IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 19 June 2023, accepted 11 July 2023, date of publication 17 July 2023, date of current version 27 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3296308

METHODS

A Stock Price Prediction Method Based on BiLSTM and Improved Transformer

SHUZHEN WANG
School of Information Science and Technology, Xiamen University Tan Kah Kee College, Zhangzhou 363105, China
Hongwang Laboratory, School of Information Science and Technology, Xiamen University, Zhangzhou 363105, China
e-mail: wangsz@xujc.com

This work was supported by the Educational Research Project of Young and Middle-Aged Teachers of Fujian Provincial Department of Education under Grant JAT191086.

ABSTRACT How to maximize shareholder returns has always been a focus of research in the financial field. In order to improve the accuracy and stability of stock price prediction, this article proposes a new method, BiLSTM-MTRAN-TCN. Improve the transformer model and introduce TCN (Temporary Revolution Network) to construct a new transformer model (MTRAN-TCN), making it suitable for stock price prediction. This method consists of BiLSTM (Bi-directional Long Short-Term Memory) and MTRAN-TCN, which can fully utilize the advantages of the three models: BiLSTM, transformer and TCN. Transformer is good at obtaining full range distance information, but its ability to capture sequence information is weak. BiLSTM can capture bidirectional information in sequences, while TCN can capture sequence dependencies and improve the model's generalization ability. Not only did the improvement effect of the transformer and the effectiveness of introducing the BiLSTM model be verified, but the effectiveness of the method was also verified using 5 index stocks and 14 Shanghai and Shenzhen stocks. Compared with other existing methods in the literature, this method has the best fit on each index stock, and the R^2 of this method is the best in 85.7% of the stock dataset. RMSE decreases by 24.3% to 93.5%, and R^2 increases by 0.3% to 15.6%. In addition, this method has relatively stable prediction performance at different time periods and does not have timeliness issues. The results indicate that the BiLSTM-MTRAN-TCN method performs better in predicting stock prices, with high accuracy and generalization ability.

INDEX TERMS Transformer, BiLSTM, TCN, stock price prediction, deep learning, hybrid neural network.

I. INTRODUCTION

For decades, stock price prediction has attracted attention from investors and researchers due to their enormous value [1]. More and more investors pay attention to the changing trend of stock price [2]. For economists, predicting stock price changes in advance is a very important task [3], [4]. It can help investors maximize their investment income. However, due to the high volatility of the stock market and the impact of random noise, its trend is complex and difficult to predict [5]. Although the financial time series is difficult to predict, it generally shows predictability is an essential task.

In the early days, the most famous is the moving average autoregressive model ARIMA [6]. Later, Narendra et al. [7] applied ARIMA model and GARCH model (autoregressive conditional heteroskedasticity model) to the NSE Indian stock market data forecast. In addition to these two models, there are Bayesian vector autoregression model and Kalman filter model. Although these techniques can be successfully used for short-term prediction, they are not suitable for nonlinear problems and have poor long-term prediction performance [8]. To solve this problem, machine learning is introduced to analyze time series, and they are successfully applied to stock price forecasting [2], [9], [10]. The advantages of machine learning in processing complex and large amounts of data have solved many limitations of traditional methods [9]. Machine learning methods include

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

support vector machine (SVM), decision tree, naive Bayes, random forest, etc. Wang et al. [11] mixed decision trees and SVM models to predict future price trends. Chen et al. [12] established feature weighted SVM and K-nearest neighbor algorithm to predict the stock market index. Experimental results have shown that the model has good short-term, medium-term, and long-term prediction capabilities. In 2021, Yan Zhengxu et al. [13] proposed a new combination model method of Random forest based on Pearson coefficient on the basis of Random forest to achieve short-term forecasting regression of stock price.

In recent years. deep learning methods that rely solely on datasets can be used to predict stock price without the need for professional knowledge. Therefore, their application in the field of stock prediction has gradually become a research hotspot for scholars. Deep learning methods include gated cyclic unit (GRU), recurrent neural network (RNN), Convolutional neural network (CNN), Long short-term memory (LSTM) and bidirectional Long short-term memory (BiLSTM). In 2017 and 2018, wavelet neural networks and CNN were successively used for stock price prediction [14], [15]. Experiments have shown that CNN is effective in predicting time series. In 2018, the authors proposed Conv1D-LSTM model, which combines one-dimensional CNN and LSTM. It can integrate the advantages of the two networks: CNN can effectively extract features, and LSTM can well process sequence data. The results indicate that the prediction results are more accurate than machine learning prediction models [16]. In 2019, Yang Qing et al. conducted predictive research on global stock indexes using BiLSTM, indicating that BiLSTM has excellent predictive accuracy and strong generalization ability [17]. In 2020, one study [10] used LSTM regression model to forecast India's NIFTY 50 index. The results showed that LSTM models based on deep learning performed better than traditional machine learning methods.

Now, in most of the current studies, the attention mechanism has begun to be the main structure to solve the problem of financial market forecasting, focusing on the key position that has a greater impact on the results. In 2020, Lu et al. introduced an attention mechanism (AM) based on CNN and BiLSTM, proposed CNN-BiLSTM-AM model, which proved to be more accurate than the existing models [8]. In 2022, a wavelet transform was used to denoise historical stock data based on LSTM and an attention mechanism [18]. Moreover, the transformer is the state-of-the-art model based on the attention mechanism, which was proposed for sequence modeling [19]. New methods based on the transformer were proposed to tackle the stock movement prediction task. In 2020, Ding et al. demonstrated that the model based on the transformer with the enhancement of Multi-Gaussian prior can be used for stock movement prediction [20]. In 2021, a transformer neural network based on self-attention was proposed, which has the special ability in forecasting time series, and the electricity consumption and traffic data were used to validate the proposed model [21].

In 2022, Zhang et al. proposed a novel transformer encoder-based attention network framework with the fusion of media text and stock price, and it has been shown to be effective to predict the rise or fall of stock price [8]. In 2022, Peng et al. used a data organization method with LSTM and transformer to predict Chinese bank stock price [22]. In 2022, Wang utilized the latest deep learning framework, transformer, to predict the stock market index and it demonstrated that the transformer can outperform other classical methods [23].

We can see that different LSTM-based and transformer-based models have been proposed for stock prediction. But so far, it is rare to use transformer-based models to predict price without considering the significance of stock data or social media text. Most of the models based on transformer are used to predict the stock trend of up and down, rather than predicting the stock price. Furthermore, now most of the proposed methods usually only target specific stocks or a single stock index, while the prediction model has timeliness issues. Therefore, there is still a lot of room for optimization in terms of accuracy and depth in the network structure of stock prediction models.

In order to improve the stability and accuracy of stock price prediction, this article proposes a novel method BiLSTM-MTRAN-TCN based on transformer. It can not only predict the index price, but also the individual stock price. This method is formed by introducing BiLSTM and TCN (Temporary Revolution Network) on the basis of transformer encoder. Transformer is a mechanism that can extract deep features of small samples to obtain key information. BiLSTM can capture bidirectional information in sequences, while TCN can capture sequence dependencies and improve the model's generalization ability.

The main work completed in this article is as follows:

- Improve the transformer model and introduce TCN to construct a new transformer model (MTRAN-TCN), making it suitable for stock price prediction.
- This method consists of BiLSTM and MTRAN-TCN, which can fully utilize the advantages of the three models: BiLSTM, transformer and TCN.
- Propose a bidirectional stock selection strategy to select stock experimental subjects.
- Compare with other existing models in the literature to verify the effectiveness of the method.
- Experimental results have shown that the proposed method has good generalization ability and solves the problem of timeliness.

