

Audit Risk Evaluation Model for Financial Statement Based on Artificial Intelligence

Yanhua Li

School of Accounting, Wuhan College, Wuhan, China

In recent years, the economy in China has been steadily improving. The financial situation of enterprises in mainstream industries has become the focus of public concern. However, financial statement frauds, which occur frequently, greatly disrupt the economic order in the country. Thus, it is of practical significance to accurately identify and evaluate the audit risks of financial statements. For this purpose, this paper proposes an audit risk evaluation model of financial statement based on artificial neural networks (ANN). Firstly, the authors designed the audit risk indices and quantified the fraud factors of financial statement. Next, an audit risk verification model was established for financial statement and used to verify the predictions on three aspects of financial statement, namely, audit violation penalty (AVP), audit violation announcement (AVA), and financial statement restatement (FSR). Finally, a feedforward neural network was constructed based on the homomorphic encryption algorithm, which was subsequently used to evaluate and predict the audit risks of financial statements. The effectiveness of our model was proved valid through experiments.

ACM CCS (2012) Classification: Applied Computing → Enterprise computing → Business process management

Keywords: artificial neural network (ANN), financial statement, audit risks

1. Introduction

In recent years, the economy in China has been steadily improving, providing a solid support to the development and growing demands of various industries. Currently, the financial situation of enterprises in mainstream industries has become the focus of public concern, such as the

Currently, the financial situation of enterprises in mainstream industries, such as the Internet, finance, industrial automation, bioengineering, medical care, and real estate [1–3], has become the focus of public concern. The public has raised stricter requirements on the auditing efficiency and quality of financial statement on each project of enterprises [4–7].

Unfortunately, financial statement frauds occurred frequently in the past few years, which greatly disrupted the economic order in the interested countries [8–11]. To prevent the risk of material misstatement, The Chinese Institute of Certified Public Accountants has issued *regular Auditing Standards for CPAs of China No. 1211 – Understanding the entity and its environment and assessing the risks of material misstatement* [12]. Therefore, it is of practical significance to accurately identify and evaluate the audit risks of financial statement.

So far, the audit risks of financial statement have been studied effectively at home and abroad [13–16]. Ziolo *et al.* [17] developed a risk evaluation model based on internal control and operating risks, and empirically explored the impact of business environment on the audit risks of financial statement. Lucarelli *et al.* [18] argued that the audit quality of financial statement mainly depends on the economic policy and the professionalism of auditors; if the auditors are staffed properly, the audit can be implemented independently at a low labor cost, greatly reducing the audit risk. Balan *et al.* [19] constructed an innovative audit

risk-oriented model for financial statement of financial enterprises, summarized the factors affecting audit risks, and provided the strategies to improve the existing audit methods. Drawing on the theory of risk-oriented auditing, Cerqueti and Spizzichino [20] identified the material misstatement risks of the target enterprises through case analysis, which focused on such aspects as macroeconomy, industry demand, internal control, and financial statement, selected 3–4 correlated items for audit risk control tests, and designed concrete steps for auditing. Chen and Khashanah [21] shifted the focus from audit risks of financial statement to the quantification of material misstatement risks, plotted the risk transmission diagram for internal risk control of key businesses in cultural media enterprises, and quantified the risks through principal component analysis (PCA). Financial risk control can be transformed into the processing of multi-source data, making it an ideal application scenario of artificial intelligence (AI). AI-based financial risk control can effectively ensure security and adapt to the high dimensionality of financial data, providing a way to implement AI in financial auditing [22–24]. Pandey and Haes [25] prepared a financial risk control scheme based on immune neural networks, developed both the network and the security model in the design phase, and verified their performance. Sharma *et al.* [26] identified the factors affecting the material misstatement risks of financial statement, analyzed the sources of material misstatement risks in listed enterprises, and predicted these risks with both a self-designed logistic regression model and a neural network model, suggesting that the material misstatement risks are affected to varying degrees by factors in four aspects: the integrity of the listed enterprise, the external business environment, the internal control structure, and key projects.

At present, studies on the audit risks of financial statement mostly concentrate on internal control of enterprises, or tackle auditing of the annual report, economic responsibility, and liquidation of a single industry. Relatively few scholars have evaluated the audit risks of financial statements. However, as the financial market is developing and improving, the information structure of domestic financial statements and the audit risk mechanism have shown certain

deficiencies, which has resulted in the fact that the financial data calculated based on these deficient financial statements cannot well reflect the true financial status of domestic enterprises; at the same time, the few researchers engaging in material misstatement risks merely qualified the influencing factors, failing to quantify or empirically analyze them. To accurately evaluate the audit risks of financial statement, this paper creates an audit risk evaluation model suitable for financial statements, by virtue of the power of artificial neural networks (ANN) in data processing.

The remainder of this paper is organized as follows: Section 2 identifies the audit risk indices of financial statement and explains the process of quantifying fraud factors; Section 3 introduces a verification model for financial statement and applies it to verify the prediction on three aspects: audit violation penalty (AVP), audit violation announcement (AVA), and financial statement restatement (FSR); Section 4 constructs a feedforward neural network based on the homomorphic encryption algorithm; Section 5 evaluates the audit risk of financial statement with the proposed encrypted neural network; Section 6 verifies the effectiveness of the constructed model through experiments; Section 7 puts forward the conclusions of the research. Compared with the SVM (Support Vector Machine) + ANN model, the SVM model, the GM (Gaussian Mixture) model and other models proposed in references [15], [16], [23], and [24], the proposed model has higher prediction accuracy, lower time complexity, and better performance.

2. Risk Index Design and Fraud Factor Quantification

Existing risk index selection methods have the following three shortcomings:

1. evaluation indexes are not collected and classified according to certain principles, which results in insufficient logicity and low reuse rate;
2. description of the audit risk assessment items of the financial statement is not detailed enough and has poor operability;

Table 1. Indices for audit risks of financial statement.

Criteria	Names	Symbols	Definitions
Fraud motivation <i>FM</i>	Consecutive years of loss	Y_{CL}	Indicates whether the net profit is negative
	Proportion of non-operating net profit and loss	P_{NPL}	Indicates whether there is a serious non-operating loss
	Proportion of guaranteed amount	P_{GA}	Indicates whether there is an illegal or set-over guarantee
	Cash flow ratio	R_{CF}	Indicates whether there is a reliable guarantee for the repayment of current liabilities
	Under special treatment (ST) or delisting risk (*ST)?	W_{ST}	Indicates whether the listed enterprise has suffered losses for one or two consecutive fiscal years, and received special treatment from the stock exchange
	State-controlled?	W_{SOH}	Indicates whether there is an excessively high pressure on performance
Fraud opportunity <i>FO</i>	Proportion of connected transactions	P_{RPT}	Indicates whether the parent enterprise has an excessively large proportion of connected transactions
	Total number of shares outstanding	TN_{CS}	Indicates whether the proportion of outstanding shares is too small
	Proportion of public shareholders	P_{PS}	Indicates whether there are public shareholders attending the general meeting
	Percentage of shares held by shareholders attending the general meeting	SR_{GM}	Indicates whether the decision-making power of the enterprise is subject to strong supervision
	Attribution of audit responsibility	O_{AR}	Indicates whether the audit of the enterprise is carried out by a large accounting firm

3. selection methods have not introduced any mathematical model of audit risk prediction, so preciseness needs to be further improved.

