International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 330–349
© Cerebration Science Publishing

Article

Research on the Application of Long and Short-term Memory Network Algorithm in Corporate Financial Forecasting Driven by Big Data

Li Nie *

Lyceum of the Philippines University, Manila 0900, Philippines; li.nie@lpunetwork.edu.ph

Received date: 1 January 2025; Accepted date: 4 April 2025; Published online: 29 May 2025

Abstract: In the current context of rapid development of digital economy, corporate financial risks occur frequently, and traditional prediction means are difficult to cope with the complex data structure. In order to improve the accuracy of corporate financial risk prediction, this paper proposes a combined model based on ARIMA-IPOA-LSTM. The study firstly constructs a dataset of A-share listed companies in Shanghai and Shenzhen from 2012 to 2024, and screens 18 financial indicators for modeling. In the methodology, ARIMA is combined to extract trend features, and Improved Pelican Optimization Algorithm (IPOA) is used for LSTM parameter optimization to achieve model accuracy. The experimental results show that the accuracy of the training samples reaches 94.02%, the accuracy of the test samples is 93.48%, and the RMSE is as low as 0.574×10^{-5}. The financial data of 15 listed companies in 2024 are further selected for simulation, and the model warning accuracy reaches up to 100%. The conclusion shows that ARIMA-IPOA-LSTM has strong adaptability under dealing with unbalanced data, can effectively improve the recall rate and F1 score, and has high practical value.

Keywords: financial risk prediction; big data; LSTM; ARIMA; IPOA; neural network

1. Introduction

In the era of big data, data-driven decision making has become the core of financial decision making in modern enterprises, which can more accurately predict market trends, assess risks and formulate strategies by collecting, analyzing and utilizing large amounts of data [1–3]. This kind of decision-making not only improves the efficiency and accuracy of decision-making, but also brings higher competitiveness for enterprises [4,5]. Financial accounting, as a core area of enterprise management, also faces new challenges in data processing and analysis.

However, it is not easy to analyze the enterprise financial data, which has a huge amount of data and changes rapidly over time [6]. Traditional financial forecasting methods, which mainly rely on historical data and expert experience, are often difficult to accurately reflect the actual financial status of an enterprise in a timely manner in the face of a complex and changing market environment [7,8]. In contrast, big data analytics, with its ability to handle massive heterogeneous data and its advantage of extracting valuable information from data, provides a new path for the change and innovation of enterprise financial management methods [9–11]. Due to the widespread use of big data platforms, the method of solving complex problems through big data analysis technology has been deeply applied in enterprise financial analysis, which is able to complete the training and analysis of high-dimensional financial data and obtain effective training results [12–15]. Based on the development of big data technology, the ability of enterprise financial forecasting has been strengthened, enabling enterprises to more comprehensively and accurately grasp their own financial situation and predict future development trends [16,17].

Enterprises face unprecedented financial challenges in the wave of economic globalization and

Copyright: © 2025 by the authors

technological change. Macroeconomic fluctuations, industrial structure upgrading and the frequent occurrence of sudden public events make the volatility of enterprise financial status increase significantly. The traditional static and linear thinking financial early warning model has been difficult to adapt to the data complexity and non-linear characteristics in the context of the new era. At the same time, the financial market is increasingly relying on big data-driven intelligent analysis means, driven by artificial intelligence and machine learning technology, prediction models based on deep learning are widely used in the field of enterprise financial risk assessment. In particular, recurrent neural networks with time-series memory capability show obvious advantages in identifying the evolutionary trend of financial risks. However, the complexity of the deep model parameter tuning process and the high computational cost have become the bottleneck restricting the model's practicality. Therefore, how to construct an efficient and interpretable financial risk prediction system is an urgent problem in theory and practice.

In order to solve the above problems, this paper proposes a corporate financial risk prediction model integrating ARIMA, LSTM and Improved Pelican Optimization Algorithm (IPOA). The specific ideas are as follows: firstly, a large-scale financial risk dataset is constructed based on A-share listed companies in Shanghai and Shenzhen, and five categories of 18 financial indicators are extracted as feature variables; secondly, ARIMA model is used to capture the time series trend and calculate the residuals, and IPOA is used to optimize the network structure and parameter combinations of LSTM, and then a fusion model is constructed to achieve accurate prediction; finally, SHAP is combined with an interpretable method to enhance the transparency of the model, and the model can be used to predict the financial risk of enterprises. Finally, the SHAP interpretability method is used to enhance the transparency of the model, which provides a quantitative basis and decision-making reference for enterprise risk identification and response.

2. Enterprise Financial Risk Dataset Construction

Risk exists in all stages of the development of enterprises, when it accumulates to a certain stage will easily lead to the occurrence of financial crisis, so the establishment of appropriate financial early warning model is of great significance for enterprises. In the international academic community on the financial early warning research can be said to be in the ascendant, scholars from various countries have done unremitting exploration in this field. The role of financial risk early warning is to determine what kind of risk the enterprise will face now or in the future through the historical operation and financial information of the enterprise, and whether this risk will cause financial risk. The causes of financial risk are complex, diverse and uncertain, and it is necessary for enterprises to recognize and warn in advance.

2.1. Determination of Data Labels

2.1.1. Definition of Enterprise Financial Risk

Enterprise financial risk early warning is through the selection of financial indicators and the use of analytical techniques, the potential financial risks of enterprises can be detected early, and can be early warning process. This process can prompt enterprises to implement timely and effective measures to reduce possible financial damage. Commonly used indicators are generally current ratio, quick ratio, gearing ratio, etc., these indicators from different perspectives to show the financial situation of enterprises. Analysis techniques methods mainly include trend analysis, ratio analysis, composition analysis and cash flow analysis, etc. To understand the financial situation of an enterprise, combine financial indicators and analysis methods to establish an early warning model of the enterprise's financial risk. According to the industry standards and the enterprise's historical data, set the early warning threshold, regularly monitor the financial indicators, the specific implementation process, once the early warning threshold is exceeded, the early warning signal will be issued in a timely manner. At this time, the enterprise can implement a series of coping strategies, such as strengthening internal management control, enhance the standardization and efficiency of financial management activities. Adjust the capital structure, rational allocation of the proportion of debt and equity, to achieve the purpose of reducing financial risk. At the same time, enhance the profitability of the enterprise through cost control and revenue enhancement. In addition, through effective cash flow management, ensure that enterprises have sufficient liquidity to cope with potential risks. By establishing an effective financial risk early warning model, the enterprise can improve the level of financial risk management, reduce the financial risk that the enterprise will face, so that the enterprise can be sound and sustainable development.

Enterprise financial risk early warning has the function of risk identification and assessment, through the monitoring of financial indicators and analysis methods to identify potential financial risks, and then assess the identified risks to determine the severity and scope of their impact. With the function of early warning signal generation, set a reasonable early warning threshold and trigger conditions directly related to the sensitivity and accuracy of the early warning system, in-depth research and accurate analysis of

the financial risk has been constructed in the early warning indicator system of each indicator, to determine a reasonable early warning threshold. Setting the warning threshold, when the financial indicators exceed the threshold, the system will issue an early warning signal, the warning signal may include warnings, alarms or emergency notification. With the function of decision support, it provides decision support for the management of the enterprise to help them develop effective risk management strategies and countermeasures, and assists the management to make wise decisions when facing financial risks [18].

2.1.2. Study Population and Data Labeling

This study refers to existing scholars' research and takes listed companies in China as the research object, meanwhile, the special treatment or not is used as the sign of whether the company is in financial difficulties. With the continuous revision of the stock listing rules, the conditions for special treatment have been changed, and the operation of special treatment has been replaced by the implementation of risk warnings. Risk warnings are mainly categorized into “delisting risk warnings” and “other risk warnings”. According to the stock listing rules of the Shanghai Stock Exchange, the Exchange issues a delisting risk warning to a company if significant abnormal financial conditions are observed, such as the company's audited net profit for the last two consecutive fiscal years being negative, or the company's audited net assets for the previous fiscal year being negative, as well as serious violations of the disclosure rules in the annual reports, etc. The paper examines a large number of research papers on this issue.

Based on the study of a large number of research literature, this paper establishes the relevant data labeling indicators that can reflect the early warning of corporate financial risk, the specific content of which is shown in Table 1, mainly including five indicators of solvency, profitability, operating capacity, development capacity, cash flow, and a total of 18 secondary indicators.

Table 1. Early warning indicators for enterprise financial risks.

Level 1 index	Level 2 index	Code
Debt-paying ability	Current ratio	FR1
	Quick ratio	FR2
	Asset-liability ratio	FR3
	The ratio of working capital to total assets	FR4
Profit ability	Return on net assets	FR5
	Earnings per share	FR6
	Return on total assets	FR7
	Net profit margin on sales	FR8
Operating ability	Accounts receivable turnover rate	FR9
	Total asset turnover rate	FR10
	Inventory turnover rate	FR11
	Current asset turnover ratio	FR12
Development ability	Total asset growth rate	FR13
	Net profit growth rate	FR14
	Cash flow growth rate	FR15
Cash flow	Cash flow ratio	FR16
	The net amount of cash flow generated by business activities per share	FR17
	Cash recovery rate of total assets	FR18

2.2. Financial Risk Dataset Construction

2.2.1. Data Sources and Processing

In this paper, the A-share listed companies in Shanghai and Shenzhen from 2012 to 2024 are selected as the research sample. First of all, according to the previous experience of the same selection of the company's recent two consecutive fiscal years of audited net profit are negative, or the company's last fiscal year audited net assets are negative, as well as the emergence of serious violations of annual report of the disclosure rules as a criterion, if the difference is greater than zero, on behalf of the enterprise that year there is a replication of the net profit, will be such enterprises labeled as financial risky enterprises, and the opposite is labeled as financial risk-free enterprises The opposite is labeled as financial risk-free firms. In addition, this paper argues that the occurrence of abnormal financial or other conditions, i.e., the listed company is issued as ST, then it indicates that the listed company has operational problems and has a higher probability of risk, so when the company is ST (or ST*) in that year, then the company is

labeled as 1, otherwise, the company is labeled as 0. Accordingly, this study has obtained 31,078 normal samples and 6,227 risky samples. The selected data are from CSMAR, Wind, and CNRDS databases.