II. ALGORITHM INTRODUCTION

A. TRANSFORMER

Transformer is a classic NLP model proposed by Google's team in 2017 [24], and Bert, which is popular now, is also based on the transformer. The transformer uses the self-attention mechanism and does not use the RNN sequential structure, so that the model can be parallelized and have

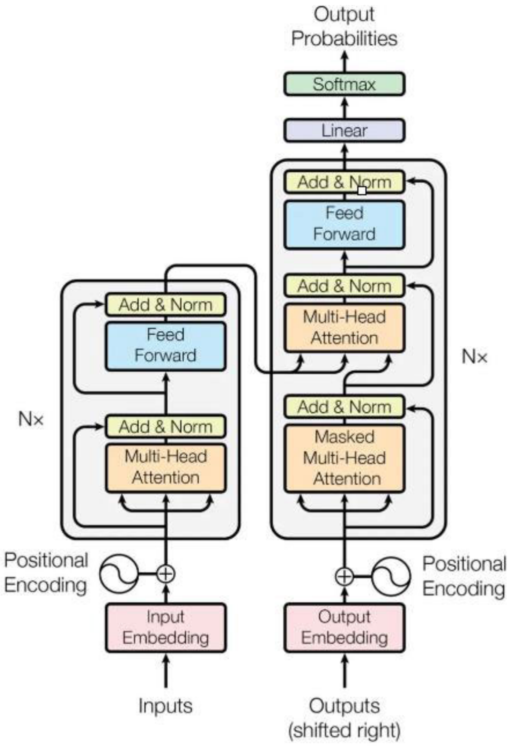

FIGURE 1. **Transformer overall structure.**

global information. Different from the recurrent networks, the transformer has no problem of gradient disappearance, and can access any point in the past, regardless of the distance between words.

Transformer consists of an encoder and a decoder section. The encoder section contains a stack of encoders, as shown in Figure 1. It encodes the input data according to a specific mode, and the decoder section decodes the output according to the encoded input to generate the required output. The most important part of the encoder section is the multi-head self-attention mechanism, through which the transformer captures long-term and short-term dependency. Different attention is focused on different aspects of the time pattern, so that more feature information can be captured. The output of self-attention is mainly calculated by Q, K and V matrices, where d_k is the dimension of K vector, as shown in Formula (1).

$$\text{Attention}\,(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Transformer has a better understanding of the context, so that it has special capabilities in predicting time series problems. However, the transformer model is designed for machine translation, and it cannot be directly used to predict time series. To predict the stock price, the encoder section of transformer is used as the basic model, as shown in Figure 3.

B. TCN (TIME CONVOLUTIONAL NETWORK)

TCN refers to a time convolutional network, a new algorithm that can be used to solve time series prediction. This

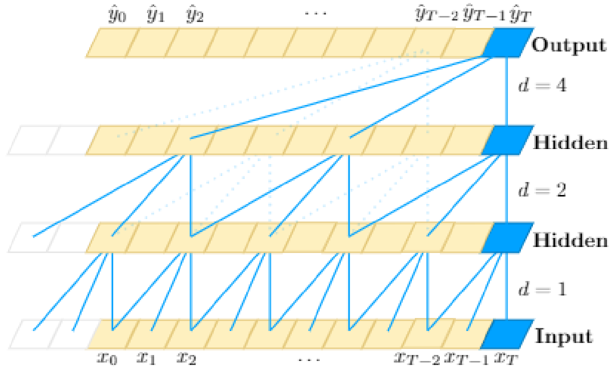

FIGURE 2. **Modified overall structure of transformer (MTRAN-TCN). Modified transformer, abbreviated as MTRAN-TCN.**

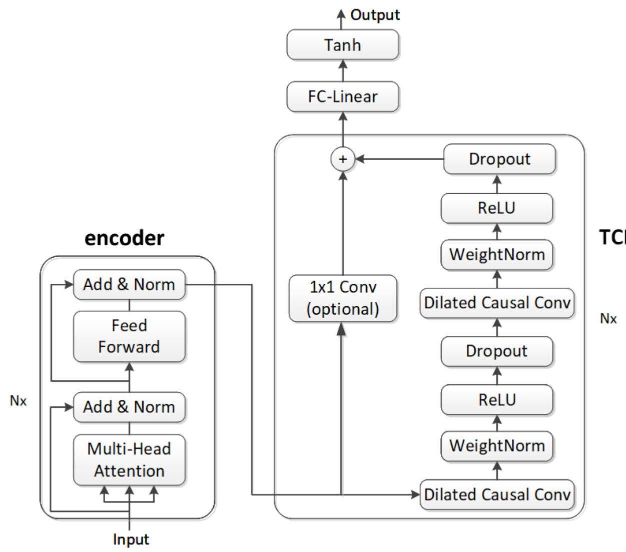

FIGURE 3. **Time convolutional neural network (TCN).**

algorithm was first proposed by Lea et al. in 2016 [25]. The network structure of TCN mainly consists of causal convolution, dilation convolution and residual connections. Each convolution layer is a causal convolution with a unidirectional structure, as shown in Figure 2. Causal convolution ensures that the output at time T is only convolved with elements that occurred before time T, which avoids future information leakage. And it can accept input sequences of any length and output the same length. The calculation of causal convolution, as shown in Formula (2).

$$F\,(s) = \sum_{i=0}^{k-1} f\,(i)\,x_{s-di} \tag{2}$$

The dilation convolution is used to capture longer dependency information without stacking more layers or adding pooling layers to obtain larger a receptive field.

III. BiLSTM-MTRAN-TCN METHOD OF THIS ARTICLE

MTRAN-TCN, which is an improved model of transformer. In this section, we first provide a detailed introduction on how to improve the transformer, and then describe the entire network structure of the proposed method.

