Research on the S&P 500 Index Based on LSTM and GRU

Zexu Feng
Business School, Central University of Finance and Economics, Beijing, 100098, China
* Corresponding Author Email: 2017311337@email.cufe.edu.cn

Abstract. Forecasts regarding the prices of stock indices are commonly observed in both academic and corporate circles. Exponential prices are difficult to predict because of uncertain noise. Due to advancements in computer science, neural network has been applied in various industrial fields. This article examines three models, Long Short Term Memory (LSTM) networks and Gate Recurrent Unit (GRU), and their combination. At the same time, it's also investigated that how the performance of the models will change when incorporating technical indicators such as the weighted moving average and Bollinger bands and macroeconomic indicators which include CBOE Volatility Index and Effective Federal Funds Rate into analysis. The experimental findings indicate that LSTM performs best in the experiment of predicting S&P500 index, and adding indicators to the model does not make the model more effective. This article provides more perspectives on the application of machine learning in stock price prediction, and provides relevant theoretical guidance for the selection of variables in stock price prediction practice.

Keywords: Long Short-Term Memory, Gate Recurrent Unit, stock price prediction, machine learning.

1. Introduction

It's known that the stock market can accurately reflect the state of the national economy and is essential to the nation's economic growth. The stock market is one of the most prominent avenues for financial investment among investors. By purchasing and selling stocks at optimal periods, investors on the stock market can maximize their profits. The primary determinant of achieving substantial returns in the stock market is the capacity to forecast the price trajectory of equities and execute trades at the most opportune moment while minimizing trading risk. Nevertheless, the prices of stocks are susceptible to a multitude of variables, including the economic environment, policy trends, public opinion, industry conditions, and enterprise operating conditions, and are therefore highly unpredictable. Some individuals believe that stock prices are unpredictable, while others believe they are predictable. Hence, the task of accurately forecasting stock prices poses a significant challenge for researchers, leading to extensive discourse regarding the predictability of stock prices. Leaving aside the fact that profits have been falling in recent years, one study posed a serious challenge to semi-strong forms of market efficiency. All the machine learning approaches considered in the study were able to beat the benchmark buy-and-hold strategy after accounting for 0.1% of transaction costs [1]. There have been multiple endeavours to forecast stock prices, and numerous advanced theories and techniques have been devised by scholars from various nations. Initial research methods are mainly applied statistics, such as regression analysis and differential analysis, as well as securities investment analysis techniques such as technical analysis, fundamental analysis, and analysis of related policies. However, the data in the financial market is huge, the data update speed is fast, and it has the characteristics of big data. These statistical methods and methods for analysing securities investments have significant limitations in the treatment and analysis of such data. The recent advancements in computer performance, coupled with the swift progress of artificial intelligence, data mining, text analysis, and related technologies, have led to a surge in the need for financial data analysis. A variety of new stock price prediction methods have arisen progressively, including analysing social media information through text analysis technology to predict the stock market. In addition, prevalent machine learning techniques such as neural networks and random forests are utilized in the investigation of the financial market. The utilization of machine learning algorithms for forecasting stock prices has surfaced as a noteworthy research domain in the financial sector.

