Journal of Global Information Management
Volume 30 • Issue 7

An Improved BPNN Algorithm Based on Deep Learning Technology to Analyze the Market Risks of A+H Shares

Yi Wu, EMLYON Business School, Écully, France

Delong Zhu, School of Management Engineering, Anhui Institute of Information Technology, Wuhu, China

Zijian Liu, University of International Business and Economics, Beijing, China

Xin Li, Economics College, Jiaxing University, Jiaxing, China*

ABSTRACT

The backpropagation neural network (BPNN) algorithm of artificial intelligence (AI) is utilized to predict A+H shares price for helping investors reduce the risk of stock investment. First, the genetic algorithm (GA) is used to optimize BPNN, and a model that can predict multi-day stock prices is established. Then, the principal component analysis (PCA) algorithm is introduced to improve the GA-BP model, aiming to provide a practical approach for analyzing the market risks of the A+H shares. The experimental results show that for A shares, the model has the best prediction effect on the price of Bank of China (BC), and the average prediction errors of opening price, maximum price, minimum price, as well as closing price are 0.0236, 0.0262, 0.0294, and 0.0339, respectively. For H shares, the model constructed has the best effect on the price prediction of China Merchants Bank (CMB). The average prediction errors of opening price, maximum price, minimum price, and closing price are 0.0276, 0.0422, 0.0194, and 0.0619, respectively.

KEYWORDS

Artificial Intelligence, BP Neural Network, Global A+H Shares, Risk Analysis

INTRODUCTION

As financial exchanges between mainland China and Hong Kong become more frequent, many investors begin to analyze and study the H-share market. However, according to the current findings, the research on A+H shares is not comprehensive enough. The so-called A+H shares refer to the stocks of companies listed on the Shanghai/Shenzhen Stock Exchange in mainland China as A shares and simultaneously the exchange in Hong Kong, China as H shares (Dimpfl & Kleiman, 2019; Loriot et al., 2020). Various factors can affect the price of A+H shares, such as the economic and political background, human manipulation, and the overall psychological quality of investors (He et al., 2018). Under the trend of globalization, numerous companies are launching H shares in Hong Kong. Therefore, the analysis of A+H shares can help solve real problems. The scholars have done many works on analyzing the market risks of A share and A+H shares. Wu and Choudhry (2018) took the Chinese A-share market as the research object and conducted empirical research on the influence of company-level information uncertainty on momentum profits from seven aspects, including company

DOI: 10.4018/JGIM.293277

*Corresponding Author

size, company age, and reported volatility. The experimental results showed that the uncertainty of information magnifies momentum profits (Wu & Choudhry, 2018). Jiang and Subramanian (2019) took the securities market of the United States as the research object and proposed an auto-regressive integrated average model to predict the future stocks' value. The experimental results proved that the Auto-regressive Integrated Moving Average (ARIMA) model could provide good results for processing time-series data (Jiang & Subramanian, 2019). However, it is worth noting that no matter which method is used in the published academic research, the results are not ideal, or there is still room for improvement. Few researchers are interested in both the A-share market and the H-share market simultaneously (Liu & Kemp, 2019; Li & Bastos, 2020).

The securities market is an investment market with high returns and high risks. How to correctly judge the development trend of the stock market to obtain higher returns while minimizing risks has become a critical issue that securities investors should pay attention to. The stock market is a complex nonlinear decision-making system affected by many factors. Among them, macroeconomic factors include economic cycles, international financial shocks, and currency changes; macroeconomic policies include regulatory policies, industrial policies, and monetary policies. Under the combined effect of these factors, the securities market has undergone violent fluctuations, and investment risks are incredibly high. Predicting the trend of stock prices and the fluctuation range of stock prices can be very helpful for securities investors to make correct securities choices.

At present, the back propagation neural network (BPNN) is the most popular artificial intelligence (AI) algorithm, whose basic concept is that the learning process is composed of forward as well as back propagation. Regarding the forward propagation, the samples are transferred from the input layer and then processed by the hidden layer to the output layer. The error will enter back propagation when the difference between the actual output and the expected output goes beyond the specified range. It means that the output error is propagated back to the input layer layer by layer, and it is distributed in all units of each layer, so that the error signal in each unit can be obtained.

To this end, an improved BPNN algorithm was proposed based on deep learning technology for the analysis of the market risks of A+H shares. First, the genetic algorithm (GA) was used to optimize BPNN, and a model that can predict multi-day stock prices was established. Then, the Principal Component Analysis (PCA) algorithm was introduced to improve the GA-BP model, aiming to provide a practical approach for analyzing the market risks of the A+H shares.

METHOD

Shares Price Prediction and AI

At present, scholars in related fields have proposed the concept of hybrid algorithm, which is the combination of artificial intelligence and traditional algorithm. While the conventional method is used to deal with the linear part of financial time series, and the intelligent algorithm is used to deal with the residuals, resulting in better analysis and prediction effect. Computational intelligence is to recognize and simulate intelligence from the perspective of biological evolution. There are some same points of all approaches mentioned above, which can be summarized as follows: adaptive structure, randomly generated or specified initial state, fitness evaluation function, operation of modifying structure, system state memory, condition of terminating calculation, method of indicating result, parameters of control process. These methods of computational intelligence are characterized by self-learning, self-organizing, self-adaptive, simple, general, as well as robust. Besides, they can be applied to parallel processing and perform well. Artificial neural network (ANN) depends on the complex system and adjusts the correlation among multiple internal nodes, so that it can process information. ANN can conduct self-learning as well as self-adapting. With the corresponding input-output data provided in advance, the potential rules among them can be analyzed and grasped. Finally, the output results with new input data can be calculated on the basis of these rules. In the analysis of financial time series, BPNN is usually used, one of which uses back propagation algorithm is multi-layer feedforward network. It is introduced into the analysis of financial time series. The residual obtained from the traditional method of analyzing historical data can be regarded as the input of neural network. Through training, the nonlinear part of the analysis value or prediction value is obtained. Then, the linear part of the analysis value or prediction value obtained by the traditional method is superimposed, and the superimposed result is the final analysis value. In this way, the accuracy of analysis and prediction is greatly improved.

An Overview of A+H Shares

The A+H shares refer to stocks simultaneously listed as A shares on Shanghai/Shenzhen Stock Exchange and as H shares on the Hong Kong Stock Exchange. The shares issued by the same company are listed on different exchanges, and there are some differences in stock prices (Iyke et al., 2021; Suzuki et al., 2020). Companies in the transportation, petrochemical, and financial services sectors in China often issue A+H shares. According to the different issuance orders of A share and H share, there are three types of issuance modes:

1) The issuance mode from A to H;
2) The issuance mode from H to A;
3) The issuance mode of A+H.

The A+H shares' mode is of great significance for the gradual connection between mainland China's domestic capital market and Hong Kong's capital market (Sarwar et al., 2020). First, it represents the continuous acceleration of the internationalization of China's financial market. Second, the mode of A+H shares can help the state-owned enterprises in China absorb foreign investment via the Hong Kong market, thereby increasing their international competitiveness and reputation, helping them expand their scale, and strengthening their examples. Third, the mode of A+H shares can further improve the structure of Hong Kong's listed companies and help them consolidate their financial status. However, the mode of A+H shares also has some hidden dangers. Due to the large scale of the company and the limited digestibility of the A-share market (Ali et al., 2020). Relevant institutions and departments must plan and arrange the speed and pace of the company's issuance and listing in the A-share market step by step. Otherwise, it will be difficult for the A-share market to undertake the expansion speed of the stock market. Coupled with investor panic, it is easy to cause a continued downturn in the A-share market. Second, although Chinese companies have been listed overseas for ten years, the H share market has problems with the equity system, corporate management mechanism, and executive incentive system (Maqsood et al., 2020).