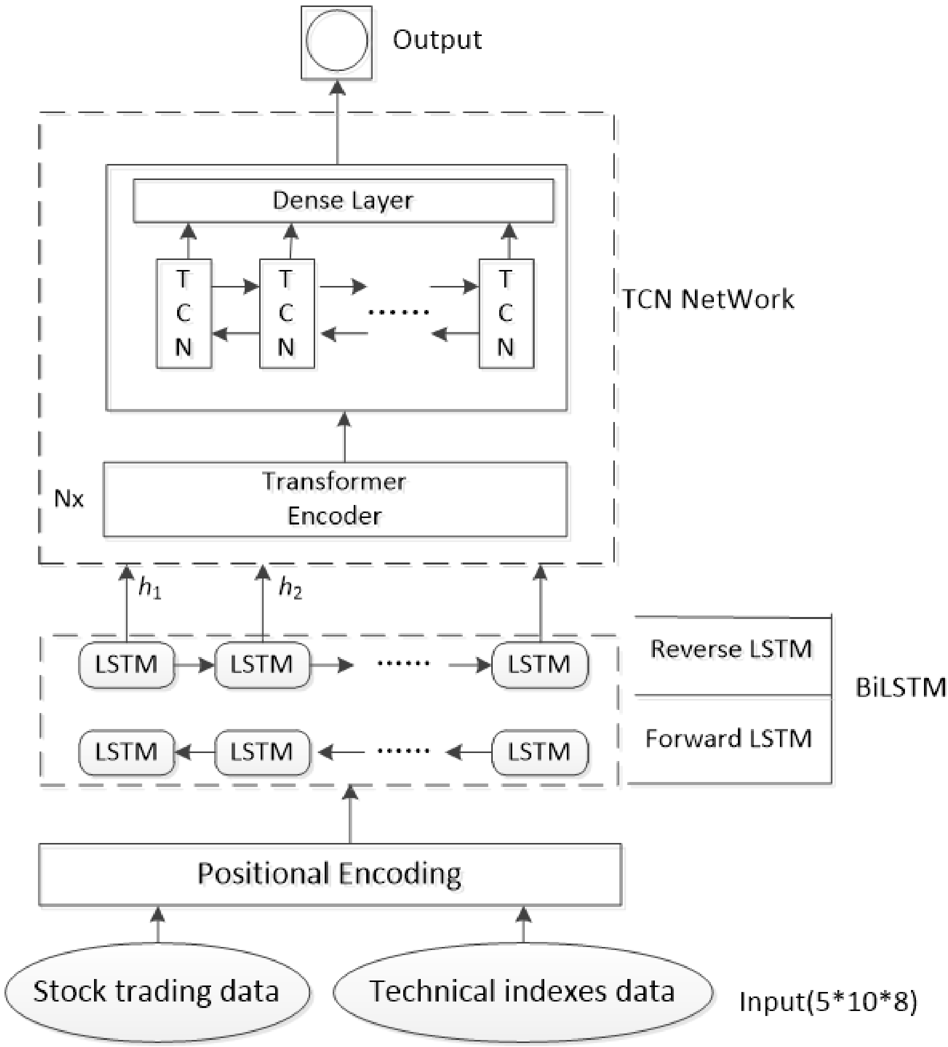

FIGURE 4. BiLSTM-MTRAN-TCN model structure diagram.

A. IMPROVE TRANSFORMER TO MTRAN-TCN

Modify the transformer model to better predict stock price, mainly by modifying the decoder of the transformer model, as shown in Figure 3.

- Delete the first module Input Embedding (see Figure 1). This module is to vectorization the language and text, which is a module required for machine language translation, while the stock price does not need to vectorization the text.
- Adjust the position of the Position Encoding module, remove it from the MTRAN-TCN, and move it to the front of BiLSTM, as shown in Figure 4.
- Replace the transformer decoder with TCN layer, full connection layer and the activation function Tanh.
- Removed other inputs of the decoder, leaving only the output of the encoder as the only input to the decoder.

TCN has been proven to be suitable for sequence data prediction. In 2022, Chen Zhe et al. proposed using time convolutional networks to mine time series features in traffic datasets [26]. In 2022, Wang Jun et al. proposed a multivariable TCN-Attention model that combines TCN and attention mechanism to predict daily average traffic [27]. In 2023, Yang Zhiyong et al. proposed a two-way short-term memory neural network integrating self-attention mechanism and TCN to predict stocks [28]. These all indicate that TCN can be used for sequence data prediction with good results. Therefore, this article introduces TCN to improve the transformer model.

The improved model is called MTRAN-TCN, and its internal structure is shown in Figure 3. The left half is the transformer encoder, and the right half is the TCN and fully connected layer. The transformer encoder consists of multiple encoder layers. Each encoder layer is composed of two sub layer blocks. The first sub layer block includes a multi-head attention, a normalization layer and a residual connection. The second sub layer block includes a feed forward fully connection layer, a normalized layer, and a residual connection.

TCN is composed of multi-layer residual blocks, with each residual block mainly composed of two layers of dilated causal convolution, weightnorm and dropout. The output of the encoder module serves as the input of the TCN first layer residual block.

B. NETWORK STRUCTURE

This study is based on the transformer model and introduces TCN to modify it to make it suitable for predicting stock series data. The transformer model can achieve parallel computing and obtain global signals well. However, the ability to capture sequence information is weak, and the effect of directly using transformer for stock prediction is not ideal. TCN can capture advanced and low-level features with stable gradients. And it can enable the model to process time series information in parallel, improving the model prediction accuracy and training efficiency. Its introduction can fully utilize the advantages of transformer and TCN to better predict sequence data. At the same time, BiLSTM has a strong ability to capture sequence signals, and BiLSTM is introduced to achieve better prediction results. Combining BiLSTM with the improved transformer to construct a hybrid network BiLSTM-MTRAN-TCN, as shown in Figure 4.

The output of each time step of BiLSTM is influenced by the input of the current time step and the previous memory, which can greatly capture sequence information. In the process of stock prediction analysis, the time range of stock series data is usually large, and BiLSTM may encounter the problem of rapid gradient decay when processing long time series. This may affect the BiLSTM network's learning of important feature information in stock data, leading to the loss of important feature information in the model. The multi-head self-attention mechanism of MTRAN-TCN allows the model to prioritize important feature information of the data, ignore other irrelevant information, and improve the prediction accuracy of the model. However, the MTRAN-TCN model lacks sequential information. Although its positional encoding uses sine and cosine to model the position, it is not sufficient for complete modeling and there is a certain lack of information. Therefore, before the data is transmitted to the multi-head self-attention of MTRAN-TCN, BiLSTM is used for processing to capture order dependencies. Therefore, combining it with MTRAN-TCN can improve the prediction efficiency of the model.

As the number of network layers and iterations increases, the weight changes too quickly, which can lead to network degradation effects in the model and poor performance when processing new data. For this reason, this article introduces the TCN network layer to handle variable length time series, capturing the dependencies of the sequence through the convolutional layer in TCN. And its residual connections can reduce the network depth and number of parameters, which improve the model's generalization ability and training efficiency.

Therefore, introducing TCN to improve the transformer model and constructing a hybrid network with BiLSTM can greatly improve the model's expression performance and training efficiency.

Use stock trading data and technical indexes data as the input, which are detailed in the dataset section, and the output is the closing price of the next day. The input is a 3D tensor, which includes samples, time_steps and features. After the input is processed by the Positional Encoding layer, the sequence features are captured by the BiLSTM, and then processed by the encoder of the transformer. After that, the features are further extracted through the TCN network layer. Finally, the full connection layer and the activation function are used for dimension reduction processing. The main structure includes Positional Encoding layer, BiLSTM layer, transformer encoder layer, TCN layer and Dense layer (Full Connection layer), as shown in Figure 4.

TABLE 1. Selected index stocks.

NUMBER	Index Name	Index Code
1	A-share	000002.XSHG
2	Shanghai Composite Index	000001.XSHG
3	Shenzhen Component Index	399001.XSHE
4	CSI 300	399300.XSHE
5	Growth Enterprise Board.Index	399006.XSHE

IV. EXPERIMENT ENVIRONMENT

In this section, we introduce the experimental dataset, evaluation indexes and experimental parameters.

A. DATASET

Previous studies [29], [30], [31], [32] only selected a few stocks from the Shanghai and Shenzhen indexes or the Shanghai and Shenzhen stock markets for experiments, and the price trends of the selected stocks were relatively stable and lacked volatility. The experimental coverage is poor, and the model experimental results lack persuasiveness.

To increase coverage, select 5 index stocks and 14 Shanghai and Shenzhen stocks for experiments. When selecting index stocks, representative index stocks were selected: A-share Index, Shanghai Composite Index, Shenzhen Component Index, CSI 300 and Growth Enterprise Board Index. The selected index stocks are shown in Table 1.

When selecting stocks in Shanghai and Shenzhen markets, a bidirectional stock selection strategy is adopted, which broadens the coverage of the experiment. Horizontally, select based on the size of the company's market value, divided into large-cap stock and small-cap stock; vertically, select according to stock classification, including 7 general categories including finance, real estate, coal, steel, non-ferrous metals, petrochemical and automotive. The selected stocks are shown in Table 2.

The data of the index stocks comes from JoinQuant (https://www.joinquant.com/research). And the data of 14 Shanghai and Shenzhen stocks is sourced from Tushare (https://tushare.pro/).

TABLE 2. Selected stocks from Shanghai and Shenzhen markets.