Aiming at these shortcomings and drawing on the existing literature, this paper has identified the fraud motives and fraud opportunities in financial statements as the most direct substitute variables for fraud behaviors in financial

reports. Therefore, this paper tries to assess whether there is fraud in the financial statement within the framework of the two substitutive factors. These two factors were broken down into detailed indices, and each index was allocated to one of the four levels: excellent, good, medium, and poor. Table 1 explains the indices for audit risks of financial statement.

The above-mentioned fraud factors need to be further quantified before evaluating the audit

risks of the financial statement for an enterprise. Therefore, this paper constructs an audit risk evaluation model for financial statement, and uses the model to calculate the controllable accrued operating net profit and the real earnings management of the enterprise. Firstly, a joint regression model was created to compute the accrued earnings quality:

$$\left\{ \begin{array}{l} \frac{ACC(t)}{TA(t-1)} = \lambda_{ACC1} \frac{1}{TA(t-1)} + \\ \quad + \lambda_{ACC2} \frac{\Delta GR(t)}{TA(t-1)} + \\ \quad + \lambda_{ACC3} \frac{FA(t)}{TA(t-1)} + \delta(t) \\ \frac{ACC(t)}{TA(t-1)} = \lambda_{ACC1} \frac{CF(t-1)}{TA(t-2)} + \\ \quad + \lambda_{ACC2} \frac{CF(t)}{TA(t-1)} + \\ \quad + \lambda_{ACC3} \frac{CF(t+1)}{TA(t)} + \\ \quad + \lambda_{ACC4} \frac{\Delta GR(t)}{TA(t-1)} + \\ \quad + \lambda_{ACC5} \frac{FA(t)}{TA(t-1)} + \delta(t) \end{array} \right. \quad (1)$$

where $ACC(t)$ is the accrued operating net profit in the period t ; $\Delta GR(t)$ is the increase in operating income in the period t compared to period $t-1$; $FA(t)$ is the book value of fixed assets in the period t ; $CF(t)$ is the cash flow of the enterprise during normal operation in the same period.

As shown in formula (1), all parameters are compared with the total assets $TA(t)$ of the enterprise in the period t , thereby reducing the interference of heteroscedasticity. During regression, the number of parameter observations is required to reach the preset value, and the controllable accrual item is the controllable accrued operating net profit $M_ACC(t)$, i.e., the regression residual of formula (1).

Before computing the real earnings management of the enterprise, it was assumed that the enterprise management promotes corporate performance by controlling the sales, production, and expenses through normal operating activities. The net operating cash flow, production cost, as well as management and sales cost

can be solved by regressing the data on discount, cost, revenue, production volume, and expenditures on advertising, and research and development (R&D) in each year and in each industry. The net operating cash flow CF can be calculated by:

$$\begin{aligned} \frac{CF(t)}{TA(t-1)} &= \lambda_{CF1} \frac{1}{TA(t-1)} + \\ &+ \lambda_{CF2} \frac{SR(t)}{TA(t-1)} + \\ &+ \lambda_{CF3} \frac{\Delta SR(t)}{TA(t-1)} + \delta(t) \end{aligned} \quad (2)$$

where $SR(t)$ is the sales revenue in the period t ; $\Delta SR(t)$ is the increase in sales revenue in the same period. Under normal sales and production circumstances, the cost $GSC(t)$ of goods sold in this period can be calculated by:

$$\begin{aligned} \frac{GSC(t)}{TA(t-1)} &= \lambda_{GSC1} \frac{1}{TA(t-1)} + \\ &+ \lambda_{GSC2} \frac{SR(t)}{TA(t-1)} + \delta(t) \end{aligned} \quad (3)$$

The corresponding inventory change $CI(t)$ in the period t can be calculated by:

$$\begin{aligned} \frac{CI(t)}{TA(t-1)} &= \lambda_{CI1} \frac{1}{TA(t-1)} + \\ &+ \lambda_{CI2} \frac{\Delta SR(t)}{TA(t-1)} + \\ &+ \lambda_{CI3} \frac{\Delta SR(t-1)}{TA(t-1)} + \delta(t) \end{aligned} \quad (4)$$

The production cost $PC(t)$ in this same period can be calculated by:

$$\begin{aligned} \frac{PC(t)}{TA(t-1)} &= \lambda_{PC1} \frac{1}{TA(t-1)} + \\ &+ \lambda_{PC2} \frac{SR(t)}{TA(t-1)} + \\ &+ \lambda_{PC3} \frac{\Delta SR(t)}{TA(t-1)} + \\ &+ \lambda_{PC4} \frac{\Delta SR(t-1)}{TA(t-1)} + \delta(t) \end{aligned} \quad (5)$$

Under normal sales and production circumstances, the management and sales cost $MSC(t)$ can be calculated by:

$$\begin{aligned} \frac{MSC(t)}{TA(t-1)} = & \lambda_{MSC1} \frac{1}{TA(t-1)} + \\ & + \lambda_{MSC2} \frac{SR(t-1)}{TA(t-1)} + \delta(t) \end{aligned} \quad (6)$$

The controllable amount of expenses $M_CF(t)$, $M_PC(t)$, and $M_MSC(t)$ can be obtained by deducting the actual net operating cash flow, production cost, and management and sales cost with the corresponding values fitted by (2), (5), and (6).

The controllable amount of the real earnings RE(t) of the enterprise can then be calculated by

$$M_RE(t) = M_PC(t) - M_CF(t) - M_MSC(t).$$

Like the calculation of accrued operating net profit, the number of parameter observations must reach the preset value in order to regress the real earnings management model.

3. Verification of the Audit Risk Model

The source of audit risks of financial statements is closely related to the involved activities of the company being audited. For corporate financial statements of different industries, fraud motives mainly comprise cover-up of business difficulties, avoidance of delisting (*i.e.*, removal from a stock exchange) risks, making false profits to obtain future loans, and anyway obtaining big money. In terms of general risk factors, potential fraud opportunities mainly include over-concentrated equity, low probability of being discovered, and mild punishment; hence, it is necessary to implement the corresponding prevention and treatment measures.