Due to the high noise of corporate financial risk data, it is necessary to pre-process the samples before inputting them into the model. This paper mainly removes some useless variables from the initial samples and performs maximum-minimum normalization on the data to make the input data more reliable.

(1) Delete the variables with too many missing values and high rate of the same value, and delete the irrelevant and repetitive variables.

(2) Maximum and minimum normalization of data. Since many variables in the corporate financial risk data set belong to different ranges of quantitative intervals, the gradient descent process makes the model spend more time on training, thus reducing the accuracy of the model. Therefore, maximum-minimum normalization of the data set can greatly accelerate the training speed and reduce the gap in the data volume range to affect the final prediction effect and prediction time of the model.

2.2.2. Financial Risk Data Set

Given that the number of normal firms is significantly higher than the number of risky firms in the actual situation, the sample sizes of the two classes in the training and test sets will be adjusted here. In the training set, we set the imbalance ratio between normal and risky firms according to the distribution of classes to about 0.0572. In addition to this, due to the presence of a large number of mean interpolated data samples in the dataset, a new method is proposed here to divide the training set and the test set. According to this method, the test set contains most of the data samples that are not mean-padded in the original data, which is more capable of confirming the model's ability to predict risky firms under the influence of missing values.

Different operations are taken for risky and normal firms. For normal firms, it is first determined whether they are mean-padded samples, and if so, they are used as training data. Otherwise, it is deposited into a temporary storage set S1, from which two samples totaling 200 are then randomly selected to be put into two sets (e.g., SubsetA and SubsetB). Meanwhile, for the risky firm sample, if this is a sample of data generated by random averaging, use it as training data. Otherwise, it is put into the temporary storage set S2, and two other sets with a sample number of 200 are randomly extracted from this storage set to be put into two other subsets (e.g., SubsetC and SubsetD), respectively. In addition to this, the remaining samples in the temporary storage set are used as the test set.

By using the above method, these four sets are randomly combined and the obtained combinations are added to the training and test sets.

3. Enterprise Financial Risk Prediction Modeling

Listed companies are the main body in the financial market, through the scientific method of enterprise financial risk timely prediction, early warning, to avoid the enterprise into financial risk, for enterprises and the whole financial market is of great significance. Today's era of information technology, data intelligence, full of opportunities but also full of challenges, the continuous extreme events of all kinds is to the world's economy and life has caused a significant impact. In order to survive in the market environment of fierce competition and increased uncertainty, enterprises need to predict financial risks in time and take measures to effectively avoid financial risks, which is of great significance and value for the stable development of enterprises, the protection of investors' interests and the stabilization of the market economy.

3.1. Theories Related to Time Series Forecasting

3.1.1. Long and Short-Term Memory Neural Networks

Neural networks are well known for their powerful ability to handle nonlinear data. Its features include large parameter dimensions, generalization, and the use of nonlinear activation functions in each layer, which makes neural networks well adapted to handle nonlinear data. The Long Short-Term Memory Neural Network (LSTM) is a structure developed from the Recurrent Neural Network (RNN), which provides the RNN with better long-term memory capabilities by introducing gating units in each neural unit of the hidden layer [19].

LSTM includes input gates, output gates, forgetting gates and memory cells. The σ and tanh activation functions become the mechanism of “gates”, i.e., filtering information with “gates” and preserving information with memory cells. Assuming that the information inside the cell state is denoted by C_t, the hidden state at the last moment is h_{t-1}, and the current input is x_t, the weight matrix W is multiplied point by point, and then the deviation b is added to the product. The formulas are as follows:

(1) The output information of the previous moment is controlled by the forgetting gate σ in the form:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{1}$$

where the value of f_t is between 0 and 1, 0 means delete all the information in the previous moment and 1 means save all the information in the previous moment.

(2) In the input gate to generate the information that needs to be updated, first the input gate decides which values to use for updating by using the σ function, and the tanh layer is used to generate new candidate values to be summed up to get the required candidate values for that moment. The original cell state C_{t-1} is updated to C_t with the following equation:

$$U_t = \sigma\left(W_u[h_{t-1}, x_t] + b_u\right) \tag{2}$$

$$\hat{C}_t = \tanh\left(W_c[h_{t-1}, x_t] + b_c\right) \tag{3}$$

$$C_t = f_t \times C_{t-1} + U_t \times \hat{C}_t \tag{4}$$

(3) The output gate completes the output of the model. An initial output is first obtained through the σ layer, and then the values are scaled to between $[-1,1]$ through tanh, and then multiplied pair by pair with the output obtained from σ to obtain the final output of the model, h_t, with the following formula:

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = O_t \times \tanh(C_t) \tag{6}$$

3.1.2. Principles of ARIMA Modeling

ARIMA is a commonly used time series forecasting algorithm, which is based on the analysis of time series data and contains three important components, namely autoregression (AR), differencing (I) and moving average (MA).

AR is a linear relationship between the current observation and its observations at a number of moments in the past, $AR(p)$ modeling indicates that the current observation is correlated with observations at p moments in the past. I is used to remove the non-stationarity of a time series, which is stabilized by performing first-order or multi-order differencing operations on the original data. MA, on the other hand, refers to the linear combination of the current observation correlated with the random error term at a number of moments in the past, and the $MA(q)$ model indicates that the current observation is correlated with the random error term at q moments in the past [20].

Combining the three parameters constitutes the $ARIMA(p,d,q)$ model with the mathematical expression:

$$\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1-L)^d x_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t \tag{7}$$

where L stands for the lag term, φ_i is the parameter for the autoregressive part of the model, θ_i is the parameter for the moving average part, and ε_t is the error term.

The ARIMA model building process usually includes the steps of smoothness test for time series data, determining the model order, model fitting and forecasting. The main work of each of these steps is as follows:

(1) Data preparation. Collect time series data and conduct preliminary data cleaning and processing

to ensure the reliability and integrity of the data.

(2) Smoothness test. Utilize statistical methods to verify the smoothness of the time series data, and if it is not smooth, then carry out appropriate differential operations. Usually test whether the time series is non-stationary method is ADF, ADF test of the core assumptions is the original hypothesis (null hypothesis) for the time series has a unit root, that is, the sequence is non-stationary. If the statistical value of the ADF test is less than a certain critical value (corresponding to the level of significance), then the original hypothesis can be rejected, i.e., the series is considered to be smooth, otherwise it is not possible to reject the original hypothesis, i.e., the series is non-smooth.

The usual form of its regression model is as follows:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \tag{8}$$

where Δy_t denotes the first-order difference of the time series, y_{t-1} denotes the value of the lagged term of the series, t denotes the time-trend term, Δy_{t-1} denotes the value of the lagged term of the first-order difference of the series, ε_t denotes the error term, α denotes the intercept, β denotes the coefficients of the time trend, and γ and δ are the weighting coefficients in the model.

(3) Model fitting. Determine the parameters p, d, and q of the ARIMA model and fit the model using historical data. Methods for determining the parameters usually include autocorrelation plots (ACF) and partial autocorrelation plots (PACF) of observed time series data, with ACF demonstrating the correlation of the data at different lag orders, and PACF demonstrating the direct correlation between the observed and the lagged values, excluding correlation at intermediate stages.

(4) Model Evaluation. The quality of model fit was diagnosed with the help of residual Q-Q plots and residual autocorrelograms to ensure the predictive performance of the model. Among them, the residual Q-Q plot is used to check whether the residuals have normal distribution characteristics, and the residual autocorrelation plot shows the correlation between the residuals with different lag coefficients. If the plots show a significant correlation of the residuals over the lag coefficients, it may indicate that there are uncaptured structures or patterns in the model.

3.2. ARIMA-IPOA-LSTM modeling

3.2.1. Improving the Pelican Optimization Algorithm

(1) Pelican Optimization Algorithm

Pelican Optimization Algorithm (POA) is a heuristic intelligent optimization algorithm, the basic idea of the algorithm is to mimic the behaviors and strategies of pelicans when attacking and hunting.The optimization process of POA algorithm is divided into the initialization phase of the population and the hunting phase, and the initialization formula of the pelican population is as follows:

$$x = lb + rand \cdot (ub - lb) \tag{9}$$

where x is the position of the pelican, $rand$ is a random number in the interval $[0,1]$, and ub and lb denote the upper and lower position boundaries of the pelican, respectively.

The hunting (iterative optimization search) process of POA algorithm is divided into two phases as follows:

Moving to the prey (exploration phase): in this phase the prey is randomly distributed in the search space, and the pelican recognizes the location of the prey and moves towards it. The mathematical model of the new position of the pelican in each iteration is:

$$x_{i,j}^{P1} = \begin{cases} x_{i,j} + rand \cdot (p_j - I \cdot x_{i,j}), & F_p < F_i \\ x_{i,j} + rand \cdot (x_{i,j} - p_j), & F_p \geq F_i \end{cases} \tag{10}$$

where $x_{i,j}^{P1}$ is the position of the i th pelican in the j th dimension after the first stage update, P_j is the position of the prey in the j th dimension, I is a random integer within $[1,2]$, F_p is the value of the objective function for the prey, and F_i is the value of the objective function for the i th candidate solution.

Skimming the water surface (development phase): this order of pelicans reaches the water surface

R, the new position of the pelican in each iteration is mathematically modeled as:

$$x_{i,j}^{P2} = x_{i,j} + R \cdot (2 \cdot rand - 1) \cdot x_{i,j} \tag{11}$$

$$R = 0.2 \cdot \left(1 - \frac{t}{T}\right) \tag{12}$$

where $x_{i,j}^{P2}$ is the position of the i th pelican in the j th dimension after the update in the second stage, t is the current iteration number and T is the maximum iteration number.