Utilizing a blend of technical indicators and computational intelligence tools is a more lucrative approach to making trading decisions, as opposed to relying solely on individual technical indicators. And Dash also built a classification model computational efficient functional link artificial neural network (CEFLANN) which performed better than other traditional models including the support vector machine (SVM), K nearest neighbour model (KNN) and decision tree (DT) model [2]. One study empirically compares the effectiveness of three machine learning methods, KNN, SVM and artificial neural network (ANN), in predicting nonlinear financial time series. And the study found that for root mean square error (RMSE), in most cases, KNN and ANN demonstrated superior performance compared to SVM [3]. By conducting a retrospective analysis of the S&P 500 index's historical data spanning the years 2014 to 2018, the present study aims to evaluate its performance, one research found that in terms of overall performance metrics, the machine learning model outperforms the benchmark market index. In addition, the empirical results after risk adjustment show that random forest has the best effect, followed by support vector machine and neural network [4]. Taking the Moving Average Convergence/Divergence (MACD) into consideration, someone has proved that the addition of machine learning techniques, such as ANN, to technical analysis strategies improves the trading performance [5]. In the context of macroeconomic analysis, the variables that exhibit the highest degree of predictive accuracy with respect to stock market recessions in the United States are term spreads and inflation rates. It was also discovered that macrovariables are superior at predicting bear markets than stock market returns [6]. By applying ANN and support vector regression (SVR) models, the paper finds that the application of fundamental indicators to predict stock prices is better than the application of technical analysis indicators, which is true in several industries. Using a mix of indicators is better than using fundamentals or technical indicators alone [7]. An attention-based (UA) model which is attention-based deep learning method performs best when predicting stock prices, comparing to multilayer perceptron (MLP) model, long short term memory (LSTM) model, convolutional neural network (CNN) model [8]. In addition, all the four models perform better than developing financial markets in developed financial markets based on the S&P 500 index. For the prediction of trading signals, a gated recurrent unit combined with convolutional neural network (GRU-CNN) model was constructed to forecast the direction of stock indexes of the Hang Seng Indexes (HSI), the Deutscher Aktienindex (DAX) and the S&P 500 Index, the model achieved an accuracy of about 56% [9]. A research study has developed a comprehensive approach to analysing the stock market's behaviour by utilizing fundamental market data, macroeconomic data, and technical indicators. The study has identified a set of nine predictors that are carefully balanced to provide a holistic view of the stock market's performance and demonstrates that the single layer LSTM model outperforms multilayer LSTM models in terms of providing a better fit and achieving higher prediction accuracy [10]. Compared to statistical methods, artificial intelligence has demonstrated superior capabilities in handling the stochastic, turbulent, and nonlinear nature of stock market information, and is widely utilized for precise forecasting of stock market indices.

This paper will examine the hot LSTM model and its related GRU model in the field of stock price forecasting, as well as the combination of the two models. On the basis of the closing price, this paper will examine whether the SP500 is more accurate when technical and macroeconomic indicators are included in the model. Among them, the technical indicators include variables such as weighted moving average (WMA), relative strength index (RSI) and average directional index (ADX), and comprehensively consider the stock price trend, transaction price and quantity. Macroeconomic indicators take into account domestic bank borrowing costs, international exchange rates and investor sentiment. This paper will further deepen the application of LSTM and other models in the field of stock price prediction and provide a certain degree of reference for similar research. This paper will additionally offer recommendations on the variable selection process within the empirical domain of forecasting stock prices.

2. Methodology

Recurrent neural network (RNN) originated from Hopefield Networks proposed by Saratha Sathasivam in 1982. In the previous CNN neural network or BP neural network, the model is successively composed of input layer, hidden layer and output layer. Each layer of the network is connected to each other, but the internal nodes of the network are not connected to each other, and the data signals are just transmitted along the network layer by layer. The hidden layer of a cyclic neural network exhibits interconnectivity among the neuron nodes of every layer within the network. This is the primary distinction between RNN neural network and cyclic neural network. In the RNN neural network, the hidden layer not only receives the input information of the input layer, while it is also affected by the input data of the hidden layer from the preceding time. Thus, the structure of the RNN neural network is capable of characterizing the correlation between the present output of a given time series and the antecedent historical data. Compared with CNN neural network, RNN neural network is more used to process and predict time series data, which is also determined by the structural characteristics of cyclic neural network.

2.1. Long short-term memory networks

LSTM network is an improvement on the topology of RNN. The LSTM model proficiently addresses the issue of gradient vanishing that arises due to the augmentation of network layers and the passage of time by introducing controllable self-loop. Its ingenious design structure is especially suitable for processing tasks with long delay and time interval. The structure of LSTM neural network is different from other deep learning algorithms because of its special neuron cell state, which can learn sequence data information that needs to be recorded and forgotten in a long-term state. The cell unit is composed of three threshold elements: forgetting gate, input gate and output gate. The diagram presented in Figure 1 illustrates the general framework.

