Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 16, No. 3, 2024

CNN-BiLSTM-Attention Based Stock Price Prediction and Quantitative Investment Strategy

Chunzhong Li *, **Weiqi Hua**

College of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China
* **Corresponding author**: Chunzhong Li (Email: czhongli@163.com)

Abstract: With the rapid development of technology and the increasing maturity of financial markets, stock price prediction has become a hot topic and an important trend in the financial field. However, the stock market has complexity and volatility, in order to reduce the investment risk and ensure the maximization of benefits, selecting the optimal stock to predict the stock price and formulating the quantitative trading strategy are extremely important issues for financial academics and investors. For the research of stock price prediction and quantitative trading, firstly, the quantitative stock selection model combining XGBoost and multi-factor stock selection model is constructed, and four stocks are screened out, and then the CNN-BiLSTM-Attention model is proposed to predict the stock price trend of the selected stocks, and it is found from the prediction results of the four stocks that the prediction accuracies all reach more than 95%, which is higher than that of the single model prediction. accuracy and passed the validity test of the fitted model. Secondly, based on the quantitative investment portfolio given in the above analysis and the quantitative stock trading strategy of MACD, the prediction results are verified, and the total return of most stocks based on the strategy is higher than 40% at the highest, and the loss is controlled within 10%, which indicates that the trading strategy in this paper is effective and feasible. The study concludes that greater returns can be obtained by calculating the total returns of stock portfolios based on different portfolio approaches.

Keywords: Stock price prediction, Quantitative trading, XGBoost multifactor stock picking, Deep learning.

1. Introduction

Stock market trends reveal the direction of national economic development, is an important indicator of economic development potential, is an essential and important part of the socialist market economy, more and more scholars in the financial sector began to study it. With the development of the economy, people's income level increases, there are more deposits for securities investment, the stock market provides investors with a huge profit space, attracting more and more people's attention, the stock price trend is one of the most concerned about the problem of investors. Therefore, how to choose the optimal stock and predict the stock return is extremely important to the financial academia and individual investors. Therefore, choosing the appropriate method to extract effective information from the huge amount of financial data to forecast the stock market has become an urgent problem in China.

In recent years, research in the field of stock price trend prediction has continued to make new progress, and a variety of methods have emerged, which can be categorized into three main categories: time series prediction methods, traditional machine learning and deep learning methods. The evolution of this field can be traced back to the initial stage of constructing models using traditional linear regression, and with the boom in artificial intelligence, new forecasting algorithms and implementation tools have emerged. Box and Jenkins created a stochastic time series analysis model Autoregressive Sliding Average Model (ARMA) in the 1970s [1]. Ariyo (2014) [2] utilized the stock data from the New York Stock Exchange and the Nigerian Stock Exchange to build an ARIMA stock price prediction model, and experiments have shown that the ARIMA model has strong short-term prediction ability. Traditional machine learning methods use different feature extraction techniques to select the most useful feature information to improve stock forecasting performance. In order to overcome the situation that traditional time series models are only limited to linear representation, Wen (2014) [3] used singular spectrum analysis (SSA) to decompose stock prices into trends, market fluctuations and noise with different economic characteristics in different time periods and input these features into support vector machine model for stock price prediction, SSA-SVM model has good prediction results. Salim (2018) [4] used a model combining Singular Spectrum Analysis (SSA), Support Vector Regression (SVR) with Particle Swarm Optimization (PSO), SSA decomposed the stock price time series into a few independent components to be used as predictors, SVR was applied to the forecasting task and PSO was used to optimize the parameters of the SVR, SSA-PSO-SVR showed a significant effect in noisy financial time series analysis and forecasting shows obvious effects. And deep learning models can better model nonlinear relationships, automatically extract features, process sequence data, utilize large-scale data and adapt to market changes, thus improving the forecasting accuracy and stability of stock movements. Abdul (2023) [5] proposed an optimization method for stock price forecasting based on a multilayer sequential long- and short-term memory (MLS-LSTM) model. The MLS-LSTM algorithm uses normalized time series data divided into time steps to determine the relationship between past and future values for accurate prediction. Ren et al. (2023) [6] proposed a novel hybrid model for stock price prediction that combines wavelet transforms, combining encoder forests with Informer. Explored the Decomposition-Prediction-Reconstruction methodology with machine learning models The impact of fusion is aimed at enhancing the predictive power of the model. Wen et al. (2024) [7] developed a hybrid model called MVL-SVM for predicting stock price trends by combining multi-view learning with Support Vector Machines (SVM).

The model reduces the loss of information by simply inputting heterogeneous multiview data at the same time. The model is used for the prediction of stock price trends by combining multi-view learning with support vector machines.

In summary, the stock price prediction algorithm has high requirements for prediction accuracy, and this paper improves the technical analysis method based on deep learning LSTM algorithm. In this paper, we take CSI 300 stocks for research, use Random Forest for factor selection, use XGBoost method to select high-quality stocks from the stock pool, and then construct CNN-BiLSTM-Attention for stock price prediction of the selected high-quality stocks, compared with CNN, LSTM and other separate models, combining the two and adding Attention mechanism has better performance, with better prediction accuracy, then the trading strategy is formulated by MACD trading strategy, and finally backtesting is performed based on the trading strategy to check the accuracy and effectiveness of the strategy. The quantitative trading strategy using this deep learning model can improve the stability and accuracy of stock return prediction, which is of some significance in maintaining market stability and strengthening the adjustment of national economic policies, and helps investors to make faster decisions and reduce investment risks when making investments.