Category	Large-cap stock	Small-cap stock
Finance	China Merchants Bank (SH600036)	Guojin Securities (SH600109)
Real Estate	Poly Development (SH600048)	Tibet Urban Investment (SH600773)
Coal	China Shenhua (SH601088)	Power Investment Energy (SZ002128)
Steel	Zhongxin Special Steel (SZ000708)	Fangda Carbon (SH600516)
Nonferrous Metal	Tianqi Lithium Industry (SZ002466)	Yunnan Copper Industry (SZ000878)
Petrochemical	China Petroleum (SH601857)	Yueyang Xingchang (SZ300164)
Automotive	BYD (SZ002594)	Dongfeng Motor (SH600006)

This experiment selected stock data from the last 2700 trading days of each stock, covering the period from February 2012 to May 2023. Index stocks use historical trading data as the dataset, including: closing price, highest price, lowest price, opening price, ups and downs, change, turnover and volume. The Shanghai and Shenzhen stocks use the first six fields above and two relevant technical indexes (5-day moving average, 10-day moving average) as the stock dataset.

There are significant differences in each feature of the stock dataset, and the data needs to be normalized [33]. The standardization method used in this article is the Z-score, as shown in formula (3).

$$y_i = \frac{x_i - \bar{x}}{s} \tag{3}$$

where y_i is the standardized value, x_i is the input data, $\bar{x}$ is the average of the input data, and s is the standard deviation of the input data.

B. PERFORMANCE EVALUATION

The mean square error (MSE), the mean absolute error (MAE), root mean square error (RMSE) and R-square (R^2) are used as the evaluation criteria of methods. The calculation method of these error evaluation indexes is shown in Formula (4).

$$\left.\begin{aligned} \mathrm{MSE} &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ \mathrm{MAE} &= \tfrac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \\ \mathrm{RMSE} &= \sqrt{\tfrac{1}{n}\textstyle\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \\ \mathrm{R}^2 &= 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (\bar{y} - \hat{y}_i)^2} \end{aligned}\right\} \tag{4}$$

where y_i is the real value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average value. The smaller the MSE, MAE, and RMSE values, the better the performance. And the value of R^2 is between 0 and 1. The closer it is to 1, the better its performance.

TABLE 3. Parameter's setting of BiLSTM-MTRAN-TCN method.

Parameters	Value
Batch size	5
Sequence length of training data	10
Hidden size of BiLSTM	64
Number of BiLSTM layer	3
Number of transformer encoder head	8
Number of transformer enconder layer	6
Number of TCN layer neurons	32
TCN layer kernel size	1
Number of TCN hidden layer	4
Kernel size of TCN layer	7
Activation function of TCN layer	RELU

TABLE 4. Comparison of seven methods.

Method	MAE	MSE	RMSE	R^2
TRAN	0.284	0.224	0.438	0.778
MTRAN	0.175	0.056	0.229	0.945
MTRAN-TCN	0.173	0.055	0.231	0.945
BiLSTM-MTRAN-TCN	**0.087**	**0.014**	**0.118**	**0.986**
BiLSTM	0.136	0.035	0.182	0.965
BiLSTM-MTRAN	0.104	0.020	0.138	0.981
BiLSTM-TCN	0.147	0.040	0.195	0.960

C. NETWORK PARAMETERS

The parameter settings of the BiLSTM-MTRAN-TCN method in this experiment are shown in Table 3. The loss function is MSE, the optimizer chooses Adam, and the learning rate is 0.00001. The window size is 10, which predicts the closing price of stocks on the following day based on stock data from 10 days ago.

V. EXPERIMENT AND ANALYSIS

In this section, we validate the effectiveness of our proposed method from four aspects: comparative analysis with mainstream methods in the existing literature, validation of the effectiveness of introducing MTRAN-TCN and BiLSTM models, whether this method has timeliness issues, and the generalization ability of the method.

A. EFFECTIVENESS OF MTRAN-TCN AND BiLSTM

Using the index stocks in Table 1 as the dataset, verify the improvement effect of the transformer model and the effectiveness of introducing BiLSTM. Compare the BiLSTM-MTRAN-TCN method with TRAN, MTRAN-TCN, MTRAN, BiLSTM-MTRAN, BiLSTM, and BiLSTM-TCN, where TRAN is the transformer model, and MTRAN refers to the transformer encoder plus a fully connected layer. The comparison results of the above seven methods are shown in Table 4.

The predictive performance of TRAN is much lower than other methods, with an R^2 of only 0.778, indicating that the transformer model is not ideal for directly predicting stock price. Compared with TRAN, MTRAN-TCN has significantly improved its performance, with R^2 increasing from 0.778 to 0.945 and RMSE decreasing from 0.438 to 0.231. This indicates that the improvement of the transformer has a significant effect.

But the predictive performance of MTRAN and MTRAN-TCN is similar, indicating that using the improved MTRAN-TCN alone did not result in much improvement in performance. However, by adding BiLSTM to the MTRAN-TCN, the prediction performance of BiLSTM-MTRAN-TCN is significantly better than that of MTRAN-TCN. Its RMSE (0.118 compared to 0.231) decreases by 49.1%, while its R^2 (compared to 0.986 and 0.945) increases by 4.28%. Its addition greatly improves the predictive performance of BiLSTM-MTRAN-TCN. This indicates that the introduction of BiLSTM is effective.

When building a hybrid network with BiLSTM and TCN, there is no significant decrease or improvement in prediction performance compared to BiLSTM. Compared with BiLSTM, the RMSE of BiLSTM-MTRAN decreased by 24.06%, while R^2 increased by 1.57%. When using BiLSTM and MTRAN-TCN to construct a hybrid network, compared to BiLSTM, its RMSE decreased by 35.22% and R^2 increased by 2.09%. It can be seen that the hybrid network constructed by BiLSTM and MTARN-TCN can achieve better results.

In addition, compared to BiLSTM-MTRAN, BiLSTM-MTRAN-TCN reduces MAE from 0.104 to 0.087, MSE from 0.020 to 0.014, RMSE from 0.138 to 0.118, and increases R^2 from 0.981 to 0.986, indicating that the introduction of TCN to improve the transformer has a significant effect.

In conclusion, this article has shown good results in improving the transformer, and using BiLSTM to extract sequence features before MTRAN-TCN processing has shown significant results.

B. COMPARISON WITH OTHER METHODS

The BiLSTM-MTRAN-TCN method is compared with five other methods: LSTM, BiLSTM in literature [35], CNN-BiLSTM in literature [34], CNN-BiLSTM-AM in literature [8], and BiLSTM-SA-TCN in literature [28]. CNN-BiLSTM-AM, is a hybrid network of CNN, BiLSTM and AM (Attention Mechanism). BiLSTM-SA-TCN is a hybrid network of BiLSTM, SA (Self-Attention) and TCN.

Comparative analysis was conducted using index stocks and individual stocks from Shanghai and Shenzhen. Each stock is tested 5 times and the average of the five results is taken. All data is first standardized by Formula (3) before training, so the experimental results are all standardized values.

1) SHANGHAI AND SHENZHEN STOCKS

To further verify the effectiveness and progressiveness of the BiLSTM-MTRAN-TCN method, compare with other four methods mentioned above: BiLSTM, CNN-BiLSTM, CNN-BiLSTM-AM and BiLSTM-SA-TCN. Select 14 stocks from the Shanghai and Shenzhen stock markets as the dataset, from 7 major categories of large and small cap stocks, as shown in Table 2. The experimental results are shown in Tables 5-11.

