Stock Market Movement Prediction: A Comparative Study Between Machine Learning and Deep Time Series Models

Leheng Sheng[✉]

Southeast University, Nanjing 211189, China
213182080@seu.edu.cn

Abstract. The stock market is one of the most important financial markets. It can not only provide financial support for listed companies, but also enable common shareholders to obtain profits through market transactions. In the literature, time series model, machine learning model and deep learning model are used for stock prediction. Because of the development of artificial intelligence, more and more advanced time series models have been proposed and achieved good results on a series of problems. However, whether these models can work on stock market movement prediction has not been studied. In this paper, based on the stock index data of the United States and China, we try to compare and predict the movement by two machine learning models and four deep learning models for different horizon lengths. Our results show that deep learning models are more prone to over fitting than SVM and XGBoost models, and the result is not as good as traditional machine learning model.

Keywords: Stock market movement prediction · Machine learning · Deep learning · Time series models

1 Introduction

People can make profits through financial transactions, so there are various trading markets around the world, such as stock market, futures trading and foreign exchange trading market. Among these financial markets, the stock market is the most widely accepted. Taking China and the United States as examples, there are Shenzhen Stock Exchange, Shanghai Stock Exchange and other trading markets in China. And the total market value of China's stock market has exceeded 11 trillion US dollars in 2020. Similarly, there are several stock markets in US. As China is one of the most representative developing countries and the United States is a typical developed country, we choose the stock markets of these two countries to study in this paper.

There are two opinions about whether the stock price can be predicted. One is the proponent of the Efficient Market Hypothesis (EMH). This hypothesis holds that if a stock market is highly regulated, transparent and competitive, the stock price trend will reflect all valuable information timely and accurately, such as the current and future value of the company. When EMH is established, there is no way to forecast the stock price of tomorrow by analyzing the past data. Another view is the opposition to the

EMH. Because of the factors of insider trading and false trading, the stock market in real world is not effective. At the same time, with the rise of mobile Internet, people can collect more and more data as input features. Different input characteristics, including historical price, technical indicators, emotional indicators of stock evaluation text, are used to achieve higher accuracy in predicting the trend.

In this study, we mainly rely on historical price to forecast the market index movement 1 day, 5 day and 10 days later. We also calculate technical indicators as input features. Based on the stock index data of the United States and China, we try to compare the movement prediction of two machine learning models and four deep learning models for different horizon lengths. Our results show that deep learning models are more prone to over fitting than SVM and XGBoost models, and the result is not as good as traditional machine learning model.

This paper is arranged as follows. In Sect. 2, previous studies concerning to this paper are discussed. In Sect. 3, we give the dataset and problem description. In Sect. 4, we give the feature engineering as well as the prediction models. In Sect. 5, we give the implementation details as well as the results. In Sect. 6, we give the conclusion.

2   Related Works

The financial market is a complex system. Traditional econometric models, although highly theoretical and explanatory, often fail to perform as well as novel machine learning models. In recent years, artificial intelligence has been more and more widely used in the financial field and showed outstanding performance.

The knowledge of experts is used for financial prediction in Misuk [1], which is incorporated in a data mining scheme with a focus on model interpretability and prediction metrics.

Wei Chen et al. [2] propose a method for trend forecasting with convolution neural network (CNN) model based on graph convolution feature. And this model achieves an accuracy over 51% on the dataset of China stock market.

Yi et al. [3] propose a new hybrid model based on long short term memory (LSTM) and particle swarm optimization (IPSO). They use adaptive mutation factor in this work as a parameter in model optimization. This technology keep model away from converging to local optimum too early. And the R^2 of this model on Australian stock market exceeds 0.94.

Sezer et al. [4] propose CNN-TA using a two-dimensional CNN based on image processing characteristics. The time series are transformed into two-dimensional images and processed by CNN. The accuracy of their model on test data (Dow-30) is 58%.

Thibaut et al. [5] use the method of deep reinforcement learning (DRL) to build a novel trading model. This algorithmic trading system works efficiently. The Sharpe Ratio of this strategy on Apple stock is 1.484 which is higher than B&H, S&H and TF methods.

