ARIMA with Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction in the US stock market

Luocheng Liang[1, *]

[1]School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, Suzhou, China, 215000

Abstract: Stock price forecasting is considered one of the most difficult tasks in financial forecasting. Combining ARIMA with neural networks helps to enhance the model's predictive capabilities when dealing with complex, nonlinear time series data. Attention-based CNN-LSTM and XGBoost hybrid model achieves the accuracy of stock prediction results. However, the predictive effect of this hybrid model has only been confirmed by Chinese stock market data. Therefore, this paper proposes to use ARIMA with Attention-based CNN-LSTM and XGBoost hybrid model to predict the stock price of five different types of companies in the US stock market. Experimental results show that this hybrid model performs well in the training phase and has small prediction errors. However, during the testing phase, the model performs better in predicting stocks with larger price fluctuations, while the prediction error for stocks with more stable prices increases significantly. This shows that this hybrid model has good predictive ability when dealing with stocks with large market fluctuations, but it still needs further optimization for stable stocks. This study further demonstrates the potential of this hybrid model in stock price prediction, especially during market fluctuations, as it can be used as an effective tool to avoid risks and obtain returns.

1. Introduction

Due to the complexity of the stock market, stock price prediction is considered one of the most difficult tasks in financial forecasting123. With the rapid development of financial markets and the increasing complexity of stock price movements, there is a growing need for more complex predictive models. In the past three decades, countless price prediction methods have been extensively studied. Most early studies used econometric models to predict stocks price, such as autoregressive integrated moving average 6(ARIMA). However, this model has strong assumptions about the data that cannot always be satisfied, and it is often difficult to satisfy and consider the impact of other factors on stock price fluctuations22. Therefore, in later studies, researchers have identified predicting stocks on nonlinear and non-stationary financial time series as one of the most challenging tasks5. Although several mathematical models have been developed, the results are still often unsatisfactory4.

In recent years, with the advancement of artificial intelligence, the application of neural networks in stock market prediction has garnered significant attention. Neural networks have the ability to model complex nonlinear relationships and learn from large datasets, providing a powerful tool for predicting stock prices. These models, including Recurrent Neural Networks13 (RNNs) and Long Short-Term Memory 12(LSTM) networks, are particularly effective in capturing temporal dependencies in stock price data.

Additionally, hybrid models that combine traditional statistical methods with neural networks, such as the Attention-based CNN-LSTM-XGBoost 17hybrid model, have shown promise in improving prediction accuracy by leveraging the strengths of multiple approaches. The integration of attention mechanisms19 further enhances these models, allowing them to focus on important features in the data. Furthermore, incorporating ensemble learning methods such as XGBoost 18enhances the robustness and accuracy of stock price predictions. And the combination of these complex models can help people better understand and predict the highly dynamic and complex behavior of the stock market and achieve reliability and accuracy of prediction results.

2. Literature review

The autoregressive integrated moving average (ARIMA) model is one of the most important and traditional time series models. Although ARIMA models are quite flexible as they can represent several different types of time series, namely pure autoregressive (AR), pure moving average (MA) and combined AR and MA (ARMA) series, their main limitation is the pre-assumed linear form of the model. That is, it is assumed that there is a linear correlation structure between time series values, and therefore, ARIMA model cannot capture nonlinear patterns7. In recent years, with the continuous

*Corresponding author: Luocheng.Liang23@student.xjtlu.edu.cn

development of artificial intelligence and intelligent algorithms, more and more artificial intelligence methods are widely used in financial markets. The most typical artificial intelligence method among them is neural network with strong nonlinear generalization ability. For the above reasons, combining ARIMA with neural networks is a common and effective practice to enhance the predictive capabilities of the model when dealing with complex and nonlinear time series data.

With the development of deep neural networks, attention mechanism has been widely used in various application fields8.It simulates human attention, that human can focus on useful information and ignore useless information9.Since CNN10mainly relies on convolution operations and can usually only capture local spatial features, introducing the attention mechanism to this model can improve the performance and expressiveness of the model. And Attention-based Convolutional Neural Networks (ACNN) are widely used in sequence modeling1115.