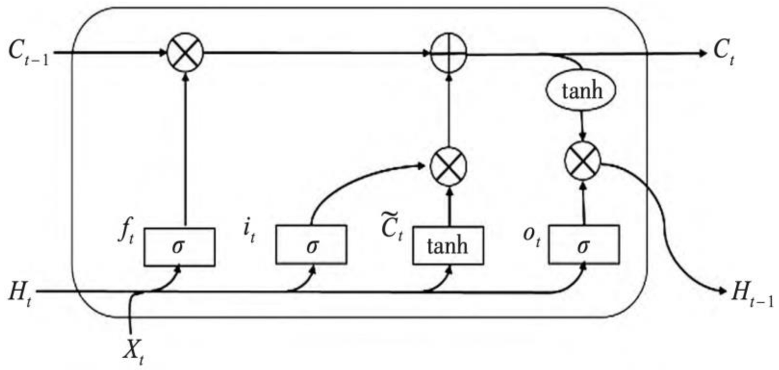

Fig. 1. LSTM unit structure diagram

2.1.1 Forget Gate

Firstly, the oblivion gate is used to determine which historical state information is discarded from the historical state. The output of the sigmoid layer is between 0 and 1. When the output of the oblivion gate is 1, it means that the historical state information in the previous step is completely retained. Otherwise, it means that all historical status information before this is cleared:

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \tag{1}$$

2.1.2 Input Gate

Input gate controls the processing degree of adding new input sequence data information into cell state. Firstly, the updating degree of information is determined according to Sigmoid function. Secondly, decide how much to update the content, and finally combine the two parts, discard the abandoned information in the forgetting door, add the updated information, and store it in the long-term state:

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \tag{2}$$

$$\tilde{C}_t = tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \tag{3}$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \tag{4}$$

2.1.3 Output Gate

LSTM structure determines the output degree of sequence information based on cell state. The sigmoid layer is responsible for determining the output component of the cell state, and the long-term state of cell is processed by tanh function and multiplied by the previous output part. Finally, the output sequence information is determined:

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \tag{5}$$

$$H_t = O_t \odot tanh(C_t) \tag{6}$$

2.2. Gate Recurrent Unit

GRUs are similar to LSTM in structure, both of which are RNN cyclic neural networks with gated units added. The difference is that the GRU merges LSTM's forget gate and output gate into an improved update gate, and further merges short-term and long-term states. Due to the reduction of parameters, GRUs converge faster in the process of network training. In some scenarios, GRUs have better performance than LSTM (Figure 2).

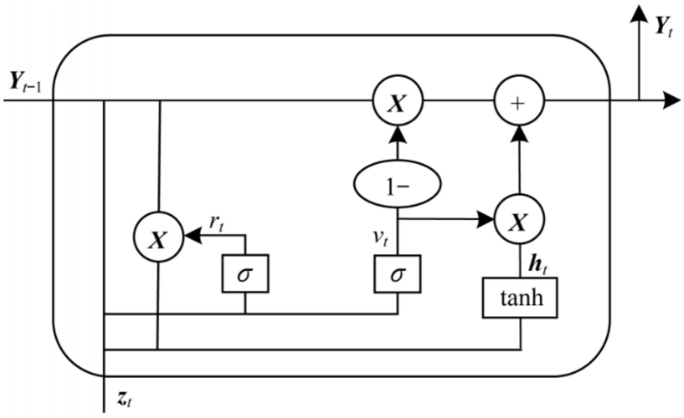

Fig. 2. GRU unit structure diagram

3. Data

In this study, the object to be predicted is the SP500 index, a prominent US stock market index. The process of feature selection involves identifying the core factors that lead to fluctuations in index values. Some features, such as fundamental data and technical indicators, are extracted directly from the underlying index. The choice of macroeconomic indicators is predicated upon their capacity to exert influence on the broader economy and related markets. Data was collected for a full 18 years, from 2006 to 2023. The chosen time frame encompasses two significant declines in markets: the 2008 financial crisis and the 2020 COVID-19 pandemic. Therefore, constructing a model that emulates the comprehensive market scenario, including bear and bull markets, may lead to better forecasts. The three categories of input features are shown in Table 1.

Table 1. List of potential features for the model.