2. Model Building and Trading Strategies

2.1. Random Forest and XGBoost Models

Random Forest is an integrated learning algorithm using decision trees as estimators, combining multiple decision trees together, each time the dataset is randomly have put back selected, while randomly selecting some of the features as inputs [8]. The random forest model will construct n different sample datasets by randomly sampling in the original stock dataset, and then construct n different decision tree models based on these datasets, and finally the final result is obtained by averaging these decision tree models [9].

XGBoost iteratively trains multiple decision trees by gradient boosting algorithm and combines them to form a powerful integrated model [10]. It features regularization, feature importance evaluation, learning rate control, and early stopping strategy, which can effectively deal with complex nonlinear relationships and high-dimensional features, as well as high performance and scalability [11]. It is known that the training dataset $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, the loss function $l(\hat{x}_i, \hat{y}_i)$, the regularization term $\Omega(f_k)$, then the overall objective function can be written as:

$$L(\phi) = \sum_t (l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)) \tag{1}$$

Where $L(\phi)$ is the expression on linear space; i is the i sample, k is the k tree; $\hat{y}_i$ is the predicted value of the i sample x_i, $\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)$.

Since: $\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) = \hat{y}_i^{t-1} + f_t(x_i)$, then $L(\phi)$ is transformed into the following form:

$$L^t = \sum_{i=1}^{n} \left(y_i, \hat{y}_i^{t-1} + f_t(x_i)\right) + \sum_k \Omega(f_k) \tag{2}$$

Definition: $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$, are the first and second order partial derivative cumulative sums, respectively. Substituting into the objective function yields: $L^t = \sum_{j=1}^{T} [G_j W_j + 1/2(H_j + \lambda) W_j^2]$.

2.2. Convolutional Neural Network

Convolutional neural network (CNN) is a deep learning model that is widely used in image processing [12]. However, it can also efficiently process sequence data such as stock price time series [13]. In CNN, automatic extraction of local features from input data is achieved by using multiple convolutional layers with pooling layers. In this paper, the CNN model is processed with a series of convolutional and pooling layers to capture the patterns and trends of stock prices in different time scales.

Input Layer: The input to the model is a stock price time series data containing feature variables, where each time step corresponds to a feature vector. Let the input data be: $X = [x_1, x_2, \cdots, x_T]$, where x_t represents the feature vector of time step$t$.

Convolutional layer: The convolutional layer contains several convolutional kernels (filters), each of which can be regarded as a feature detector, which utilizes convolutional operations to extract local features from the input data in order to capture patterns and trends in stock prices.

$$y_i = f(\sum_{j=0}^{k-1} \omega_j x_{i+j} + b) \tag{3}$$

Where y_i is the output of the convolutional layer, x_{i+j} is the $i + j$ element of the input sequence, ω_j is the weight of the convolutional kernel, b is the bias, and $f(\cdot)$ is the activation function.

Pooling layer: Used to reduce the dimensionality of the time series and extract higher-level features. Common pooling operations include Max Pooling and Average Pooling, which extract local maxima and averages, respectively.

$$y_i = \max(x_{i-s}, x_{i-s+1}, \cdots, x_{i-s+s-1}) \tag{4}$$

Where y_i is the output of the pooling layer, x_{i-s} to $x_{i-s+s-1}$ are a set of elements of the input sequence, and s is the size of the pooling window.

2.3. Short- and Long-term Memory Neural Networks

Long Short-Term Memory Network (LSTM) is a recurrent neural network (RNN) variant for sequence data processing [14]. Compared to traditional RNNs, LSTM can effectively overcome the problems of gradient vanishing and gradient explosion in long time sequences through the gating mechanism to better capture long term correlations [15]. BiLSTM, on the other hand, adds backpropagation to LSTM and is able to utilise past and future contextual information at the same time, which allows the model to better understand the temporal dependencies and patterns in sequence data [16]. The core of the LSTM network is the Memory Cell, which controls the input, output and forgetting of information through a series of gates to achieve effective modelling of sequential data, and is mainly composed of four parts: the forgetting gate, the input gate, the output gate and the internal memory cell [17].

The forgetting gate decides which information in the memory cell should be forgotten, and its calculation process can be expressed as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{5}$$

The input gate determines which parts of the input data should be updated into the memory cell, and its computation can be expressed as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (6)$$

The internal memory unit updates the current memory state based on the candidate memory state information, which may be represented as:

$$\tilde{C}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (7)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (8)$$

The output gate determines which information in the memory cell should be output to the hidden states and outputs at the next moment and can be expressed as follows:

$$O_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (9)$$

$$h_t = O_t * \tanh(C_t) \quad (10)$$

Where W_f, W_i, W_c, W_o are the weight matrices of the forgetting gate, the input gate, the update unit link and the output gate, b_f, b_i, b_c, b_o are their offsets, f_t, i_t, O_t are the outputs of the forgetting gate, the input gate, and the output gate, respectively, σ is the activation function Sigmoid, h_t is the hidden state within the unit, and U is the corresponding weight matrices of the three gates.

The Bidirectional Long Short-Term Memory Neural Network (BiLSTM) model splices the hidden states of the forward LSTM and the inverse LSTM, which can refer to both historical and future data on the prediction results, and can be expressed as:

$$H = [\vec{h}_1, \vec{h}_2, \cdots, \vec{h}_T, \overleftarrow{h}_T, \cdots, \overleftarrow{h}_2, \overleftarrow{h}_1] \quad (11)$$

Where H is the output of BiLSTM.

2.4. Attention Mechanism

Attention Mechanism is a technique used to enhance the ability of neural networks when processing sequential data [18]. Attention Mechanism allows the model to assign different weights and attention to different parts of the input when processing sequential data. In this paper, the attention mechanism can be used to dynamically select and weight the importance of different feature variables, allowing the model to pay more attention to features that have a greater impact on stock price forecasting.

The attention weights can be obtained by calculating the similarity between the input sequences and the hidden states of the BiLSTM. Commonly used calculation methods include dot product, additive and multiplicative attention mechanisms.