Long Short-Term Memory (LSTM) model is the most commonly used model in RNN. Although LSTM can handle long-term sequence data, it is still an RNN structure, which lacks extraction of global information14 and may still encounter challenges on particularly long sequences. Combining ACNN and LSTM is a self-attention-based sequence-to-sequence (seq2seq) 16 model for encoding and decoding sequential data. This model not only can solve long-term dependency problem in LSTM, allowing better modeling of long sequences, but also make the architecture more flexible and robust, since ACNN can capture local and global communication, while LSTM can capture specific long-distance communication that fits the structure of LSTM itself 17.

XGBoost is a scalable end-to-end tree boosting system18 that be widely used for stock price prediction because of its robustness, flexibility and high prediction performance. Based on the above model, this article uses a new hybrid deep learning model called Attention-based CNN-LSTM and XGBoost hybrid model (AttCLX) to predict stocks. Based on the results of Shi and Hu (2022) using this model to predict the price of Bank of China in the Chinese stock market, it is shown that this hybrid model is beneficial to helping institutions or investors achieve the purpose of avoiding risks and expanding returns. And in the experiment, they used the ARIMA model in the preprocessing stage of the data, which made the prediction results more accurate. However, Shi and Hu (2022) only used Bank of China’s stock in the Chinese stock market in their article. In summary, due to the good performance of the ARMIA with Attention-based CNN-LSTM and XGBoost hybrid model and differences between the Chinese and American stock markets, this article will explore the predictive effect of ARMIA with Attention-based CNN-LSTM and XGBoost hybrid model on the US stock market.

3. Materials and methods

3.1. ARIMA

Classical stock forecasting methods are based on ARMA6(Auto Regressive Moving Average) model and ARIMA (Auto Regressive Integrated Moving Average) model. An ARMA (p, q) mode.

$$s_t = a_0 + \sum_{i=1}^{p} a_i s_{t-i} + w_t + \sum_{i=1}^{q} b_i w_{t-i}, \quad (1)$$

where a, b are parameters, w is noise. ARMA model can be used when sequence $s_{1:N}$ is stationary, which means

$$E[s_t] = Constant. \quad (2)$$

$$Cov(s_t, s_{t-k}) = Constant. \quad (3)$$

where t = 1, 2, ..., N and k = 1, 2, ..., t.

When the sequence is non-stationary, ARIMA (p, q, d) applies a d-order differencing to it. In stock forecasting, the first-order difference $x_{1:N} = \dot{s}_{1:N}(i.e. x_k = s_k - s_{k-1})$ is generally regarded as a stationary sequence.

3.2. Deep learning on sequential data

In basic feed-forward neural network (FFNN), the output at the current moment o_t is determined only by the input at the current moment i_t, which inhibits the ability of the FFNN to model time-series data. In recurrent neural network (RNN), a delay is used to save the latent state from the previous moment h_{t-1}, then the latent state at the current moment h_t is influenced by both h_{t-1} and i_t. RNNs can lead to errors due to vanishing or exploding gradients during backpropagation through time, which makes it difficult to learn long-term dependencies12. Human can selectively remember information, and the LSTM (long short-term memory) model mimics this through gated activation functions, allowing it to selectively remember new information and forget old information.

Sequence-to-sequence (seq2seq) model uses an encoder-decoder architecture for analyzing sequential data, which enhances the LSTM's ability to learn hidden information from noisy data. In seq2seq, the encoder LSTM encodes the inputs into a context (usually the hidden state at the last step h_N), then the context is decoded by the decoder. In the decoder, the output from the previous moment becomes the input for the next moment.