Variables	Details
close	Daily close price
volume	Daily volume
WMA	Weighted moving average
MACD	Moving average convergence/divergence
RSI	Relative strength index
ADX	Average directional index
BBands	Bollinger Bands
OBV	On-balance volume
VIX	CBOE Volatility Index
FFER	Effective Federal Funds Rate
USDI	Nominal Broad U.S. Dollar Index
IFR	10-Year Breakeven Inflation Rate
T10Y3M	10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity

3.1. Fundamental data

The first two variables in Table 1 come directly from the trading market, which furnishes fundamental information essential for stock trading. The close price refers to the final transaction price at the conclusion of a trading day. Volume represents the volume of a full day. The entirety of the historical trading data that is obtained from Yahoo is comprised solely of daily data.

3.2. Technical indicators

The Weighted Moving Average (WMA) is a technical indicator utilized by traders to ascertain optimal buying or selling opportunities. The method assigns greater significance to the most recent data points while reducing the significance of older data points. The computation of the weighted moving average involves the multiplication of each individual data point with a pre-established weighting factor.

The MACD, also known as MAC-D, is a momentum indicator that adheres to trend-following principles. It depicts the relationship between two exponential moving averages (EMAs) pertaining to the price of a given security. The calculation of the MACD line involves the mathematical operation of subtracting the 26-period EMA from the 12-period EMA. The momentum oscillator is predominantly employed for trend trading purposes. Despite being an oscillator, it is not commonly employed for the purpose of identifying overbought or oversold market conditions.

The Relative Strength Index (RSI) evaluates overvalued or undervalued conditions in the price of a security by measuring the speed and amplitude of its recent price changes. The Relative Strength Index (RSI) is depicted as an oscillating indicator in the form of a line graph, which spans from 0 to 100. The RSI possesses a wider scope beyond the mere identification of securities that are overbought or oversold. Additionally, it could suggest securities that exhibit potential for a reversal in trend or a correction in price. It has the potential to signal optimal timing for buying and selling.

Average directional index (ADX) is used to measure the intensity of a trend. The computation of ADX is founded on a mobile average of price expansion within a designated duration of time. The default configuration has 14 bands.

BBands is the BOLL index, whose full name is “Bollinger bands”. BOLL was created by Mr. John Brin. The methodology employs statistical principles in order to compute the standard deviation and confidence interval of stock prices, with the aim of ascertaining the range of fluctuations and prospective trends in stock prices and displays the safe and low levels of stock prices using wave bands. Additionally known as the Bollinger belt. The variability of the upper and lower limits is contingent upon the current market value of rolling stock.

On Balance Volume (OBV) is a momentum technical analysis tool. The OBV index is to compile the trend line based on the volume data and cooperate with the stock price trend to judge the trend of stock price from the perspective of volume energy.

3.3. Macroeconomic data

The CBOE Volatility Index (VIX) is a mathematical computation intended to produce a quantification of consistent, anticipated 30-day instability of the American stock market. The measure in question is widely acknowledged as a significant indicator of volatility on a global level. It receives extensive coverage in financial media and is closely monitored on a daily basis. It has been demonstrated to be a reliable singular predictor [1].

The calculation of Effective Federal Funds Rate (EFFR) involves the identification of the median value of federal funds transactions conducted overnight, as reported in the FR 2420 Report of Selected Money Market Rates, with the weights of each transaction being based on their respective volumes. The EFFR for the preceding business day is published by the Federal Reserve Bank of New York at around 9:00 a.m. on its official website.

The Nominal Broad U.S. Dollar Index refers to a measure of the value of the United States dollar against a basket of other major currencies. The broad index, or trade-weighted US dollar index, is a metric that gauges the worth of the US dollar in comparison to various global currencies. The aforementioned is a weighted index that enhances the previous U.S. Dollar Index by integrating a greater number of currencies and undergoing annual rebalancing.

The breakeven inflation rate is calculated by utilizing the 10-Year Treasury Constant Maturity Securities and 10-Year Treasury Inflation-Indexed Constant Maturity Securities as a gauge for anticipated inflation. The most recent value indicates what market participants anticipate inflation to average over the next decade.