Althelaya et al. [6] study how the combination of deep learning technology and multi-resolution analysis improves the prediction accuracy. And their method is evaluated with the S&P500 index.

Yldrm et al. [7] merge two LSTM models together to predict the direction of the EUR and USD currency pair. One LSTM is based on the data of financial factors, while the other is based on technical indicators. Their hybrid model achieves an accuracy of about 73.09% on the prediction of EUR and USD currency pair.

Adesola and Michael [8] use machine learning to predict the trend reverse. They apply this model on data from 20 foreign exchange market and find out this model is widely applicable.

Yilin et al. [9] study how RF, SVR, LSTM, DMLP and CNN advance the mean-variance and omega models. They find out RF + MVF has the best result and SVR based omega model performs best among OF models.

Sarbjit et al. [10] use a soft computing model of discrete wavelet transform, wavelet denoising and autoregressive model to predict the weekly closing price and daily closing price of the BSE100 S&P Sensex index.

Pooja et al. [11] propose a fusion supervision framework SDCF. The framework relies on a novel CTL model. The experiment is based on dataset from several markets and this novel model achieves an average F1 score of 0.62 on 15 Indian stocks. This result is better than 1-D CNN approach.

Gite et al. [12] propose an efficient machine learning technology using LSTM and XAI to forecast the movement of stock market and explain why AI would make this choice. They settle the model on the data of National Stock Exchange and news headlines. They not only give the prediction of stock movement, but also present the explanation that why AI will make this prediction by using XAI.

More related work can be referred to some recent reviews [13, 14].

3   Dataset and Problem Description

3.1   Dataset

The stock indices we use cover both the US and China markets, with the following four indices:

1. SSE Composite: It is a stock market index which contains all stocks trading at the Shanghai Stock Exchange, including A shares and B shares traded in the China market. The currency of this index is CNY.
2. S&P 500: It contains 500 large companies listed in the US market. The company list may change and may not necessarily have the exact number of 500. As one of the most commonly used equity indices, S&P 500 has been widely used in previous studies. The currency of this index is USD.
3. NASDAQ Composite: It is an index containing all stocks listed in a specific US stock market, i.e., the Nasdaq stock market. The currency of this index is USD.
4. NYSE Composite: It is also a stock market index of all stocks listed in a specific US market, i.e., the New York market. The currency of this index is USD.

The original daily data we get from Yahoo Finance include the following fields: the open/high/low/close prices, the adjusted close price, and the trading volume. The adjusted prices reflect the stock's value after considering the corporate actions,

including the stock's dividends, stock splits and new stock offerings. We adjust the open/high/low prices accordingly, e.g., the adjusted open price = open price * (adjusted close price/close price). In the following parts, we use the adjusted prices unless specifically noted.

We obtain the daily data for more than ten years, from Jan 1, 2009 to Dec 12, 2020. We show the market situations of the four indices in Fig. 1, 2, 3 and 4.
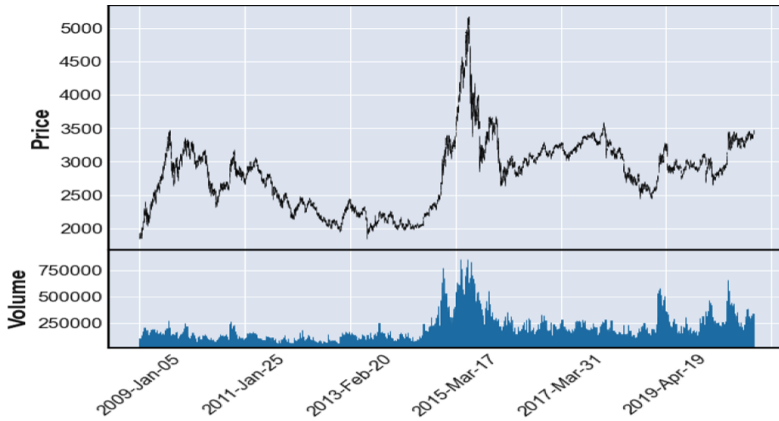