Multiplicative Attention Mechanism: $e_t = softmax(W_a \cdot \tanh(W_h \cdot H + b_h))$, where W_a and W_h are the learned weight matrices, b_h is the bias vector, and H is the output of the BiLSTM.

Attention weights are applied to the outputs of BiLSTM, which are weighted and summed to obtain an attention-weighted representation:

$$A = \sum_{t=1}^{T} e_t \cdot H_t \quad (12)$$

Finally, the attention-weighted representation A can be fed into a subsequent layer (e.g., a fully connected layer) for further processing and prediction.

2.5. CNN-BiLSTM-Attention Model

In this paper, three models, CNN, BiLSTM and Attention, are combined to form a stock prediction framework CNN-BiLSTM-Attention. The network structure of the stock price prediction model CNN-BiLSTM-Attention is shown in Fig. 1. Firstly, the model extracted five stock features in historical time through CNN, secondly, BiLSTM trained and predicted daily individual stock data through two LSTM layers, forward and backward, which captured the forward and reverse information of the sequence, respectively, and based on this, the Attention mechanism was introduced, which gave a different weight to the features of each stock, to improve the neural network's learning ability of the neural network to achieve effective mining of individual stock features.

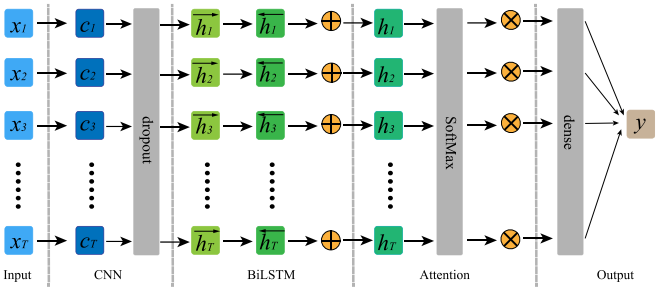

Figure 1. Network structure of stock price prediction model CNN-BiLSTM-Attention

2.6. Quantitative Trading Strategies

This article is a quantitative trading strategy based on the MACD indicator. MACD (Moving Average Convergence Divergence) is an indicator commonly used in technical analysis to identify the strength, direction and possible reversal points of a trend [19]. MACD consists of two moving averages: a fast line (MACD line) and a slow line (signal line), and a bar chart that indicates the difference between the two. When the MACD line crosses the signal line from below, creating a 'golden cross', it indicates the possible start of an uptrend, so consider buying; when the MACD line crosses the signal line from above, creating a 'dead cross', it indicates the possible start of a downtrend, so consider selling.

3. Empirical Analysis

3.1. Data Sources and Selection of Indicators

This paper collects the daily frequency data of CSI 300 constituents from January 2020 to December 2022, which includes the stock market indicators and fundamental indicators, which are gain, amplitude, yesterday's closing price, opening price, closing price, high price, low price, average price, number of shares traded, turnover, turnover, and the total market capitalisation, market capitalisation in circulation, total share capital, outstanding share capital, price-earnings ratio, and price-to-capitalisation ratio, etc. This paper gives 17 indicators that can evaluate both the value and the risk of a stock, which are mainly used to reflect the change of the price and the market activity, as well as to assess the investment value of a stock.

This paper gives data of 17 indicators capable of evaluating both the value and risk of a stock, which are mainly used to reflect the changes in stock prices and the activity of the market, as well as to assess the investment value of a stock. Yang et al. integrated three types of factors, which are the price factor, the hard-coded factor, and the factor based on the rolling arithmetic [20]. In addition to these three types of factors, based on the original data, this paper constructs the sentiment indicator ARBR, which measures the power of buyers and sellers in the market, and some commonly used technical indicators, such as moving averages and deviation ratios. The types of the above factors are shown in Table 1.

Table 1. Factor description table

style	name	formulas	hidden meaning
price factor	OPEN0	open/close	Percentage of opening price relative to closing price
	HIGH0	high/close	Highest price as a percentage of closing price
hard coding factor	KMID	(close-open)/open	Percentage increase (decrease) of closing price over opening price
	KSFT	(2*close-high-low)/open	Percentage of the difference between the closing price and the high and low prices relative to the opening price
Factorization based on rolling operators	ROC5	Ref (close, 5)/close	Ratio of the closing price on day t-5 to the closing price on that day
	ROC10	Ref (close, 10)/close	Ratio of closing price on day t-10 to closing price on that day
technical factor	5D_MA	Sum of closing prices of the previous 5 days/5	Average of the closing prices of the previous 5 days
	BIAS	(Closing price - 5D_MA)/5D_MA	Percentage of difference between closing price and 5-day moving average
emotional factor	ARBR	(high-open)/(open-low)-(high-lastclose)/(lastclose-low)	Difference between popularity indicator AR and willingness indicator BR

Including the original indicators, there are 25 factors in this paper, and it is found by correlation analysis that many feature factors are highly correlated, i.e., generating the problem of multiple covariance, which will make the model unstable and the reliability of the prediction results reduced. So, in this paper, after feature extraction, the number of redundant features is reduced by feature selection, and the correlation coefficient of 0.9 is used as the threshold to remove some features.

After screening, a total of 8 factors are removed as yesterday's closing price, closing price, high price, low price, average price, total equity, KMID, 5D_MA, and the remaining 17 factors are used as candidate factors.