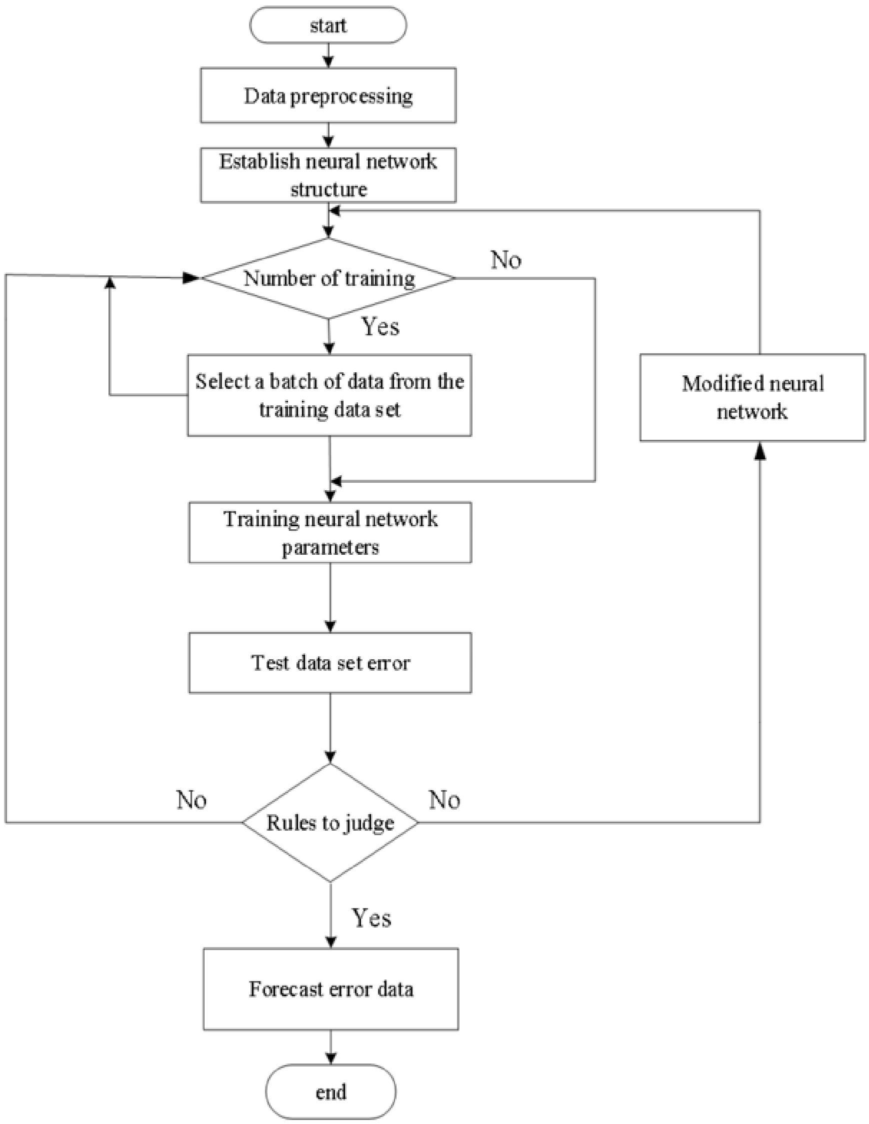
An Introduction to BPNN

A complete BPNN includes one input layer, one output layer, as well as many hidden layers. In a BPNN, the neurons between hidden layers are connected (Batten et al., 2019; Zhang & Wang, 2019); however, the input layer, together with the hidden layer, the hidden layer, together with the hidden layer, as well as the hidden layer, together with the output are fully connected. The algorithm of BPNN consists of the following three processes.

1) When the input signal enters BPNN, it flows through the input layer to the hidden layer. Then, it passes to the output layer. When the signal passes to the neurons on the hidden layer, it passes to the next hidden layer after the neurons, and finally, to the output layer.
2) The error signal refers to the difference between the actual output and the expected output of the network, and the output layer is adopted to modify the connection weights layer by layer through the hidden layer to the output layer.

3) Process 1) and Process 2) are iterated repeatedly until the global error of the network reaches the threshold (Gupta & Banerjee, 2019).

Figure 1. The Algorithm Flowchart of BPNN



BPNN has advantages in stock price prediction. BPNN is currently the most widely used and most successful neural network. It has the following advantages when predicting stock prices (Akbar et al., 2019; Narayan, 2019).

1) BPNN has good nonlinear mapping ability. According to relevant data, there is a nonlinear relationship between stock price and its influencing factors. The three-layer BPNN has the ability of approximating any nonlinear continuous function with any accuracy to meet the needs of stock price prediction.
2) BPNN has good self-learning and self-adaptive abilities. BPNN can automatically get the relationship between input data and output data during training, and this relationship can be strengthened and memorized through the weights and thresholds of each layer of the network. Therefore, the relationship between stock price's influence factors and itself does not need to be characterized by a clear nonlinear function expression (Kim et al., 2019).
3) Its generalization ability donates the ability of a trained neural network to solve practical problems. In terms of stock price prediction, the trained data can predict the untrained data, or the past data can predict the future data. BPNN has strong fault tolerance. An error in a neuron in BPNN will not fundamentally affect the neural network's performance. Compared with the valuation and statistical regression prediction methods commonly used in the financial industry, the fault tolerance of a BPNN dramatically improves the accuracy and stability of stock price prediction (Agarwal et al., 2019; Wen et al., 2019).

The Principles of GA

GA is a meta-heuristic natural selection process, a kind of evolutionary algorithms. Generally, it uses biological heuristics, like mutation, crossover, as well as selection, for the optimization of better performance as well as the searching of problem solutions (Mishra et al., 2019; Jimenez-Rodriguez, 2019). GA encodes the feasible solution of the problem into chromosomes to form the initial population and generates new populations through operations such as selection, crossover, and mutation. While generating a new generation of populations, individuals with more excellent fitness are retained, and individuals with lower fitness are eliminated. Each new generation of the new population preserves the information in the previous generation and has better performance than the previous generation. Therefore, GA can be regarded as an initial process of a group composed of feasible solutions. GA consists of four essential elements: coding mechanism, fitness function, genetic operation, and control parameters (Feldman et al., 2019).

1) Coding. When GA is used to solve a problem, the objective function and variables of the problem are determined first; then, the variables are encoded. Standard encoding methods include binary encoding and actual number encoding. If the decoding equation that converts binary numbers to decimal is used, it can be expressed as the form shown in Eq. (1):

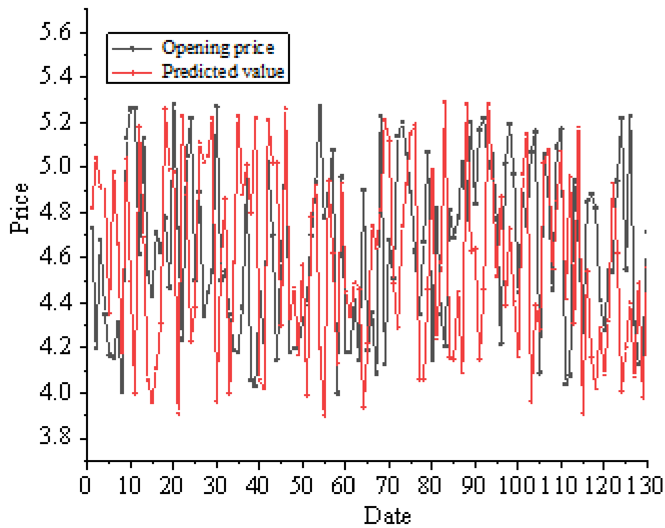
$$F(b_{i1}, b_{i2}, b_{il}) = R_i + \frac{T_i - R_i}{2^l - 1} \sum_{j=1}^{l} b_{ij} 2^{j-1} \tag{1}$$

where, $(b_{i1}, b_{i2}, \ldots, b_{il})$ refers to the i-th segment of each individual whose segment length is l, and T_i and R_i are the left and right endpoints of the domain of the i-th component.