The seq2seq model uses beam search 21 to optimize the decoding sequence. Both beam search and Viterbi algorithm in hidden Markov model (HMM) are based on dynamic programming. In HMM, solving the optimal estimate of the current state based on observations and previous states is called decoding or inference. Solving $p(x_k|y_{1:k})$ and $p(x_k|y_{1:N})$ respectively, which is equivalent to the forward-backward algorithm in HMM. The optimal bidirectional estimation can be achieved through the distribution of x_k This is the probabilistic perspective that proposes a bidirectional LSTM that combines forward and backward20.

3.3. Attention mechanism

Humans usually focus on salient information. Attention mechanism is a deep learning technology based on the human cognitive system. It allows a model to focus on specific parts of the input sequence when generating each part of the output, improving the handling of long-range dependencies in sequential data. For input $X = (x_1, x_2, \dots, x_N)$,given the query vector q, use attention z=1, 2, …, N to describe the index of the selected information, and then describe the distribution of attention

$$\alpha_i = p(z = i|X, q) = \frac{exp\big(s(x_i, q)\big)}{\sum_{j=1}^{N} exp\left(s(x_j, q)\right)}. \quad (4)$$

i.e.

$$\alpha_i = softmax\big(s(x_i, q)\big) \quad (5)$$

Here

$$s(x_i, q) = \frac{x_i^T q}{\sqrt{d}} \quad (6)$$

is attention score through scaled dot product, d is the dimension of input. Assume that the input key-value pairs $(K, V) = [(k_1, v_1), \dots, (k_{N,} v_N)]$, for given q, attention function.

$$att\big((K, V), q\big) = \sum_{i=1}^{N} \alpha_i v_i = \sum_{i=1}^{N} \frac{exp\big(s(k_i, q)\big)}{\sum_j exp\left(s(k_j, q)\right)} v_i. \quad (7)$$

Multi-head mechanism usually uses multi-query $Q = [q_1, \dots, q_M]$ for attention function computation. $att\big((K, V), Q\big) = \big(att\big((K, V), q_1\big) \| \dots \| att\big((K, V), q_M\big)\big)$. Here, ‖ denotes Concatenate operation. This is called multi-head attention (MHA).

Attention mechanism can be employed to generate data-driven different weights. Here, Q, K, V are all obtained through linear transform of X, and W_Q, W_K, W_V can be dynamically adjusted.

$$Q = W_Q X, K = W_K X, V = W_V X. \quad (8)$$

This is called self-attention. Similarly, output

$$h_i = att\big((K, V), q_i\big). \quad (9)$$

Hence

$$h_i = \sum_{j=1}^{N} \alpha_{ij} v_j = \sum_{j=1}^{N} softmax\left(s\big(k_j, q_i\big)\right) v_j \quad (10)$$

Adopting scaled dot product score, the output

$$H = V\, softmax\left(\frac{K^T Q}{\sqrt{d}}\right). \quad (11)$$

4. experiment analysis

4.1. Data source

The names of these stocks are Microsoft Corporation (MSFT), Johnson & Johnson (JNJ), JPMorgan Chase & Co. (JPM), The Walt Disney Company (DIS) and Exxon Mobil Corporation (XOM). The data is downloaded from Yahoo (www.yahoo.com). The stock price data on Yahoo is with public availability. The data is selected from the data from July 16, 2019 to July 16, 2024, the data in one day denotes a point of the sequence, and each stock contains 1258 points. The data includes the stock's date, opening price, highest price, lowest price, closing price, adjusted close price and volume.

4.2. Experimental results

The prediction results of ARMIA with Attention-based CNN-LSTM and XGBoost hybrid model for the stock prices of MSFT, JNJ, IPM, XOM, and DIS are shown in Figures 1, 2, 3, 4, and 5, respectively.

