in predicting risk.

Table 12: Regression results of the LPM risk prediction model

Model	$R^2 = 0.674$		$\overline{R}^2 = 0.642$	F=21.64	F=0.000
	Parameter value	Error	BETA	T	P
Nodal increment	0.429	0.061	-	6.458	0.000
A_1	0.475	0.002	0.264	3.271	0.002
A_2	0.269	0.002	0.211	2.984	0.003
A_3	-0.071	0.002	-0.109	-1.473	0.119
A_4	-1.072	0.003	-0.631	-4.451	0.000
A_5	-0.014	0.001	-0.162	-2.135	0.043
A_6	-0.521	0.000	-0.335	-3.427	0.002
A_7	-0.252	0.000	-0.321	-1.268	0.056
A_8	1.267	0.004	-0.342	-1.959	0.001

Table 13: Based on the regression results of the BP neural network risk prediction model

Model	$R^2 = 0.674$		$\overline{R}^2 = 0.642$	F=21.64	F=0.000
	Parameter value	Error	BETA	T	P
Nodal increment	-0.453	0.438	0.247	4.138	0.000
A_1	0.337	0.000	0.374	3.124	0.002
A_2	0.226	0.000	0.135	2.408	0.001
A_3	-0.135	0.000	-0.029	-1.538	0.025
A_4	1.274	0.001	0.625	-2.405	0.000
A_5	-0.028	0.001	-0.121	-1.273	0.015
A_6	-0.649	0.001	0.132	3.044	0.000
A_7	0.163	0.001	0.417	0.249	0.032
A_8	1.374	0.002	-0.335	-0.631	0.000

In order to facilitate an easier comparison between error rate, misclassification rate, and the accuracy rate of both models, the related data are tabulated below in Table 14. Upon comparing the results of the sample data, it can be seen that both risk error detection rate and incorrect classification rate of the BP neural network–based financial information prediction model is much lower compared to the LPM model, demonstrating a definite advantage in terms of predictive precision. It should be pointed out that in 2023, its value achieves 98.27%, which exceeds the LPM model by 6.90%.

Table 14: Comparison of Indicators between the two models

Year	Risk identification error rate		Risk misjudgment rate		Accuracy rate of risk prediction	
	LPM model	BP model	LPM model	BP model	LPM model	BP model
2021	2.42%	1.33%	4.62%	0.14%	92.34%	97.43%
2022	2.56%	1.32%	3.35%	0	93.45%	98.24%
2023	3.15%	1.12%	4.10%	0.04%	91.37%	98.27%

The results of the comprehensive analysis of the sample data from the test results for the chosen firms from the list are demonstrated in Figures 9 to 12. These findings show that the financial risk management model based on a BP neural network meets all requirements of a reliable predictive tool. Therefore, it allows for identifying and predicting financial management situations at industrial companies in future accounting periods.