To verify the assumptions and calculations of audit risks and fraud factors for financial statement, a verification model was designed to test the predictions of AVP, AVA, and FSR:

$$\begin{aligned} S_{VP}(t+1) = & \lambda_{VP1} M_ACC(t) + \\ & + \lambda_{VP2} M_RE(t) + \\ & + \lambda_{VP3} FM(t) + \\ & + \lambda_{VP4} FO(t) + \\ & + \lambda_{VP5} S_{VP}(t) + \delta(t), \end{aligned} \quad (7)$$

where S_{VP} is the predictor variable of AVP. If the financial statement disclosed in the period t involves one or several kinds of frauds, namely, fictitious profit, fictitious debt, and fictitious asset, $S_{VP} = 1$; otherwise, $S_{VP} = 0$. Let S_{VN} be the predictor variable of AVA. If the regulator issues a violation announcement about the fraud(s) in the financial statement disclosed in the period t , then $S_{VN} = 1$; otherwise, $S_{VN} = 0$.

$$\begin{aligned} S_{VN}(t+1) = & \lambda_{VN1} M_ACC(t) + \\ & + \lambda_{VN2} M_RE(t) + \\ & + \lambda_{VN3} FM(t) + \\ & + \lambda_{VN4} FO(t) + \\ & + \lambda_{VN5} S_{VN}(t) + \delta(t) \end{aligned} \quad (8)$$

Let S_{FR} be the predictor variable of FSR. If the enterprise corrects the disclosed financial statement under external regulatory pressure, $S_{FR} = 1$; otherwise, $S_{FR} = 0$.

$$\begin{aligned} S_{FR}(t+1) = & \lambda_{FR1} M_ACC(t) + \\ & + \lambda_{FR2} M_RE(t) + \\ & + \lambda_{FR3} FM(t) + \\ & + \lambda_{FR4} FO(t) + \\ & + \lambda_{FR5} S_{FR}(t) + \delta(t) \end{aligned} \quad (9)$$

The above models share the same main parameters. As can be seen from expressions (7)–(9), the prediction of $S_{VP}(t+1)$, $S_{VN}(t+1)$, and $S_{FR}(t+1)$ in the period $t+1$ needs to consider two kinds of eigenvariables at the enterprise level *i.e.*, (controllable accrued operating net profit $M_ACC(t)$, and real earnings management $M_RE(t)$), while controlling the effects of fraud motivation $FM(t)$ and fraud opportunity simultaneously, as well as the actual values of $S_{VP}(t)$, $S_{VN}(t)$, and $S_{FR}(t)$ in the period t . Note that $\delta(t)$ is the residual value of the regression model in this same period.

4. Construction of a Feedforward Neural Network

To accurately evaluate the audit risks of financial statement, this section proposes an audit risk evaluation method for financial statement based on encrypted neural networks, in view of

the privacy and security of relevant data. The method was designed with reference to various audit risk evaluation strategies, AI techniques, and encryption algorithms at home and abroad. The calculation process of the proposed feed-forward neural network is illustrated in Figure 1.

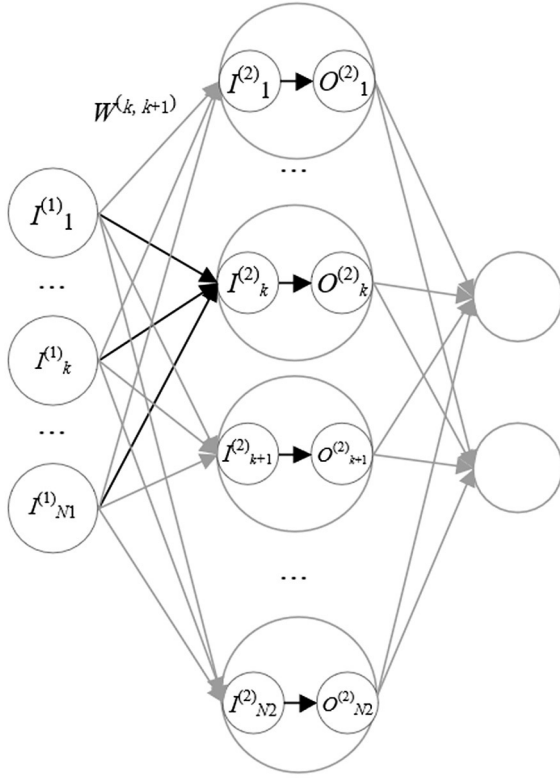


Figure 1. Feedforward neural network calculation.

As shown in Figure 1, the audit risk indices of financial statement are imported to the input layer of the neural network. The input $I_j^{(2)}$ of the first hidden layer in the network can be calculated by:

$$I_j^{(2)} = \sum_{i=1}^{N_1} \omega_{ij}^{(1,2)} O_i^{(1)} + b_j^{(1,2)} \quad (10)$$

where $O_i^{(1)}$ is the i -th output data of the input layer; $\omega_{ij}^{(1,2)}$ is the connection weight between the i -th input layer node and the j -th hidden layer node; $b_j^{(1,2)}$ is a bias. To facilitate parameter calculation and network operation, each layer of the neural network is usually characterized in the form of the matrix operation:

$$I^{(k+1)} = W^{(k, k+1)} O^{(k)} + B^{(k, k+1)} \quad (11)$$

where $I^{(k)}$ is the input matrix of the $(k+1)$ -th layer of the neural network; $O^{(k)}$ is the output matrix of the k -th layer; $W^{(k, k+1)}$ and $B^{(k, k+1)}$ are the connection matrix and bias matrix between the k -th and $(k+1)$ -th layers of the neural network, respectively. The input-output relationship of nodes in each network layer can be expressed as:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (12)$$

The sigmoid function was chosen as the activation function for the evaluation of the audit risks for the financial statement. The reason for choosing this function is that, during the verification on the predictions of AVP, AVA, and FSR, neither ultralow prediction accuracy nor vanishing gradients appeared when the neural network had three layers and was trained by mini-batch. For the binary classification of the final prediction, the softmax function was selected as the activation function:

$$\begin{cases} p_0 = \frac{e^{O_0}}{e^{O_0} + e^{O_1}}, \text{ the output is no} \\ p_1 = \frac{e^{O_1}}{e^{O_0} + e^{O_1}}, \text{ the output is yes} \end{cases} \quad (13)$$

where O_0 and O_1 are the activation values corresponding to the prediction results yes and no, respectively; p_i is the probability that the prediction result is yes or no.

The homomorphic encryption algorithm adopted in this paper can construct the mapping relationship between ciphertext and plaintext for different data encrypted by the same key. Let $Encrypt(*)$ and $Decode(*)$ be the encryption and decryption operations of the encryption algorithm, respectively. Then, the operation of two plaintexts x and y in the ciphertext domain can be constructed as:

$$\begin{cases} Decode(Encrypt(x \oplus y)) = x + y \\ Decode(Encrypt(x \otimes y)) = x \cdot y \end{cases} \quad (14)$$

If the formula (14) holds, the encryption algorithm is additive and multiplicatively homomorphic. Homomorphic encryption algorithms are widely used in distributed computing sce-

narios. During the audit risk evaluation for financial statement, if the enterprise wants to rely on the neural network to complete the evaluation $EP(*)$ of the data on audit risk indices $FR Data$, without leaking the data to the neural network, then the encryption and decryption need to be implemented in the following steps:

- Step 1:** Generate the public key K_{pub} and private key K_{pr} required by the algorithm.
- Step 2:** The enterprise uses K_{pub} to encrypt the $FR Data$, and sends the encrypted result $FRE Data$ to the neural network for evaluation $EP(*)$.
- Step 3:** The neural network evaluates $EP(FRE Data)$, and sends the results to the regulator.
- Step 4:** The regulator encrypts $FRE Data$ with K_{pr} to obtain the evaluation results.