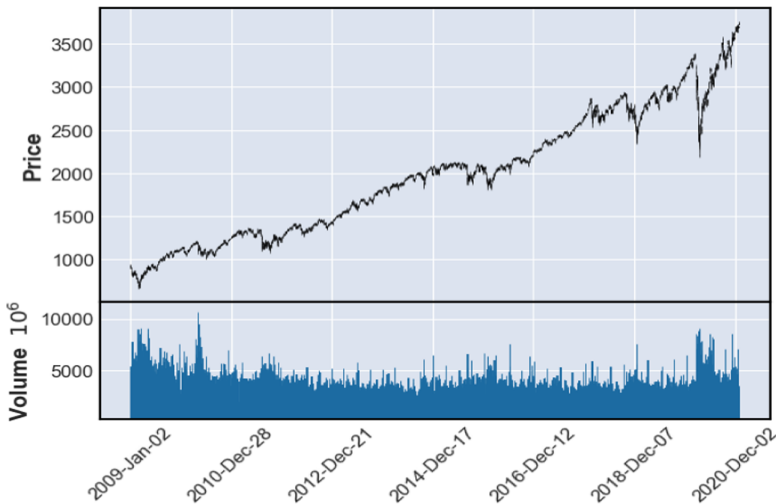**Fig. 1.** The market situation of SSE Composite.

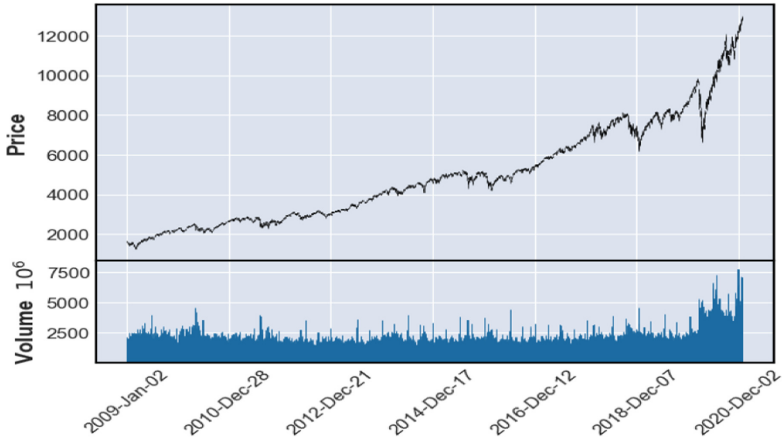
Fig. 2. The market situation of S&P 500.

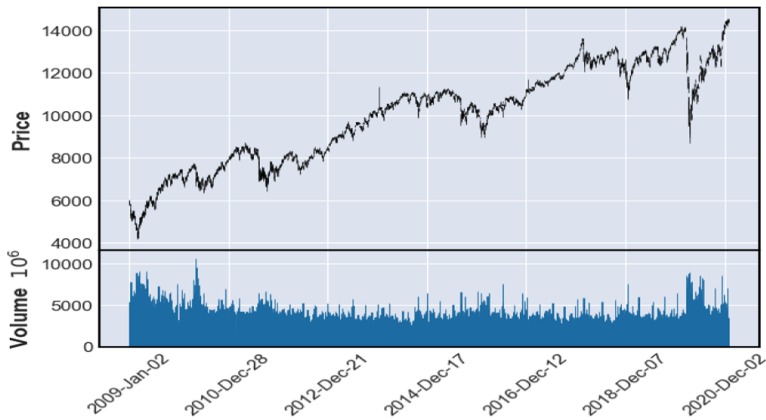**Fig. 3.** The market situation of NASDAQ Composite.

Fig. 4. The market situation of NYSE Composite.

Following a standard machine learning scheme, the raw data is divided into three subsets, namely, training, validation and testing sets. The data from 2009 to 2016 is for training models. The data from 2017 to 2018 is the validation set. The data from 2019 to 2020 is the testing set. The models would be trained on the training set, the hyper parameters would be chosen according to the validation set, and the final performance would be evaluated in the testing set in the following parts.