It can be observed that the evaluation error indexes of CNN-BiLSTM and CNN-BiLSTM-AM methods are significantly lower than those of BiLSTM, BiLSTM-SA-TCN, and BiLSTM-MTRAN-TCN methods. This indicates that these two methods can only roughly fit the stock trend, and there is a significant error, resulting in a low fitting degree of the model. Compared with CNN-BiLSTM and CNN-BiLSTM-AM, BiLSTM has a smaller error, indicating that using CNN to extract stock data features first and then learning through the BiLSTM network does not improve the prediction accuracy, but rather reduces the prediction accuracy. How to improve the accuracy of prediction models requires neural network models to focus on feature information related to stock prices. The use of transformer's multi-head attention mechanism can effectively solve this problem.

Although BiLSTM-SA-TCN also uses a self-attention mechanism, its predictive performance is lower than BiLSTM-MTRAN-TCN in 100% of datasets. Compared with BiLSTM-SA-TCN, BiLSTM-MTRAN-TCN has a 40.2% reduction in MAE, 34.6% reduction in MSE, 55.6% reduction in RMSE, and 1.2% increase in R^2. This indicates that the multi-head attention mechanism of the transformer can better improve prediction accuracy than SA (Self-Attention).

Overall, for the BiLSTM-MTRAN-TCN method, the experimental results show that the R^2 evaluation index is the best in 85.7% of the stock dataset, and the RMSE evaluation index is the best in 78.6% of the dataset. Compared with the CNN-LSTM, CNN-BiLSTM-AM, BiLSTM, and BiLSTM-SA-TCN methods, its RMSE decreases by 93.2%, 93.5%, 55.6%, and 24.3%, respectively. And R^2 increases by 0.3% to 15.6%. This indicates that the BiLSTM-MTRAN-TCN method has more accurate expression ability than other methods and is in line with the fundamental trend of stock price fluctuations.

2) INDEX STOCKS

In addition to the Shanghai and Shenzhen stocks, this article also uses index stocks to further compare and analyze the above five methods and the method of LSTM, further verifying that the proposed method is superior to other methods and enhancing the coverage of the experiment. Using the 5 index stocks in Table 1 as the test dataset, conduct 5 experiments on each stock and each method. Then, calculate the average value of five stocks for each method, as shown in Table 12.

The prediction ability of the six methods is ranked in descending order: BiLSTM-MTRAN-TCN, BiLSTM, BiLSTM-SA-TCN, LSTM, CNN-BiLSTM-AM and CNN-BiLSTM. The MAE, MSE, and RMSE values of BiLSTM-MTRAN-TCN are the smallest among all methods, while the R^2 value is the largest among all methods. This shows that the prediction accuracy of the BiLSTM-MTRAN-TCN method is

TABLE 5. Comparison of evaluation indexes of different methods in China Merchants Bank and Guojin Securities.

Method	China Merchants Bank (SH600036)				Guojin Securities (SH600109)			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
BiLSTM	0.081	0.012	0.109	0.988	0.081	0.012	0.109	0.988
CNN-BiLSTM	0.265	0.134	0.366	0.861	0.264	0.132	0.364	0.862
CNN-BiLSTM-AM	0.277	0.143	0.378	0.851	0.240	0.121	0.348	0.874
BiLSTM-SA-TCN	0.110	0.020	0.142	0.979	0.116	0.021	0.146	0.978
BiLSTM-MTRAN-TCN	**0.042**	**0.004**	**0.064**	**0.996**	**0.062**	**0.007**	**0.083**	**0.993**

TABLE 6. Comparison of evaluation indexes of different methods in poly development and tibet urban investment.

Method	Poly Development (SH600048)				Tibet Urban Investment (SH600773)			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
BiLSTM	0.081	0.012	0.108	0.988	0.080	0.011	0.107	0.988
CNN-BiLSTM	0.265	0.134	0.366	0.860	0.257	0.129	0.359	0.866
CNN-BiLSTM-AM	0.293	0.149	0.387	0.845	0.241	0.123	0.351	0.872
BiLSTM-SA-TCN	0.119	0.022	0.148	0.977	0.107	0.018	0.136	0.981
BiLSTM-MTRAN-TCN	**0.056**	**0.009**	**0.096**	**0.990**	**0.096**	**0.017**	**0.131**	**0.982**

TABLE 7. Comparison of evaluation indexes of different methods in Shenhua and power investment energy in China.

Method	China Shenhua (SH601088)				power Investment Energy (SZ002128)			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
BiLSTM	0.084	0.012	0.111	0.987	0.083	0.012	0.111	0.987
CNN-BiLSTM	0.260	0.132	0.363	0.863	0.264	0.133	0.365	0.861
CNN-BiLSTM-AM	0.243	0.122	0.349	0.873	0.272	0.138	0.372	0.856
BiLSTM-SA-TCN	0.110	0.020	0.140	0.980	0.115	0.021	0.145	0.978
BiLSTM-MTRAN-TCN	**0.093**	**0.013**	**0.112**	**0.987**	**0.079**	**0.010**	**0.098**	**0.990**

TABLE 8. Comparison of evaluation indexes of different methods in Zhongxin special steel and Fangda Carbon.

Method	Zhongxin Special Steel (SZ000708)				Fangda Carbon (SH600516)			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
BiLSTM	0.082	0.012	0.109	0.988	0.081	0.012	0.109	0.988
CNN-BiLSTM	0.256	0.129	0.359	0.866	0.263	0.133	0.365	0.862
CNN-BiLSTM-AM	0.255	0.131	0.362	0.864	0.265	0.135	0.367	0.860
BiLSTM-SA-TCN	0.109	0.019	0.137	0.981	0.122	0.023	0.153	0.976
BiLSTM-MTRAN-TCN	**0.075**	**0.010**	**0.099**	**0.990**	**0.104**	**0.016**	**0.128**	**0.983**

optimal. And the prediction performance of CNN-BiLSTM or CNN-BiLSTM-AM is not very ideal.

Compared with other methods, BiLSTM-MTRAN-TCN reduces the error by 27.36% to 88.4%, and increases R^2 by 1.5% to 12.4%. Compared to LSTM, BiLSTM has smaller MAE and RMSE, while R^2 is larger. Its MAE (0.121 vs 0.138) decreases by 12.3%, RMSE (0.162 vs 0.185) decreases by 12.2%, and R^2 increases by 0.9%. Therefore, BiLSTM is superior to LSTM. When compared to BiLSTM-SA-TCN, the MAE of BiLSTM-MTRAN-TCN decreases from 0.139 to 0.098, while RMSE decreases from 0.185 to 0.118, and R^2 increases from 0.963 to 0.986, an increase of 2.4%. This also indicates that the transformer's self-attention mechanism performs better than SA's.

TABLE 9. Comparison of evaluation indexes of different methods in Tianqi Lithium Industry and Yunnan Copper Industry.

Method	Tianqi Lithium Industry (SZ002466)				Yunnan Copper Industry (SZ000878)			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
BiLSTM	0.081	0.012	0.108	0.988	0.081	0.012	0.109	0.988
CNN-BiLSTM	0.258	0.131	0.361	0.864	0.269	0.137	0.370	0.857
CNN-BiLSTM-AM	0.277	0.141	0.376	0.853	0.303	0.157	0.396	0.837
BiLSTM-SA-TCN	0.113	0.020	0.140	0.980	0.114	0.020	0.142	0.979
BiLSTM-MTRAN-TCN	**0.059**	**0.007**	**0.082**	**0.993**	**0.057**	**0.008**	**0.088**	**0.992**

TABLE 10. Comparison of evaluation indexes of different methods in China Petroleum and Yueyang Xingchang.