3.2. Data Preprocessing

3.2.1. Data Normalisation

Data dimensionlessness can eliminate the incomparability between feature variables, which usually includes z-score normalization and min-max normalization, where normalization processing is also a frequently used method in neural networks to improve the model prediction performance. From the data set can be seen that the number of traded shares and other variables are of large magnitude, so before the data analysis, this paper normalizes the sample set, and under the premise of keeping the data distribution unchanged, the range is scaled to [0, 1], which is expressed by the formula:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{13}$$

3.2.2. Random Forest Selection Factor Results

Random Forest is a machine learning algorithm based on decision trees, which can select the most important feature variables for regression prediction among a large number of features, and rank the importance of each feature, so as to achieve the effect of dimensionality reduction and eliminate redundant features, which helps to improve the performance of the learning algorithm. Figure 2 is based on the importance score of each factor of the Random Forest, the larger the score value, the higher the degree of importance of the influence factor, it can be seen that the closing price and yesterday's closing price has the greatest impact on the stock gain. In this paper, we take the indexes ranked in the top 10 as the influence factors of the rate of increase, which are used in the construction of the next prediction model.

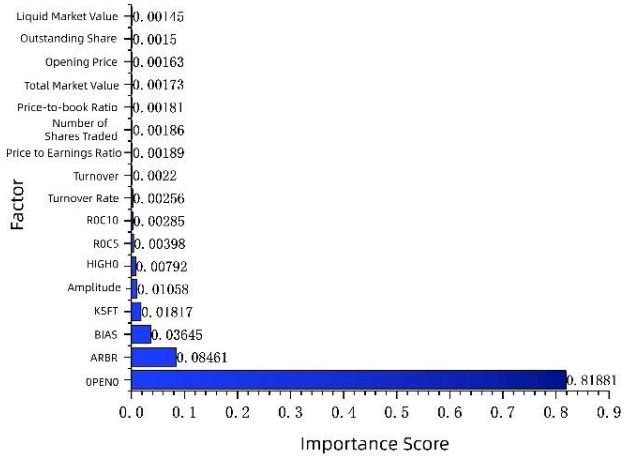

Figure 2. Factor importance score

3.2.3. XGBoost Stock Picks Results

In the above 10 influencing factors of the increase were determined by random forest regression, excluding variables such as market capitalisation, outstanding share capital, opening price, total market capitalisation, P/B ratio, number of shares traded, P/E ratio, etc. XGBoost is the integrated model with high predictive accuracy and operational efficiency, which can make full use of a large amount of data in the financial field, and at the same time, has very good stability and explanatory power, which is more and more frequently applied in the modern multi-factor stock selection models. Therefore, in this paper, we will divide the training set and test set according to the ratio of 8:2, and construct the multifactor regression model through XGBoost algorithm and influence variables, and the parameters are set as follows: the number of trees n_estimators is taken to be 200, the learning rate learning_rate is taken to be 0.01, the base classifiers booster is the decision tree gbtree, the objective function objective is the squared error reg: squarederror.

From the analysis of the evaluation metrics in Table 2, it can be seen that the MAE, MSE, and RMSE are small, and the R-square of both the training set and the test set reaches more than 0.95, all of which indicate that XGBoost fits the rise better. According to the stock rise trend of part of the test set3 it can be seen that the local prediction accuracy of the model decreases when the stock rise fluctuates a lot, but from the overall trend, the predicted value is roughly the same as the real value.

Table 2. XGBoost model prediction results

Evaluation indicators	MAE	MSE	RMSE	r2_score_train	r2_score_test
Projected results	0.2730	0.3263	0.5713	0.9801	0.9604

In this paper, the XGBoost regression model is constructed based on the influencing factors after filtering out the redundant features, that is, the XGBoost model will take into account the weights of the influencing factors in the prediction, and get the final prediction results through the weighted summation of each factor. Therefore, compared with the real value, the predicted value of stock gain takes into account the market performance and investment risk of the stock, and is more representative of the daily return of each stock under the influence of important factors. In this paper, the geometric mean return of each stock in the range of July-December 2022 is calculated, and then the four stocks with the highest ranking are selected. The results are shown in Table 3, and four stocks, China Unicom, Kingsoft Office, China Life, and China Mobile, are finally selected for the portfolio.

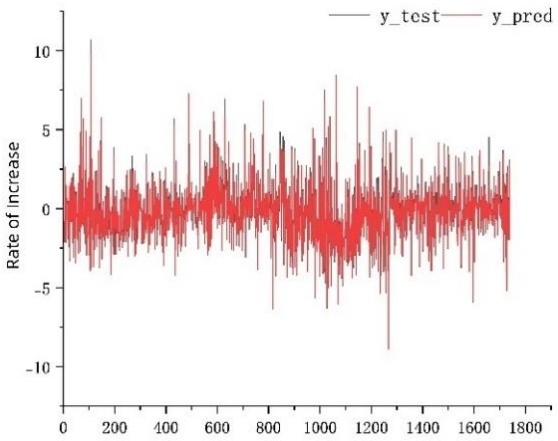

Figure 3. Partial test set fitting plots

Table 3. Ranking of the average return of each stock

Stock code	600050	688111	601628	600941	601808	601021	600570	600036	601728	600150
Stock Name	China Unicom	Kingsoft Office	China Life	China Mobile	China National Offshore Oil Service	Spring Airlines	Hang Seng Electronics	China Merchants Bank	China Telecom	China Shipbuilding
average rate of return /%	1.3113	1.2602	1.1819	1.1737	1.1625	1.1612	1.1539	1.1474	1.1447	1.1417

3.3. Analysis of Results

3.3.1. CNN-BiLSTM-Attention Stock Price Prediction

Based on the correlation analysis, this paper selects the top five factors with the highest correlation with the closing price: yesterday's closing price, opening price, high price, low price and average price for the next model construction. After the stock selection model selection in the previous chapter, this paper will 688111 Kingsoft Office, 600050 China Unicom, 601628 China Life and 600941 China Mobile total four stocks of daily frequency data set for experiment.