2) The fitness function aims to measure the fitness of individuals in the population, which can accurately reflect the gap between the optimal fitness of a feasible solution and the chromosome. The objective function is set as the fitness function (Xiao et al., 2019; Xiang et al., 2021; Yi, 2021).
3) The task of genetic operation is to impose operations on individuals according to their fitness to complete the evolutionary process of survival of the fittest. Regarding optimization search,

genetic operations can optimize the solution of the problem generation by generation and gradually approach the optimal solution (Hamdi et al., 2019; Niu et al., 2021). The genetic operation contains three genetic operators: selection, crossover, and mutation. Selection and crossover complete the search function of GA. The mutation operator improves GA's ability to find the optimal solution. The basic process of GA is presented in Figure 2 below:

Figure 2. The Flowchart of GA

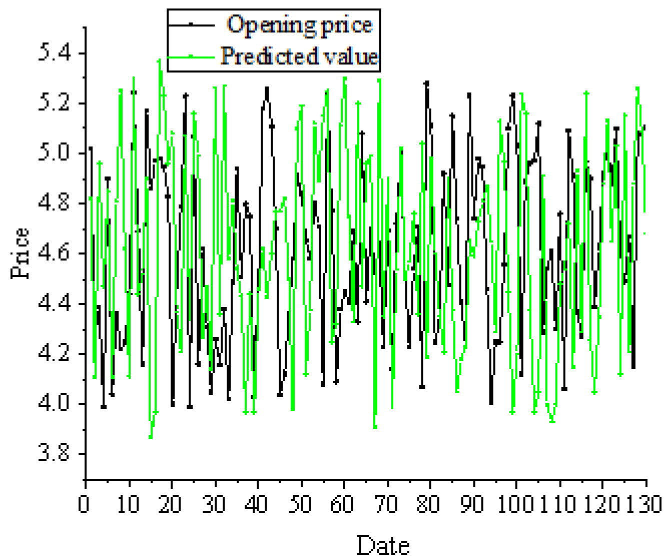


As shown in Figure 2, the GA process can be categorized as the following three steps:

1) The credible solution of the problem is encoded;
2) The fitness function is constructed, and the fitness values of individuals are calculated;
3) The genetic operations are performed on the selected individuals, and new individuals are generated by selection, crossover, and mutation operations based on the pre-selection probability, crossover probability, as well as mutation probability.

Regarding the common problems of BPNN in analyzing the market risks of the A+H shares, GA is adopted for the optimization of the initial weights and thresholds of BPNN to reduce the limitations of the BPNN, reduce the prediction error of BPNN, and improve its prediction accuracy ((Uddin et al., 2019; Wang et al., 2021; Liu et al., 2021). Figure 3 presents the basic workflow of the improved BPNN. The workflow can be divided into nine steps:

Figure 3. The Algorithm Flowchart of the Improved BPNN



1) An initial population is randomly generated.
2) BPNN in the initial state is established.
3) BPNN is trained, and the total error of BPNN is calculated.
4) The total error of BPNN is taken as the total fitness function, and the fitness value of each individual in the group is calculated.
5) The genetic operations are performed on the current population to generate a new population.
6) If the maximum evolutionary algebra is not reached, the current population is used to build a new BPNN, and steps 3), 4) and 5) are repeated.
7) It is essential to calculate the fitness values of all individuals in the current population. At this time, the optimal individual is the one with minor fitness. Then, the chromosomes of the individual with the smallest fitness are transformed into BPNN's initial weights as well as thresholds.
8) The remaining data are utilized for training a network with optimal initial weights and thresholds, and the output error of the network is calculated.
9) The algorithm stops until the termination condition is met.

PCA Method

PCA is a statistical analysis method, which uses the concept of dimensionality reduction to transform multiple single variables into few comprehensive variables (Misra&Chaurasia, 2020; Shahrestani & Rafei, 2020). It uses the orthogonal approach to transform the original component-related matrix into a new component-unrelated matrix so that fewer comprehensive variables are employed for replacing the original variables, thereby reducing the dimensionality (Wu & Hu, 2019). Assuming that X_1, X_2, …, X_P represent P independent variables, and X is linearly changed, the linear combination can be obtained as shown in Eq. (2):

$$\begin{cases} F_1 = a_{11}X_1 + a_{12}X_2 + \ldots \mathrm{a}_{1p}\, X_P \\ F_2 = a_{21}X_1 + a_{22}X_2 + \ldots a_{2p}X_P \\ \qquad \ldots \\ F_{\mathrm{m}} = a_{\mathrm{m}1}X_1 + a_{\mathrm{m}2}X_2 + \ldots a_{mp}X_P \end{cases} \tag{2}$$

If F satisfies the conditions shown in equations (3) and (4; it means that F_1, F_2, …, F_m are the principal components of the original variables.

$$Conv(F_iF_i)=0 \tag{3}$$

$$Var(F_i)=\mathrm{A}_i^{'}\sum A_i \tag{4}$$

where, Σ means the covariance of X.

The calculation of the PCA method can be divided into two steps:

1. The covariance matrix $\Sigma(s_{ij})p*p$ of the sample data is calculated, where s_{ij} can be written in the form shown in Eq. (5):

$$s_{ij} = \frac{1}{n}\sum_{k=1}^{n}(\mathrm{x}_{ki} - \overline{x}_i)(\mathrm{x}_{kj} \text{-} \overline{\mathrm{x}}) \tag{5}$$

The eigenvalue λ_i of the covariance matrix is found, and the first m more considerable eigenvalue λ_i corresponding to the eigenvector of the orthogonal unitized eigenvector a_i is the variance corresponding to the first m principal components. The unit eigenvector *ai* corresponding to λ_i is the coefficient of F_i concerning the original variable, and the i-th principal component F_i of the original variable has the form shown in Eq. (6):

$$F_i = a_i^{'}X \tag{6}$$

The principal component variance's contribution rate α_i shows how much information is extracted, which is calculated according to Eq. (7):

$$\alpha_i = \lambda_i \,/ \sum_{i=1}^{p}\lambda_i \tag{7}$$

2. The principal component is selected. The cumulative contribution rate can generally decide the selection of the principal component. The equation for calculating the cumulative contribution rate is shown in Eq. (8):

$$G(m) = \sum_{i=1}^{m} \lambda_i \Big/ \sum_{k=1}^{p} \lambda_k \tag{8}$$

In general, a cumulative contribution rate higher than 85% is considered sufficient to reflect the information of the original variable. The corresponding *m* is the first *m* principal component extracted. For specific problems, sometimes more than 70% can meet the requirements.

BPNN Model for Predicting Multi-Day Prices of A+H Shares

To ensure that the selection of data samples is representative, among seven banks listed in the A-share and H-share markets, the Bank of China (BC) and China Merchants Bank (CMB) with the issuance mode of H to A and the Industrial and Commercial Bank of China (ICBC) with the issuance mode of A+H are selected to evaluate the risks of A shares and H shares and conduct comparative analysis. The five characteristic values of opening price x_1, highest price x_2, lowest price x_3, closing price x_4, and turnover rate x_5 for each trading day are selected as quantitative indicators of the sample. The output of the network is determined by the information that needs to be predicted. The securities markets of mainland China and Hong Kong implement the T+n trading system, which means that investors can sell it at least on the next trading day after buying a stock. Due to the correlation between stock prices before and after trading days, the multi-day and multi-indicator data will predict the future stock price with continuous dates. Two days are taken as an example to build a model. The opening price of a stock x_1, the highest price x_2, the lowest price x_3, the closing price x_4, and the turnover rate x_5 for three consecutive days are used as a set of input variables. The opening price x_1, highest price x_2, lowest price x_3, and closing price x_4 for the next two days are used as a set of output variables. The prediction of opening price can be expressed as Eqs (9), (10) and (11):

$$\mathrm{x}_{11},\mathrm{x}_{21},\mathrm{x}_{31},\mathrm{x}_{41},x_{51},x_{12},x_{22},x_{32},x_{42},x_{52},\mathrm{x}_{13},x_{23},x_{33},x_{43},x_{53} \rightarrow x_{14},x_{15} \tag{9}$$

$$x_{11},x_{22},x_{32},x_{42},x_{52},x_{13},x_{23},x_{33},x_{43},x_{53},x_{14},x_{24},x_{34},x_{44},x_{54} \rightarrow x_{15},x_{16} \tag{10}$$

$$x_{1,1211},x_{2,1211},x_{3,1211},\mathrm{x}_{4,1211},x_{5,1211},x_{1,1212},x_{2,1212},x_{3,1212},x_{4,1212},x_{5,1212},x_{1,1213},x_{2,1213},x_{4,1213},x_{5,1213\rightarrow}x_{1,1214},x_{1,1215} \tag{11}$$

where, x_{ij} refers to *j*-th day data of the *i*-th indicator.