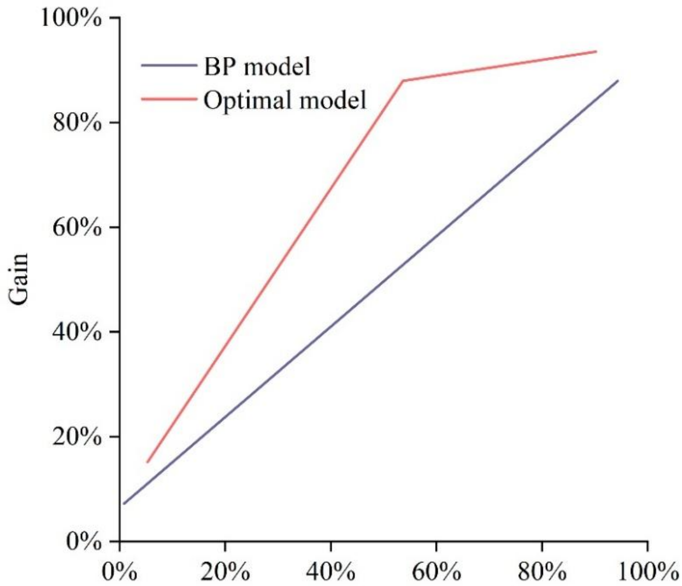

Figure 9: Risk model gain based on BP neural network algorithm

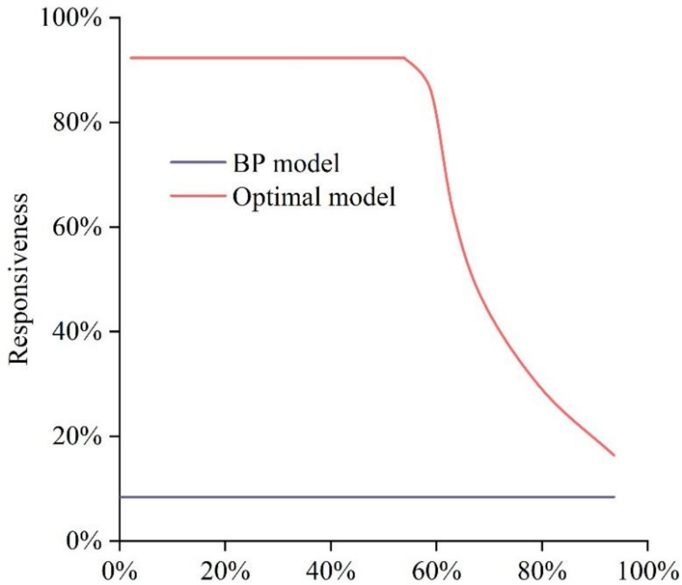

Figure 10: Risk model response based on BP neural network algorithm

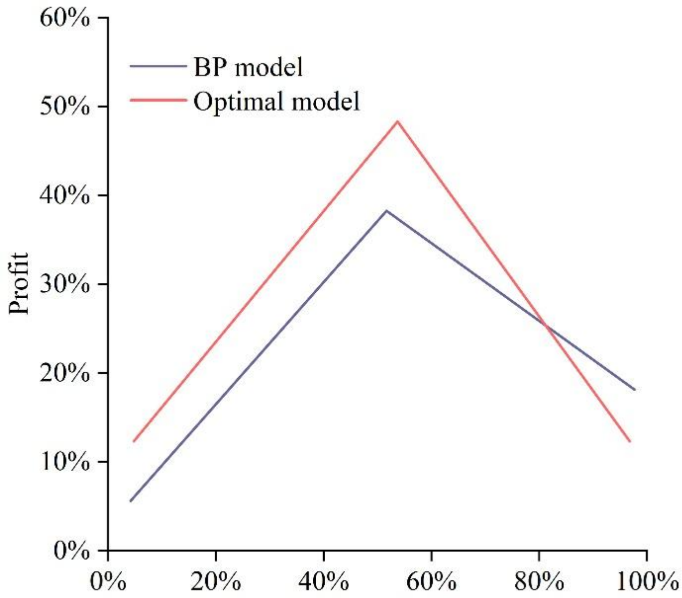

Figure 11: Profit changes based on BP neural network algorithm

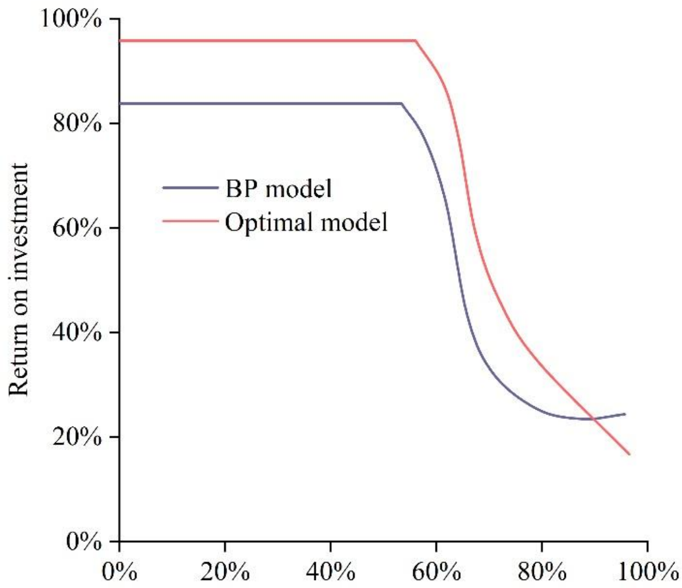

Figure 12: The change in return on investment based on BP neural network algorithm