In the process of stock price prediction, in order to improve the learning ability of the model and the prediction ability of the stock price change, this paper divided the data set into two parts, in which the first 70% of the time series was used as a training set to train the network model, and the remaining 30% of the data set was used as a test set, and the feature data of the test set was used to predict the four stock prices. Based on this, the prediction results in the test set were compared with the real results and the fitting effect was analyzed.

In this paper, model construction and experiments are carried out. The ReLU activation function is used for the one-dimensional CNN model, the Tanh activation function is used for the bidirectional LSTM model, and the Sigmoid activation function is used for the Attention model. In this model, the mean square error is used as the loss function, and the optimiser is specified as Adam, and the evaluation metric is specified as the mean absolute error. The parameters are set as follows: Window, Kernel_size and Pool_size is all set to 1, Lstm_units and Filters are set to 128, Dropout is set to 0.01, Epochs are set to 200, and the number of training rounds is set to 200. Epochs) is 200, the number of feature dimensions (Input_dim) is 5, and the number of samples per training round (Batch_size) is 256.

For this stock price prediction framework, the daily dataset of each stock is pre-processed and fed into the model for learning and training respectively, and through continuous experimentation and observation of the results, the knowledge and experience are used for constant parameter adjustment and optimisation training, finally forming a set of parameter values that are better for the prediction of each stock, and then inverse normalisation is performed on the output stock price prediction values to obtain the final prediction value of the model. The results of the model are shown in Table 4.

Table 4. Model prediction results

stock	688111 Kingsoft Office	600050 China Unicom	601628 China Life	600941 China Mobile
MSE	3.4486	0.0023	0.0415	0.5458
MAPE	0.69%	0.99%	0.55%	0.58%
R^2score	0.9968	0.9788	0.9952	0.9479

From the results of each evaluation index in Table 4, it can be found that the MAPE of the four stocks is less than 1%, which indicates that the prediction accuracy of the stocks is more than 99%. Except for Kingsoft Office the MSE of the other 3 stocks is not more than 0.6. For the fit of this regression model, it can be seen that the coefficients of determination of the 4 stocks are greater than 0.94, which indicates that the predictive performance of the model is better.

The experimental results show that in the stock dataset tested in this paper, the stock price prediction model used, CNN-BiLSTM-Attention model, has better generalisation ability and robustness, which is beneficial for the subsequent quantitative trading strategies and backtesting experiments.

In order to more intuitively observe the prediction effect of the model on the test set, the fitting plots of 4 stocks, 688111 Kingsoft Office, 600050 China Unicom, 601628 China Life, and 600941 China Mobile, are next plotted on the test set, which are sequentially shown in Figures 4-Figure 7, respectively.

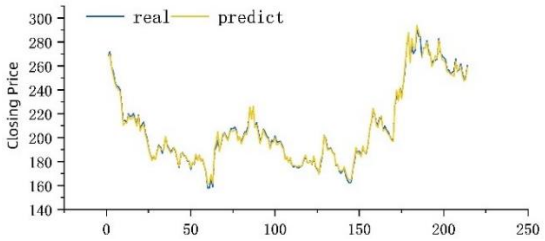

Figure 4. 688111 Kingsoft Office Test Set Fitting Chart

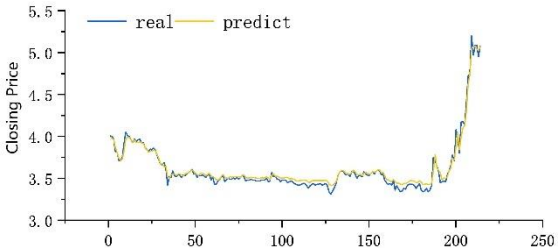

Figure 5. 600050 China Unicom Test Set Fit Chart

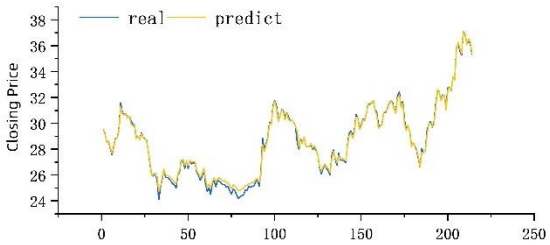

Figure 6. 601628 China Life Test Set Fit Chart

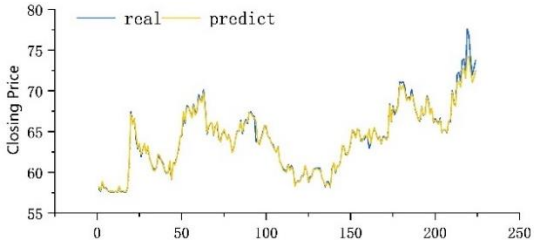

Figure 7. 600941 China Mobile Test Set Fit Chart

From the prediction graphs of these 4 stocks on the test set, the model can roughly predict the trend of the stock price, for the local peaks with the accuracy is lacking, but in general, the generalisation ability and stability of the model are good. Moreover, when the model reaches the highest point or the lowest point, its prediction curve has almost no offset compared with the real curve, which indicates that the model can eliminate the phase difference without time delay phenomenon. And it can be seen from the results of the above evaluation indexes that the overall performance of the model is better and the error of prediction is smaller.

3.3.2. Quantitative Stock Trading Strategies and Backtesting

In this paper, the MACD trading strategy is applied to 4 stocks, 688111 Kingsoft Office, 600050 China Unicom, 601628 China Life, 600941 China Mobile, and 688111 Kingsoft Office, respectively, for June-December 2022, based on the data test set obtained from the CNN-BiLSTM-Attention model prediction, and the respective stock The MACD strategy backtest total return results are shown in Figure 8-Figure 11. cumulative_returns denote the total returns obtained based on the strategy and tcumulative_returns denote the actual total returns.