Data from January 2, 2019 to July 20, 2019 were selected as the sample data of BPNN. Data from January 2, 2019 to July 1, 2019 were training samples, and data from July 2, 2019 to July 20, 2019 were prediction samples. The map min-max function in Matlab was used to eliminate the influences of variables' different units and significant differences in values between variables, as shown in Eq. (12). The input variables are normalized and mapped to the interval [-1,1].

$$[\mathrm{Y}, \mathrm{PS}] = \mathrm{mapminmax}(\mathrm{X}) \tag{12}$$

where, *Y* refers to the standardized data, as shown in Eq. (13), and *PS* indicates the structure that records the standardized mapping.

$$Y = 2 * (\mathrm{X} - \mathrm{Xmin}) / (\mathrm{X\,max\text{-}\,Xmin}) - 1 \tag{13}$$

Since the input variables of the constructed network are stock-related information and their form is simple, the number of neurons in the input layer as well as output layer of the constructed network is set to 15 and 2, respectively, while that in the hidden layer can be determined by Eq. (14):

$$\mathrm{N}_2 = \sqrt{N_1 + \mathrm{N}_2} + \alpha \tag{14}$$

where, N_1, N_2, and N_3 are the neurons in the hidden layer, input layer, as well as output layer, respectively; α is an arbitrary constant.

Experiments can reveal that in the constructed neural network model, the convergence speed of the network is faster, and the error is minor when the number of neurons in the input layer, hidden layer as well as output layer is 15, 16, and 2, respectively. In terms of GA parameter selection, since the purpose of establishing the model is to minimize the error between the actual output and the expected output, the mean square error (MSE) between the actual output and the expected output is used as the fitness evaluation value, and the equation is shown in Eq. (15):

$$F = mse(\mathrm{Y} - O) = \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{y}_i - o_i)^2 \tag{15}$$

where, y_i and o_i are the expected output value as well as the actual output value on the i-th day, respectively, and n is the number of samples taken during the daily test. The fitness ratio method is applied to the probability calculation. The selection probability of individual i is shown in Eq. (16):

$$sp_i = \frac{1 / \mathrm{F}_i}{\sum_{i=1}^{N} (1 / \mathrm{F}_i)} \tag{16}$$

where, F_i refers to the fitness of the i-th individual, and N indicates the number of populations.

For actual number coding, the length of the code obtained by the coding equation is shown in Eq. (17):

$$\mathrm{l} = \mathrm{l}_{\mathrm{in}} * \mathrm{l}_{\mathrm{hidden}} + \mathrm{l}_{\mathrm{hidden}} + \mathrm{l}_{hidden} * + l_{out} \tag{17}$$

where, l_{in}, l_{hidden}, and l_{out} are the neuro quantity in the input layer, hidden layer, as well as output layer.

Relevant parameter settings are as follows: the evolutionary algebra of the population: 60, the scale: 40, the crossover probability: 0.95, the opening price mutation probability: 0.03900009, the highest price mutation probability: 0.0399811, the lowest price mutation probability: 0.03099081, and the closing price mutation probability: 0.034555052. The training data are reused to train the model. Afterward, the test data are utilized to test the prediction effect of the model. The prediction error is shown in Eq. (18), and the MSE in Eq. (19) is employed to judge the performance of the model.

$$precision = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - o_i}{o_i}\right| \tag{18}$$

$$\mathrm{m}se = \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{y}_i - o_i\right)^2 \tag{19}$$

where, y_i and o_i represent the expected output as well as the actual output value on the i-th day, respectively. n refers to the quantity of test samples per day.

BP-GA-PCA Neural Network Model for Predicting Multi-Day Prices of A+H Shares

In this section, the historical transaction data of the same period as in Section 2.3 were used to train the model, but the indicators used were different. The original indicators used in this section are opening price, closing price, lowest price, highest price, transaction volume, and transaction amount. On this basis, five technical indicators are added, including the moving average indicator, the stochastic indicators (K, D, and J), and the relative strength indicator (RSI), to construct a PCA-based GA-BPNN, written as BP-GA-PCA. In this section, the variables of opening price x_1, highest price x_2, lowest price x_3, closing price x_4, transaction volume x_5, transaction amount x_6, five-day closing average x_7, K x_8, D x_9, J x_{10}, RSI x_{11} are used as the input variables of the network. The Z-score method is employed to standardize the above 11 technical indicators to eliminate the influence of dimensions. According to previous studies, these 11 (Husain et al., 2019; Yu et al., 2021) variables are linearly related, and the information carried by the linearly related variables overlaps. Thus, inputting these variables directly into the neural network will cause the network to be very complicated; consequently, the network running speed is slow, and the generalization ability is low. In this regard, the PCA method is used first to remove redundant information and reduce input variables. Second, the original data information is filled in, the correlation moment is calculated, and the scree plot is obtained. It is found that only three principal components of the 11 variables can replace the original variables by observing the singular value decomposition of the scree plot correlation matrix. After the above data preprocessing, the principal components and 12 data for four consecutive days are input into the neural network as the variables of the neural network. In the output layer of the network, the opening price, the highest price, the lowest price, as well as the closing price for two consecutive days are still used as output variables.

Empirical Analysis Parameters and Algorithm Performance Evaluation Parameters

A package in the R software can download the daily closing prices of the Hong Kong Hang Seng Index and the Shanghai Composite Index from January 2, 2019 to July 20, 2019. These data are the research samples. The remuneration utilization rate R_t is adopted for data analysis, which can be expressed as:

$$R_t = \left[1n\left(P_{t-1}\right)\right]*100 \tag{20}$$

where, P_t refers to the closing price on the $t -_{th}$ day.

The last 300 returns of the sample data (Hong Kong Hang Seng Index and the Shanghai Composite Index) are retained as the prediction evaluation of the out-of-sample Value at Risk (VaR) model. Intra-sample data use the rolling window to predict VaR one step forward. MSE as well as Mean

Absolute Error (MAE) are adopted for the evaluation of the model's effectiveness. These two metrics are calculated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(f_1 - y_i)^2 \tag{21}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(f_i - y_i) \tag{22}$$

where, f_i represents the prediction result of the model, and y_i represents the actual value.

RESULTS AND DISCUSSION

The Comparison of Share Opening Price's Actual Value and the Value Predicted by BP-GA

Figures 4, 5, and 6 show the parameters and real values of BC, CMB, and ICBC predicted by AI.

As shown in Figure 4, from July 20, 2019 to July 21, 2019, the average prediction errors of the opening price, highest price, lowest price, and closing price of BC's A share are 2.36%, -2.62%, 2.94%, and 3.39%, respectively. The average prediction errors of the opening price, highest price, lowest price, and closing price of BC's H share in these two days are -3.86%, 5.72%, -3.64%, and 7.59%, respectively. It is concluded that the prediction error of H share is greater than that of A share.

According to Figure 5, from July 20, 2019 to July 21, 2019, the average prediction errors of the opening price, highest price, lowest price, and closing price of CMB's A share are -1.36%, 2.22%, -2.64%, and 3.39%, respectively. The average prediction errors of the opening price, highest price, lowest price, and closing price of CMB's H share in these two days are 3.36%, -5.12%, 3.64%, and -7.29%, respectively.

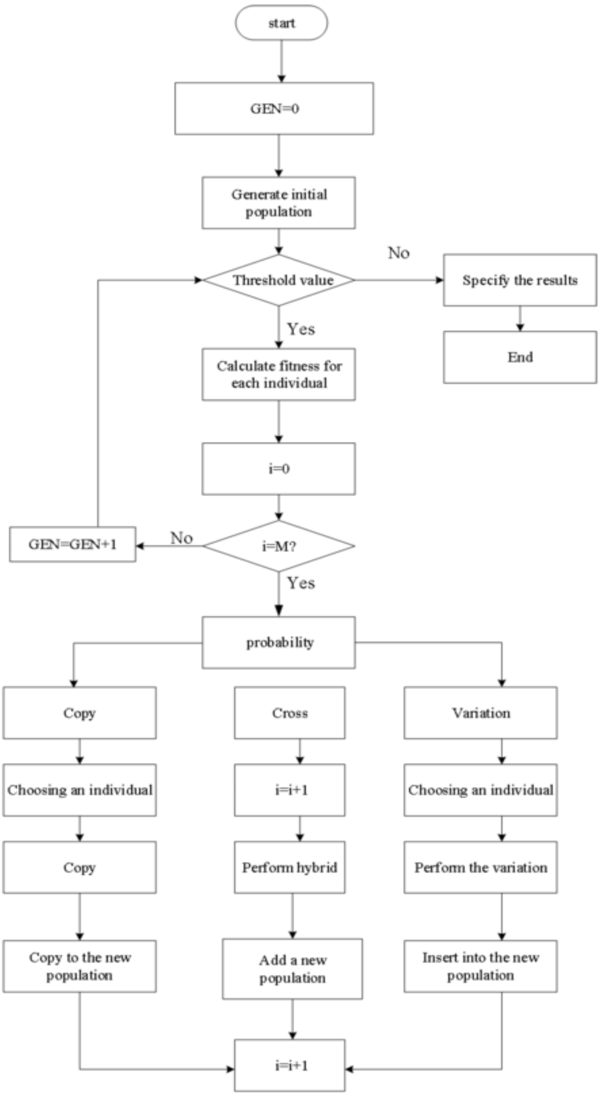
As shown in Fig. 6, from July 20, 2019 to July 21, 2019, the average prediction errors of the opening price, highest price, lowest price, and closing price of ICBC's A share are 1.16%, 1.22%, -2.34%, and 3.39%, respectively. In these two days, the average prediction errors of the opening price, highest price, lowest price, and closing price of ICBC's H share are 3.46%, -5.15%, 3.74%, and -7.19%, respectively. The prediction error of H share is smaller than that of A share.