From the above steps, it can be learned that, during the audit risk evaluation, the neural network cannot know the details about the data without K_{pri} , for the data on the audit risk indices of the financial statement are encrypted. This effectively protects the security of the internal financial data of the enterprise. Thanks to the homomorphism of the encryption algorithm, the enterprise can process the data on the audit risk indices of its financial statement, by virtue of the computing power of the neural network. The execution of our algorithm is detailed as follows, see Figure 2:

- randomly generate two relatively large prime numbers a and b , so that the greatest common divisor between ab and $(a - 1)(b - 1)$ equals 1;
- calculate the value of ab , as well as the least common multiple between $a - 1$ and $b - 1$;
- randomly select an integer γ that satisfies formula (15) from the reduced residue system of $(ab)^2$:

$$g\left(\frac{\gamma^\tau \bmod(ab)^2 - 1}{(ab)^2}, (ab)^2 \gamma\right) = 1 \quad (15)$$

where g is the function of the greatest common divisor; the meaning of (15) is to ensure that the modulo ab of $\gamma^\tau \bmod(ab)^2$ has an inverse element $D = (ab)2/[\gamma^\tau \bmod(ab)^2 - 1] \bmod(ab)$; after the above steps, it is possible to obtain the necessary public key (ab, γ) and private key (τ, D) ;

- let $FR Data$ be the data on the audit risk indices of the enterprise, waiting to be encrypted; $0 \leq FR Data < ab$;
- randomly select a positive integer ρ smaller than ab from the reduced residue system of $(ab)^2$;
- the ciphertext $FRE Data$ of $FR Data$ can be calculated by:

$$FRE Data = \gamma^{FR Data} \cdot \rho^{ab} \bmod(ab)^2; \quad (16)$$

- let $FRE Data$ be the ciphertext to be decrypted by the regulator, which satisfies that $FRE Data$ belongs to the reduced residue system of $(ab)^2$;
- $FR Data$ can be restored by:

$$FR Data = \frac{FRE Data^\tau \bmod(ab)^2 - 1}{(ab)^2} \cdot D \bmod ab. \quad (17)$$

5. Model Construction and Optimization

Based on the encrypted neural network, the audit risk evaluation model aims to evaluate the audit risks in the encrypted data on audit risk indices $FR Data$, and provides the evaluated results to the regulator. As shown in Figure 3, the model consists of an offline phase and an online phase.

In the *offline* phase, the neural network is trained with the original data on audit risk indices. During training, the connection weights and thresholds are updated continuously to improve the evaluation accuracy.

In the *online* phase, the part of the dataset on audit risk indices not used in network training is encrypted with the public key, and then imported to the trained neural network for evalu-

ation. After evaluation, the regulator can obtain the evaluation results by decrypting the output results.

The encrypted neural network, which is based on the homomorphic encryption algorithm, cannot directly encrypt decimals. To solve this problem, the real numbers of the decimal part were converted into integers with fixed precision. Specifically, a polynomial was constructed, whose coefficients are expanded by binary coding. When the polynomial is taken at 2 places, the coded value will be returned. After the decimal part has been encrypted, the complete data are encrypted. The decryption steps are the same as described in the previous section.

Since nonlinear operation does not satisfy the homomorphic operation of the homomorphic encryption algorithm, the nonlinear operations in the neural network must be optimized. As

mentioned in the previous section, the neural network involves such operations as weighted summation, activation functions, and loss functions.

The weighted summation, involving addition and multiplication, can add and multiply ciphertexts in line with the homomorphism requirement of the homomorphic encryption algorithm. The activation functions sigmoid and softmax, both containing the exponential function, are nonlinear and pending optimization. The sigmoid function was approximated by the McLoughlin series:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n \quad (18)$$

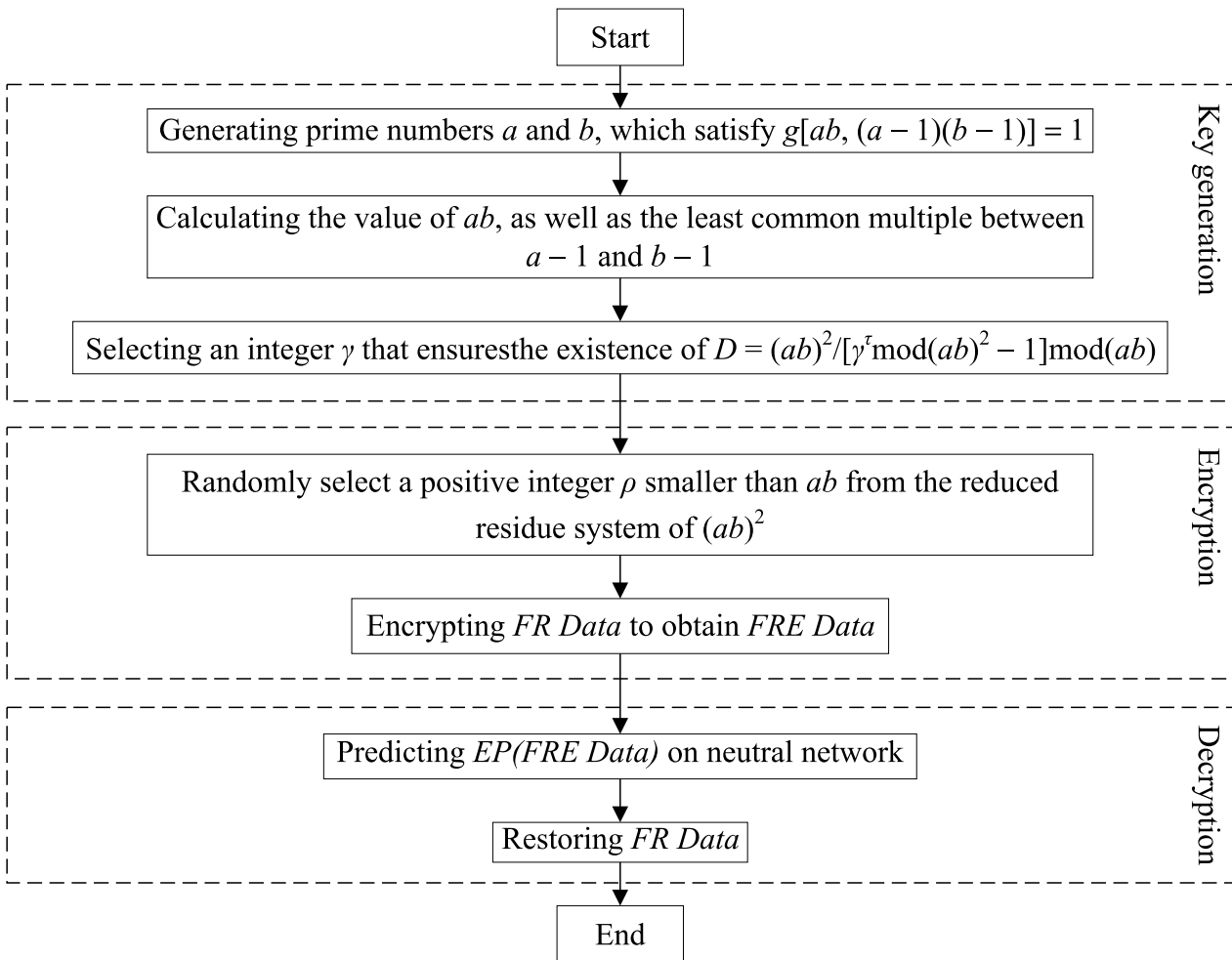


Figure 2. Algorithm flow.

while expanding the sigmoid function by (18) we obtain:

$$f(x) = \frac{1}{2} + \frac{1}{4}x - \frac{1}{48}x^3 + \frac{1}{480}x^5 + \dots \quad (19)$$

Formula (19) contains only multiplication and addition, and supports ciphertext addition and multiplication, as per the homomorphism requirement of the homomorphic encryption algorithm.