The metric known as the '10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity' is derived by computing the difference between the 10-Year Treasury Constant Maturity and the 3-Month Treasury Constant Maturity. In general, the yield curve is a curve with a positive slope at each position. This is due to the fact that lenders desire funds to be withdrawn as soon as possible in order to increase portfolio liquidity. To attract lenders, extended maturities would logically necessitate higher interest rates, ignoring all other variables. Furthermore, it is worth noting that as the term of a loan increases, the borrower's creditworthiness at the time of maturity becomes increasingly uncertain, making it more challenging to mitigate the risk of default. As a result, it is imperative to offer a risk premium as a form of compensation. One of the most accurate predictors of future growth, inflation, and recession is the inclination of the yield curve.

4. Results

4.1. Visualization of model results

Since the research object is a time series, the methodology employed by the author involves utilizing the first 70% of the data as the training set, while allocating the remaining 30% as the test set to assess the effectiveness of the model. The author wants to further study the classical LSTM model. In most previous studies, only the past closing price was used as the input feature. However, in addition, this paper also constructs a model through technical indicators and macroeconomic indicators to contrast with the single independent variable model. This paper constructs three types of models: LSTM, GRU and LSTM-GRU, and makes the above comparison for each type of model. So, a total of six models are built to compare their performance. Given the input feature of solely the closing price, the LSTM model is being contemplated. When epochs=100, the validation loss decreases first and then increases, reaching the lowest point at epochs=40(Figure 3). Therefore, this paper adopted epochs=40 to train the model (Figure 4).

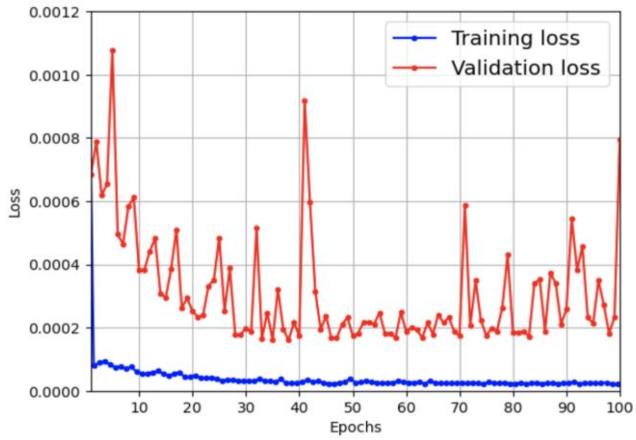

Fig. 3. LSTM with just close: epochs=100

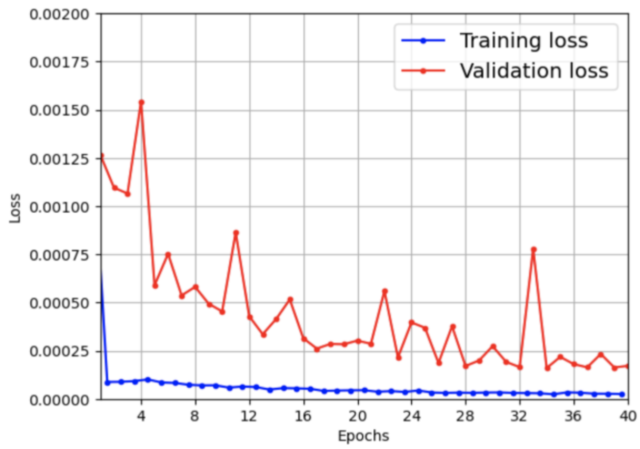

Fig. 4. LSTM with just close: epochs=40

However, when all features were input into the LSTM model, validation loss showed an obvious trend of first decreasing and then increasing when epochs=100, indicating that our model had the problem of overfitting. As can be seen from the Figure 5, epochs=20(Figure 6) or 30(Figure 7) is a more appropriate point. By comparison, epochs=30 is adopted to train the model.