Scree Plot Analysis

Figure 7 shows the scree plot of 11 variables in BC, CMB, and ICBC.

According to Figure 7, whether it is A share or H share, in its 11 variables, the eigenvalue of the third principal component is already minimal, and its downward trend has begun to slow down. Therefore, the first three principal components can substitute the original variables and input the corresponding neural network. Furthermore, calculation reveals that the cumulative contribution rate of the first two principal components of BC is only 75.12%, the original information carried cannot meet the requirements. The cumulative contribution rate of the first three principal components of BC reaches 90.25%, reflecting enough original information. The cumulative contribution rate of the first two components of ICBC is only 72.12%, and the original information carried cannot meet the requirements. The cumulative contribution rate of the first three principal components of ICBC reaches 80.25%, which can reflect enough original information. The cumulative contribution rate of the first two principal components of CMB is only 82.12%, and the original information carried cannot meet

Figure 4a. Comparison of Share Opening Price's Actual Value and Predicted Value of BC [(a) shows the H share



the requirements. The cumulative contribution rate of the first three principal components of CMB reaches 92.25%, reflecting enough original information.

The Comparison of Share Opening Price's Actual Value and the Value Predicted by BP-GA-PCA

Figures 8, 9 and 10 show the comparison results between the predicted value and the real value of the parameters of BC, CMB, and ICBC predicted by AI.

As shown in Fig. 8, from July 20, 2019, to July 21, 2019, the average prediction errors of the opening price, highest price, lowest price, and closing price of BC's A share are 1.36%, -1.62%, 1.94%, and 3.39%, respectively. In these two days, the average prediction errors of the opening price, highest price, lowest price, and closing price of BC's H share are 2.86%, 4.72%, -1.64%, and -6.59%, respectively. The experimental results reveal that the improved algorithm produces a slight decrease

Figure 4b. Comparison of Share Opening Price's Actual Value and Predicted Value of BC (b) shows the A share]

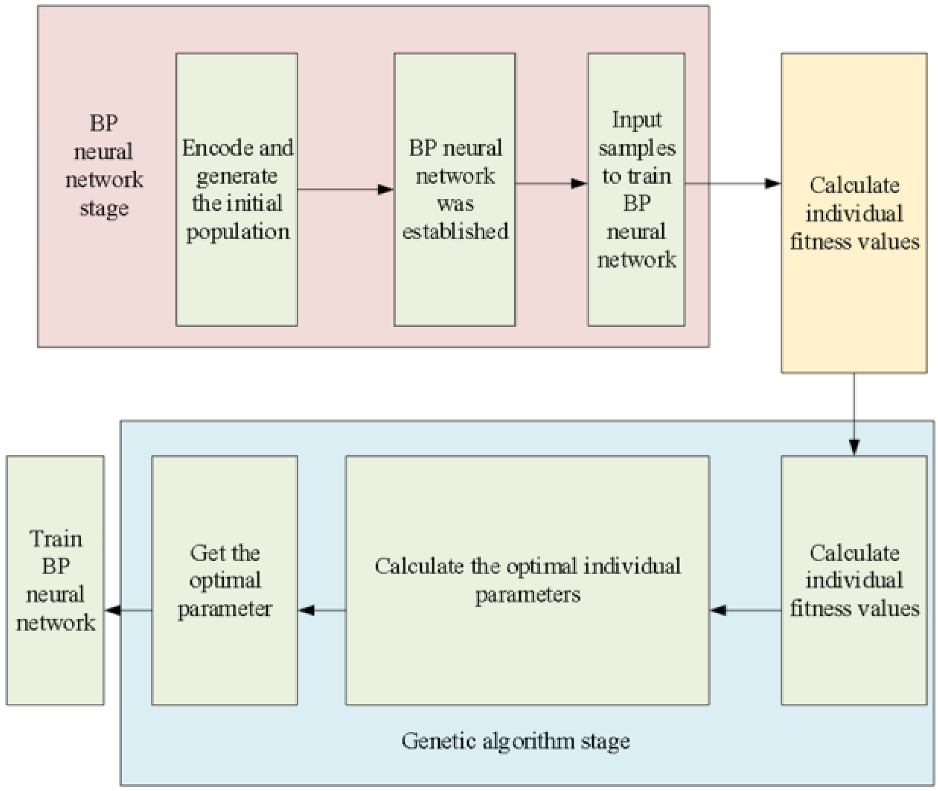


Figure 5a. Comparison of Share Opening Price's Actual Value and Predicted Value of CMB [(a) shows the H share

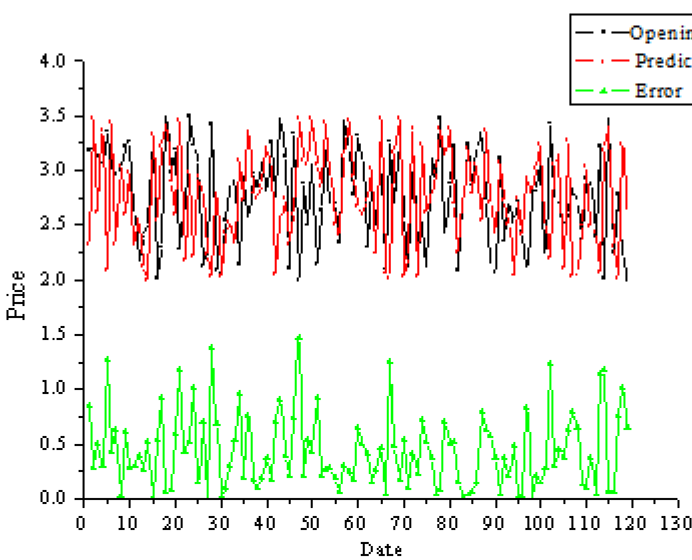


Figure 5b. Comparison of Share Opening Price's Actual Value and Predicted Value of CMB (b) shows the A share)]

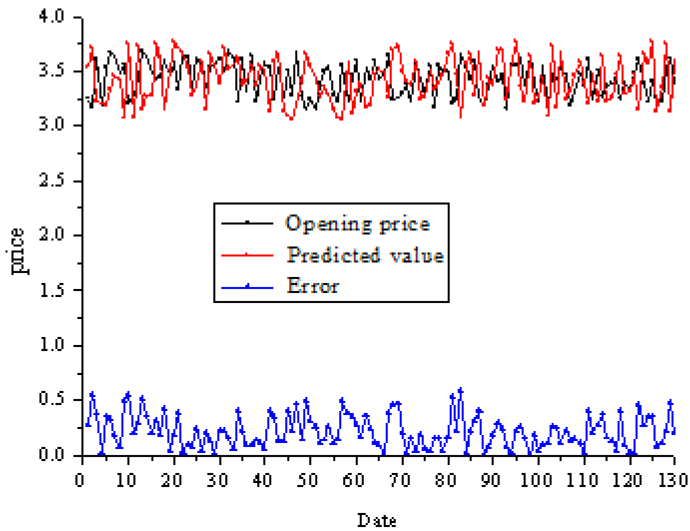


Figure 6a. Comparison of Share Opening Price's Actual Value and Predicted Value of ICBC [(a) shows the H share