Figure 4 shows the overlap between sigmoid function curves before and after the approximation. It can be seen that the nonlinear sigmoid function was converted into a linear operation. However, the approximation effect was not desired when the data fell in $[-\infty, -1.5]$ or $[1.5, +\infty]$. This problem can be solved by normalizing the data on audit risk indices to $[-1, 1]$ before network training.

The above analysis shows that the sigmoid function is close to a linear function in $[-1, 1]$.

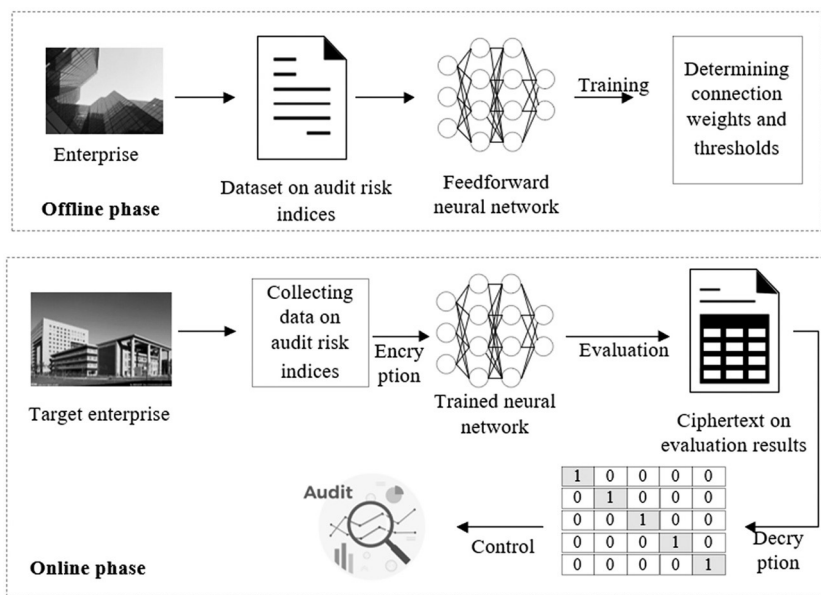


Figure 3. System structure.

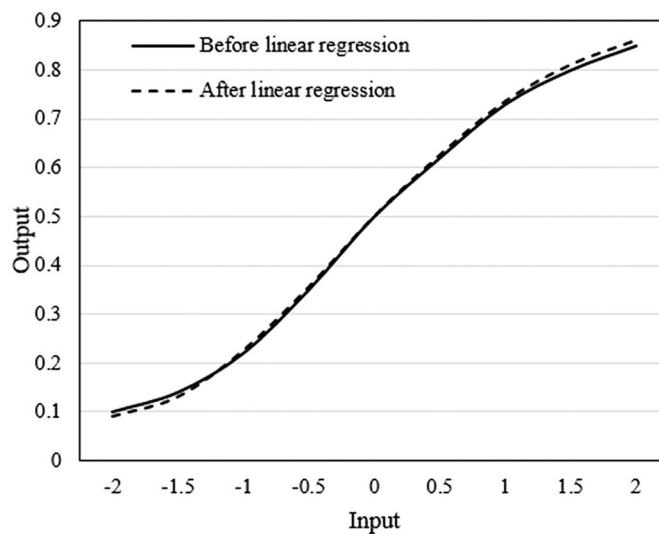


Figure 4. Overlap between sigmoid function curves before and after the approximation.

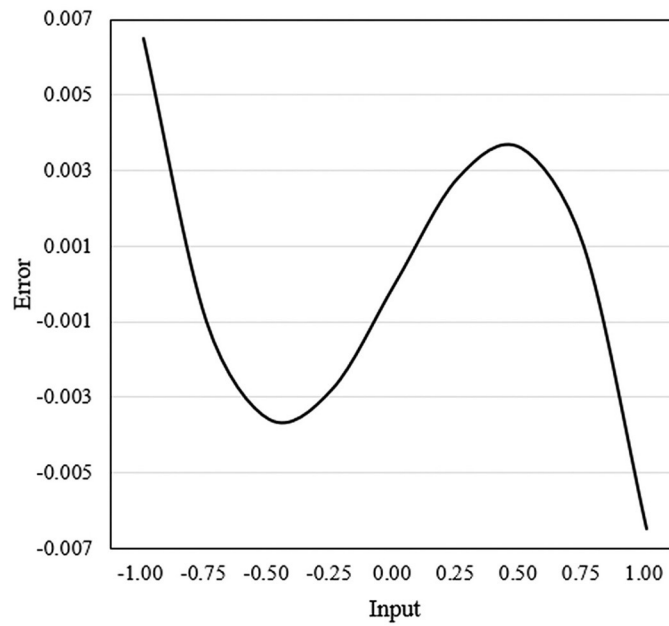


Figure 5. Functional error before and after linear regression.

Figure 5 shows the linear regression results of the function in $[-1, 1]$. The linear function thus obtained can be expressed as:

$$f(x) = 0.271x + 0.486. \quad (20)$$

As shown in Figure 5, the absolute error of the function in $[-1, 1]$ peaked at 0.0067, which is a desired level.

From the perspective of evaluation, the activation function softmax picks only the largest probability, and requires no additional processing. The loss function only works in the training of the neural network, without affecting the entire evaluation process. Hence, there is no need to consider whether it satisfies the homomorphism of the homomorphic encryption algorithm.

6. Experiments and Results Analysis

Table 2 records the descriptive statistics on evaluation indices for audit risks of financial statement. The sample enterprises were state-controlled and audited by large accounting firms. The experiments were performed based on a sample set of the audit data of 1500 corporate financial statements from 2015 to 2019. After performing the statistical analysis, it was found

that the company had suffered continuous loss for 5 to 6 months, but it didn't get ST or *ST warnings. In terms of controllable accrued operating net profit and real earnings management, the average of the accrued operating net profit of all sample enterprises was greater than zero, and their real earnings were also greater than zero. The standard deviation of real earnings management was larger than that of accrued operating net profit, reflecting strong profit controllability of the sample enterprises.

Table 3 lists the Pearson correlation coefficients between the evaluation indices. It can be seen that P_{NPL} is positively correlated with R_{CF} , TN_{CS} , and SR_{GM} . Thus, an enterprise with a strong growth capability and high net cash flow is unlikely to have audit risks in its financial statement. This result is in line with the previous hypothesis.

Moreover, the correlation coefficient of W_{SOH} has opposite signs with respect to that of P_{NPL} , R_{CF} , and SR_{GM} , suggesting that state-owned enterprises with a low performance pressure are not very likely to have audit risks in financial statement.