Method	China Petroleum (SH601857)				Yueyang Xingchang (SZ300164)			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
BiLSTM	0.081	0.012	0.108	0.988	0.083	0.012	0.110	0.987
CNN-BiLSTM	0.261	0.133	0.365	0.861	0.262	0.132	0.364	0.862
CNN-BiLSTM-AM	0.284	0.146	0.382	0.849	0.303	0.157	0.396	0.837
BiLSTM-SA-TCN	0.117	0.021	0.145	0.978	0.119	0.022	0.147	0.977
BiLSTM-MTRAN-TCN	**0.059**	**0.006**	**0.080**	**0.993**	**0.051**	**0.006**	**0.076**	**0.994**

TABLE 11. Comparison of evaluation indexes of different methods in BYD and Dongfeng Motor.

Method	BYD (SZ002594)				Dongfeng Motor (SH600006)			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
BiLSTM	0.082	0.012	0.109	0.988	0.081	0.012	0.109	0.988
CNN-BiLSTM	0.271	0.137	0.370	0.857	0.256	0.130	0.360	0.865
CNN-BiLSTM-AM	0.258	0.130	0.361	0.865	0.257	0.131	0.362	0.863
BiLSTM-SA-TCN	0.111	0.019	0.140	0.980	0.121	0.022	0.149	0.977
BiLSTM-MTRAN-TCN	**0.068**	**0.008**	**0.087**	**0.992**	**0.048**	**0.006**	**0.076**	**0.994**

Therefore, this method can better fit the stock trend, including the stock index and individual stock trend, to better predict the stock price.

C. VALIDATION OF GENERALIZATION ABILITY

Based on the above 14 Shanghai and Shenzhen stocks experimental data, conduct in-depth analysis on the generalization ability and accuracy of each method. It can be seen from Figure 5 that the R^2 of all six methods are concentrated between 0.8 and 1.0, and there are no singular values. The R^2 of BiLSTM is the most concentrated, with most concentrated at 0.988. The R^2 of BiLSTM is between 0.987 and 0.988, with a median of 0.988 and an average of 0.988. But the R^2 of BiLSTM-MTRAN-TCN is the closest to 1.0, indicating that its prediction accuracy is higher.

The R^2 value of the BiLSTM-MTRAN-TCN method is between 0.983 and 0.996, with a median of 0.992 and an average of 0.991. The R^2 value of the BiLSTM-SA-TCN method is between 0.976 and 0.981, with a median of 0.977 and an average of 0.979. The R^2 value of CNN-BiLSTM method is between 0.857 and 0.866, with a median of 0.862 and an average of 0.862. The R^2 value of CNN-BiLSTM-AM method is between 0.837 and 0.874, with a median of 0.858 and an average of 0.857. It can be seen from the distribution of R^2 that the BiLSTM-MTRAN-TCN method has no singular values and the values are relatively concentrated, indicating its good generalization ability. In addition, it has a higher mean and median, and is closer to 1.0, indicating better accuracy.

D. VALIDATION OF TIMELINESS ISSUES

Experiments were conducted on five index stocks in Table 1, and stock data from four different periods (2009.1∼2020.12, 2010.1∼2021.12, 2011.1∼2022.12, 2012.1∼2023.5) were used to verify whether there is a problem with timeliness. That is to verify the stability of the method in the time dimension.

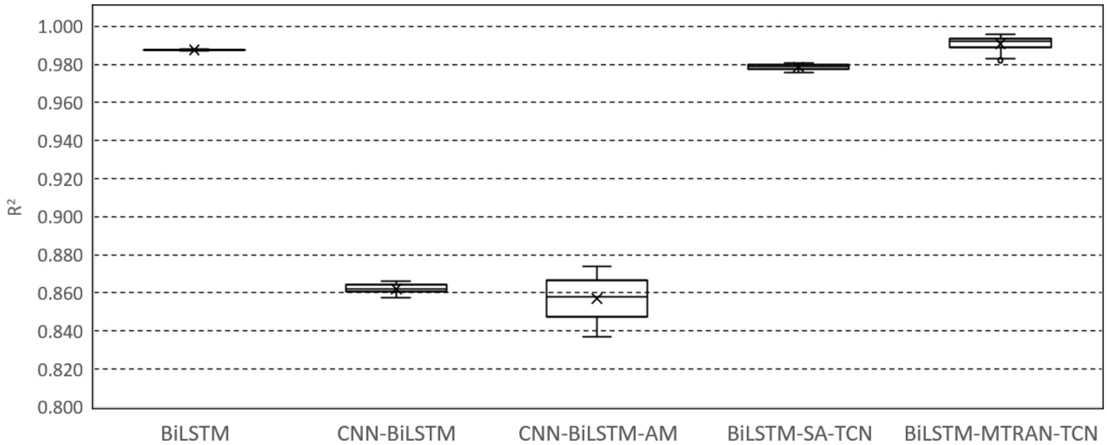

FIGURE 5. Box diagram of R^2.for each method. The graph drawn based on the analysis of the test results of 14 Shanghai and Shenzhen stocks.

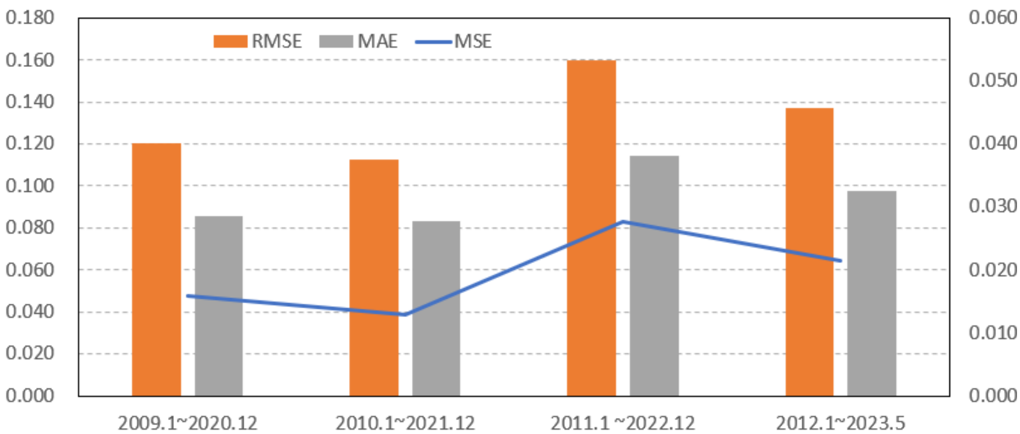

FIGURE 6. Error indicator values for each time period. MSE uses the secondary coordinates on the right.

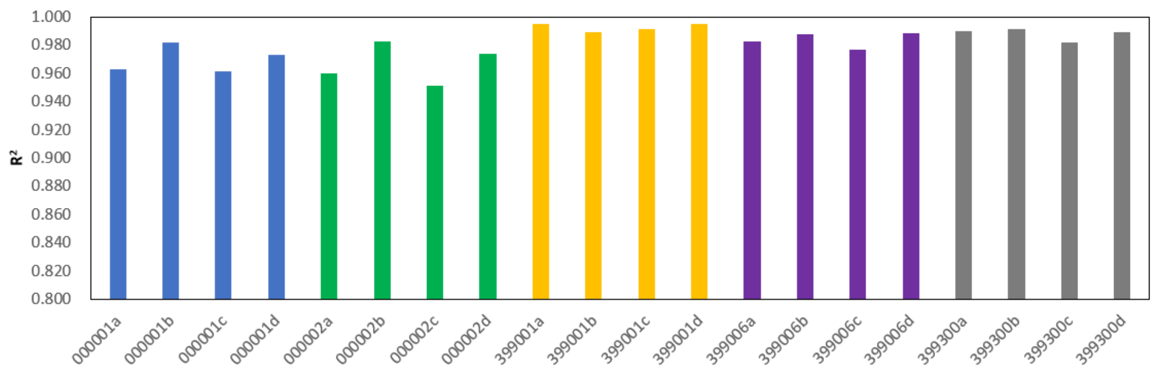

FIGURE 7. The R^2 of different stocks at different time periods. Each time period is marked with 'a', 'b', 'c','d'for differentiation, and the same identifier is used for the same time period.