3.2 Problem

According to the future adjusted close price relative to today's change, we label 0 or 1 distribution to represent the decline and rise. Therefore, we can define a classification problem. We use three different horizons, i.e., 1 day, 5 days and 10 days to label the

data. For the label on the day t with horizon of n days, the specific formula to calculate the label y_t is as follows.

$$y_t = \begin{cases} 1, & C_{t+n} > C_t \\ 0, & otherwise \end{cases}$$

4   Models

4.1   Feature Engineering

Technical indicators are mainly calculated from the data of stock price, trading volume or price index. In this paper, we use 10 popular technical indicators as follows.

Table 1. Selected technical indicators.

Name	Formula
Simple n-day moving average (SMA)	$\frac{C_t + C_{t-1} + ... + C_{t-n-1}}{n}$
Weighted n-day moving average (WMA)	$\frac{nC_t + (n-1)C_{t-1} + ... + C_{t-n-1}}{n + (n-1) + ... + 1}$
Momentum	$C_t - C_{t-(n-1)}$
Stochastic K%	$\frac{C_t - LL_{t-(n-1)}}{HH_{t-(n-1)} - LL_{t-(n-1)}} \times 100$
Stochastic D%	$\frac{\sum_{i=0}^{n-1} K_{t-i}}{10}\%$
Relative strength index (RSI)	$100 - \frac{100}{1 + \left(\sum_{i=0}^{n-1} UP_{t-i}/n\right)/\left(\sum_{i=0}^{n-1} DW_{t-i}/n\right)}$
Moving average convergence divergence (MACD)	$MACD(n)_{t-1} + \frac{2}{n+1} \times (DIFF_t - MACD(n)_{t-1})$
Larry William's R%	$\frac{H_n - C_t}{H_n - L_n} \times 100$
Accumulation/distribution (A/D) oscillator	$\frac{H_t - C_{t-1}}{H_t - L_t}$
Commodity channel index (CCI)	$\frac{M_t - SM_t}{0.015 D_t}$

C_t is the closing price of day t, L_t is the low price and L_t is the high price at time t, LL_t and HH_t are lowest low and highest high price in the last t days, respectively. UP_t means upward price change while DW_t is the downward price change at time t.

Some other indicators are calculated as follows:

$$DIFF_t = EMA(12)_t - EMA(26)_t$$

$$M_t = \frac{H_t + L_t + C_t}{3}$$

$$SM_t = \frac{\sum_{i=1}^{n} M_{t-i+1}}{n}$$

$$D_t = \frac{\sum_{i=1}^{n} |M_{t-i+1} - SM_t|}{n}$$

4.2   Machine Learning Models

We adopt two machine learning models which are also often considered in previous studies, namely, Support Vector Machine (SVM) and eXtreme Gradient Boosting (XGBoost) [15].

SVM is a widely used supervised machine learning, which is suitable for the classification problems we are dealing with. A hyperplane with the largest amount of margin is built iteratively by SVM, which aims to separate data of different types. SVM has the advantages of both the elegant mathematical formulation as well as the theoretical performance guarantee.

XGBoost is an ensemble machine learning model that is based on the framework of gradient boosting. The weak learner used as the base estimator is usually the decision tree, which is prone to overfitting. By boosting individual tree models, all the trees are built sequentially with the aim to reduce the errors in previous ones. XGBoost also makes many engineering improvements and algorithmic optimizations than the standard boosting process.

4.3   Deep Learning Models

We adopt four models, namely, Gated Recurrent Units (GRU) [16], Long Short Term Memory (LSTM) [17], Temporal Convolutional Network (TCN) [18], and Time Series Transformer (TST) [19].

GRU and LSTM are both RNN (Recurrent Neural Network) variants, which are designed for time series. RNNs allow previous outputs to be used as inputs, so that the information from previous time steps to be kept. While the idea is simple, RNNs is troubled by the problems of the vanishing gradient problem, which makes training the neural network weights impossible. To solve the problem of vanishing gradient, GRU uses the two gates to control what information to be passed to the output, i.e., update gate and reset gate. LSTM is similar to GRU, by introducing three gates to control how much and which information to retain.

TCN extends the parallel processing ability of convolutions in the image processing field to the temporal area. CNN has been extremely successful in the image processing problems, which are usually two dimensional. By reducing the convolution operation to the one dimensional time series, TCN manages to be applied to the time series problems and has been proven effective in time series classification and prediction, while retaining the parallel processing ability of convolutions.