Furthermore, the probability of audit risks is low when external audit of the enterprise is performed by a large auditing firm.

Table 2. Descriptive statistics of evaluation indices.

Index	Mean	Minimum	Median	Maximum	Standard deviation	
Y_{CL}	0.562	0	0.873	3	0.187	Consecutive years of loss
P_{NPL}	0.361	0.012	0.424	0.6	0.224	Proportion of non-operating net profit and loss
P_{GA}	0.483	0.125	0.491	0.55	0.118	Proportion of guaranteed amount
R_{CF}	0.321	0.051	0.481	0.5	0.189	Cash flow ratio
W_{ST}	0	0	0.5	1	0.578	Under special treatment (ST) or delisting risk (*ST)
W_{SOH}	1	0	0.5	1	0.412	State-controlled
P_{RPT}	0.716	0.219	0.769	1	0.513	Proportion of connected transactions
TN_{CS}	0.435	0.200	0.498	0.6	0.321	Total number of shares outstanding
P_{PS}	0.035	0.010	0.047	0.1	0.209	Proportion of public shareholders
SR_{GM}	0.012	0.010	0.017	0.1	0.211	Percentage of shares held by shareholders attending the general meeting
O_{AR}	1	0	0.5	1	0.589	Attribution of audit responsibility

Table 3. Correlation test results.

	Fraud?	Y_{CL}	P_{NPL}	P_{GA}	R_{CF}	W_{ST}	W_{SOH}	P_{RPT}	TN_{CS}	P_{PS}	SR_{GM}	O_{AR}
Y_{CL}	0.047	1										
P_{NPL}	0.0561	0.036	1									
P_{GA}	0.0169	0.037	0.120	1								
R_{CF}	0.159	0.097	0.308	0.152	1							
W_{ST}	0.137	0.022	0.176	0.183	0.398	1						
W_{SOH}	-0.095	0.129	-0.177	0.009	-0.172	-0.116	1					
P_{RPT}	0.014	0.092	-0.019	0.297	-0.023	0.011	0.178	1				
TN_{CS}	0.043	0.171	0.298	0.153	0.198	0.126	0.046	0.068	1			
P_{PS}	-0.127	0.045	-0.086	-0.078	-0.064	-0.087	0.073	-0.141	-0.036	1		
SR_{GM}	0.187	0.046	0.192	0.269	0.239	0.419	-0.146	0.089	0.179	-0.092	1	
O_{AR}	0.038	0.028	0.041	0.079	0.074	0.039	-0.027	0.019	0.047	-0.004	0.016	1

In addition, the absolute value of the Pearson correlation coefficient between all indices was smaller than 0.5. This means that the multicollinearity has little effect on the relationship between the indices, *i.e.*, audit risk indices data can be subject to multiple regression analysis. Table 4 lists the regression results of the audit risk indices of financial statement.

Table 4 displays the multiple logistic regression (MLR) results of S_{VP} , S_{VN} , and S_{FR} . The results show that the same statistical effect was achieved when the financial statement of the target enterprise was measured by any of the three evaluation items. Hence, the proposed model can derive the probabilities of AVP, AVA, and FSR of financial statement at a level of 9.27%, provided that the eigenvectors (fraud motivations and fraud opportunities) and effects (controllable accrued operating net profit and real earnings management) have been controlled.

Through calculation, it can be observed that the controllable accrued operating net profit and

the real earnings management of the enterprise are significantly positively correlated with the probability of audit risks for financial statement at the level of 15%.

As for the other indices, P_{PS} is significantly negatively correlated with the three evaluation items, indicating that the public shareholders externally supervise the financial statement of the enterprise. P_{NL} is significantly negatively correlated with the three evaluation items at the level of 1.5%, suggesting that an enterprise with low non-operating net profit and loss has low audit risks. As expected, the financial statement faces low audit risks when the enterprise is state-owned, and the external audit is carried out by a large accounting firm.

For the adjustment of neural network parameters, an important step is to determine the number of hidden layer nodes. The number must be determined through careful weighing between evaluation accuracy and time consumption. Without changing other parameters, several

Table 4. Regression results of the audit risk indices of financial statement.

	S_{VP}		S_{VN}		S_{FR}	
	Coefficient	Z-score	Coefficient	Z-score	Coefficient	Z-score
Y_{CL}	-0.072	(-2.79)	-0.031...	(-1.76)	-0.029	(-2.56)
P_{NPL}	-0.207	(-2.42)	-0.247	(-3.41)	-0.305	(-3.94)
P_{GA}	0.375	(1.47)	-0.129	(3.42)	0.026	(0.57)
R_{CF}	-0.018	(-3.70)	-0.021.	(-3.25)	-0.024	(-1.22)
W_{ST}	-0.136	(-1.88)	-0.133	(-2.11)	0.068	(0.69)
W_{SOH}	-0.178	(-0.89)	-0.987...	(-3.43)	-0.351	(-0.35)
P_{RPT}	-0.079	(-0.28)	-0.028	(-0.21)	0.029	(0.89)
TN_{CS}	0.571	(1.79)	1.295..	(6.10)	1.073	(1.94)
P_{PS}	-0.379	(-1.02)	-0.259	(-2.07)	0.459	(0.86)
SR_{GM}	-0.196	(-0.49)	-0.148	(-0.59)	-0.011	(-0.24)
O_{AR}	-1.004	(-4.19)	-0.412	(-2.03)	-0.279	(-0.88)

contrastive experiments were conducted with different number of hidden layer nodes. From the experimental results (Table 5), it can be seen that the evaluation accuracy of the network during training gradually increased and then tended to be stable, with the growing number of hidden layer nodes; meanwhile, the training time, and encryption and decryption time both increased linearly. This paper sets the number of hidden layer nodes to 20, aiming to minimize the computing time without sacrificing much of evaluation accuracy.

Table 6 records the influence of different learning rates (0.01, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, and 0.5) on computing time and evaluation accuracy. The data in the table are the computing time and the highest evaluation accuracy in each scenario after 10,000 iterations. It can be seen that the greater the learning rate, the shorter the training time, and the higher the evaluation accuracy. After 10,000 iterations, the evaluation accuracy averaged at about 88%. Through comprehensive consideration, the most reasonable value of the initial learning rate was determined as 0.4-0.5.

Table 5. Influence of the number of hidden layer nodes on computing time and evaluation accuracy.

Test number	Number of hidden layer nodes	Training time	Encryption and decryption time	Evaluation accuracy
1	5	80 s	37.8 s	88.79%
2	10	114 s	45.7 s	89.15%
3	15	158 s	56.7 s	90.27%
4	20	277 s	67.4 s	90.45%
5	25	326 s	80.1 s	90.57%
6	30	421 s	96.2 s	91.03%
7	35	569 s	120.5 s	91.16%
8	40	670 s	147.8 s	91.58%

Table 6. Influence of learning rate on computing time and evaluation accuracy.