From Figure 6, it can be observed that the error values of the BiLSTM-MTRAN-TCN method vary little at different time periods. The average values of RMSE during different periods were 0.120, 0.112, 0.160 and 0.137, respectively. The average values of MAE were 0.086, 0.083, 0.114 and 0.098, respectively. The average values of MSE were 0.016, 0.013, 0.028 and 0.021, respectively. Their R^2 differences are also very small. From Figure 7, it can be seen in detail that the R^2 of each index stock in the four time periods have very small changes (each time period is marked with 'a', 'b', 'c', 'd' for

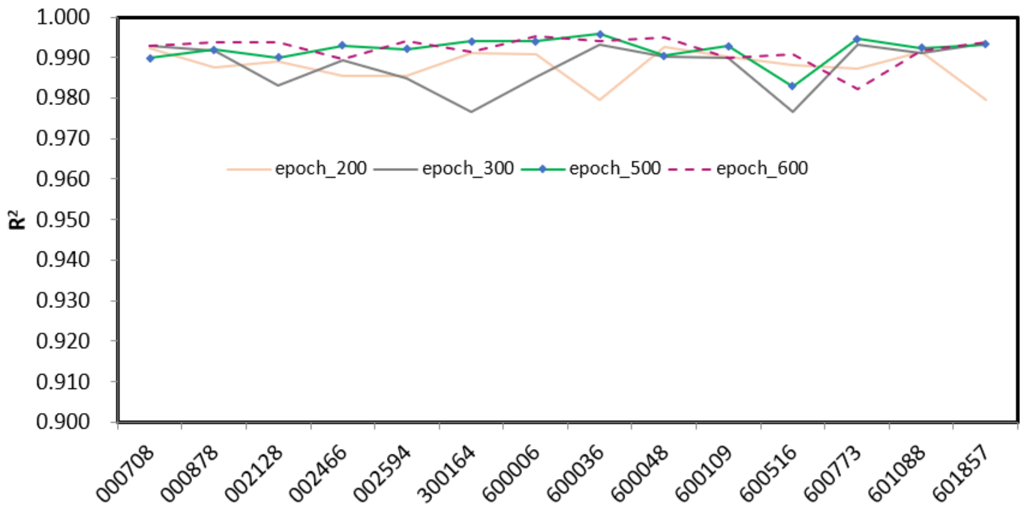

FIGURE 8. The R^2 of 14 Shanghai and Shenzhen stocks in epoch=200, 300,500 and 600.

TABLE 12. Average of evaluation indexes of the six methods.

Method	MAE	MSE	RMSE	R^2
CNN-BILSTM	0.268	0.120	0.342	0.877
CNN-BILSTM-AM	0.259	0.116	0.335	0.882
LSTM	0.138	0.037	0.185	0.962
BILSTM-SA-TCN	0.139	0.036	0.185	0.963
BILSTM	0.121	0.028	0.162	0.971
BILSTM-MTRAN-TCN	0.087	0.014	0.118	0.986

TABLE 13. Variance of evaluation indexes for 5 index stocks (over four time periods).

INDEX CODE		MAE	MSE	RMSE	R^2
000001	S^2	0.0006	0.0001	0.0009	0.0001
000002	S^2	0.0008	0.0002	0.0016	0.0002
399001	S^2	0.0002	0.0000	0.0003	0.0000
399006	S^2	0.0002	0.0000	0.0005	0.0000
399300	S^2	0.0002	0.0000	0.0004	0.0000

differentiation, and the same identifier means the same time period). The average values of R^2 at different time periods are 0.984, 0.987, 0.973 and 0.978, respectively. It can be seen that the R^2 fluctuations of different stocks during different time periods are not significant and relatively stable.

In addition, by calculating the variance of each index stock over four time periods, the dispersion of different time periods can be reflected, as shown in Table 13. The variance of MAE is between 0.0002 and 0.0008, the variance of MSE is between 0.0000 and 0.0002, the variance of RMSE is between 0.0003 and 0.0016, and the variance of R^2 is between 0.0000 and 0.0002. The variance of each index approaches 0, indicating that the deviation of the five index stocks at different time periods is very small.

According to the above analysis, the error indexes values of different stocks at different time periods only show small fluctuations, and the prediction results are relatively stable. This indicates that the BiLSTM-MTRAN-TCN method performs well in processing new data, has high accuracy and generalization ability, and does not have timeliness issues.

E. OPTIMAL EPOCH VALUE

The training performance of neural network models usually increases with the number of training rounds, but when it reaches a certain level, the performance tends to stabilize or actually decreases. In order to obtain the optimal epoch value and achieve the best prediction effect of this method, 14 Shanghai and Shenzhen stocks from Table 2 were used for experiments. Comparative analysis was conducted on 5 different scenarios as shown in Figure 8. The closer the R^2 is to 1, the better the performance of the method. When epoch=500, there is a significant increase in R^2 compared to epoch=200 and epoch=300. However, when epoch=600, the R^2 of most stocks is basically the same as when epoch=500, and it will not increase due to the increase in rounds. So it can be seen that epoch=500 is the optimal parameter value.

VI. CONCLUSION

The BiLSTM-MTRAN-TCN method proposed in this article is used to predict the closing price of stocks. This method modifies the transformer model, removes Input Embedding, replaces the original decoder part with TCN layer and fully connected layer, treats the output of the encoder as the only input of the decoder, and cancels other inputs. After Position Encoding Layer processing, the data is first processed by BiL-

STM to capture sequence dependent signals, and then sent to the modified transformer (MTRAN-TCN) for processing. Mixing multiple models can fully utilize the advantages of each model while avoiding their drawbacks and improving prediction accuracy.

In this article, experiments were conducted on the effectiveness of introducing BiLSTM, the improvement effect of transformer, the accuracy of the method, the generalization ability and timeliness issues. Experiments have shown that the improved transformer (MTRAN-TCN) needs to be mixed with BiLSTM to achieve optimal performance, and the introduction of BiLSTM has a significant effect on stock prediction.

Compared with LSTM, BiLSTM, CNN-BiLSTM, CNN-BiLSTM-AM, BiLSTM-SA-TCN in the literature, the BiLSTM-MTRAN-TCN method has the best predictive performance. Five representative index stocks and 14 Shanghai and Shenzhen stocks were selected for the experiments, among which 14 stocks were selected from 7 major categories using a bidirectional stock selection strategy. In the index stock experiment, the results show that the R^2 of BiLSTM-MTRAN-TCN method is increased by 1.5% to 12.4% compared to other methods. In the Shanghai and Shenzhen stock experiments, the R^2 value of this method is the best in 85.7% of the stock dataset, and RMSE is the best in 78.6% of the dataset. RMSE decreases by 24.3% to 93.5%, and R^2 increases by 0.3% to 15.6%.

In the timeliness experiment of 5 index stocks over four different time periods, the error indexes values of different stocks at different time periods only show small fluctuations, and the prediction results are relatively stable. This indicates that the BiLSTM-MTRAN-TCN method performs well in processing new data, has high accuracy and generalization ability, and does not have timeliness issues.