6 Conclusion

Based on the algorithms of system clustering, factor analysis, C4.5 decision tree and BP neural network, this paper implements the construction of FSSC operation and management optimization model based on RPA, and then realizes the optimization design of intelligent financial internal control system of industrial enterprises.

(1) Through systematic clustering, this paper divides the 75 sample industrial enterprises into six major categories. The five discriminant functions explain 67.88%, 16.05%, 9.48%, 4.99%, and 1.60% of the data, respectively, with a total explanation rate of 100% and high typical correlation coefficients. Meanwhile, the five discriminant functions directly have significant differences and discriminatory power, all of which are statistically significant. And the accuracy of the clustering results is as high as 97.33%, the sample enterprises are centered

around six groups, and the clustering analysis results are reasonable. In addition, three principal components are extracted from the enterprise credit risk factors, and the enterprises can be subdivided by calculating their factor scores: the first category belongs to high-risk enterprises, and the second, third and fourth categories belong to medium-high risk enterprises. Among them, the second category is weak overall strength, the third category is excessive debt pressure, and the fourth category is limited profitability. The fifth and sixth categories belong to enterprises with better development level and healthy corporate finance.

(2) By constructing the C4.5 decision tree, the management can employee's relevant situation and accurately predict whether the employee is disciplinary or not. In addition, the root cause of internal employee discipline risk can be eliminated by strengthening the ideological and moral construction of employees. If an employee once violates the discipline, he/she should be strictly held responsible to reflect the discipline.

(3) BP neural network-based financial information prediction system proposed in this paper has significantly low risk error identification rate and risk misclassification rate compared to the LPM model indicating a significant benefit in predictive accuracy. In particular, the classification accuracy of this framework in 2023 is 98.27 per cent, which exceeds the LPM model by 6.90 per cent and also proves that it can accurately identify and forecast the financial management situation of the industrial companies in the subsequent accounting periods.

Funding

This research was supported by the Ministry of Education Industry-Education Cooperation Collaborative Talent Training Project: Construction and Practice of the Innovation Center for Digital and Intelligent Education Transformation of "One Heart, Two Uses" in Business and Trade Majors (Project Number: 241004978181156).

About the Author

Chengbing He, Associate Professor, he was born in 1972, Anqing city, Anhui Province, China. he obtained his Master's degree in Business Administration from Lanzhou Jiaotong University. Currently, employed at Nanchang Vocational University, where he teaches finance and economics courses. His research in business management for small and medium-sized enterprises.

Huanxia Hao, Associate Professor, was born in 1981 in Yuanping City, Shanxi Province, China. She obtained her Master's degree in Business Administration from Shijiazhuang Tiedao University. Currently, she works at the School of Economics and Management, Xinjiang University of Political Science and Law. Her research interests focus on Business Administration and Regional Economic Development.

Yuanyuan Zhou, Associate Professor, she was born in 1986 in Zhumadian, Henan Province, China. She obtained a Master's degree in Software Engineering from Jilin University. Currently employed at Nanchang Vocational University, she teaches courses in economics and management, with her research focus on enterprise management.

Jiaqi Peng, born in January 1993 in Yichun, Jiangxi Province, China, graduated from Jiangxi University of Finance and Economics. She is currently the Service Director of XinDao Technology Co., Ltd.

Xiongjun Wang, Affiliation: Institute of Economic Development, Nanchang Vocational University, Fenghuang Development Zone, Anyi County, Nanchang, Jiangxi Province, 330500, China.