Test number	Learning rate	Training time	Evaluation accuracy
1	0.01	493.2 s	82.41%
2	0.05	280.9 s	85.74%
3	0.1	65.7 s	87.24%
4	0.15	59.2 s	88.11%
5	0.2	45.7 s	89.25%
6	0.3	32.6 s	89.97%
7	0.4	21.3 s	90.53%
8	0.5	14.88 s	91.05%

Table 7 brings the evaluation performance of our neural network. It is intuitive that the precision, recall, and F1 score of our network were around 90% for the three items (SVP, SVN, and SFR), which is a sign of good evaluation performance.

In our model, the audit risks are evaluated after processing the encrypted data on audit risk indices with the Paillier homomorphic encryption algorithm and feedforward neural network. To verify the effectiveness of our model, the evaluation effect of the model was compared with that of four other evaluation models.

As shown in Table 8, the traditional classification models, which require manual feature extraction, cannot achieve the evaluation accuracy of our model based on feedforward neural network. This result verifies the reasonability of our model. Compared with the CNN algorithm based on the fully homomorphic encryption, our model based on the Paillier homomorphic encryption algorithm achieved a high accuracy, a low time complexity, and a superior performance, under the same conditions.

Figure 6 displays the training error curve and training accuracy curve of the proposed neural network. As shown in Figure 6(a), the training error decreases and tends to be stable with the growing number of training cycles; conversely, as shown in Figure 6(b), the test accuracy gradually increases and tends to be stable, with the growing number of training cycles.

Finally, Table 9 summarizes the evaluation performance on encrypted datasets with different number of training cycles. It can be seen that the evaluation accuracy of test samples stabilized after around 200 training cycles.

While the above empirically verified the evaluation and prediction methods of financial statement audit risks, there are still deficiencies related to the methods in terms of research scope, variable definition, and sample selection. Further studies could further seek more accurate index measurement methods for evaluation indexes of financial statement audit risks, also exploring how to test the scoring characteristics of the evaluation indexes of listed companies on GEM and SME boards.

Table 7. Evaluation performance.

Metric \ Item	S_{VP}	S_{VN}	S_{FR}	Comprehensive evaluation
Precision	95.51%	87.81%	87.91%	88.18%
Recall	87.46%	89.92%	92.36%	88.27%
F1 score	92.14%	88.92%	89.18%	87.61%

Table 8. Evaluation effects of different models.

Model	Processing method	Encryption method	Two-class/multi-class classification	Accuracy
SVM+ANN	Discrete wavelet transform	/	Two-class classification	75.82%
SVM	Filter bank co-space mode	/	Two-class classification	85.61%
GMM	Joint factor analysis	/	Four-class classification	84.39%
CNN	Fuzzy comprehensive evaluation	Fully homomorphic encryption	Six-class classification	89.77%
Our model	Pearson correlation coefficient analysis	Semi-homomorphic encryption	Six-class classification	90.26%

Note: SVM, GMM, and CNN are short for support vector machine, Gaussian mixture model, and convolutional neural network, respectively.

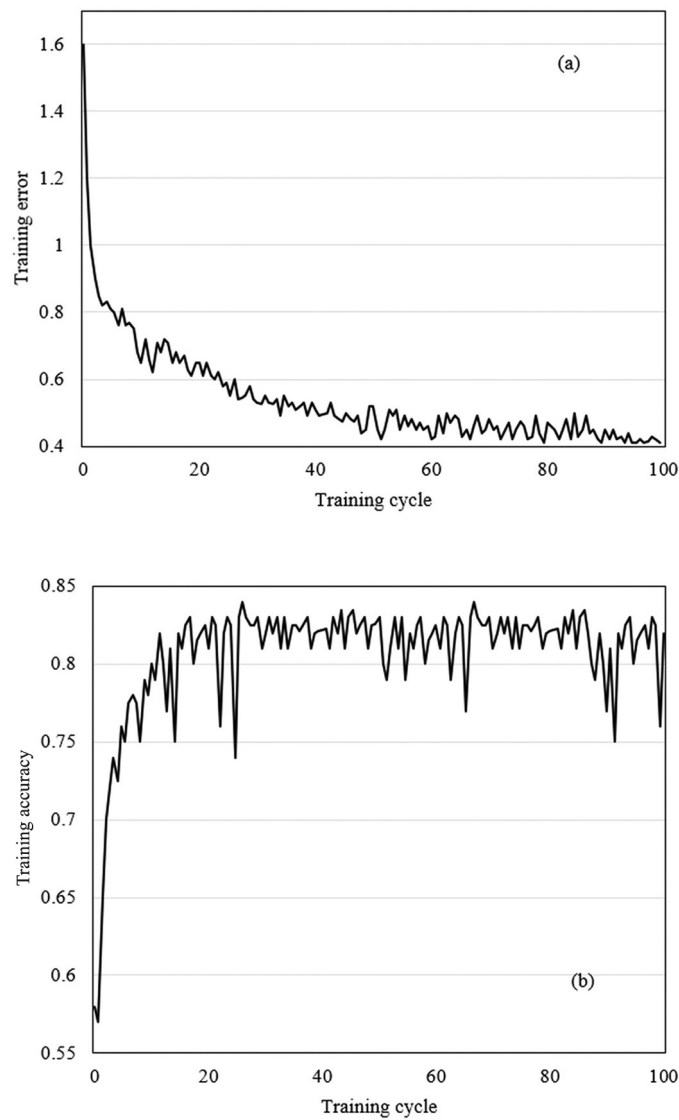


Figure 6. Training error curve and training accuracy curve of proposed neural network.

Table 9. Evaluation performance of encrypted datasets with different number of training cycles.

Type	Number of training cycles	Training accuracy	Test accuracy	Training time
Encrypted dataset of audit risk indices of financial statement	50	79.21%	80.21%	20.41m
	100	81.69%	81.35%	31.86m
	150	83.52%	85.91%	46.24m
	200	84.21%	89.42%	53.21m
	300	85.18%	90.12%	66.42m
	400	85.24%	90.21%	87.02m

7. Conclusion

This paper presents an ANN-based audit risk evaluation model for financial statement. Firstly, the audit risk indices and the fraud factors for financial statement were identified. Then, an audit risk verification model was developed and used to test the predictions on AVP, AVA, and FSR. Through experiments, the descriptive statistics of all audit risk indices were calculated. In this way, the Pearson correlation coefficients between the indices were obtained, as well as the MLR results on three items S_{VP} , S_{VN} and S_{FR} . Finally, a feedforward neural network was constructed based on the homomorphic encryption algorithm and the audit risks of financial statement were evaluated by the encrypted neural network. Experimental results show that our model achieved a good evaluation performance, with higher accuracy, lower time complexity and better performance than other algorithms.