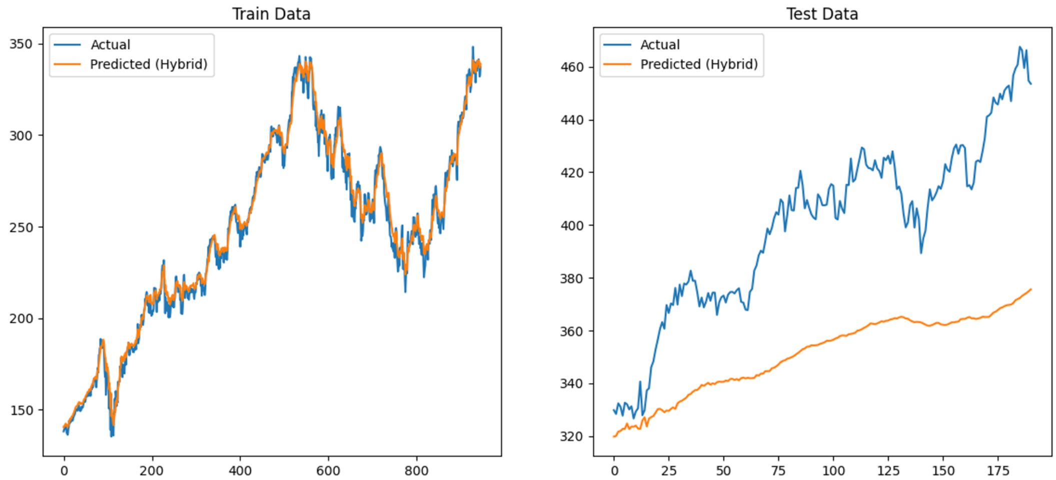

Fig.1. The stock price prediction results of MSFT

 https://doi.org/10.1051/shsconf/202419602001

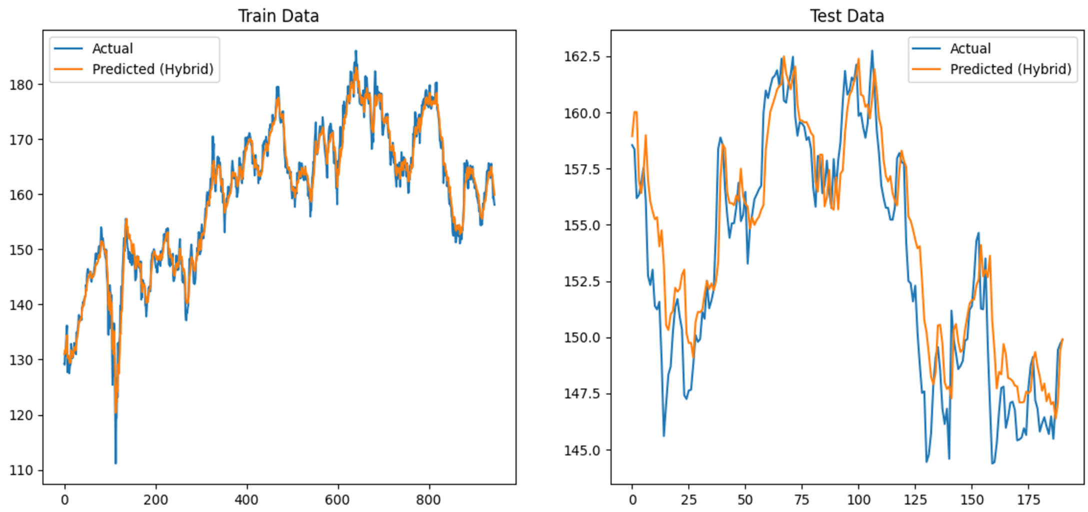

Fig.2. The stock price prediction results of JNJ

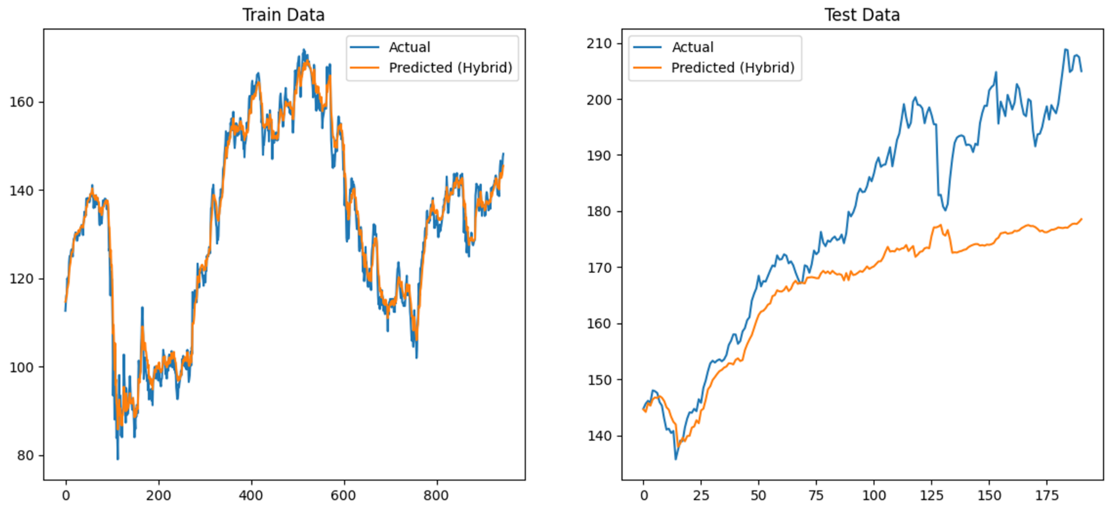

Fig.3. The stock price prediction results of JPM

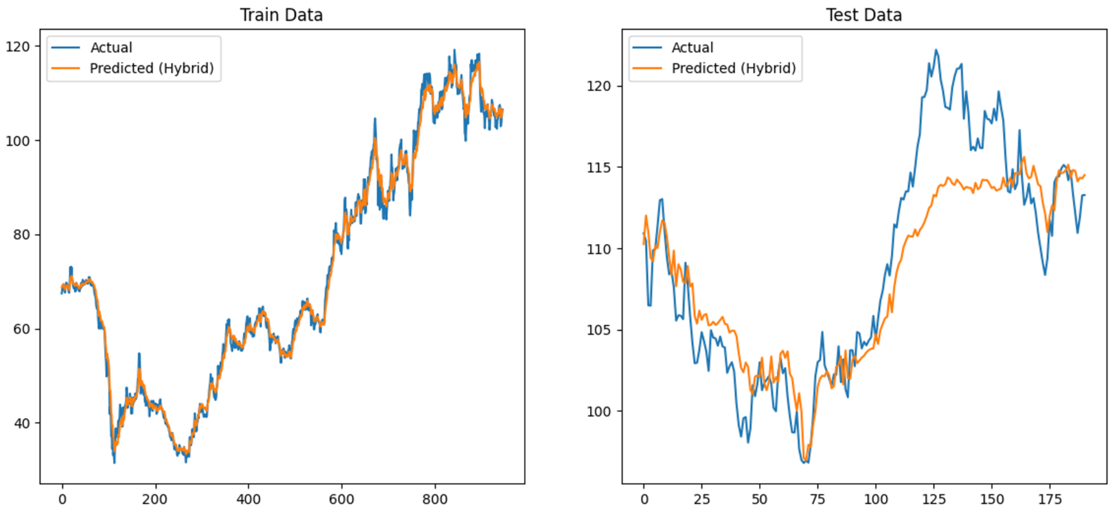

Fig.4. The stock price prediction results of XOM

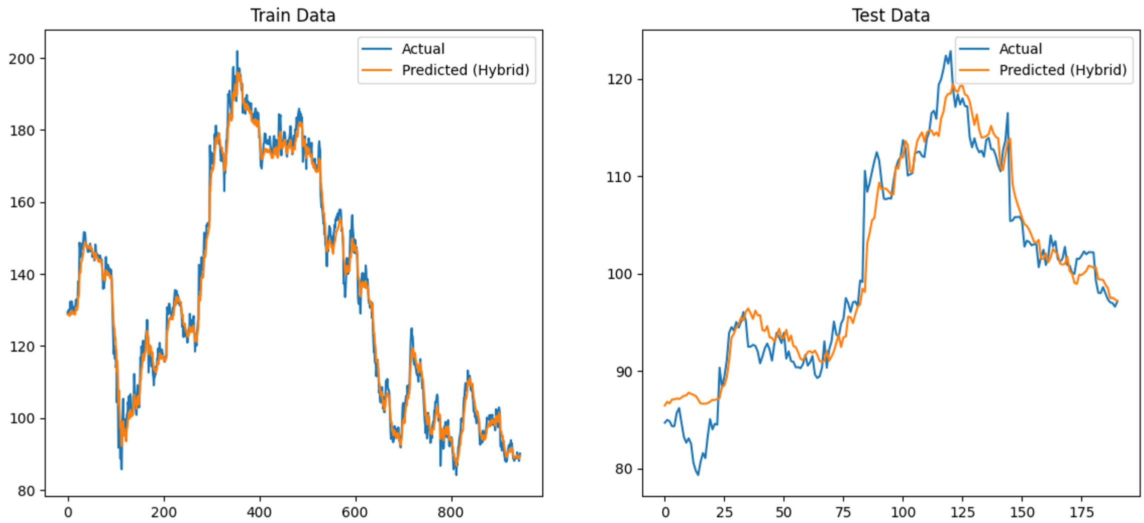

Fig.5. The stock price prediction results of DIS

Based on the observation of the images of the above experimental results, it can be concluded that ARIMA with Attention-based CNN-LSTM and XGBoost hybrid model has a good overall prediction effect during the training period. However, during the test period, the hybrid model has a poor prediction effect on stocks with relatively stable growth, such as MSFT and JPM. And it has a better prediction effect on stocks with obvious price fluctuations, such as JNJ and DIS. This shows that the hybrid model has good predictive ability when dealing with stocks with large market fluctuations, but it still needs further optimization for stable stocks.