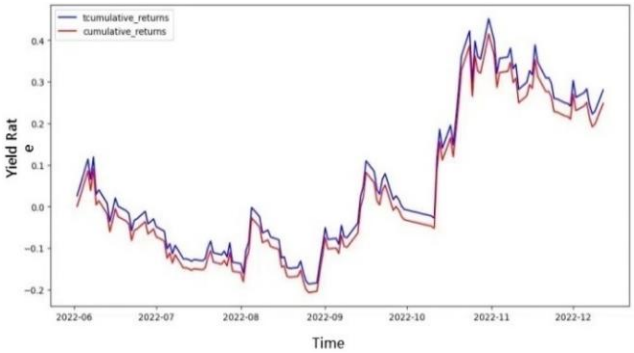

Figure 8. 688111 Kingsford Office based on strategy and true cumulative returns

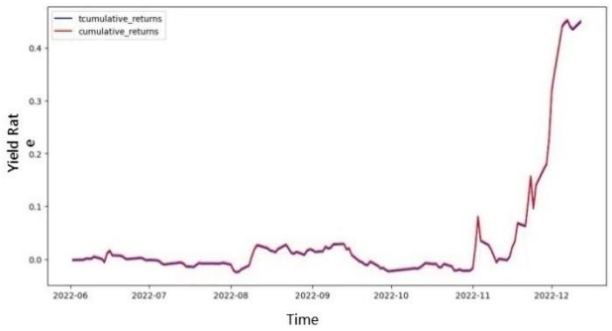

Figure 9. 600050 China Unicom Based on Strategy and True Cumulative Returns

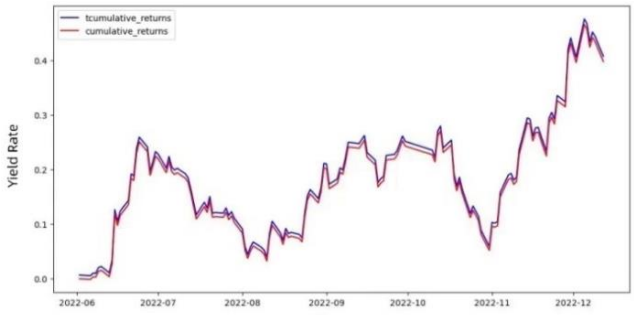

Figure 10. 601628 China Life Based on Strategy and True Cumulative Returns

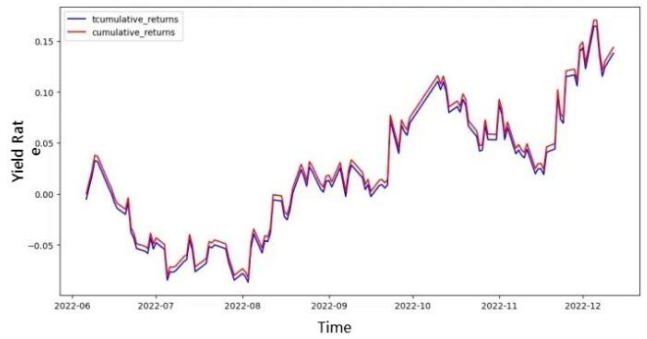

Figure 11. 600941 China Mobile Based on Strategy and True Cumulative Returns

It can be observed that the backtesting of the four stocks selected above using the MACD trading strategy yields that most of the stocks' total returns based on the strategy are lower than their actual total returns, with the exception of China Unicom, a stock whose own closing price base is relatively small, resulting in the total return based on the strategy nearly overlapping with its actual total return curve, and most of the remaining stocks' returns based on the strategy curve are all below the actual return curve, indicating that when we can get a return based on the trading strategy, we can actually get a higher return, which can be traded at that time. And from Table 5, we can see that the total return based on the strategy is more than 40% for most of the stocks, and the loss is controlled within 10%, which indicates that the strategy is still effective and feasible.

Table 5. Total return on four stocks

stock	total return	Maximum Gross Yield	Minimum total return
688111 Kingsford Office	28.00%	45.19%	-18.66%
600050 China Unicom	44.75%	44.98%	-2.52%
600941 China Mobile	13.80%	16.46%	-8.66%
601628 China Life	44.39%	47.61%	0.56%

The above introduces the quantitative trading strategy based on MACD single stock investment, but in real life, we may invest in more than one stock when investing, at this time the amount of investment in each stock has become a difficult point in our investment, based on this we continue to take the above four stocks as an example, select four kinds of investment portfolio, through the calculation of the total return on each portfolio to determine the portfolio mode, the amount of investment in the stock is determined by its corresponding weights. The amount of stock investment is determined by its corresponding weight. The four portfolio styles will be introduced next:

(1) Equal weighted portfolio: the weight of each stock is assigned equally, in this paper the weight of each stock is assigned as 0.25, this investment method is the simplest.

(2) Market capitalisation weighted portfolio: the weight of this portfolio is determined by the percentage of total market capitalisation, which is positively proportional.

(3) Investment Risk Minimising Portfolio: this strategy is to select the portfolio with the lowest investment risk and the highest return under that condition. In this paper, according to Monte Carlo simulation Markowitz model, the weight of each group is randomly generated, calculate the investment return and return standard deviation (risk standard deviation) of each group, the process is repeated a number of times, this paper repeated 10,000 times, the investment return and risk standard deviation of each group is plotted as a scatterplot ultimately as shown in Figure 12, and then find out the point of the portfolio with the minimum risk, as shown in Figure 13.