References

[1] Zeng, H., Shao, Y., Ding, N., Zheng, L., & Zhao, J. (2025). Internal and External Cultivation to Drive Enterprises’ Green Transformation: Dual Perspectives of Vertical Supervision and Environmental Self-Discipline. Sustainability, 17(15), 7062.

[2] Rahman, M. M., & Ashfaq, S. (2021). DATA-DRIVEN DECISION SUPPORT IN INFORMATION SYSTEMS: STRATEGIC APPLICATIONS IN ENTERPRISES. International Journal of Scientific Interdisciplinary Research, 2(2), 01-33.

[3] Subramanyam, S. V. (2024). Transforming financial systems through robotic process automation and AI: The future of smart finance. International Journal of Artificial Intelligence Research and Development (IJAIRD), 2(1), 203-223.

[4] Wensheng, D. (2020). Rural financial information service platform under smart financial environment. Ieee Access, 8, 199944-199952.

[5] Gao, M., He, Q., & Ruan, X. (2023). Financial Big Data Intelligent Service System Based on Cloud Computing of Internet of Things. Mobile Information Systems, 2023(1), 6093197.

[6] Ardian, Z., Abdullah, D., Bintoro, A., Yusra, M., Akbar, A. H., & Lubis, F. A. (2024). A Smart Accounting System for Real-Time Mosque Financial Fund Management Based On Android and Web Mobile through The Implementation of The Agile Development Scrum Method. JINAV: Journal of Information and Visualization, 5(2), 220-228.

[7] Marzhan, Y., Talshyn, K., Kairat, K., Belginova, S., Karlygash, A., & Yerbol, O. (2022). Smart technologies of the risk-management and decision-making systems in a fuzzy data environment. Indonesian Journal of Electrical Engineering and Computer Science, 28(3), 1463-1474.

[8] Ionescu, S. A., & Diaconita, V. (2023). Transforming financial decision-making: the interplay of AI, cloud computing and advanced data management technologies. International Journal of Computers Communications & Control, 18(6).

[9] Syed, R., Suriadi, S., Adams, M., Bandara, W., Leemans, S. J., Ouyang, C., ... & Reijers, H. A. (2020). Robotic process automation: contemporary themes and challenges. Computers in industry, 115, 103162.

[10] Patrício, L., Varela, L., & Silveira, Z. (2025). Framework for Integrating Requirements Engineering and DevOps Practices in Robotic Process Automation with a Focus on Optimizing Human–Computer Interaction. Applied Sciences, 15(7), 3485.

[11] Chakraborti, T., Isahagian, V., Khalaf, R., Khazaeni, Y., Muthusamy, V., Rizk, Y., & Unuvar, M. (2020, September). From Robotic Process Automation to Intelligent Process Automation: –Emerging Trends–. In International Conference on Business Process Management (pp. 215-228). Cham: Springer International Publishing.

[12] Xu, L., Wang, C., Luo, X., & Shi, Z. (2006). Integrating knowledge management and ERP in enterprise information systems. Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research, 23(2), 147-156.

[13] Ratchatawetchakul, Y., Bansri, B., Ratchatawetchakul, K., Satchawatee, N., Ratchatawetchakul, W., & Eampoonga, I. (2024, June). Integrating Robotic Process Automation (RPA) with Hybrid Cloud Enterprise Resource Planning (ERP). In 2024 5th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON) (pp. 1-4). IEEE.

[14] Zhan, J. X., Ling, Z., Xu, Z., Guo, L., & Zhuang, S. (2024). Driving efficiency and risk management in finance through AI and RPA. Journal of Advanced Computing Systems, 4(5), 1-9.

[15] Mamidala, V., Yallamelli, A. R. G., & Yalla, R. K. M. (2022). Leveraging robotic process automation (RPA) for cost accounting and financial systems optimization—A case study of ABC company. ISAR International Journal of Research in Engineering Technology, 7(6).

[16] Tang, W., Cao, H., Ye, S., Yang, L., & Chen, F. (2024, August). Intelligent Financial Decision Support System Based on RPA Financial Robot and Artificial Intelligence. In 2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC) (pp. 943-948). IEEE.