TST extends the successful structure of Transformer in the NLP (Natural Language Processing) field to the problems of time series. This is the first time that TST is applied to both US and China markets in the literature. Transformer abandons the common structures of RNNs and CNNs. Instead, Transformer is fully based on the attention mechanism. At first, this specific structure is used for NLP and contributes many successful language models. Then the idea is extended to all sequential data, including time series, by taking these data as the languages. This extension has been proven effective unexpectedly, even in the visual field by taking image patches from a big picture and organize them in a sequential manner.

5 Results

5.1 Settings

We used Tsai[1] as the programming platform. Tsai is an open-source deep learning package based on Pytorch[2] & fastai.[3] This framework focuses on most advanced techniques for time series classification, regression and forecasting. Also, we use the package of hyperopt[4] to search for the best hyperparameter.

We use the accuracy and weighted F1 score for evaluation metrics.

Accuracy is the proportion of the correct sample number to the total sample number. However, it is not a good evaluation index when the model is overfitting. It is possible to have the highest accuracy when the model is predicted to be 1 or 0. As the harmonic average of precision and recall, F1 score is widely used as the final evaluation method in many machine learning competitions. Therefore, we also introduce F1 score as the evaluation index. Also, we draw the confusion matrix. Confusion matrix is a visual method for comparing the true and predicted labels. The diagonal elements are the sample numbers that are correctly classified.

5.2 Results

We first collate the results of different models on 1-day, 5-day and 10 day scales. The accuracies and F1 scores of the models we compare are presented in Table 1, 2, 3, and 4 (Table 5).

---

[1] https://timeseriesai.github.io/tsai/.

[2] https://pytorch.org/.

[3] https://docs.fast.ai/.

[4] http://hyperopt.github.io/hyperopt/.

Table 2. The accuracies and F1 scores for SSE Composite.

Model	1-day movement		5-days movement		10-days movement	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
SVM	0.530	**0.530**	0.553	**0.554**	0.579	**0.586**
XGBoost	0.520	0.520	0.538	**0.540**	0.588	**0.586**
GRU	0.524	0.451	0.578	0.423	0.627	0.483
LSTM	0.527	0.390	0.577	0.422	0.627	0.483
TCN	0.534	0.397	0.577	0.423	0.627	0.483
TST	0.507	0.410	0.542	0.469	0.594	0.582

Table 3. The accuracies and F1 scores for S&P 500.

Model	1-day movement		5-days movement		10-days movement	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
SVM	0.477	0.480	0.531	0.521	0.578	0.588
XGBoost	0.519	0.520	0.669	**0.537**	0.685	**0.641**
GRU	0.582	0.428	0.669	**0.537**	0.717	0.607
LSTM	0.582	0.428	0.669	**0.537**	0.727	0.612
TCN	0.582	0.428	0.669	**0.537**	0.727	0.612
TST	0.544	**0.540**	0.451	0.448	0.455	0.464

Table 4. The accuracies and F1 scores for NASDAQ Composite.

Model	1-day movement		5-days movement		10-days movement	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
SVM	0.523	**0.509**	0.521	0.521	0.545	0.560
XGBoost	0.489	0.494	0.642	**0.539**	0.760	0.660
GRU	0.600	0.478	0.665	0.532	0.764	**0.662**
LSTM	0.598	0.448	0.665	0.532	0.764	**0.662**
TCN	0.598	0.448	0.665	0.532	0.746	0.660
TST	0.598	0.448	0.661	0.533	0.764	**0.662**

From Table 1, although the accuracy of deep learning method on some data sets is higher than that of traditional machine learning method, the F1 score is often lower than that of traditional machine learning method. This is because the deep learning model is too complex, in many cases, there is a phenomenon of overfitting, that is, for any data input data on any test set, the output of the deep learning model is 1. In contrast, the over fitting problem of machine learning is not so serious. Specifically, we draw the confusion matrix for each machine learning model. It is clear to see that in most cases, the machine learning model does not only predict 1 or 0 (Figs. 5, 6, 7, 8, 9, 10, 11 and 12).