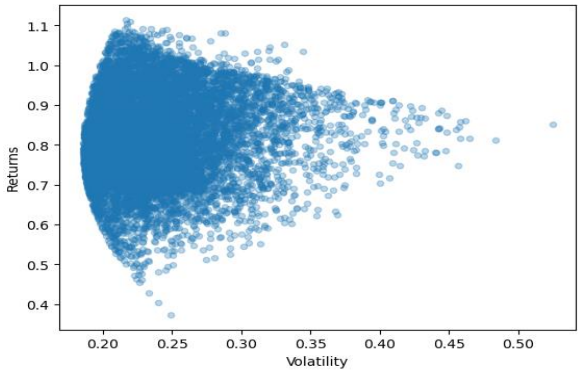

Figure 12. Markowitz simulated standard deviation of risk and scatterplot of returns for the 10,000 group

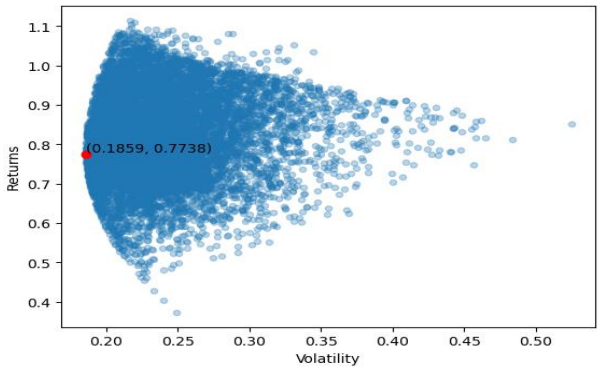

Figure 13. The point in the return-risk scatterplot that minimizes risk

(4) Sharpe Ratio Portfolio: The Sharpe Ratio is a commonly used risk-adjusted rate of return metric designed to measure the amount of excess reward that can be achieved for each unit of total risk taken.

Higher Sharpe ratios indicate higher excess returns per unit of total risk taken and better risk-adjusted returns. Again based on Monte Carlo simulation of the Markowitz model, the Sharpe ratio corresponding to each group is calculated and plotted as the third variable in a scatter plot of return-risk, as shown in Figure 14. The portfolio corresponding to the largest Sharpe ratio is then found, as shown in Figure 15.

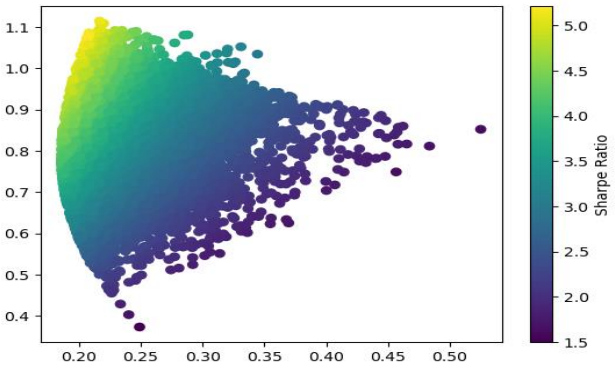

Figure 14. Sharpe Ratio Depicts Scatterplot of Returns-Standard Deviation

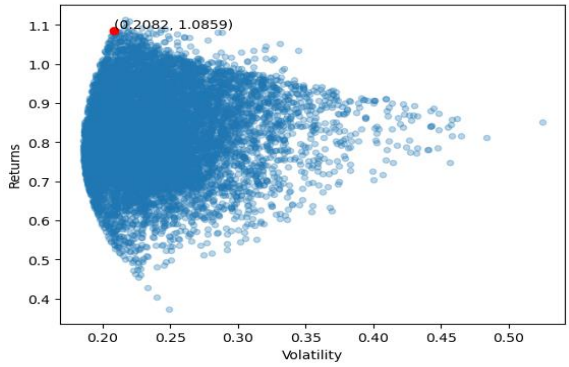

Figure 15. The point in the return-risk scatterplot with the largest Sharpe ratio

The final stock weights and total returns corresponding to the four portfolio approaches are obtained as shown in Table 6 and Figure 16:

Table 6. Equity weights corresponding to the four portfolio approaches

stock	600050 China Unicom	600941 China Mobile	601628 China Life	688111 Kingsoft Office
Equal Weighted Portfolio	0.25	0.25	0.25	0.25
Market capitalization-weighted portfolio	0.04591965	0.56786866	0.34554871	0.04066298
Investment risk minimization portfolio	0.43547867	0.36178263	0.11778557	0.08495312
Sharpe Ratio Portfolio	0.67685089	0.00755021	0.21619762	0.09940129

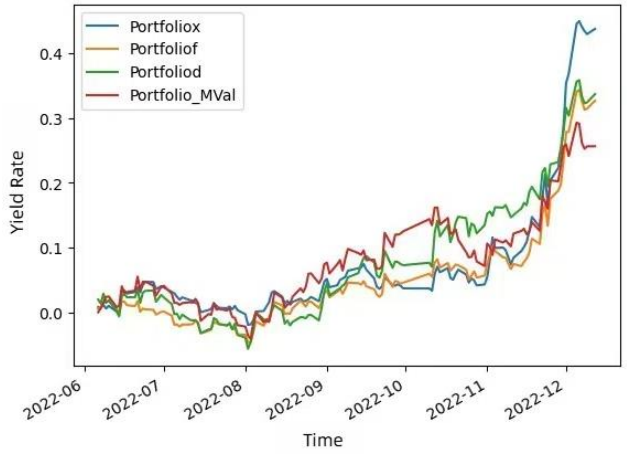

Figure 16. Total return graphs corresponding to the four portfolio approaches

Table 7. Portfolio projected total return

stock	total return	Minimum total return	Maximum Gross Yield
Equal weighted portfolio	33.64 %	-5.60%	35.80%
Market capitalization-weighted portfolio	25.62 %	-4.07%	29.27%
Investment risk minimization portfolio	32.64 %	-4.59%	34.33%
Sharpe Ratio Portfolio	43.72 %	-1.93%	44.90%

As can be seen from Table 7, the highest total returns obtained through the four portfolio approaches are basically above 30%, with losses controlled within 5%, indicating that the returns that can be obtained from these four portfolios are all relatively stable and less risky. Returns from Figure 16 can be seen, in these four kinds of investment portfolio approach, Portfoliof investment risk of the smallest portfolio in these four portfolio approach to obtain the return has been relatively small, in the long run, Portfolioiox Sharpe ratio investment portfolio is relatively more potential, in the later stage can obtain a relatively high return. Investors can therefore choose the right portfolio to trade at the right time depending on the length of time they have been buying shares.