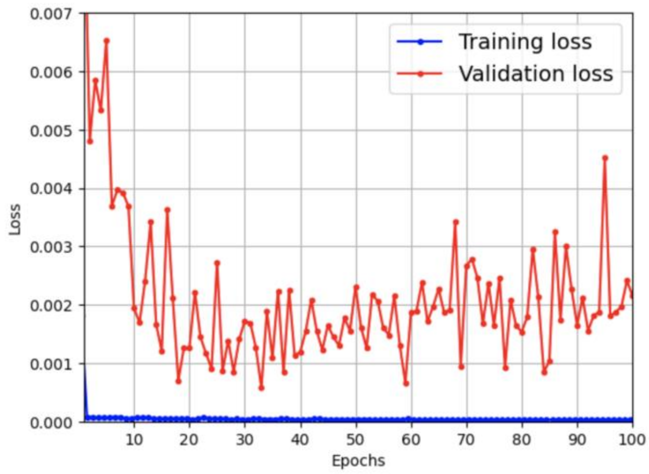

Fig. 5. LSTM with all indicators: epochs=100

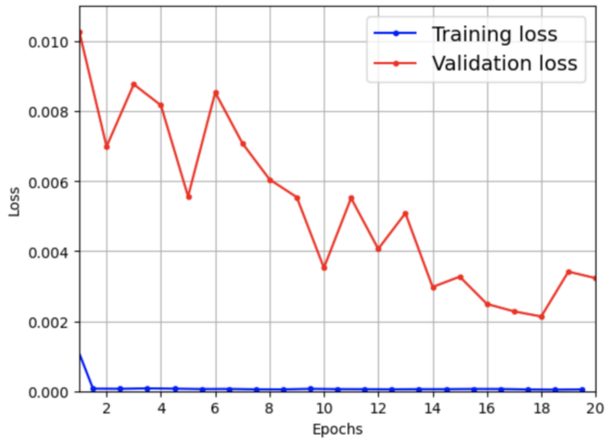

Fig. 6. LSTM with all indicators: epochs=20

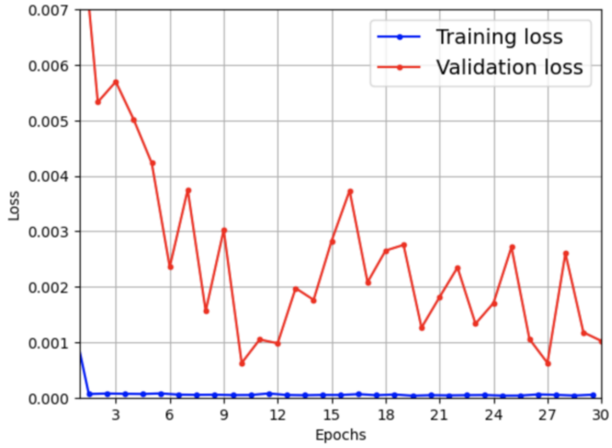

Fig. 7. LSTM with all indicators: epochs=30

Again, this optimization process is performed in the remaining four models.
The six models' forecasts for the S&P500 index are shown in the Figures 8-13:

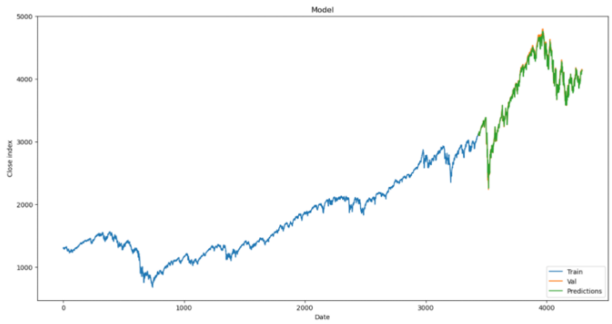

Fig. 8. LSTM with just close

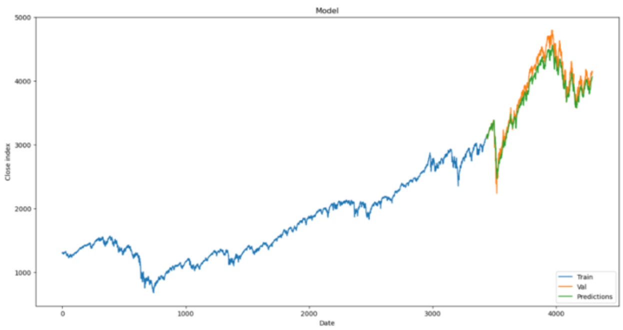

Fig. 9. LSTM with all indicators

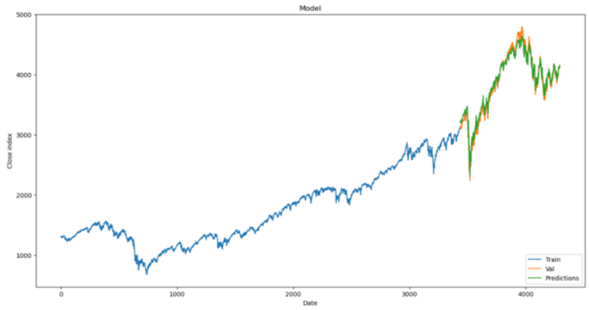

Fig. 10. GRU with just close

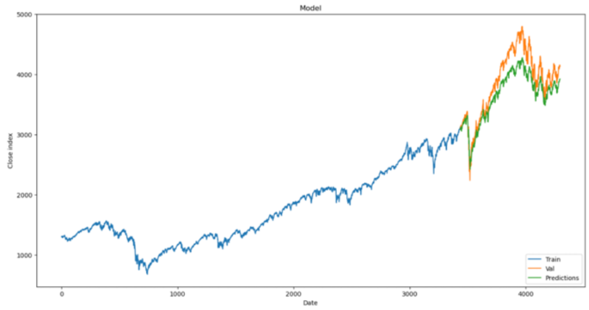

Fig. 11. GRU with all indicators

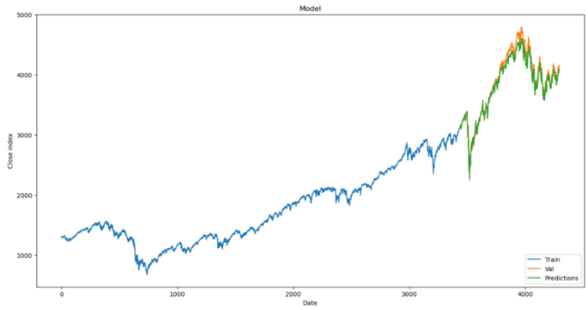

Fig. 12. LSTM-GRU with just close

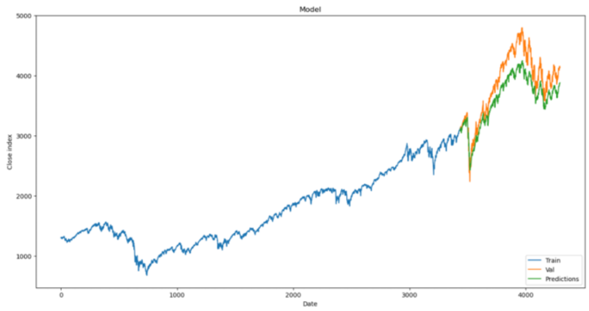

Fig. 13. LSTM-GRU with all indicators

4.2. Model evaluation

The assessment of the predictive accuracy and reliability of the models was conducted through the computation of five distinct performance indicators, namely root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), R square, and accuracy, for the test data. The MSE is a statistical metric that evaluates the average of the squared discrepancies between the anticipated and factual values in a specified dataset. The variance of the residuals is quantified by the measurement. The RMSE or root mean square deviation (RMSD) is a commonly used metric in various fields for assessing the accuracy of predictions. Using Euclidean distance, it demonstrates the distance between the predictions and the measured genuine values. The term MAE denotes the average of the absolute discrepancies between the observed and estimated values in a particular set of data. It quantifies the mean residuals in the data set. The R-squared denotes the proportion of the observed variability in the response variable that can be explained by the linear regression model. The score is characterized as scale-free, indicating that the R squared value will remain below one, irrespective of the magnitude of the values. Accuracy is an indicator meaning the degree to which predictions about the rise or fall of the index are accurate. A regression model's accuracy is indicated by a higher degree of precision when the MAE, MSE, and RMSE values are lower. Nevertheless, it is deemed desirable to have elevated values of R square and Accuracy. In addition, the performance scores are calculated by inverting the predicted results obtained from the normalized data.