In both phases of the pelican hunting process, if the new position after updating is better, the position update is carried out, otherwise it remains unchanged, and the expression is as follows:

$$X_i = \begin{cases} X_i^{P_k}, F_i^{P_k} < F_i \\ X_i, else \end{cases} \tag{13}$$

where k is 1 or 2 and $X_i^{P_k}$ denotes the new position of the i th pelican after the k th stage update.

(2) Improved Pelican Optimization Algorithm (IPOA)

In order to solve the problems of slow optimization search in the late stage of the pelican algorithm and easy to fall into the local optimum, the detection region of each pelican is expanded from the local neighborhood to the whole solution space, and the search is carried out in a wider solution space, which significantly reduces the risk of falling into the local optimum, and improves the global search capability of the pelican algorithm. The modified equation is as follows:

$$x_{i,j}^{P2} = x_{i,j} + R \cdot \left(1 - \frac{t}{T}\right) \cdot (2 \cdot rand - \partial) \cdot x_{i,j} \cdot \Delta \tag{14}$$

where ∂ is the introduced probability factor and Δ is the increased optimization search factor, i.e.,:

$$\Delta = \left(\left\| \frac{F(x'_{i,j}) - F(x_{i,j})}{x'_{i,j} - x_{i,j}} \right\| \right) \cdot (x_{best,j} - x_{i,j}) \tag{15}$$

where $x'_{i,j}$ is the j -dimensional position of the i th pelican of the last iteration of $x_{i,j}$ (the current position), $F(x'_{i,j})$ is the objective function of $x'_{i,j}$, and the j th dimensional feature of $x_{best,j}$ is the current globally optimal position.

The main innovation of introducing Δ as the optimization search factor is to use the slope of the position around the current search point as the speed of the Pelican algorithm search, which solves the disadvantage of the Pelican algorithm's slow search for the optimum at a later stage. And the introduction of the probability factor ∂ enhances the probability distribution of the population to the optimal solution area aggregation, this strategy allows individuals to jump out of the current position with a certain probability, thus effectively avoiding the algorithm to prematurely converge on the local optimal solution, the algorithm not only maintains the ability of local development, but also enhances the stochasticity of the global exploration.

3.2.2. Enterprise Financial Risk Prediction Models

LSTM is a special kind of recurrent neural network that can deal with long-term dependency problems, and it has a wide range of applications in the fields of time series data processing and natural language processing. However, the performance of LSTM is highly dependent on the parameter settings, and finding the optimal parameter combination is a challenging problem. In terms of goal consistency, LSTM pursues better processing of time series data, and the IPOA algorithm aims to find the optimal solution, both of which ultimately aim to optimize performance. In terms of complementary advantages, LSTM is

sensitive to initial parameters, while the chaotic mapping of the IPOA algorithm allows for a more uniform distribution of the initial population, providing LSTM with a wide range of initial search points. In terms of balanced search ability, LSTM needs to balance global and local search, and the adaptive adjustment of inertia weight factors of IPOA algorithm can meet this need. Therefore, the combination of IPOA and LSTM is fully feasible.

The specific implementation steps of optimizing ARIMA-LSTM neural network based on IPOA algorithm are as follows:

Step1 Data preprocessing. Clean the data, remove noise, missing values or abnormal data, carry out data normalization, and convert the data to a suitable range.

Step2 Determine the parameters in the ARIMA (p, d, q) model, and use the ARIMA model to get the predicted and residual values.

Step3 Initialize the algorithm parameters. Determine the number of hidden layer neurons, the learning rate, the number of iterations, the batch size in the LSTM model, and set the pelican population size as N, the spatial dimension as D, and the maximum number of iterations as T.

Step4 Input the residual values into the LSTM model optimized based on the improved pelican algorithm to get the predicted values of the residuals. And combine the predicted value of ARIMA model with the predicted value of residuals of IPOA-LSTM to get the predicted value of ARIMA-IPOA-LSTM combined model.

The process of constructing the combined prediction model based on ARIMA-IPOA-LSTM is shown in Figure 1.

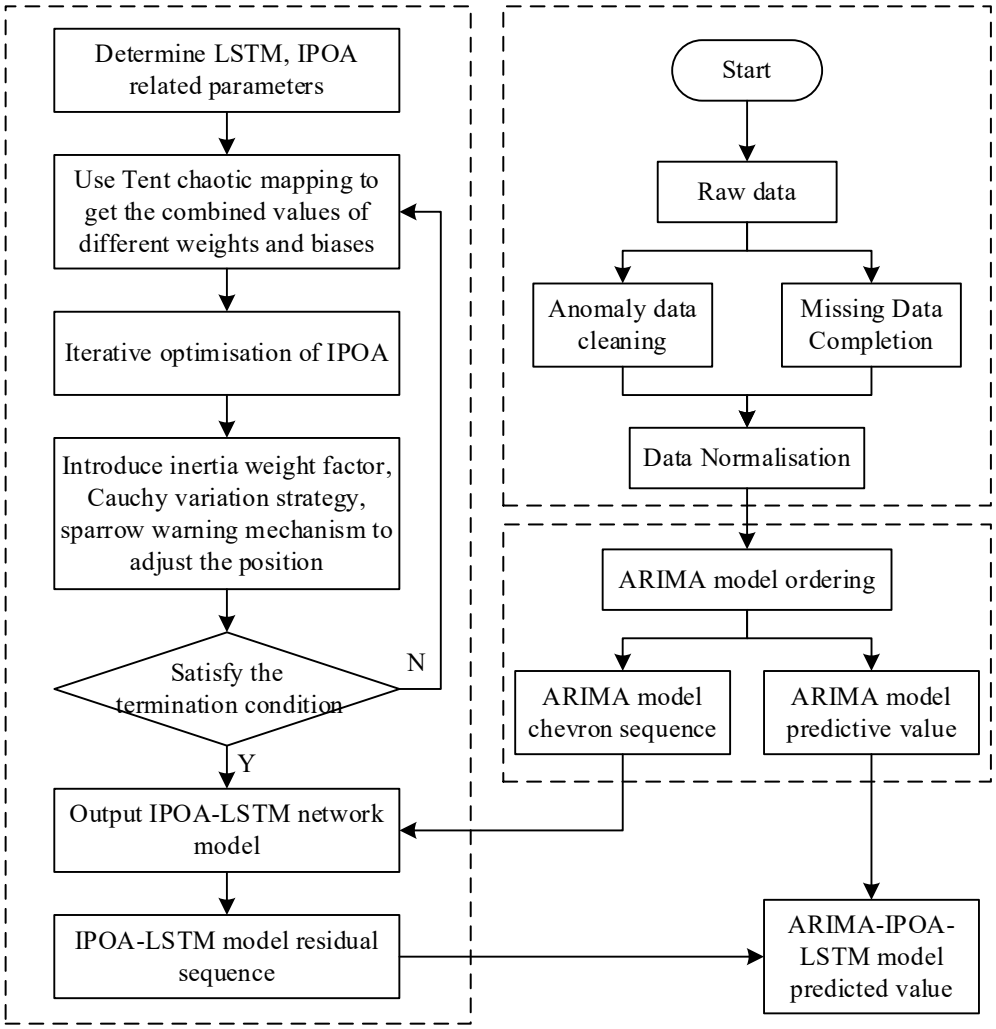


Figure 1. The combined prediction model of ARIMA-IPOA-LSTM.

3.2.3. Interpretability of Financial Forecasting Models

Since there are 18 statistical characteristic variables used in this paper for enterprise financial risk

prediction, it is obviously difficult to use the traditional mathematical empirical formula fitting method to build the estimation model, and it is necessary to screen the key characteristic variables based on the correlation between the characteristic variables and the characteristics of the financial risk, so as to be able to construct a reasonable empirical estimation model. In this section, SHAP theory will be used to build an analytical model with interpretability. SHAP theory can reveal the degree of contribution to the output of the target variable under the interaction of single and multiple feature variables in the prediction model, and select the features that have the greatest impact on the model prediction results, so as to reduce the complexity of the model and improve the performance of the model.

The core idea of SHAP theory is that the contribution of each feature to the whole game should be related to the order of its inclusion in different combinations of features, i.e., the marginal benefit of each feature added to the game should be considered. The calculation of Shapley values is the key to SHAP theory, which usually requires considering all possible combinations of feature subsets and calculating the marginal contribution of each feature in each combination. The Shapley value of each feature is then obtained by averaging over all possible combinations. Assuming that the total output value in the game is:

$$R = f(x) - f\left(\varnothing\right) \tag{16}$$

where $f(x)$ denotes the output obtained with all inputs and $f\left(\varnothing\right)$ denotes the output obtained with empty inputs. In order to fairly calculate the degree of contribution of each feature, the effect on the output value when that member participates or does not participate in the game under different circumstances needs to be taken into account. From this, the Shapley values for each different input cell can be calculated as:

$$\phi_i(f,x) = \sum\nolimits_{S\subseteq\{N\setminus i\}} \frac{|S|!(M-|S|-1)!}{M!}\left[f(x_{S\cup i}) - f(x_S)\right] \tag{17}$$

where $N = \{1,2,\ldots,M\}$ represents the subscripts of the feature variables in the dataset, i represents the i th feature, and M is the total number of feature variables. S is a subset of the set $\{1,2,\ldots,M\}$ with 2^{M-1} possibilities and represents the total number of elements in S. $f(x_S)$ denotes the model output value when there are S features in the sample, and $f(x_{S\cup i}) - f(x_S)$ denotes the marginal contribution of feature i under the set S.

The previous weights are obtained from permutations and represent the probability of an element being in the set.Shapley values are linearly additive for importance values computed for different models in the same input cell. The importance values are also the same for two input units that always play the same role. The importance of a redundant unit in the sample is equal to its output in the model for its individual inputs (a redundant unit is one that does not interact with any other unit), and the sum of the importance of all the input units is equal to the output of the model.