References

- [1] C. B. Bao *et al.*, "Optimization of Integrated Risk in Commercial Banking Based on Financial Statements", *Procedia Computer Science*, vol. 31, pp. 501–510, 2014.
<http://dx.doi.org/10.1016/j.procs.2014.05.295>
- [2] A. L. Hamilton *et al.*, "Managing Financial Risk Trade-Offs for Hydropower Generation Using Snowpack-Based Index Contracts", *Water Resources Research*, vol. 56, no. 10, 2020.
<http://dx.doi.org/10.1002/essoar.10502068.1>
- [3] N. Kuznietsova and P. Bidyuk, "Intelligence Information Technologies for Financial Data Processing in Risk Management", *Communications in Computer and Information Science*, vol. 1158, pp. 539–558, 2020.
http://dx.doi.org/10.1007/978-3-030-61656-4_36
- [4] X. P. Song *et al.*, "Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China", *Journal of Forecasting*, vol. 33, no. 8, pp. 611–626, 2014.
<https://doi.org/10.1002/for.2294>
- [5] P. Leonov *et al.*, "Visual Analysis in Identifying a Typical Indicators of Financial Statements as an Element of Artificial Intelligence Technology in Audit", *Procedia Computer Science*, vol. 169, pp. 710–714, 2020.
<https://doi.org/10.1016/j.procs.2020.02.174>
- [6] M. Tirea *et al.*, "Stock Market Trading Strategies Applying Risk and Decision Analysis Models for Detecting Financial Turbulence" in *Proc. of the 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing', Synasc*, 2015, pp. 216–223.
<http://dx.doi.org/10.1109/SYNASC.2015.42>
- [7] M. Kerste *et al.*, "Systemic Risk in the Energy Sector-Is there Need for Financial Regulation", *Energy Policy*, vol. 78, pp. 22–30, 2015.
<https://doi.org/10.1016/j.enpol.2014.12.018>
- [8] S. Howison and D. Schwarz, "Risk-Neutral Pricing of Financial Instruments in Emission Markets: A Structural Approach", *SIAM Review*, vol. 57, no. 1, pp. 95–127, 2015.
<https://doi.org/10.1137/140987365>
- [9] S. R. Sahu *et al.*, "Hybrid CRO Based FLANN for Financial Credit Risk Forecasting. Smart Innovation", *Systems and Technologies*, vol. 31, pp. 57–65, 2015.
http://dx.doi.org/10.1007/978-81-322-2205-7_6
- [10] A. Pilkova *et al.*, "Assessment of the Pillar 3 Financial and Risk Information Disclosures Usefulness to the Commercial Banks Users", *Lecture Notes in Computer Science*, vol. 9227, pp. 429–440, 2015.
https://doi.org/10.1007/978-3-319-22053-6_46
- [11] A. Guslawa *et al.*, "Problem Transformation Methods for Prediction of Opinion and Exceptions in Financial Statements Audit Reports: Case for Financial Statements Audit in Central Kalimantan Province", in *Proc. of the 2018 International Conference on Information and Communications Technology*, 2018, pp. 747–752.
<http://dx.doi.org/10.1109/ICOIACT.2018.8350755>
- [12] A. Rehwinkel, "Corporate Financial Risk Analysis According to the Constructal Law: Exploring the Composition of Liabilities to Assets", *International Journal of Heat and Technology*, vol. 34, no. 1, pS133–S140, 2016.
<http://dx.doi.org/10.18280/ijht.34S117>
- [13] M. Karwanski and U. Grzybowska, "Propensity Score Matching and Its Application to Risk Drivers Detection in Financial Setting", *Acta Physica Polonica A*, vol. 129, no. 5, pp. 945–949, 2016.
<http://dx.doi.org/10.12693/APhysPolA.129.945>
- [14] A. Dziadosz *et al.*, "Financial Risk Estimation in Construction Contracts", *Procedia Engineering*, vol. 122, pp. 120–128, 2015.
<https://doi.org/10.1016/j.proeng.2015.10.015>
- [15] Y. G. Shan and I. Troshani, "Does XBRL Benefit Financial Statement Auditing", *Journal of Computer Information Systems*, vol. 54, no. 4, pp. 11–21, 2015.
<https://doi.org/10.1080/08874417.2014.11645718>
- [16] S. Bhadani *et al.*, "Mining Financial Risk Events from News and Assessing their Impact on Stocks", *Lecture Notes in Computer Science*, vol. 11985, pp. 85–100, 2020.
http://dx.doi.org/10.1007/978-3-030-37720-5_7

- [17] M. Ziolo *et al.*, "Environmental, Social and Governance Risk in Public and Private Financial Systems: Fuzzy Cognitive Mapping Comparative Analysis", in *Proc. of the ACM International Conference*, 2020, pp. 31–38.
<http://dx.doi.org/10.1145/3396743.3396752>
- [18] C. Lucarelli *et al.*, "Misclassifications in Financial Risk Tolerance", *Journal of Risk Research*, vol. 18, no. 4, pp. 467–482, 2015.
<https://doi.org/10.1080/13669877.2014.910678>
- [19] C. B. Balan *et al.*, "The Statistical Assessment of Financial Distress Risk in the Case of Metallurgical Companies", *Metallurgija*, vol. 54, no. 3, pp. 575–578, 2015.
<https://hrcaj.srce.hr/131626>
- [20] R. Cerqueti and F. Spizzichino, "Signatures of Systems with Non-Exchangeable Lifetimes: Some Implications in the Analysis of Financial Risk", *Springer Proceedings in Mathematics and Statistics*, vol. 136, pp. 361–375, 2015.
https://doi.org/10.1007/978-3-319-18029-8_27
- [21] K. H. Chen and K. Khashanah, "The Reconstruction of Financial Signals Using Fast ICA for Systemic Risk", in *Proc. of the 2015 IEEE Symposium Series on Computational Intelligence*, pp. 885–889, 2015.
<http://dx.doi.org/10.1109/SSCI.2015.130>
- [22] E. M. Mosconi, "Insights into Congestion Costs and Financial Risk Management: The Electricity Market", *Procedia Computer Science*, vol. 52, no. 1, pp. 738–745, 2015.
<https://doi.org/10.1016/j.procs.2015.05.087>
- [23] J. Wang, "Study on the Financial Risk of the Artwork Investment Market", *Open Cybernetics and Systemics Journal*, vol. 9, no. 1, pp. 1868–1873, 2015.
<http://dx.doi.org/10.2174/1874110X01509011868>
- [24] J. Linnerooth-Bayer and S. Hochrainer-Stigler, "Financial Instruments for Disaster Risk Management and Climate Change Adaptation", *Climatic Change*, vol. 133, no. 1, pp. 85–100, 2015.
<https://doi.org/10.1007/s10584-013-1035-6>
- [25] P. Pandey and S. D. Haes, "A Variable Payout Information Security Financial Instrument and Trading Mechanism to Address Information Security Risk", in *Proc. of the 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, vol. 2015, pp. 17–25, 2015.
<https://doi.org/10.1109/3PGCIC.2015.109>
- [26] B. Sharma *et al.*, "Clabacus: A Risk-Adjusted Cloud Resources Pricing Model Using Financial Option Theory", *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 332–344, 2015.
<http://dx.doi.org/10.1109/TCC.2014.2382099>

Received: January 2021

Revised: March 2021

Accepted: March 2021

Contact addresses:

Yanhua Li

School of Accounting

Wuhan College

Wuhan

China

e-mail: wuhancollyh@126.com

YANHUA LI was born in Tongren City, Guizhou Province, China. She is currently a lecturer at the School of Accounting, Wuhan College. She graduated from Zhongnan University of Economics and Law with a MSc degree. Her research directions are financial risk management, financial statement quality analysis, certified public accountants auditing practice behavior and quality.