Table 5. The accuracies and F1 scores for NYSE Composite

Model	1-day movement		5-days movement		10-days movement	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
SVM	0.521	0.521	0.487	0.497	0.521	0.534
XGBoost	0.525	**0.526**	0.537	0.547	0.584	**0.591**
GRU	0.556	0.390	0.650	**0.551**	0.669	0.537
LSTM	0.556	0.397	0.670	0.511	0.669	0.537
TCN	0.556	0.390	0.650	**0.551**	0.669	0.537
TST	0.556	0.390	0.648	0.518	0.669	0.537
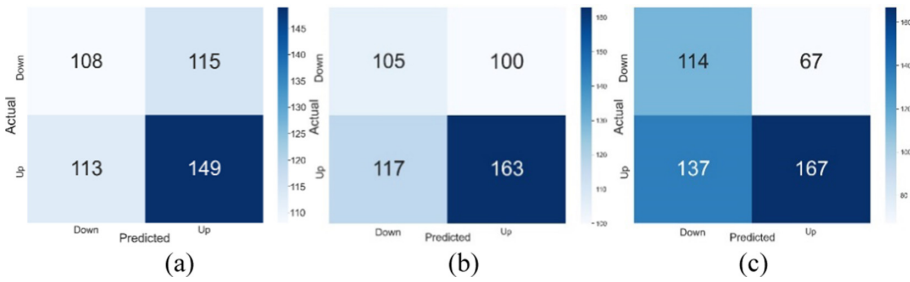
Fig. 5. Confusion matrices of SVM for SSE Composite for 1, 5, and 10 days.
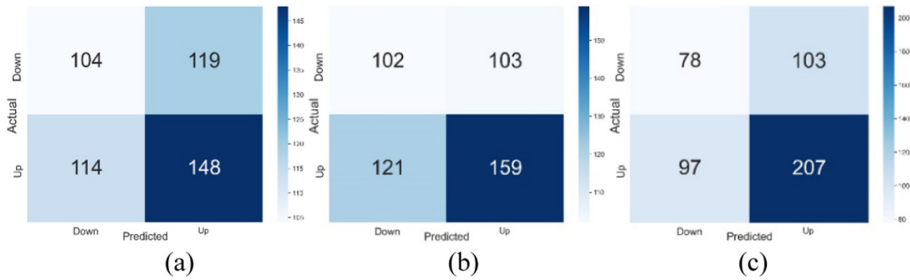
Fig. 6. Confusion matrices of XGBoost for SSE Composite for 1, 5, and 10 days.
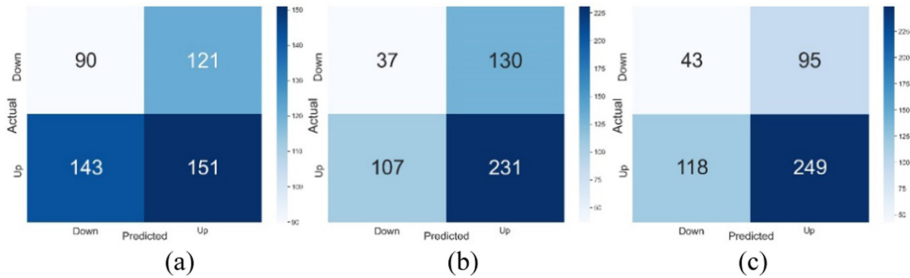
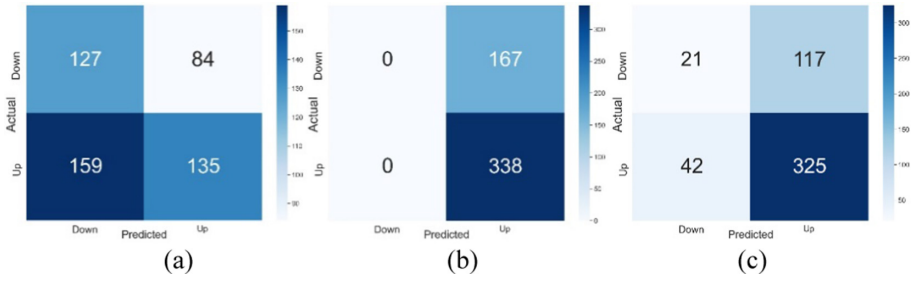**Fig. 7.** Confusion matrices of SVM for S&P 500 for 1, 5, and 10 days.