4. Validation of the Enterprise Financial Risk Prediction Model

With the development of computer science and technology, the currently used empirical modeling methods for financial risk early warning have also evolved from mathematical models to machine learning methods, when the neural network methods have entered an explosive stage. Financial risk data because of its time series characteristics, the recurrent neural network and other neural network models with time series characteristics applied to financial risk early warning has been imminent. Accurate financial risk prediction model can better and more accurately help the government, banks and other investors to monitor the financial risk of listed companies, but also to a certain extent to help listed companies to avoid financial risks and increase the financial and economic benefits of enterprises.

4.1. Tests of the ARIMA Model

4.1.1. Model ADF Smoothness Test

Based on the data provided by CSMAR, Wind, and CNRDS databases, this paper collects the historical datasets of Shanghai and Shenzhen A-market listed companies from 2012 to 2024, in which the vacant values are supplemented by manual linear interpolation, and each piece of data should contain

the specific date as well as the financial data situation.

In this paper, before using the ARIMA-IPOA-LSTM model for financial risk prediction, the smoothness of the financial risk time series data is tested. When the series is smooth, the mathematical expectation of the series at each moment should be a constant, the series should be consistent with the relative degree of probability distribution, such as non-smooth, the mean, variance and covariance will be dynamically changing without regularity suppression. Based on the time series data of corporate financial risk, it is not possible to directly and accurately determine with the naked eye whether the data is smooth, for this reason, this paper uses the ADF test. Figure 2 shows the time series graph after the first-order difference, and Table 2 shows the ADF test of the original time series and the first-order difference time series.

From the ADF test value of the original time series data, it can be seen that the test result value is -0.675 and greater than the confidence level value, the significance of the test level P value of 0.832>0.05, obviously did not pass the test of significance, which further illustrates that the sequence is not smooth. In the first-order difference ADF test, its ADF test result is -3.796, which is lower than the three confidence level values, and the significance test level P-value is 0.001<0.01, which passes the test of significance, i.e., the first-order time series has smoothness. Any two random variable most of the white noise series are uncorrelated and cannot extract effective features, so the white noise test needs to be performed in advance. The P-value of the test statistic is less than 0.05, which indicates that the time series of corporate financial risk data meets the requirements, so the first-order difference series can be modeled.

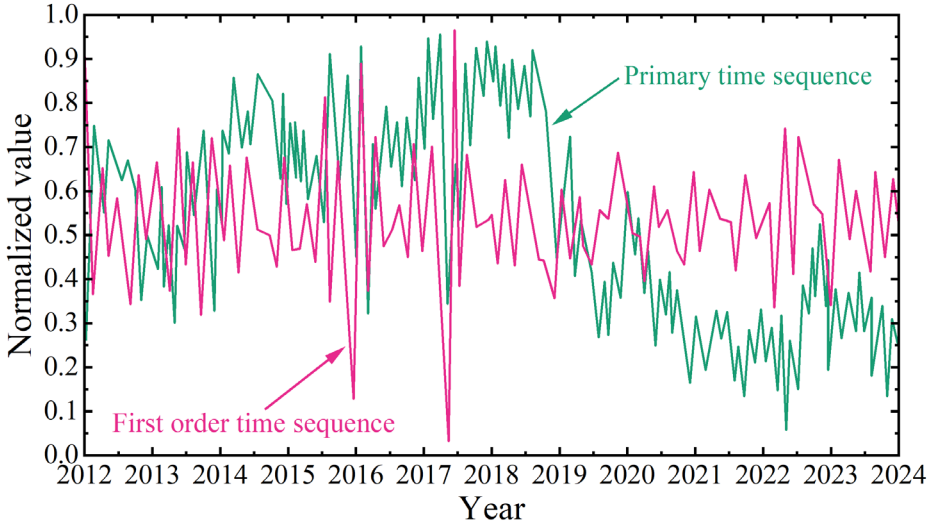
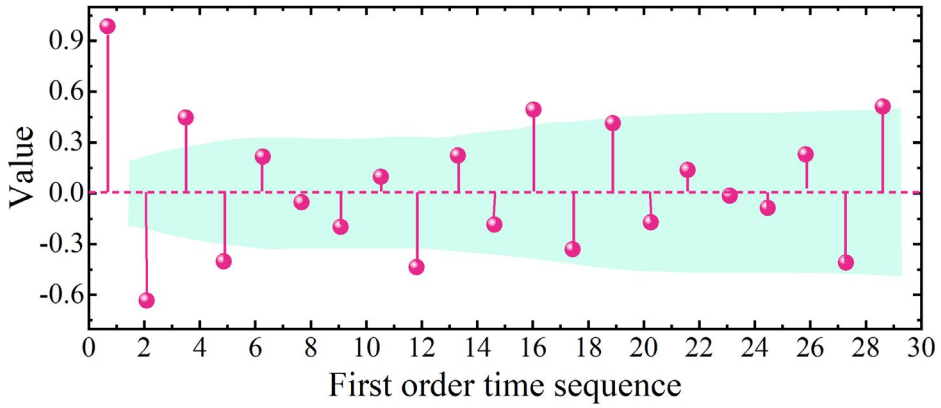


Figure 2. The time sequence diagram after first-order difference.

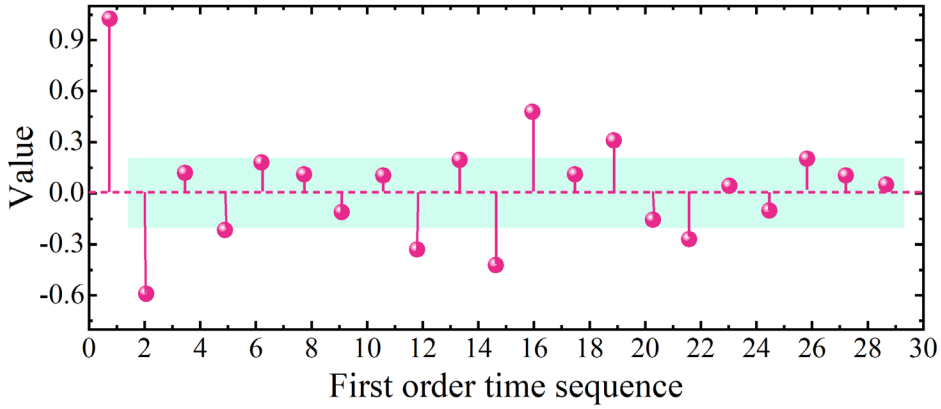
Table 2. ADF test results of time series.

Original timing sequence	Test result	*P*	95% CI of the difference		
			1%	5%	10%
	-0.675	0.832	-4.172	-2.931	-2.463
First-order difference time series	Test result	*P*	95% CI of the difference		
			1%	5%	10%
	-3.796	0.001	-3.407	-2.893	-2.381

The results of the first-order difference serial autocorrelation test for corporate financial risk data are shown in Figure 3, where Figure 3(a)~(b) show the test results of ACF and PACF, respectively. As can be seen from the figure, the ACF and PACF plots are trailing manifestations, and therefore satisfy the conditions for using the ARIMA algorithm in corporate financial risk prediction.



(a) The ACF diagram of order difference sequence



(b) The PACF diagram of order difference sequence

Figure 3. Autocorrelation test of first-order difference sequences.

4.1.2. Model Fitting Performance Tests

The fitting performance test can effectively verify the performance effect of the model training set, when the training set effect is better, the model complexity is higher, and can effectively learn the law. The fitting degree of the model is divided into three forms: fitting, underfitting and overfitting, which indicates that the model is in the optimal effect, underfitting indicates that the model has a high error, and overfitting indicates that the model complexity is higher than the actual problem, and can not have a good performance on the test set. Wavelet Support Vector Machine (WSVM) and gray GM(1,1) model are used as a control to test the fitting effect of ARIMA-IPOA-LSTM method, and the test results are shown in Figure 4. Where Figure 4(a)~(c) shows the fitting results of WSVM, GM(1,1) and the method of this paper, respectively.

Analyzing the fitting performance test results, it can be seen that the ARIMA-IPOA-LSTM model proposed in this paper is in the fitting state, and the financial risk prediction model has good computational arithmetic, and it can stabilize the output data and accurately predict the positive samples. This is due to the fact that the proposed method in this paper constructs a combined model, which in turn improves the generalization ability and convergence effect of the model. In contrast, the prediction results of the wavelet support vector machine method and the GM(1,1) model exhibit underfitting and overfitting states, respectively, and thus cannot obtain the prediction results efficiently. This shows that the corporate financial risk prediction model proposed in this paper has better applicability.