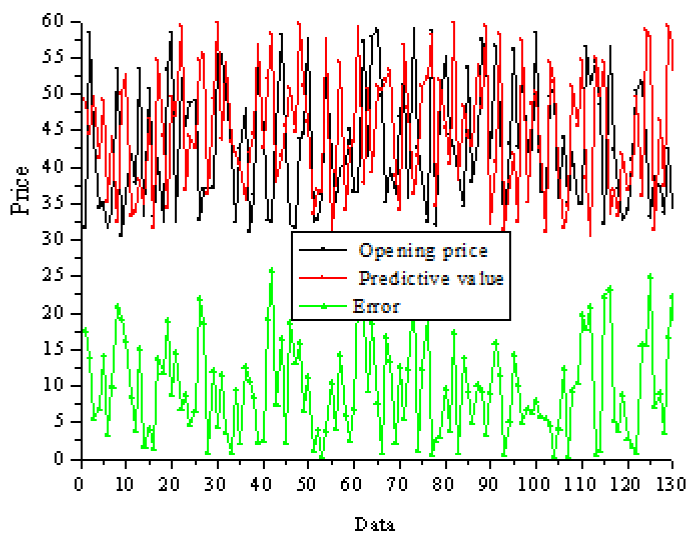


Figure 6b. Comparison of Share Opening Price's Actual Value and Predicted Value of ICBC (b) shows the A share]

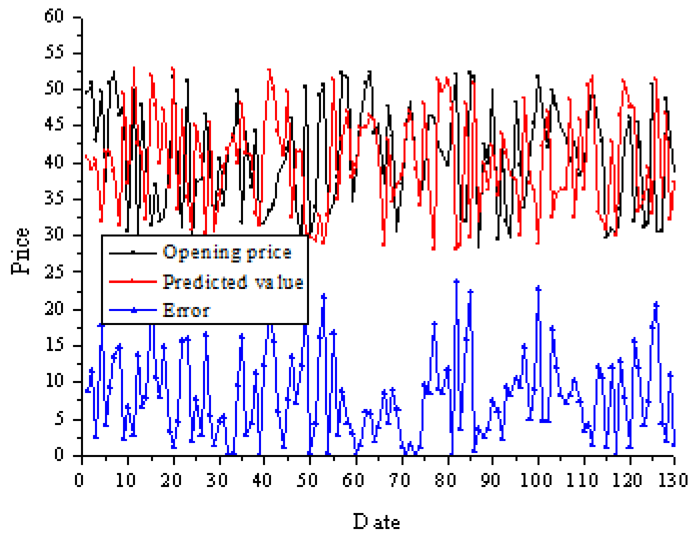


Figure 7. Scree Plot of Bank Variables

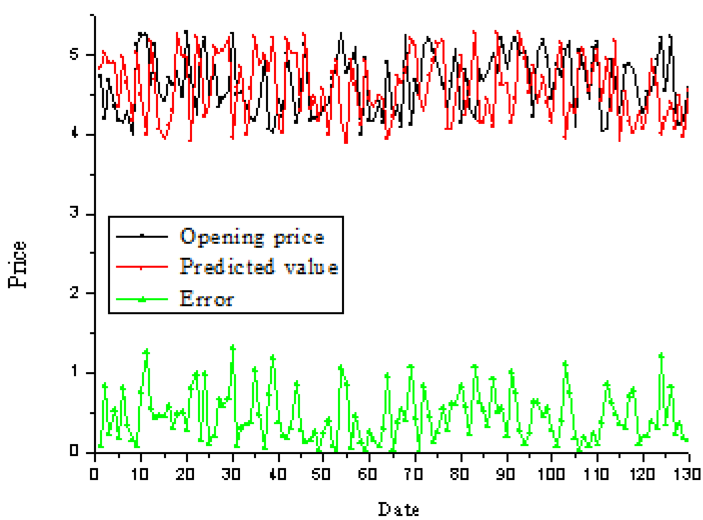


Figure 8a. Comparison of share Opening Price's Actual Value and Predicted Value of BC [(a) shows the H share

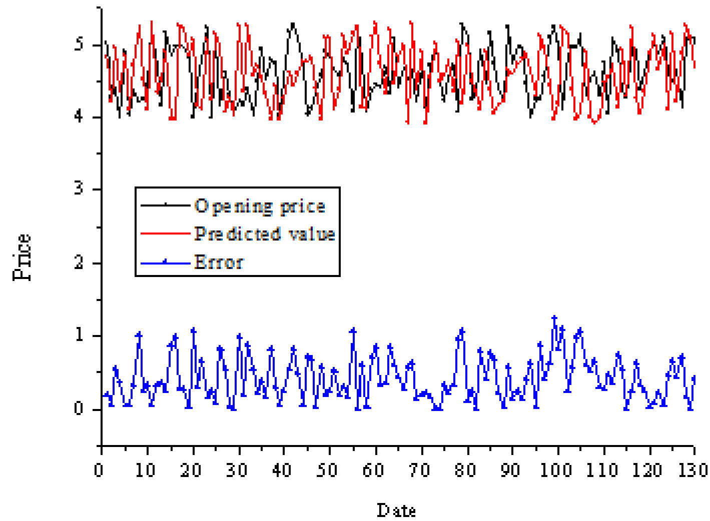
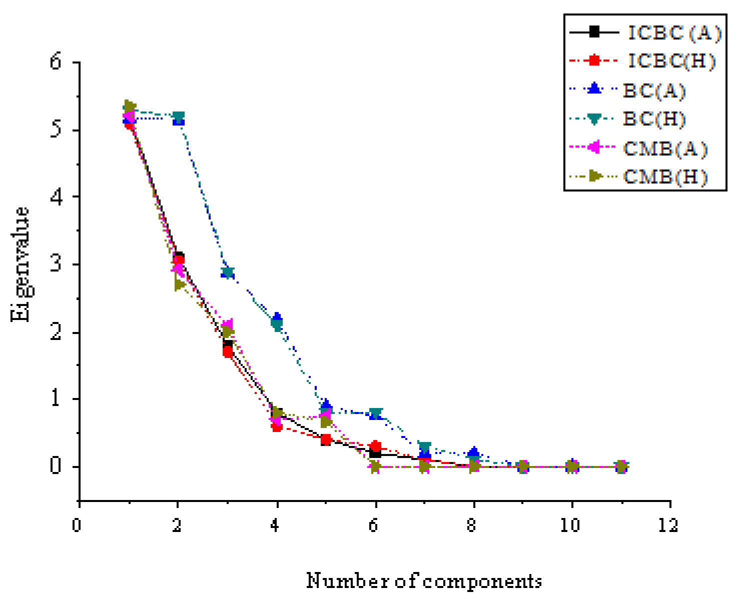


Figure 8b. Comparison of share Opening Price's Actual Value and Predicted Value of BC (b) shows the A share]



in the prediction error of the opening price, the highest price, and the lowest price compared with the previous algorithm.

Figure 9a. Comparison of Share Opening Price's Actual Value and Predicted Value of CMB [(a) shows the H share