VII. DICUSSION

However, this method can further improve the prediction performance of some individual stocks. Future research work will continue in the following three aspects:

- Further optimize the neural network structure.
- Integrating multiple data sources for prediction: stock prices, various indexes data or fundamental information.
- Considering multiple time scale information: Currently, only one type of time window length data is considered for its impact. In the future, not only should we consider the impact of stock data from 10 days ago, but we can also consider the impact of stock data from 7, 30 and 150 days ago.

ACKNOWLEDGMENT

The author would like to thank the Hongnet Laboratory, Xiaomen University Tan Kah Kee College, for providing support for GPU server resources.

REFERENCES

[1] Z. Hu, W. Liu, J. Bian, X. Liu, and T.-Y. Liu, "Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, Marina Del Rey, CA, USA, Feb. 2018, pp. 261–269.

[2] J. Li, S. Pan, and L. Huang, "A machine learning based method for customer behavior prediction," *Tehnicki Vjesnik-Tech. Gazette*, vol. 26, no. 6, pp. 1670–1676, 2019.

[3] C. Xiao, W. Xia, and J. Jiang, "Stock price forecast based on combined model of ARI-MA-LS-SVM," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 5379–5388, May 2020.

[4] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Comput. Appl.*, vol. 32, no. 6, pp. 1609–1628, Mar. 2020.

[5] Q. Zhang, C. Qin, Y. Zhang, F. Bao, C. Zhang, and P. Liu, "Transformer-based attention network for stock movement prediction," *Exp. Syst. Appl.*, vol. 202, Sep. 2022, Art. no. 117239.

[6] T. S. Rao and M. M. Gabr, *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. New York, NY, USA: Springer, 2012.

[7] C. N. Babu and B. E. Reddy, "Prediction of selected Indian stock using a partitioning-interpolation based ARIMA-GARCH model," *Appl. Comput. Informat.*, vol. 11, no. 2, pp. 130–143, Jul. 2015.

[8] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Comput. Appl.*, vol. 33, no. 10, pp. 4741–4753, May 2021.

[9] H. Xu, B. Xu, and K. Xu, "Overview of the application of machine learning in stock forecasting," *Comput. Eng. Appl.*, vol. 56, no. 12, pp. 19–24, 2020.

[10] S. Mehtab, J. Sen, and A. Dutta, "Stock price prediction using machine learning and LSTM-based deep learning models," in *Machine Learning and Metaheuristics Algorithms, and Applications* (Communications in Computer and Information Science), vol. 1366, S. M. Thampi, S. Piramuthu, K. C. Li, S. Berretti, M. Wozniak, and D. Singh, Eds. Berlin, Germany: Springer, 2021, pp. 88–106.

[11] D. Wang, X. Liu, and M. Wang, "A DT-SVM strategy for stock futures prediction with big data," in *Proc. IEEE 16th Int. Conf. Comput. Sci. Eng.*, Sydney, NSW, Australia, Dec. 2013, pp. 1005–1012.

[12] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Exp. Syst. Appl.*, vol. 80, pp. 340–355, Sep. 2017.

[13] Z. Yan, C. Qin, and G. Song, "Stock price prediction of stochastic forest model based on Pearson feature selection," *Comput. Eng. Appl.*, vol. 57, no. 15, pp. 286–296, 2021.

[14] P. Wang, Y. Lou, and L. Lei, "Research on stock price prediction based on BP wavelet neural network with Mexico hat wavelet basis," in *Proc. Int. Conf. Educ., Econ. Manag. Res. (ICEEMR)*, Singapore, 2017, pp. 99–102.

[15] Y. Hu, "Stock market timing model based on convolutional neural network—A case study of Shanghai composite index," *Finance Economy*, vol. 26, no. 4, pp. 71–74, 2018.

[16] S. Jain, R. Gupta, and A. A. Moghe, "Stock price prediction on daily stock data using deep neural networks," in *Proc. Int. Conf. Adv. Comput. Telecommun. (ICACAT)*, Bhopal, India, Dec. 2018, pp. 1–13.

[17] Q. Yang and C. Wang, "Research on global stock index prediction based on deep learning LSTM neural network," *Stat. Res.*, vol. 36, no. 3, pp. 65–77, 2019.

[18] J. Qiu, B. Wang, and C. Zhou, "Forecasting stock prices with long-short term memory neural network based on attention mechanism," *PLoS ONE*, vol. 15, no. 1, Jan. 2020, Art. no. e0227222.

[19] A. Vaswani, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 1–11.

[20] Q. Ding, S. Wu, H. Sun, J. Guo, and J. Guo, "Hierarchical multi-scale Gaussian transformer for stock movement prediction," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 4640–4646.

[21] R. M. Farsani and E. Pazouki, "A transformer self-attention model for time series forecasting," *J. Electr. Comput. Eng. Innov.*, vol. 9, no. 1, pp. 1–10, 2021.

[22] Z.-Y. Peng and P.-C. Guo, "A data organization method for LSTM and transformer when predicting Chinese banking stock prices," *Discrete Dyn. Nature Soc.*, vol. 2022, pp. 1–8, Jan. 2022.

[23] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, "Stock market index prediction using deep transformer model," *Exp. Syst. Appl.*, vol. 208, Dec. 2022, Art. no. 118128.
[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[25] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," in *Computer Vision—ECCV* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2016, pp. 47–54.
[26] Z. Chen, J. Liu, and B. Zhao, "Multi factor urban road network taxi demand prediction based on GCN and TCN," *Control Decis.*, vol. 38, no. 4, pp. 1031–1038, 2023.
[27] J. Wang, Z. Gao, and C. Shan, "Multivariate prediction of the Yellow River runoff based on the TCN-attention model," *People's Yellow River*, vol. 11, pp. 20–25, 2022.
[28] Y. Zhiyong, Y. Yuxi, and Z. Yu, "Application of BiLSTM-SA-TCN time series model in stock forecasting," *J. Nanjing Inf. Eng. Univ.*, pp. 1–12, Apr. 2023.
[29] J. Deng, F. Zhao, and X. Wang, "MTICA-AEO-SVR model for stock price forecasting," *Comput. Eng. Appl.*, vol. 58, no. 8, pp. 257–263, 2022.
[30] X. Li, C. Cui, and G. Song, "Stock trend prediction method based on temporal hypergraph convolutional neural network," *J. Comput. Appl.*, vol. 42, no. 3, pp. 797–803, 2022.
[31] T. Zhang, Z. Tang, and J. Wu, "Forecasting method of stock index based on optimized KELM model," *Statist. Decis.*, vol. 37, no. 13, pp. 148–150, 2021.
[32] Y. Gao, R. Wang, and E. Zhou, "Stock prediction based on optimized LSTM and GRU models," *Sci. Program.*, vol. 2021, pp. 1–8, Sep. 2021.
[33] A. Yadav, C. K. Jha, and A. Sharan, "Optimizing LSTM for time series prediction in Indian stock market," *Proc. Comput. Sci.*, vol. 167, pp. 2091–2100, Jan. 2020.
[34] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-based model to forecast stock prices," *Complexity*, vol. 2020, pp. 1–10, Nov. 2020.
[35] A. Zeng and W. Nie, "Stock recommendation system based on deep bidirectional LSTM," *Comput. Sci.*, vol. 46, no. 10, pp. 84–89, 2019.

SHUZHEN WANG received the B.S. degree in electrical and automation from Changan University, Xi'an, China, in 2006, and the M.S. degree in detection and automation device from Xiamen University, Xiamen, China, in 2009.

From 2009 to 2018, she worked in data analysis and processing for an internet company. Since 2018, she has been with Xiamen University Tan Kah Kee College. She has also been researching the deep learning networks for stock prediction. Her research interests include facial detection and applications, data processing, and stock prediction.