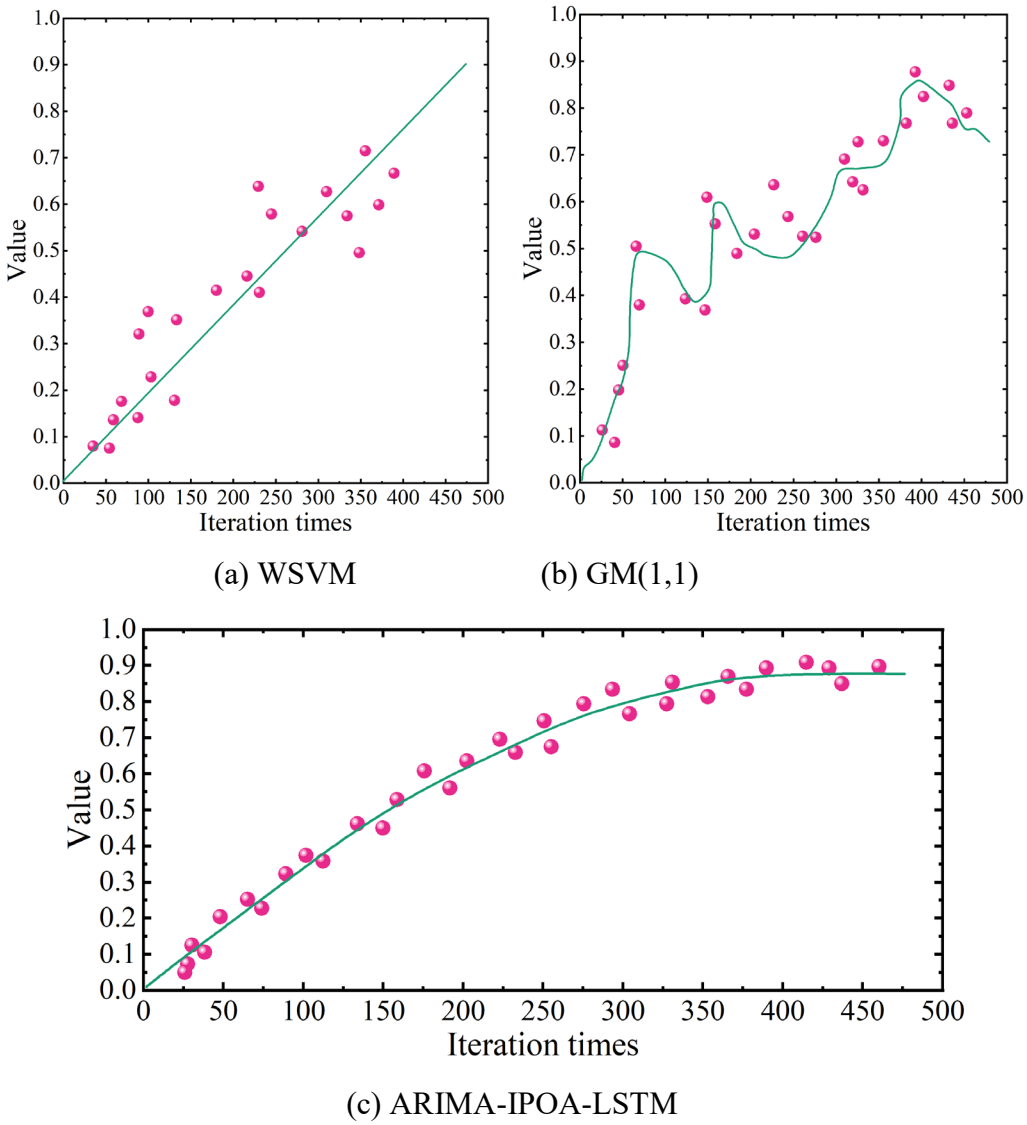


Figure 4. Model fitting performance test.

4.2. Effectiveness of the IPSO Algorithm

4.2.1. Comparison of Average Convergence Curves

In order to verify the superiority and robustness of the IPOA algorithm, this paper selects optimization algorithms published in recent years, including Black-winged Kite Optimization Algorithm (BKA), Hippopotamus Optimization Algorithm (HO), Parrot Optimization Algorithm (PO), and Benchmark Pelican Optimization Algorithm (POA), etc., to conduct a comparison and analysis with the IPOA algorithm to discern whether the improved algorithm can compare with other time comparison algorithms within a certain time range compared to other time algorithms, with superior performance.

In order to ensure the fairness and accuracy of the experiment, the population size of all the algorithms is set to 50, and the maximum number of iterations is also fixed to 300, and eight different types of test functions are selected based on the CEC2005 test set function, of which FT1~FT4 are single-peak test functions, and FT5~FT8 are multi-peak test functions. To avoid the chance of the algorithm results, each algorithm is run independently for 50 times, and the average convergence curves of IPOA and comparison algorithms in different types of test functions are plotted. The results of the average convergence curves of different algorithms are shown in Figure 5, where Figure 5(a)~(h) show the convergence curves of different algorithms in FT1~FT8 test functions, respectively. In the figure, the horizontal axis is the number of iterations in the operation of the algorithm, and the vertical axis is the average fitness value of the function under the corresponding number of iterations.

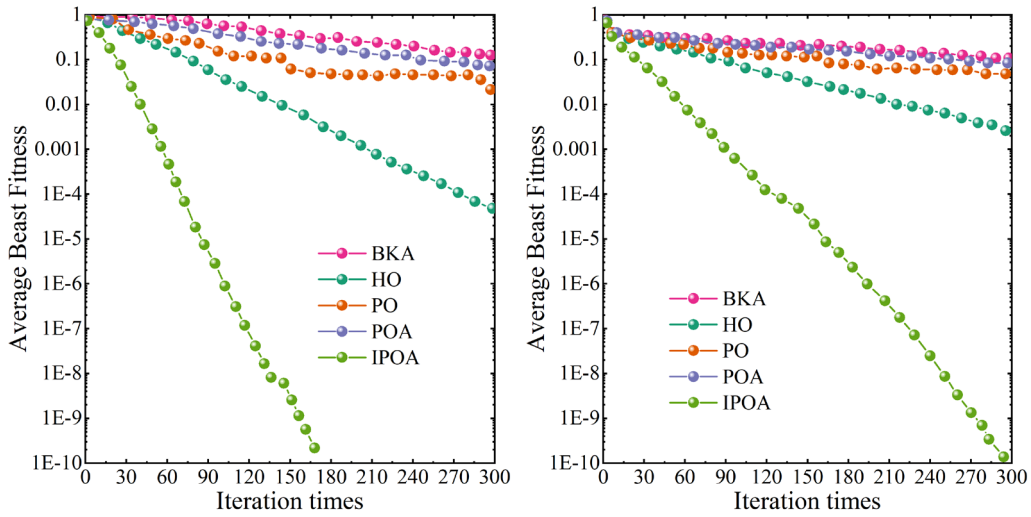
A detailed comparison and analysis of the convergence curves of different algorithms for the

FT1~FT8 test functions shows that:

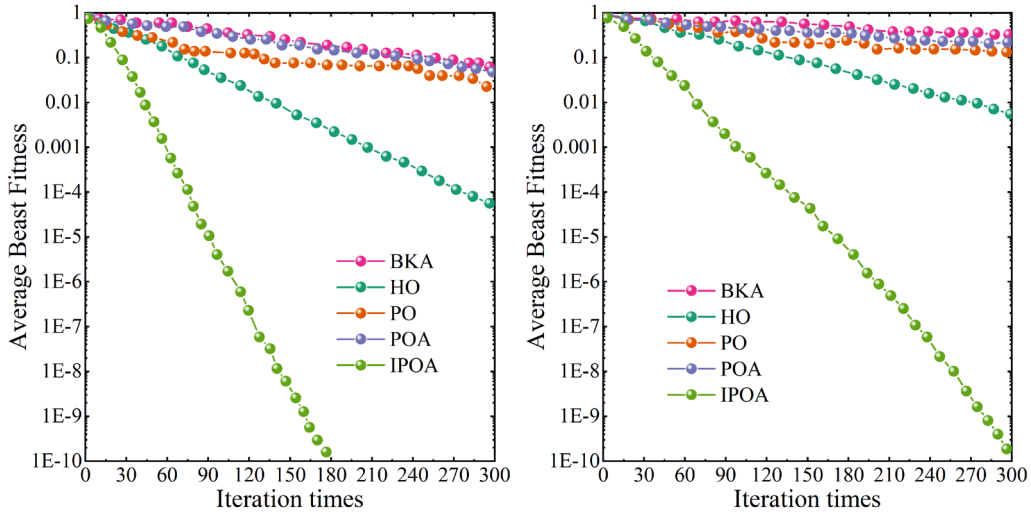
(1) For the single-peak test functions (FT1~FT4), the IPOA algorithm shows a higher convergence speed and an obvious advantage in the degree of searching for superiority. Compared with BKA, HO, POA, and PO algorithms, IPOA has improved the superiority by more than 200 orders of magnitude in the FT1~FT4 functions. Meanwhile, compared with the four comparative algorithms of BKA, POA, PO, and HO, the IPOA algorithm performs fewer iterations to achieve high optimality, which fully demonstrates its higher convergence speed.

(2) For the multi-peak test functions (FT5~FT8), the IPOA algorithm possesses higher convergence speed and the ability to jump out of local extremes. On functions FT5~FT7, IPOA only needs less than 30 iterations to converge to the theoretical optimum, which is better than the other compared algorithms. In the convergence curve graph of function FT8, it can be found that BKA and HO both finish convergence after about 30 iterations and fall into the local extreme value situation, while IPOA can continue to iterate and obtain a higher degree of finding the optimal English.

By analyzing the convergence curves above, IPOA has superior convergence speed and superiority over other comparative algorithms, which proves that the improved strategy improves the original POA algorithm in terms of jumping out of local extremes and accelerating the convergence speed, and verifies the superiority of the IPOA algorithm.