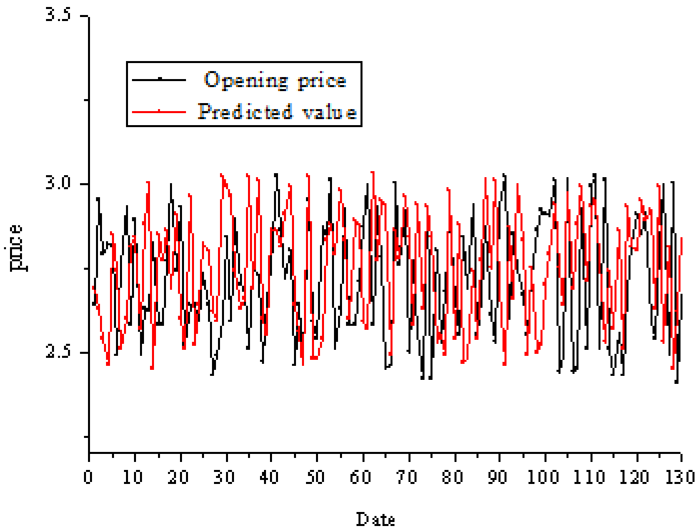
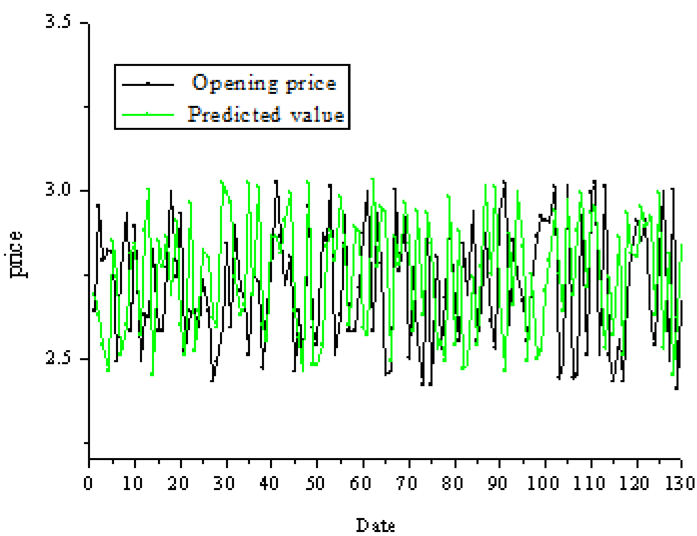
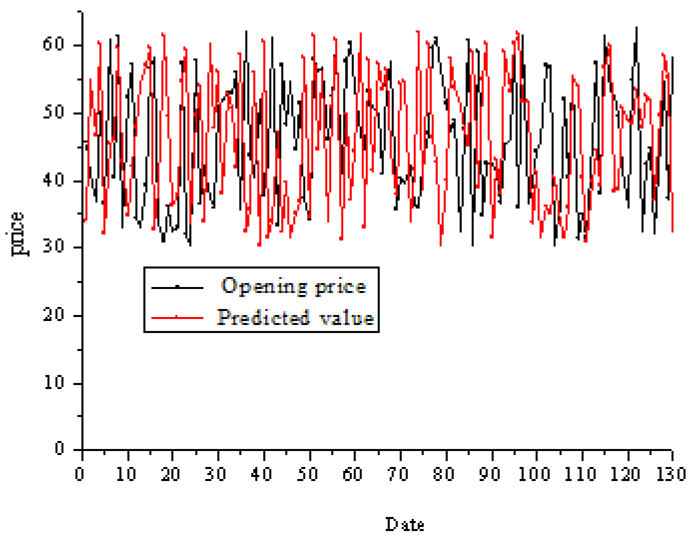


Figure 9b. Comparison of Share Opening Price's Actual Value and Predicted Value of CMB (b) shows the A share]



According to Fig. 9, from July 20, 2019, to July 21, 2019, the average prediction errors of the opening price, highest price, lowest price, and closing price of CMB's A share are1.16%, -1.12%, 1.74%, and -3.29%, respectively. In these two days, the average prediction errors of the opening price, highest price, lowest price, and closing price of CMB's H share are 2.76%, -4.22%, 1.94%, and - 6.19%, respectively. The prediction error of H share is smaller than that of A share.

Figure 10a. Comparison of Share Opening Price's Actual Value and Predicted Value of ICBC [(a) shows the H share



From July 20, 2019, to July 21, 2019, prediction errors on ICBC A share's opening price, highest price, lowest price, and closing price are -2.16%, 2.12%, -2.74%, and 1.29%, respectively; prediction errors on ICBC H share's opening price, highest price, lowest price, and closing price are 1.76%, -3.22%, 2.94%, and -5.19%. The prediction error of the H share is smaller than that of the A share.

Empirical Analysis

Table 1 presents the basic descriptive statistics of the Hong Kong Hang Seng Index and Shanghai Composite Index.

Table 1 suggests that both Hong Kong Hang Seng Index and Shanghai Composite Index have positive average returns. Regarding the maximum and minimum values, the return range of the Hong Kong Hang Seng Index is apparently more considerable than that of the Shanghai Composite Index, indicating that extreme losses or substantial gains are prone to occur in the Hong Kong stock market without daily limit restrictions. The return rate skewness of both samples is negative, indicating that the main body of the Hong Kong Hang Seng Index and Shanghai Composite Index in the sampling interval is concentrated on the left side of the distribution function, with a risk of loss. Moreover, their kurtosis presents a "high narrow peak" shape. Results of the JB test suggest that the returns of the two samples are non-normally distributed. Excessively high peaks can carry the risk of a "fat tail". Nowadays, risk management focuses on the fat tail phenomenon on the left side of the distribution function, indicating a higher probability of loss. Regarding the auto-correlation test results of Q(5) and Q(5)2, only the Hong Kong Hang Seng Index return rate does not have five auto-correlation

Figure 10b. Comparison of Share Opening Price's Actual Value and Predicted Value of ICBC b) shows the A share]

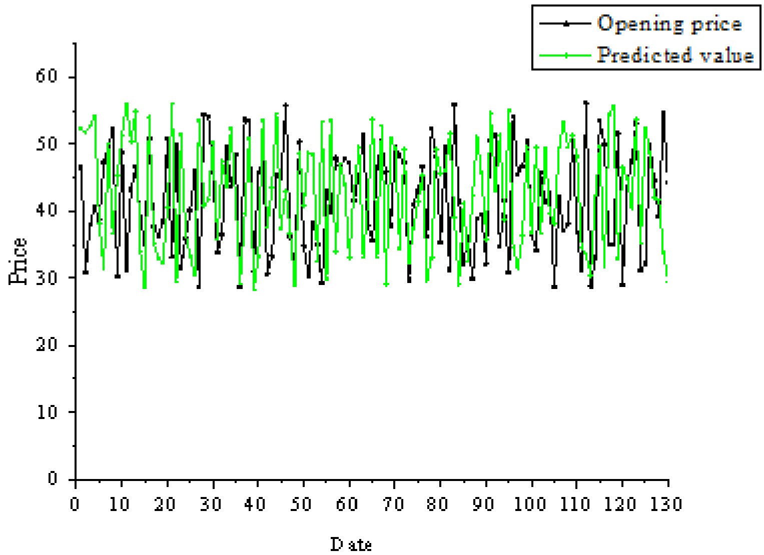


Table 1. Basic descriptive statistics of Hong Kong Hang Seng Index and Shanghai Composite Index

Rate of return	Mean	Minimum	Maximum	Skewness	Kurtosis	P-value of Jarque-Bera (JB) statistic test	Q(5) P-value	$Q(5)^2$ P P-value
Hong Kong Hang Seng Index	0.0088	-13.5	13.4	-0.1	8.31	0	0.41	0
Shanghai Composite Index	0.0186	-9.2	9.4	-0.38	4.61	0	0.0009	0

lags. In contrast, the rest all have auto-correlations, which also verifies that the yield of the securities market is generally volatility clustering.

Comparison of Prediction Results of AI Algorithms

The prediction errors are demonstrated in Table 2.

According to Table 2, both MSE and MAE of BP, PCA-BP, and the proposed BP-GA-PCA gradually decrease. Compared with BP and PCA-BP, the proposed BP-GA-PCA can provide higher prediction accuracy.

Table 2. MSE and MAE results

Algorithms	BP	PCA-BP	BP-GA-PCA
MSE	0.002	0.00211	0.00066
MAE	0.043	0.034	0.021

CONCLUSION

In this research, the representative stock opening price, the highest price, the lowest price, as well as the closing price forecast were taken as the starting point in the market risk analysis of A+H shares. Then, BPNN was optimized by introducing GA in AI, and the parameters of the neural network were fine tuned. Afterwards, the PCA algorithm in AI method was introduced into the problem of reducing the dimension of the input of the network. It is found that BPNN using the AI method has a high accuracy in the market risk analysis of A+H shares; GA can improve the performance of BPNN; PCA algorithm can effectively screen variables input into the network, and then provide a feasible algorithm for the market risk analysis and prediction of A+H shares. However, due to the limitation of the existing conditions, the experiment has some shortcomings. The study only predicts the stocks in three banks, but lacks universality to some extent. In the problem of GA parameter debugging, the manual debugging is adopted, leading to the wastes of resources and time. These problems all need to be improved in the next step.