[17] Vishnu, S., Agochiya, V., & Palkar, R. (2017). Data-centered dependencies and opportunities for robotics process automation in banking. Journal of Financial Transformation, 45, 68-76.

[18] Willcocks, L., Lacity, M., & Craig, A. (2017). Robotic process automation: strategic transformation lever for global business services?. Journal of Information Technology Teaching Cases, 7(1), 17-28.

[19] Lyamin, B. M., & Voronova, O. V. (2023). RPA Technology as a Tool for Boosting the Efficiency of an Industrial Enterprise under Digital Transformation. Technoeconomics, 2(2), 5.

[20] Jasińska, K., Lewicz, M., & Rostalski, M. (2023). Digitization of the enterprise-prospects for process automation with using RPA and GPT integration. Procedia Computer Science, 225, 3243-3254.

[21] Saha, R., Kulahri, V., Kumar, G., Rai, M., & Lim, S. J. (2019). Analyzing the tradeoff between planning and execution of robotics process automation implementation in it sector. Int. J. Control Autom., 12(1), 1-10.

[22] Bheemarpu, N. S. U. K. (2025). RPA in Salesforce: Bridging Automation Gaps in Enterprise Systems. Journal of Computer Science and Technology Studies, 7(3), 866-872.

[23] Jangam, S. K., & Karri, N. (2024). Hyper Automation, a Combination of AI, ML, and Robotic Process Automation (RPA), to Achieve End-to-End Automation in Enterprise Workflows. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 5(1), 92-103.

[24] Salmen, A. (2022). Employing RPA and AI to automize order entry process with individual and small-sized structures: a SME business case study. Acta Academica Karviniensia, 22(2), 78-96.

[25] Chennupati, N. (2025). Cognitive RPA: A Framework for Hybridizing Artificial Intelligence with Robotic Process Automation in Enterprise Systems. European Journal of Computer Science and Information Technology, 13(19), 64-78.

[26] Ma, J., & Jia, H. (2022, December). Application of financial robots based on RPA technology in small and medium-sized enterprises. In 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES) (pp. 1-7). IEEE.

[27] Ma, J., & Wang, R. (2023, April). RPA Financial Robot Boosts the Digital and Intelligent Transformation of Enterprise Finance. In 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE) (pp. 1-7). IEEE.

[28] Guo, Y., Wang, J., Yao, X., & Wang, W. (2023, July). Application of RPA-AI technology in enterprise financial intelligence. In 2023 19th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 1-6). IEEE.

[29] Zhang, S. (2024, November). Reflections on the Application of RPA in the Digital Transformation of Enterprise Finance. In 2024 5th International Conference on Management Science and Engineering Management (ICMSEM 2024) (pp. 618-631). Atlantis Press.

[30] Lan, L., & Chen, H. (2025). RPA Financial Automation Process System. Innovative Computing 2025, Volume 3: Proceedings of IC 2025: Innovative System, 1442, 1.

[31] Zhang, C., Issa, H., Rozario, A., & Soegaard, J. S. (2023). Robotic process automation (RPA) implementation case studies in accounting: A beginning to end perspective. Accounting Horizons, 37(1), 193-217.

[32] Zhang, Y., Yi, A., & Li, S. (2025). Development and application of financial statement filing robot based on RPA technology. Journal of Intelligent & Fuzzy Systems, 49(1), 153-162.

[33] Qiu, Y. L., & Xiao, G. F. (2020). Research on cost management optimization of financial sharing center based on RPA. Procedia Computer Science, 166, 115-119.

[34] Zhou, S. (2021, April). Research on banking process optimization based on RPA in financial sharing mode. In Journal of Physics: Conference Series (Vol. 1865, No. 2, p. 022069). IOP Publishing.

[35] Wang, L., Shen, Y., Liang, S., Zhou, N., Jia, C., & Shang, P. (2023, April). Application Design of RPA Financial Robot Integrating Financial Big Data and Financial Sharing Services. In International Conference on Computational Finance and Business Analytics (pp. 63-72). Cham: Springer Nature Switzerland.