4. Conclusion

This paper addresses the time series stock price prediction problem. Firstly, the most important 10 factors affecting stock returns are selected through random forest, and the constructed XGBoost regression model is used for four high-quality stocks selection. Secondly, for the prediction of stock closing price, this paper proposes a new framework CNN-BiLSTM-Attention model based on deep learning method, which extracts the historical information features of the stock through CNN, and trains and predicts the stock dataset by using the bidirectional LSTM network, meanwhile, an Attention mechanism is added in the process of training to give every stock features, so that the network can learn the feature information more effectively and improve the prediction accuracy. Finally, the cumulative returns of the four selected stocks are calculated and backtested using the single stock quantitative trading strategy of MACD, and the highest total returns of most stocks are over 40%, and the losses are controlled within 10%, which indicates that the trading strategy based on this model is effective and feasible. Moreover, when the stock portfolio returns are studied, the Portfoliox Sharpe Ratio portfolio has a relatively higher potential to achieve higher returns at a later stage. This shows that the proposed model has a certain predictive ability for stock prices, which can be used as a reference for investors to make faster decisions when investing and reduce the risk of investment.

Acknowledgment

This work is supported by the Natural Science Fund for Colleges and Universities, Department of Education of Anhui Province (KJ2021A0481), and Anhui University of Finance and Economics Graduate Student Research and Innovation Fund Project (ACYC2023173).

References

[1] Box G E P, Jenkins G M. Time Series Analysis: Forecasting and Control [M]. San Francisco: Holden-Day, 1970.

[2] Ariyo A A, Adewumi A O, Ayo C K. Stock Price Prediction Using the ARIMAModel [C]//2014 UKSim-AMSS 16th International Conference on ComputerModelling and Simulation. IEEE, 2014:106-112.

[3] Wen F, Xiao J, Zhifang H E, et al. Stock Price Prediction Based on SSA and SVM [J]. Procedia Computer Science, 2014, 31: 625-631.

[4] Lahmiri S. Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression [J]. Applied Mathematics and Computation, 2018, 320: 444-451.

[5] Md A Q, Kapoor S, AV C J, et al. Novel optimization approach for stock price forecasting using multi-layered sequential LSTM [J]. Applied Soft Computing, 2023, 134: 109830.

[6] Ren S, Wang X, Zhou X, et al. A novel hybrid model for stock price forecasting integrating encoder forest and informer [J]. Expert Systems with Applications, 2023, 234: 121080.

[7] Long W, Gao J, Bai K, et al. A hybrid model for stock price prediction based on multi-view heterogeneous data [J]. Financial Innovation, 2024, 10(1): 48.

[8] BREIMAN L. Random Forests [J]. Machine Learning, 2001,45(1):5-32.

[9] Iranzad R, Liu X. A review of random forest-based feature selection methods for data science education and applications [J]. International Journal of Data Science and Analytics, 2024: 1-15.

[10] CHEN T, GUESTRIN C. XGBoost: A scalable tree boosting svstem [C]//In proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 785-794.

[11] Niazkar M, Menapace A, Brentan B, et al. Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023) [J]. Environmental Modelling & Software, 2024: 105971.

[12] Joshi A, Vishnu C, Mohan C K, et al. Application of XGBoost model for early prediction of earthquake magnitude from waveform data [J]. Journal of Earth System Science, 2023, 133(1): 5.

[13] Kang M. Stock Price Prediction with Heavy-Tailed Distribution Time-Series Generation Based on WGAN-BiLSTM [J]. Computational Economics, 2024: 1-20.

[14] Bukhari A H, Raja M A Z, Sulaiman M, et al. Fractional Neuro-se-quential ARFIMA-LSTM for Financial Market Forecasting [J]. IEEE Access,2020,(8).

[15] Khetarpal P, Nagpal N, Siano P, et al. Power quality disturbance signal segmentation and classification based on modified BI-LSTM with double attention mechanism [J]. IET Generation, Transmission & Distribution, 2024, 18(1): 50-62.

[16] Ye Z, Zuo T, Chen W, et al. Textual emotion recognition method based on ALBERT-BiLSTM model and SVM-NB classification [J]. Soft Computing, 2023, 27(8): 5063-5075.

[17] Thumilvannan S, Balamanigandan R. A novel adaptive weight bi-directional long short-term memory (AWBi-LSTM) classifier model for heart stroke risk level prediction in IoT [J]. PeerJ Computer Science, 2024, 10: e2196.

[18] Li H, Wu X J. CrossFuse: A novel cross attention mechanism based infrared and visible image fusion approach [J]. Information Fusion, 2024, 103: 102147.

[19] Agudelo Aguirre A A, Duque Méndez N D, Rojas Medina R A. Artificial intelligence applied to investment in variable income through the MACD (moving average convergence/divergence) indicator [J]. Journal of Economics, Finance and Administrative Science, 2021, 26(52): 268-281.

[20] Yang X, Liu W, Zhou D, et al. Qlib: An AI-oriented Quantitative Investment Platform [J]. arXiv preprint arXiv:2009.11189, 2020.