REFERENCES

Agarwal, S., Kumar, S., & Goel, U. (2019). Stock market response to information diffusion through internet sources: A literature review. *International Journal of Information Management*, *45*, 118–131. doi:10.1016/j.ijinfomgt.2018.11.002

Akbar, M., Iqbal, F., & Noor, F. (2019). Bayesian analysis of dynamic linkages among gold price, stock prices, exchange rate and interest rate in Pakistan. *Resources Policy*, *62*, 154–164. doi:10.1016/j.resourpol.2019.03.003

Ali, S., Bouri, E., Czudaj, R. L., & Shahzad, S. J. H. (2020). Revisiting the valuable roles of commodities for international stock markets. *Resources Policy*, *66*, 101603. doi:10.1016/j.resourpol.2020.101603

Batten, J. A., Kinateder, H., Szilagyi, P. G., & Wagner, N. F. (2019). Time-varying energy and stock market integration in Asia. *Energy Economics*, *80*, 777–792. doi:10.1016/j.eneco.2019.01.008

Dimpfl, T., & Kleiman, V. (2019). Investor pessimism and the German stock market: Exploring Google search queries. *German Economic Review*, *20*(1), 1–28. doi:10.1111/geer.12137

Feldman, E. R., Amit, R., & Villalonga, B. (2019). Family firms and the stock market performance of acquisitions and divestitures. *Strategic Management Journal*, *40*(5), 757–780. doi:10.1002/smj.2999

Gupta, K., & Banerjee, R. (2019). Does OPEC news sentiment influence stock returns of energy firms in the United States? *Energy Economics*, *77*, 34–45. doi:10.1016/j.eneco.2018.03.017

Hamdi, B., Aloui, M., Alqahtani, F., & Tiwari, A. (2019). Relationship between the oil price volatility and sectoral stock markets in oil-exporting economies: Evidence from wavelet nonlinear denoised based quantile and Granger-causality analysis. *Energy Economics*, *80*, 536–552. doi:10.1016/j.eneco.2018.12.021

He, F., Liu, Z., & Chen, S. (2018). Industries return and volatility spillover in Chinese stock market: An early warning signal of systemic risk. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 9046–9056. doi:10.1109/ACCESS.2018.2888522

Husain, S., Tiwari, A. K., Sohag, K., & Shahbaz, M. (2019). Connectedness among crude oil prices, stock index and metal prices: An application of network approach in the USA. *Resources Policy*, *62*, 57–65. doi:10.1016/j.resourpol.2019.03.011

Iyke, B. N., Tran, V. T., & Narayan, P. K. (2021). Can energy security predict energy stock returns? *Energy Economics*, *94*, 105052. doi:10.1016/j.eneco.2020.105052

Jiang, L. C., & Subramanian, P. (2019). Forecasting of Stock Price Using Autoregressive Integrated Moving Average Model. *Journal of Computational and Theoretical Nanoscience*, *16*(8), 3519–3524. doi:10.1166/jctn.2019.8317

Jiménez-Rodríguez, R. (2019). What happens to the relationship between EU allowances prices and stock market indices in Europe? *Energy Economics*, *81*, 13–24. doi:10.1016/j.eneco.2019.03.002

Kim, J. H., Rahman, M. L., & Shamsuddin, A. (2019). Can energy prices predict stock returns? An extreme bounds analysis. *Energy Economics*, *81*, 822–834. doi:10.1016/j.eneco.2019.05.029

Li, A. W., & Bastos, G. S. (2020). Stock market forecasting using deep learning and technical analysis: A systematic review. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 185232–185242. doi:10.1109/ACCESS.2020.3030226

Liu, J., & Kemp, A. (2019). Forecasting the sign of us oil and gas industry stock index excess returns employing macroeconomic variables. *Energy Economics*, *81*, 672–686. doi:10.1016/j.eneco.2019.04.023

Liu, Z., Lang, L., Li, L., Zhao, Y., & Shi, L. (2021). Evolutionary game analysis on the recycling strategy of household medical device enterprises under government dynamic rewards and punishments. *Mathematical Biosciences and Engineering*, *18*(5), 6434–6451. doi:10.3934/mbe.2021320 PMID:34517540

Loriot, B., Hutson, E., & Au Yong, H. H. (2020). Equity-linked executive compensation, hedging and foreign exchange exposure: Australian evidence. *Australian Journal of Management*, *45*(1), 72–93. doi:10.1177/0312896219830158

Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Selim, M. M., & Muhammad, K. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management*, *50*, 432–451. doi:10.1016/j.ijinfomgt.2019.07.011

Mishra, S., Sharif, A., Khuntia, S., Meo, M. S., & Khan, S. A. R. (2019). Does oil prices impede Islamic stock indices? Fresh insights from wavelet-based quantile-on-quantile approach. *Resources Policy*, *62*, 292–304. doi:10.1016/j.resourpol.2019.04.005

Misra, P., & Chaurasia, S. (2020). Data-driven trend forecasting in stock market using machine learning techniques. *Journal of Information Technology Research*, *13*(1), 130–149. doi:10.4018/JITR.2020010109

Narayan, P. K. (2019). Can stale oil price news predict stock returns? *Energy Economics*, *83*, 430–444. doi:10.1016/j.eneco.2019.07.022

Niu, Z., Zhang, B., Li, D., Ji, D., Liu, Y., Feng, Y., Zhou, T., Zhang, Y., & Fan, Y. (2021). A mechanical reliability study of 3-dB waveguide hybrid couplers in submillimeter and terahertz bands. *Frontiers of Information Technology & Electronic Engineering*, *22*(8), 1104–1113. doi:10.1631/FITEE.2000229

Sarwar, S., Tiwari, A. K., & Tingqiu, C. (2020). Analyzing volatility spillovers between oil market and Asian stock markets. *Resources Policy*, *66*, 101608. doi:10.1016/j.resourpol.2020.101608

Shahrestani, P., & Rafei, M. (2020). The impact of oil price shocks on Tehran Stock Exchange returns: Application of the Markov switching vector autoregressive models. *Resources Policy*, *65*, 101579. doi:10.1016/j.resourpol.2020.101579

Suzuki, M., Sakaji, H., Izumi, K., Matsushima, H., & Ishikawa, Y. (2020). Forecasting Net Income Estimate and Stock Price Using Text Mining from Economic Reports. *Information (Basel)*, *11*(6), 292. doi:10.3390/info11060292

Uddin, G. S., Rahman, M. L., Hedström, A., & Ahmed, A. (2019). Cross-quantilogram-based correlation and dependence between renewable energy stock and other asset classes. *Energy Economics*, *80*, 743–759. doi:10.1016/j.eneco.2019.02.014

Wang, F., Lu, Y., Li, J., & Ni, J. (2021). Evaluating environmentally sustainable development based on the psr framework and variable weigh analytic hierarchy process. *International Journal of Environmental Research and Public Health*, *18*(6), 2836. doi:10.3390/ijerph18062836 PMID:33802188

Wen, D., Wang, G. J., Ma, C., & Wang, Y. (2019). Risk spillovers between oil and stock markets: A VAR for VaR analysis. *Energy Economics*, *80*, 524–535. doi:10.1016/j.eneco.2019.02.005

Wu, C. M., & Hu, J. L. (2019). Can CSR reduce stock price crash risk? Evidence from China's energy industry. *Energy Policy*, *128*, 505–518. doi:10.1016/j.enpol.2019.01.026

Wu, Y., & Choudhry, T. (2018). Information Uncertainty and Momentum Phenomenon Amidst Market Swings: Evidence From the Chinese Class A Share Market. *Asia-Pacific Financial Markets*, *25*(2), 111–136. doi:10.1007/s10690-018-9241-x

Xiang, D., Zhang, Y., & Worthington, A. C. (2021). Determinants of the use of fintech finance among chinese small and medium-sized enterprises. *IEEE Transactions on Engineering Management*, *68*(6), 1590–1604. doi:10.1109/TEM.2020.2989136

Xiao, J., Hu, C., Ouyang, G., & Wen, F. (2019). Impacts of oil implied volatility shocks on stock implied volatility in China: Empirical evidence from a quantile regression approach. *Energy Economics*, *80*, 297–309. doi:10.1016/j.eneco.2019.01.016

Yi, H. (2021). Secure social internet of things based on post-quantum blockchain. *IEEE Transactions on Network Science and Engineering*, 1. Advance online publication. doi:10.1109/TNSE.2021.3095192

Yu, H., Zhao, Y., Liu, Z., Liu, W., Zhang, S., Wang, F., & Shi, L. (2021). Research on the financing income of supply chains based on an E-commerce platform. *Technological Forecasting and Social Change*, *169*, 120820. doi:10.1016/j.techfore.2021.120820

Zhang, Y. J., & Wang, J. L. (2019). Do high-frequency stock market data help forecast crude oil prices? Evidence from the MIDAS models. *Energy Economics*, *78*, 192–201. doi:10.1016/j.eneco.2018.11.015