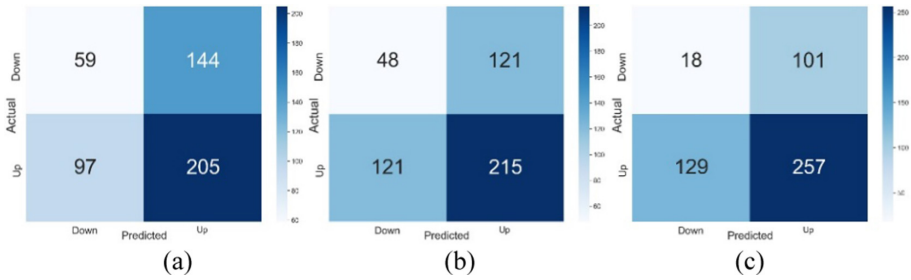**Fig. 8.** Confusion matrices of XGBoost for S&P 500 for 1, 5, and 10 days.

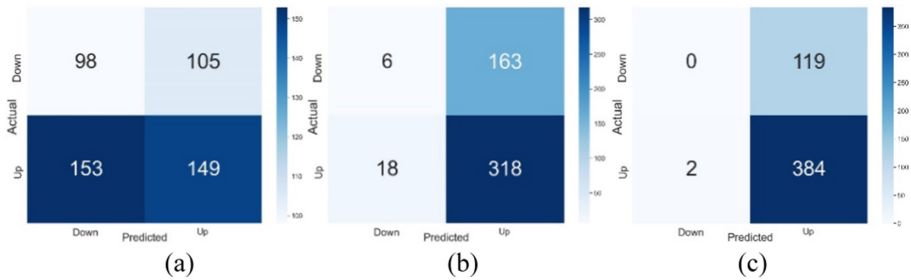**Fig. 9.** Confusion matrices of SVM for NASDAQ Composite for 1, 5, and 10 days.

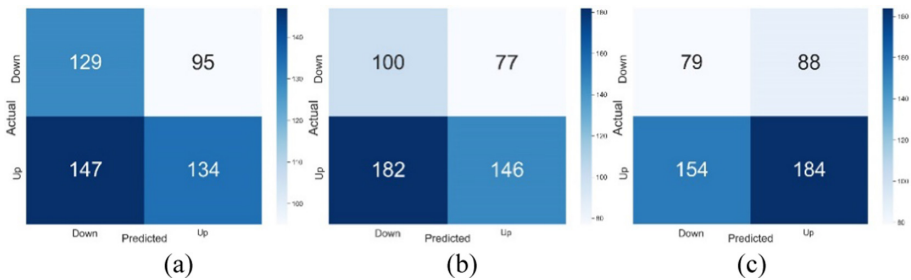**Fig. 10.** Confusion matrices of XGBoost for NASDAQ Composite for 1, 5, and 10 days.

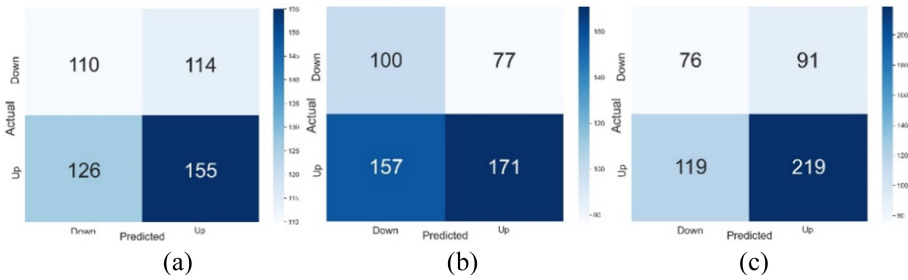**Fig. 11.** Confusion matrices of SVM for NYSE Composite for 1, 5, and 10 days.

Fig. 12. Confusion matrices of XGBoost for NYSE Composite for 1, 5, and 10 days.