(a) FT1 (b) FT2



(c) FT3 (d) FT4

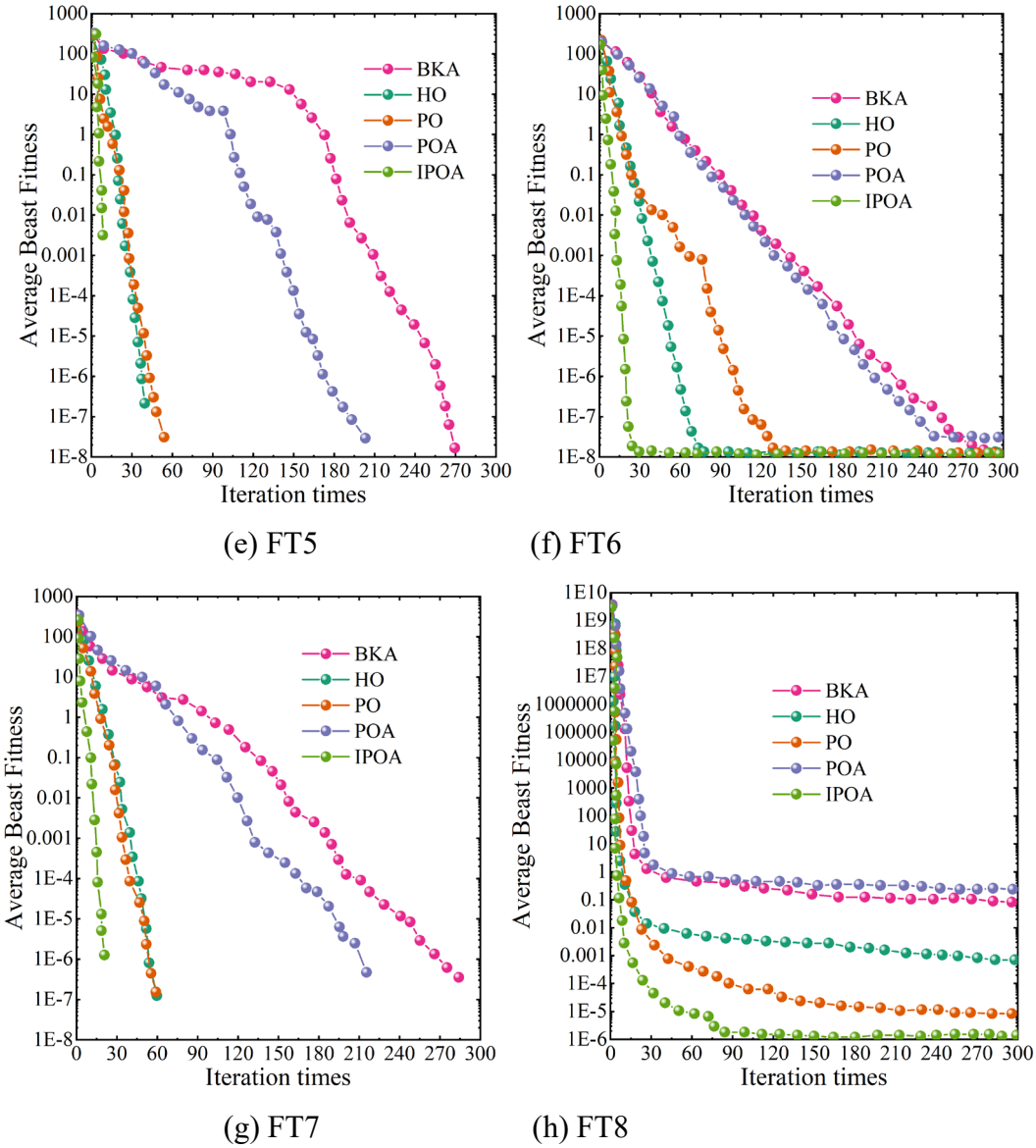


Figure 5. The average convergence curves of different algorithms.

4.2.2. Algorithm Optimization Accuracy Comparison

In order to perform meaningful statistical analysis, the relevant metrics of different algorithms such as average (Avg), standard deviation (Std), optimal value (Bst) and worst value (Wst) are counted, and their specific results are shown in Table 3 for individual performance comparison.

From the comparison results of the optimization accuracy of individual algorithms in the table, IPOA algorithm performs better in FT1 to FT8 functions. All statistical indicators are better than other algorithms in FT1 to FT8 functions. In addition, the IPOA algorithm is able to find the global optimum, i.e., the ideal value, in the desertification of each benchmark function.The performance of the IPOA algorithm in the single-peak and high-dimensional multi-peak functions shows its strong solving ability and its ability to avoid falling into the local optimum. The performance of the IPOA algorithm on the fixed-dimensional multi-peak function again demonstrates its ability to avoid the local optimum and its high search performance. In conclusion, the IPOA algorithm not only converges to most functions but also has good accuracy and stability. Therefore, IPOA algorithm can solve the optimization problem of objective function more effectively.

Table 3. Comparison of algorithm optimization accuracy.

-	Index	IPOA	POA	BKA	HO	PO
FT1	Avg	3.16E-08	7.71E-05	7.88E-05	1.06E-05	9.92E+03
	Std	0.00E+00	2.59E-02	9.13E-06	3.11E-04	1.24E+04
	Bst	3.76E-11	3.45E-10	3.81E-07	5.51E-06	2.05E+02
	Wst	6.35E-06	1.12E-05	3.02E-03	1.63E-02	4.48E+03
FT2	Avg	6.21E-09	1.16E-08	5.11E+02	1.55E+04	5.51E+04
	Std	0.00E+00	5.16E-05	1.92E+02	5.93E+02	6.03E+04
	Bst	3.41E-12	3.35E-10	3.15E+02	5.72E+03	0.00E+00
	Wst	1.26E-08	2.33E-05	1.14E+02	2.67E+04	2.29E+05
FT3	Avg	6.83E-08	3.02E-05	9.46E-01	1.45E+02	2.06E+02
	Std	2.63E-06	6.33E-06	2.16E-02	4.75E+01	9.95E+01
	Bst	5.57E-11	4.96E-09	7.06E-02	7.41E+01	6.81E+01
	Wst	1.25E-07	2.25E-04	1.46E-01	2.33E+02	4.45E+00
FT4	Avg	2.43E-07	2.76E-04	6.03E-02	1.05E+03	2.06E+06
	Std	1.85E-02	1.69E-04	3.26E+02	2.05E+03	3.41E+06
	Bst	2.43E-10	2.69E-06	1.75E+01	2.58E+02	2.21E+03
	Wst	2.48E-05	2.88E-01	1.42E+03	9.55E+02	1.15E+05
FT5	Avg	3.05E-05	3.12E-03	8.62E-04	2.75E-04	1.76E-03
	Std	1.94E-11	2.62E-06	1.42E-04	6.03E-04	1.62E-02
	Bst	3.06E-05	3.03E-02	6.03E-04	3.95E-04	5.18E-04
	Wst	3.02E-04	4.25E-03	1.11E-03	2.02E-02	8.36E-02
FT6	Avg	3.92E-02	3.92E-02	3.46E-01	3.94E-01	3.92E-01
	Std	0.00E+00	0.00E+00	0.00E+00	8.25E-13	3.66E-15
	Bst	3.91E-03	3.85E-02	3.58E-01	3.66E-01	3.98E-01
	Wst	2.94E-02	2.73E-01	3.27E-01	3.48E-01	3.51E-01
FT7	Avg	3.15E+00	3.12E+00	3.64E+00	3.78E+00	3.89E+01
	Std	9.27E-13	9.16E-10	1.42E-16	3.35E-12	4.19E-06
	Bst	3.16E+00	3.11E+00	3.48E+00	3.76E+01	3.81E+01
	Wst	3.13E+00	3.10E+00	3.51E+00	4.92E+02	5.16E+02
FT8	Avg	2.94E-05	2.91E-04	2.64E-02	3.17E-03	3.37E-04
	Std	4.02E-13	4.08E-10	6.15E-02	5.61E-04	5.51E-06
	Bst	2.35E-07	2.31E-06	2.76E-02	3.08E-02	3.06E-05
	Wst	2.31E-02	2.30E-01	2.53E-01	2.99E-01	3.74E-01

4.3. Effectiveness of Financial Risk Prediction

4.3.1. Validation of Financial Risk Projections

In order to better illustrate the effect of the financial risk prediction model proposed in this paper, the model of this paper and the prediction model based on LSSVM are compared in terms of both training sample test and test sample test. The number of iterations of the algorithm is set to 200, and five nodes are inputted into the model, corresponding to the five principal component indicators, i.e., the relevant data on solvency, profitability, operating ability, development ability, and cash flow in the data label of the enterprise's financial risk collected in the previous section, in order to carry out the prediction of the enterprise's financial risk. Figure 6 shows the model's fitness curve and Figure 7 shows the model's financial risk prediction accuracy, where Figure 7(a)~(b) shows the accuracy of the training samples and test samples, respectively.

From the model's fitness curve graph, it can be found that the best fitness value of the ARIMA-LSTM model optimized based on the IPOA algorithm is significantly higher than the average fitness value, which indicates that the model has a high prediction ability in corporate financial risk. From Figure 7(a), it is found that the accuracy of the training sample mainly reflects the comparison of the model based on the established data on the difference between the predicted value and the actual value before, the accuracy is usually used for the effect of in-sample detection, the training sample includes 50 non-financial risk listed enterprises and 25 financial risk enterprises paired with them, the model of this paper is stable compared to the curve of the LSSVM model, and the model of this paper has an overall accuracy of 94.02%. Figure 7(b) demonstrates 30 non-financial risky listed firms and 15 financial risky firms paired with them, in which the overall discrimination accuracy of this paper's model is 93.48%. The above validation results get that the ARIMA-IPSO-LSTM model designed in this paper is feasible to be used for the prediction of financial risk of listed enterprises, and it is found from the data that the financial

risk of listed enterprises is predicted more accurately.

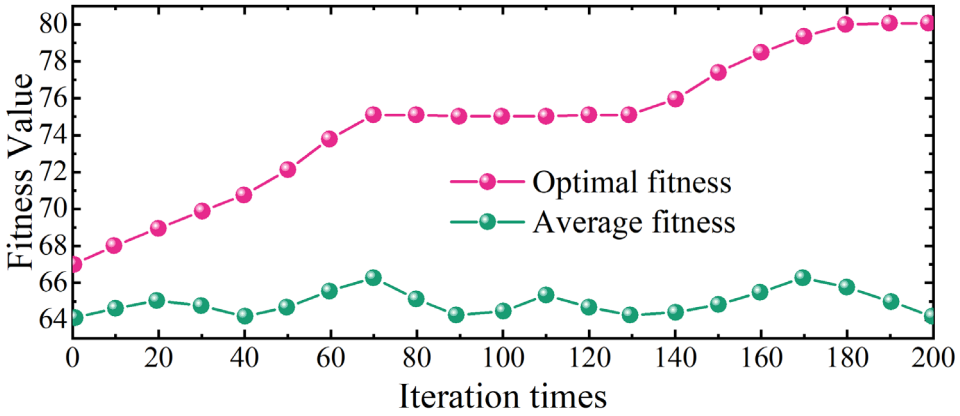


Figure 6. Fitness curve.