Table 2. The performance metrics of the models with just 'close'.

	$LSTM_1$	GRU_1	$LSTM\text{-}GRU_1$
RMSE	53.91	83.57	86.99
MSE	2906.47	6984.79	7567.37
MAE	39.79	67.45	67.73
R square	0.99	0.97	0.98
Accuracy	0.525	0.530	0.524

Table 2 displays the model's performance, and table 3 ($model_1$ means that only close is used as the input feature of the model, and $model_2$ indicates that the input features of the model are all variables). Among all models, the LSTM model with just 'close' feature has the lowest RMSE: 53.91 and the highest R square: 0.99. In addition, the accuracy of models with 'close' as the only indicator in predicting future stock price rises or falls is higher than 50%, which is the probability of flipping a coin. However, the probabilities of the models considering other variables are all lower than 50%, which has no reference value in this respect.

Table 3. The performance metrics of the models with all indicators.

	$LSTM_2$	GRU_2	$LSTM\text{-}GRU_2$
RMSE	131.67	258.99	295.58
MSE	17337.19	67077.27	87371.50
MAE	112.29	223.39	263.07
R square	0.97	0.93	0.93
Accuracy	0.489	0.470	0.473

5. Conclusion

This study examined the effectiveness of different RNN models for index prediction by applying SP500 data. At the same time, the author also examines whether the performance of such models becomes better when they include technical and macroeconomic indicators. Firstly, compared with GRU model and LSTM-GRU model, LSTM model has the best performance in SP500 index forecast. According to Table 2 and Table 3, LSTM model has lower RMSE, MSE and MAE, and higher R square compared with the other two models. In addition, taking technical and macroeconomic indicators into account does not improve the model's performance. When technical indicators and macroeconomic indicators are taken into account, the performance of all the three types of models decreases. Comparing Table 2 with Table 3, it can be found that the RMSE, MSE and MAE of the models all rise when other economic variables excluding the closing price are considered. At the same time, the R square and Accuracy of the models decreased significantly.

This paper provides some theoretical support for stock investors' decision, including the selection of stock price prediction model and the determination of model variables. Given that the S&P 500 is a reliable representation of the American equity market, market regulators can also determine future policies through better forecasting and achieve more effective supervision and guidance of the market.

In the near future, further consideration will be given to other indicators to enhance the model's performance and interpretability. In addition, there may still be room for improvement in the structure of the model, including but not limited to the principal component analysis (PCA) to screen indicators with stronger explanatory power for the index, and the introduction of attention mechanism to give greater weight to the timing data with greater influence on the prediction effect, so as to improve the reliability of the prediction results.

References

[1] L. Nevasalmi, J. financ. data sci, **6**,86 (2020)

[2] R. Dash, P.K. Dash, J. financ. data sci, **2**,42 (2016)

[3] D. Ersan, C. Nishioka, A. Scherp, J. Comput. Soc. Sci, **3**,103 (2020)

[4] C.C. Chen, C.H. Chen, T.Y. Liu, Int. J. Econ. Financial Issues, **10**,59 (2020)

[5] J. Ayala, M. Garc´ıa-Torres, J.L. V´azquez Noguera, F. G´omez-Vela, F. Divina, Knowl Based Syst, **225**,107119 (2021)

[6] S.S. Chen, J Bank Financ, **33**, 211 (2009)

[7] E. Beyaz, F. Tekiner, X.J. Zeng, J. Keane, C*omparing technical and fundamental indicators in stock price forecasting. 2018 IEEE 20th International Conference on High Performance Computing and Communications*, 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 1607 (2018)

[8] P.L. Gao, R. Zhang, X. Yang, Math Comput Appl, **25**,53 (2020)

[9] M. Zulqarnain, R. Ghazali, M.G. Ghouse, Y.M. Mohmad Hassim, I. Javid, Int. J. Intell. Syst. Appl, **12**, 21 (2020)

[10] H.N. Bhandari, B. Rimal, N.R. Pokhrel, R. Rimal, K.R. Dahal, R.K. Khatri, Mach Learn, **9**, 100320 (2022)