The evaluation metrics are mean absolute error (MAE) and mean squared error (MSE).

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\hat{X}_t - X_t\right| \quad (12)$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(\hat{X}_t - X_t\right)^2 \quad (13)$$

Mean Squared Error (MSE) and Mean Absolute Error (MAE) are two common metrics used to evaluate the performance of predictive models, particularly in regression tasks such as stock price prediction. MSE is the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value. MAE is the average of the absolute errors—that is, the average absolute difference between the estimated values and the actual value.

Here, $\widehat{X}_t$ is the actual value, X_t is the predicted value, and n is the number of observations.

The training and testing indicators of five different types of stocks are compared, as shown in Table 1 and Table 2.

Table.1. Comparison of training indicators

Stock Name	MSE	MAE
MSFT	16.97	3.25
JNJ	3.12	1.38
JPM	5.35	1.78
XOM	3.26	1.37
DIS	7.94	2.21

Table.2. Comparison of testing indicators

Stock Name	MSE	MAE
MSFT	2826.46	49.33
JNJ	5.72	1.84
JPM	262.28	13.11
XOM	13.02	2.90
DIS	7.16	1.99

From the training data shown in Table 1, the training MSE and MAE values for all stocks are relatively low, such as JNJ has the lowest MSE (3.12) and MAE (1.38), indicating that the model fits the training data very well. However, during testing, the MSE and MAE values increase significantly, especially for MSFT and JPM. The MSE of MSFT surged from 16.97 to 2826.46 during training, and its MAE increased from 3.25 to 49.33, which indicating a large gap between training and testing performance. Stocks such as JNJ and DIS show relatively consistent MSE and MAE values between training and testing. For example, the MSE and MAE of DIS during training are 7.94 and 2.21, and during testing they are 7.16 and 1.99, indicating that the model is robust in predicting these stocks.

In summary, the hybrid model has better prediction accuracy for JNJ and DIS stocks compared to other stocks such as MSFT and JPM.