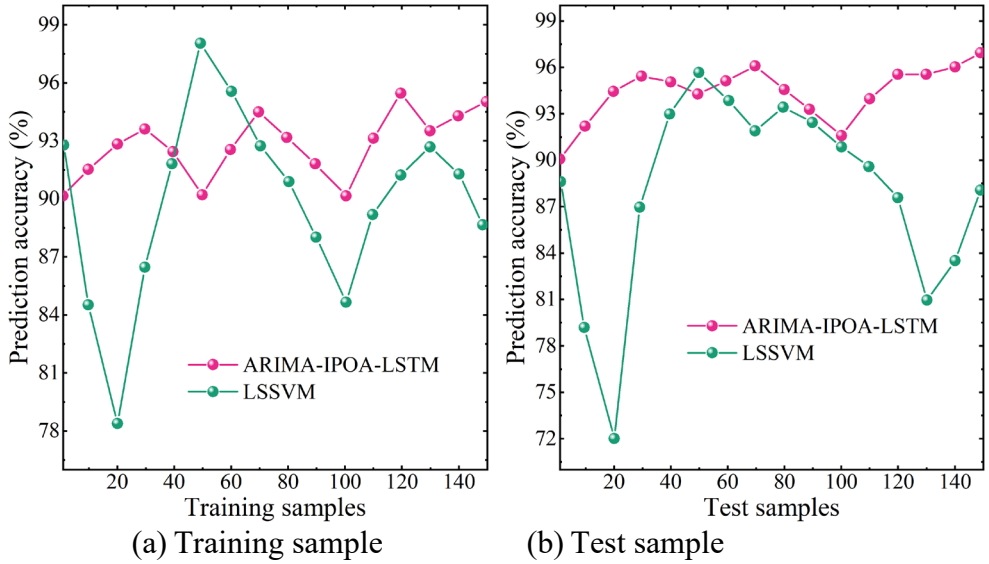


Figure 7. Verification of financial risk prediction.

4.3.2. Actual Model Prediction Performance

Based on the financial data of Shanghai and Shenzhen A-share listed enterprises collected in this paper from 2012 to 2024, the financial data of 15 listed enterprises in 2022 and 2023 are selected as the basis, and the ARIMA-IPOA-LSTM model is used to simulate the financial samples of the 15 listed enterprises in 2024, and to predict the corporate financial risk of the 15 listed enterprises in 2024. The corresponding warning output is obtained according to the predicted values of important indicators and the set thresholds, and the warning accuracy is determined according to the comparison between the warning output and the industry standardized values of financial indicators as shown in Table 4.

According to the simulation results of the financial samples of 15 listed companies in 2024 by the ARIMA-IPOA-LSTM model, the minimum and maximum warning accuracy of the proposed model reached 88.89% and 100%, respectively, and the financial risk prediction performance of the first listed company was the worst, while the ARIMA-IPOA-LSTM model could provide early warning for all risks. In terms of RMSE, the ARIMA-IPOA-LSTM model is more stable in making corporate financial risk predictions, with values ranging from 0.574*10-5 to 1.166*10-5. The industry normative values of the financial indicators are quoted from the annual reports of the industry analysis of the securities companies, specifically, the predicted value exceeds the set threshold, which is judged to have financial risk.

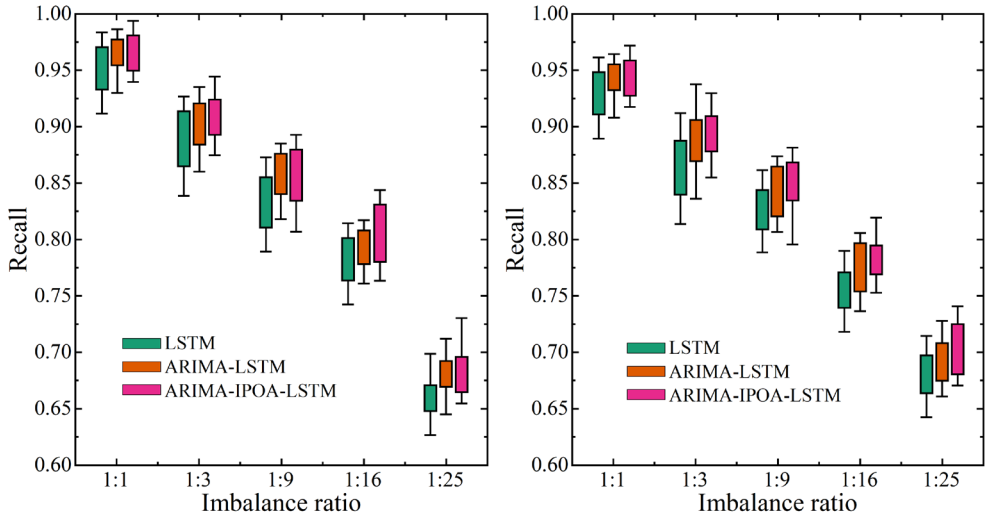
Table 4. Early warning results of enterprise financial risks.

No.	Actual value	Predictive value	Accuracy/%	*RMSE*/10^{-5}
1	18	16	88.89	0.841
2	11	10	90.91	1.043
3	14	13	92.86	0.638
4	9	9	100.00	0.574
5	6	6	100.00	0.959
6	13	13	100.00	1.095
7	3	3	100.00	1.166
8	8	8	100.00	0.787
9	17	16	94.12	1.125
10	8	8	100.00	0.778
11	12	11	91.67	0.792
12	6	6	100.00	0.761
13	15	14	93.33	1.075
14	9	9	100.00	0.846
15	12	12	100.00	0.794

4.3.3. Model Performance Comparison Results

Based on the data analysis in the previous section, this paper conducts experiments on a self-constructed financial dataset of listed firms and analyzes the results in detail. The reported results are the mean and standard deviation of five experiments. This study delves into the variation of each metric to get a comprehensive understanding of the performance of different models under different imbalance ratios. In this section of experiments, the data of 2012, 2016, 2020 & 2024 are selected as benchmarks to compare with LSTM, ARIMA-LSTM and ARIMA-IPOA-LSTM models, and samples under different imbalance ratios are selected to develop the validation. The recall rate is chosen as the evaluation index, and the results of the recall comparison of different models under different imbalance ratios are obtained as shown in Figure 8, where Figure 8(a)~(d) show the recall rate of financial risk prediction in 2012, 2016, 2020 & 2024, respectively.

It can be seen that the ARIMA-IPOA-LSTM model obtains the highest recall up to 96.72% in all 20 sets of experiments. As the imbalance ratio of the sample data increases, the ARIMA-IPOA-LSTM model shows a more significant advantage in terms of recall, especially when the imbalance ratio is 1:16 and 1:25. The significant improvement observed in this study can be attributed to the use of the improved Pelican optimization algorithm, which allows the model to allocate more performance to a small number of categories. As a result, this improvement improves the categorization performance on unbalanced data. In addition, ARIMA-LSTM also outperforms LSTM in 20 sets of comparison experiments, indicating that the introduction of the ARIMA model significantly improves the model's ability to process time-series data, which can effectively enhance the predictive recall of corporate financial risks.



(a) 2012 (b) 2016

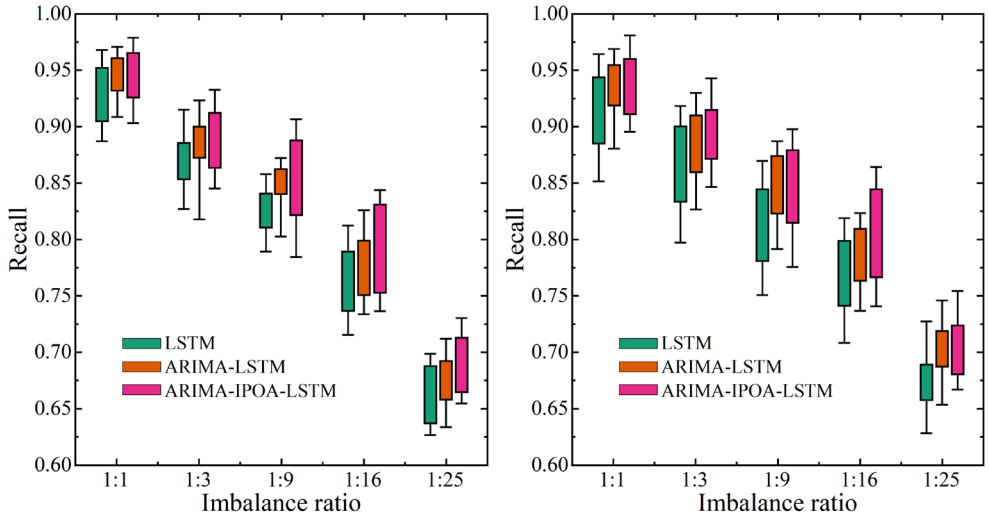


Figure 8. The recall rate of different uneven scales.

For the F1 scores, the performance of each model decreases as the disequilibrium rate increases, and although the F1 scores of ARIMA-IPOA-LSTM and ARIMA-LSTM are very close under different experimental conditions, the model in this study performs best overall. In addition, the same regularity exists between ARIMA-LSTM and LSTM. This shows that the IPOA algorithm and the ARIMA model can improve the F1 scores of the models well. In the 20 sets of comparison experiments, the ARIMA-IPOA-LSTM model has the highest accuracy of 88.75%. Although the accuracy of each model decreases as the percentage of imbalance increases, the inclusion of ARIMA models decreases the accuracy more slowly, indicating that they are more resistant to data imbalance.