6   Conclusion

We use four deep time series methods and two traditional machine learning methods to build the stock index movement prediction model. The paper also studies the market fluctuation after 1 day, 5 days and 10 days based on the financial time series data of China SSE, S & P500, NASDAQ and NYSE. The results show that the traditional machine learning method tend to beat the deep time series method. Financial market is an extremely complex system, which is more complex and changeable than image prediction. So it is very important to choose input features. At this stage, the work only stays in the analysis of historical data. In future work, we will consider adding more input features, such as sentiment factors and market fundamentals, to improve the accuracy further.

References

1. Kim, M.: A data mining framework for financial prediction. Expert Syst. Appl. **173**, 114651 (2021). https://doi.org/10.1016/j.eswa.2021.114651
2. Chen, W., Jiang, M., Zhang, W.-G., Chen, Z.: A novel graph convolutional feature based convolutional neural network for stock trend prediction. Inf. Sci. **556**, 67–94 (2021). https://doi.org/10.1016/j.ins.2020.12.068
3. Ji, Y., Liew, A.W.-C., Yang, L.: A novel improved particle swarm optimization with long-short term memory hybrid model for stock indices forecast. IEEE Access **9**, 23660–23671 (2021). https://doi.org/10.1109/access.2021.3056713
4. Sezer, O.B., Ozbayoglu, A.M.: Algorithmic financial trading with deep convolutional neural networks: time series to image conversion approach. Appl. Soft Comput. **70**, 525–538 (2018). https://doi.org/10.1016/j.asoc.2018.04.024
5. Théate, T., Ernst, D.: An application of deep reinforcement learning to algorithmic trading. Expert Syst. Appl. **173**, 114632 (2021). https://doi.org/10.1016/j.eswa.2021.114632
6. Althelaya, K.A., Mohammed, S.A., El-Alfy, E.S.M.: Combining deep learning and multiresolution analysis for stock market forecasting. IEEE Access **9**, 13099–13111 (2021). https://doi.org/10.1109/access.2021.3051872
7. Yıldırım, D.C., Toroslu, I.H., Fiore, U.: Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. Financ. Innov. **7**(1), 1–36 (2021). https://doi.org/10.1186/s40854-020-00220-2

8. Adegboye, A., Kampouridis, M.: Machine learning classification and regression models for predicting directional changes trend reversal in FX markets. Expert Syst. Appl. **173**, 114645 (2021). https://doi.org/10.1016/j.eswa.2021.114645

9. Ma, Y., Han, R., Wang, W.: Portfolio optimization with return prediction using deep learning and machine learning. Expert Syst. Appl. **165**, 113973 (2021). https://doi.org/10.1016/j.eswa.2020.113973

10. Singh, S., Parmar, K.S., Kumar, J.: Soft computing model coupled with statistical models to estimate future of stock market. Neural Comput. Appl. **33**(13), 7629–7647 (2021). https://doi.org/10.1007/s00521-020-05506-1

11. Gupta, P., Majumdar, A., Chouzenoux, E., Chierchia, G.: SuperDeConFuse: a supervised deep convolutional transform based fusion framework for financial trading systems. Expert Syst. Appl. **169**, 114206 (2021). https://doi.org/10.1016/j.eswa.2020.114206

12. Shilpa, G., Hrituja, K., Priyam, M., Ketan, K., Shilpi, S., Neerav, P.: Explainable stock prices prediction from financial news articles using sentiment analysis. PeerJ Comput. Sci. **7**, e340 (2021). https://doi.org/10.7717/peerj-cs.340

13. Jiang, W.: Applications of deep learning in stock market prediction: recent progress. arXiv preprint arXiv:2003.01859 (2020)

14. Thakkar, A., Chaudhari, K.: A comprehensive survey on deep neural networks for stock market: the need, challenges, and future directions. Expert Syst. Appl. **177**, 114800 (2021). https://doi.org/10.1016/j.eswa.2021.114800

15. Chen, T., Guestrin, C.: 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)

16. Cho, K., Van Merriënboer, B., Gulcehre, C., et al.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)

17. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

18. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 (2018)

19. Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., Eickhoff, C.: A transformer-based framework for multivariate time series representation learning. arXiv preprint arXiv:2010.02803v2 (2020)