5. Conclusions

This paper experimentally verifies the prediction effect of ARMIA with Attention-based CNN-LSTM and XGBoost hybrid model in the US stock market. Experimental results show that the mean square error (MSE) and mean absolute error (MAE) of the model on the training data are low, and the model fitting effect is good. However, the error on the test data increase significantly, especially the prediction effect for Microsoft (MSFT) and JPMorgan Chase (JPM) stocks is poor, while the prediction effect for Johnson & Johnson (JNJ) and Disney (DIS) stocks is better. This shows that the hybrid model has good robustness when dealing with stocks with large price fluctuations, but still needs improvement on stocks with relatively stable prices. Overall, ARMIA with Attention-based CNN-LSTM and XGBoost hybrid model shows good potential in stock price prediction, especially in the case of large market fluctuations, and can be used as an effective tool for institution or investor to avoid risks and obtain returns. Since the hybrid model has a poor prediction effect on stocks with relatively stable stock prices, in future learning and research, the model can be further optimized for these relatively stable stocks to improve the robustness and prediction accuracy of the model. In addition, the experimental data in this article only select stocks of five different types of American companies, the sample size is small, so it fails to fully verify the universality and stability of the model. Besides, this paper lacks comparative experiments with other advanced prediction models, and it is impossible to fully evaluate the advantages and disadvantages of the hybrid model. Therefore, in future research, I will further explore the differences between this model and other prediction models and further improve the accuracy of the prediction of this model.

References

1. P. Pai and C. Lin, “A hybrid ARIMA and support vector machines model in stock price prediction”, Omega, vol.33, pp. 497-505, 2005.
2. J.J. Wang, J.Z. Wang, Z.G. Zhang and S.P. Guo, “Stock index forecasting based on a hybrid model”, Omega, vol.40, pp.758-766, 2012.
3. L. Y. Wei, “A hybrid model based on ANFIS and adaptive expectation genetic algorithm to forecast TAIEX”, Economic Modelling, vol.33, pp. 893-899, 2013.
4. J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques”, Expert Systems with Applications, vol. 42, pp. 259-268, 2015.
5. R. Singh and S. Srivastava, “Stock prediction using deep learning”, Multimedia Tools and Applications, vol. 76, pp. 18569-18584, 2017.
6. G. E. P. Box and G. M. Jenkins, Time Series Analysis, Forecasting and Control. San Francisco, CA, USA: Holden-Day, 1976
7. G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159-175, 2003.
8. Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," Neurocomputing, vol. 452, pp. 48-62, 2021.
9. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in NeurIPS, 2017.
10. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.
11. C. Jin, J. Gao, Z. Shi, and H. Zhang, "Attcry: Attention-based neural network model for protein crystallization prediction," Neurocomputing, vol. 463, pp. 265–274, 2021.
12. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
13. J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proceedings of the National Academy of Sciences, vol. 79, no. 8, pp. 2554-2558, 1982.
14. L. Mou, C. Zhou, P. Zhao, B. Nakisa, M. N. Rastgoo, R. Jain, and W. Gao, "Driver stress detection via multimodal fusion using attention-based CNN-LSTM," Expert Systems with Applications, vol. 173, p. 114693, 2021.
15. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in NeurIPS, 2017.
16. I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in NeurIPS, 2014.
17. Z. Shi, Y. Hu, G. Mo, and J. Wu, "Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction," arXiv preprint arXiv:2204.02623, 2022.

18. T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.
19. C. Jin, Z. Shi, K. Lin, and H. Zhang, “Predicting mirna-disease association based on neural inductive matrix completion with graph autoencoders and self-attention mechanism,” Biomolecules, vol. 12, no. 1, p. 64, 2022.
20. C. Jin, Z. Shi, W. Li, and Y. Guo, “Bidirectional lstm-crf attention-based model for chinese word segmentation,” arXiv preprint arXiv:2105.09681, 2021.
21. S. Wiseman and A. M. Rush, “Sequence-to-sequence learning as beam-search optimization,” in EMNLP, 2016
22. L. Le and Y. Xie, "Recurrent embedding kernel for predicting stock daily direction," in 2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT), A. Sill and J. Spillner, Eds., 2018, pp. 160-166.