The AUC value is an important criterion for evaluating the performance of classification models, especially in the case of positive and negative sample imbalances. The results of the study found that as the imbalance rate increases, the AUC rate generally remains stable and is not greatly affected, which is consistent with the results of previous studies. However, the AUC value of the ARIMA-IPOA-LSTM model in this paper has a significant advantage, indicating a greater ability to distinguish between positive and negative samples and superior performance.

In conclusion, the ARIMA-IPOA-LSTM model proposed in this study outperforms the ARIMA-LSTM and LSTM models in terms of their ability to perform corporate financial risk prediction, especially in terms of a few samples, suggesting that the integration of the IPOA algorithm and the ARIMA model is very effective in enhancing the predictive ability of the model.

4.3.4. Predictive Model Interpretability

Comprehensive analysis of the above shows that the ARIMA-IPOA-LSTM model established in this paper has better performance in enterprise financial risk prediction, but the complex neural network model has poor interpretability, and the neural network is often used as a “black box”. People only know what is input in the neural network and what is output, but they cannot understand what is happening inside the black box, that is, they cannot get the causal relationship between the input index data and the output probability value of financial risk. In this paper, with the help of SHAP theory to compare the size and direction of the influence of each indicator on the model output, to solve the visualization problem of the model, and to make the complex ARIMA-IPOA-LSTM model with interpretability.

In the financial data of Shanghai and Shenzhen A-share listed companies collected in this paper from 2012 to 2024, five main features are included, namely, solvency, profitability, operating ability, development ability, and cash flow, which include 18 different types of features. In order to further analyze the overall contribution of each feature in the sample to the prediction structure, the SHAP of the sample data of the test set is counted, and the SHAP of each feature after aggregation is shown in Figure 9. A point in the figure represents a sample, and dense places indicate a large number of samples aggregated. The horizontal coordinate is the SHAP value, which expresses the influence of features on the model output, and the vertical coordinate is different features, and the color indicates the magnitude of the value of the feature itself, the redder the color indicates the higher the value, and the greener the value indicates the lower the value.

As can be seen in the figure, total assets cash recovery ratio (FR18), net cash flow from operating

activities per share (FR17), total assets turnover (FR10), accounts receivable turnover (FR9), gearing ratio (FR3), total assets growth rate (FR13), current assets turnover (FR12), current ratio (FR1), net sales margin (FR8), The differences in the characteristics of the indicators such as total return on assets (FR7) have a more significant effect on the predicted values of the ARIMA-IPOA-LSTM model. Analyzing these indicators, it can be found that among the top 10 characteristics of the importance of ARIMA-IPOA-LSTM model, there are indicators of cash flow, profitability, operating capacity, etc., which indicates that the financial indicators under different types all have an important impact on the prediction of corporate financial risk. In addition, among all the predictive indicators, the feature that has the greatest impact on the prediction results of ARIMA-IPOA-LSTM model is the cash recovery rate of total assets (FR18), and the financial optimization of the enterprise can focus on the top five indicators in order to improve the stability of the enterprise's financial development.

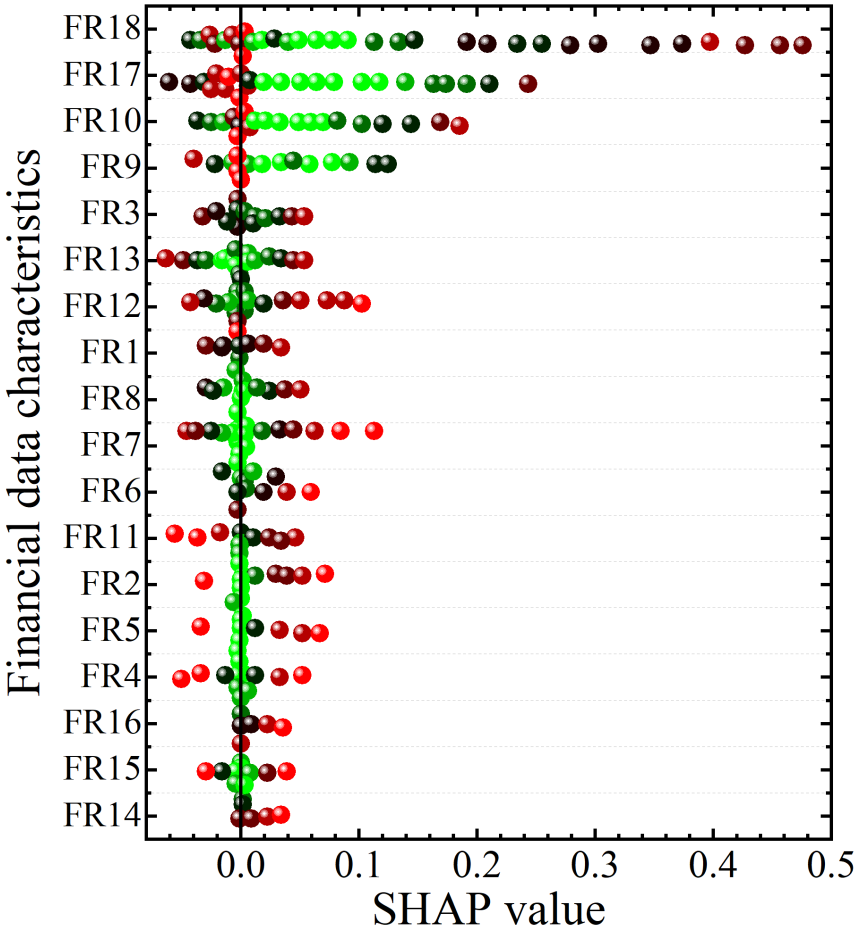


Figure 9. Features of SHAP summary results.

5. Conclusion

The ARIMA-IPOA-LSTM combined model shows excellent performance in corporate financial risk prediction. In the model's prediction of 15 listed enterprises in FY2024, the accuracy of early warning ranges from 88.89% to 100%, among which 9 enterprises are successfully and completely identified as risky subjects, which reflects the model's sensitivity and responsiveness to high-risk events. Regarding the prediction error, the RMSE value is at a minimum of 0.574 × 10-5 and at a maximum of no more than 1.166 × 10-5, which indicates that the model has the ability to stabilize the output. The model's recall under handling 1:16 and 1:25 unbalanced sample proportions is up to 96.72%, further validating its effectiveness under extreme sample structures. In terms of AUC, the model remains highly stable and outperforms the traditional model, indicating its ability to distinguish between positive and negative samples. Combined with SHAP interpretability analysis, key indicators such as cash recovery rate of total assets and net cash flow from operating activities per share contribute significantly to the model output, which provides a scientific basis and direction for enterprises to implement forward-looking financial governance.

Funding
This research received no external funding.

Conflict of Interest Statement
The author declare no conflicts of interest.

Data Availability Statement
Not applicable.

References

1. Al-Okaily, M., & Al-Okaily, A. (2024). Financial data modeling: an analysis of factors influencing big data analytics-driven financial decision quality. Journal of Modelling in Management.
2. Rauf, M. A., Shorna, S. A., Joy, Z. H., & Rahman, M. M. (2024). Data-driven transformation: Optimizing enterprise financial management and decision-making with big data. Academic Journal on Business Administration, Innovation & Sustainability, 4(2), 94-106.
3. Thanasas, G. L., & Kampiotis, G. (2024). The role of Big Data Analytics in Financial Decision-Making and Strategic Accounting. Technium Business and Management, 10, 17-33.
4. Singh, V., Chen, S. S., Singhania, M., Nanavati, B., & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda. International Journal of Information Management Data Insights, 2(2), 100094.
5. D'Acunto, F., & Rossi, A. G. (2023). IT meets finance: financial decision-making in the digital era. In Handbook of financial decision making (pp. 336-354). Edward Elgar Publishing.
6. Gao, J. (2023). Importance of introducing big data into financial management. Journal of Science, 2(1).
7. Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. Journal of Big Data, 7(1), 21.
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. Elumilade, O. O., Ogundeji, I. A., Ozoemenam, G. O. D. W. I. N., Omokhoa, H. E., & Omowole, B. M. (2023). The role of data analytics in strengthening financial risk assessment and strategic decision-making. Iconic Research and Engineering Journals, 6(10).
10. Begenau, J., Farboodi, M., & Veldkamp, L. (2018). Big data in finance and the growth of large firms. Journal of Monetary Economics, 97, 71-87.
11. Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. Accounting horizons, 29(2), 423-429.
12. Bao, Q. (2024). Advancing Corporate Financial Forecasting: The Role of LSTM and AI in Modern Accounting. Transactions on Computational and Scientific Methods, 4(6).
13. Yang, A. (2025). Big data-driven corporate financial forecasting and decision support: a study of CNN-LSTM machine learning models. Frontiers in Applied Mathematics and Statistics, 11, 1566078.
14. Chu, H. (2021). An Empirical analysis of corporate financial management risk prediction based on associative memory neural network. Computational intelligence and neuroscience, 2021(1), 4383742.
15. Huang, Y., Gao, Y., Gan, Y., & Ye, M. (2021). A new financial data forecasting model using genetic algorithm and long short-term memory network. Neurocomputing, 425, 207-218.
16. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European journal of operational research, 270(2), 654-669.
17. Wang, S. (2021). An interview with Shouyang Wang: research frontier of big data-driven economic and financial forecasting. Data Science and Management, 1(1), 10-12.
18. Zhongsheng Zhou & Jingyao Zhang. (2025). Manufacturing enterprise digital transformation, financial flexibility, and financial risk—Evidence from China. International Review of Financial Analysis,104(PA),104279-104279.
19. Jing Chen& Bo Sun. (2024). Enhancing Financial Risk Prediction Using TG-LSTM Model: An Innovative Approach with Applications to Public Health Emergencies. Journal of the Knowledge Economy,16(1),1-21.
20. Qian Cao. (2023). An enterprise financial data leakage risk prediction based on ARIMA-SVM combination model. International Journal of Applied Systemic Studies,10(3),169-181.