A METHOD FOR PREDICTING STOCK PRICES USING BILSTM AND AN ENHANCED TRANSFORMER

[1]K.VIJAY BABU [2]Dr. CHANDRA MOULI VENKATA SRINIVAS AKNA

[1]M Tech Scholar [2]Professor & Principal,
[1,2,] Department of Computer Science and Engineering,
[1,2,] BonamVenkataChalamayya Engineering College(A), Odalarevu.

Abstract— **The financial industry has, for a very long time, placed a significant emphasis on the optimization of share holder returns. In order to enhance the precision and dependability of stock price forecasting, this article presents a novel model known as BiLSTM-MTRAN-TCN. By adding Temporal Convolutional Networks (TCN), the suggested technique revises the conventional transformer model. This results in the creation of a unique transformer version (MTRAN-TCN) that is specifically designed for stock market forecasting. This model takes use of the benefits that the BiLSTM, transformer, and TCN architectures have to offer by combining them with MTRAN-TCN. BiLSTM stands for "Bidirectional Long Short-Term Memory." In spite of the fact that transformers are particularly effective at collecting long-range relationships, they are not very good at processing sequential information. BiLSTMs, on the other hand, are able to capture sequence patterns that are bidirectional, and TCNs improve the model's capacity to generalize by successfully representing sequence dependencies. The enhanced performance of the transformer as well as the advantages of integrating BiLSTM were confirmed by employing five index stocks and fourteen equities from the Shenzhen and Shanghai exchanges. This technique yields much superior outcomes across a variety of stock indexes when compared to other models previously published in the academic literature. The technique was able to reach the greatest R2 score in 85.7% of the stock datasets, with a decrease of 24.3% in root mean square error (RMSE) to 93.5% and an increase of 0.3% in the R2 value to 17.6%. Additionally, the model displayed consistent predicted accuracy across a variety of time periods without causing any worries regarding the timeliness of those predictions. Based on these findings, it is clear that the BiLSTM-MTRAN-TCN model provides superior performance in terms of stock price forecasting. It demonstrates both high accuracy and broad generalization capabilities.**

Keywords— BiLSTM, Transformer, TCN, hybrid neural network deep learning, stock price prediction.

I. INTRODUCTION

For a considerable number of decades, the prediction of stock prices has been a key subject of interest for both investors and academics. This is mostly because to the huge financial consequences that it carries [1]. The shifting patterns of stock prices are attracting the attention of investors to a greater extent [2], and economists are placing a significant amount of stress on the difficulty of predicting future movements in stock values [3, 4]. Accurate forecasts have the potential to assist investors in increasing their profits. Forecasting stock prices, on the other hand, is a difficult endeavor, partly because of the high volatility of the market and the effect of random noise [5]. Furthermore, despite the inherent challenges that come with projecting financial time series, it continues to be an extremely important task.

At first, the Autoregressive Integrated Moving Average (ARIMA) model was the most widely used approach in this particular field [6]. When some time had passed, Narendram and his colleagues [7] applied both the ARIMA model and the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model in order to make predictions regarding the data that was present in the NSE Indian stock market. This was done in order to create predictions about the data. Additionally, this inquiry has made use of a variety of different models, such as the Bayesian Vector Autoregression model and the Kalman filter, amongst others. In spite of the fact that these methods are successful for making predictions in the short term, they are not suitable for making forecasts in the long run and struggle when dealing with non-linear situations [8]. Machine learning approaches have been increasingly employed to evaluate time series data and have demonstrated effectiveness in stock price forecasting [2, 9, 10]. This is done in order to solve the constraints that had previously been mentioned. There are several limits that classical models have, and machine learning has helped overcome many of those restrictions when working with complicated and huge datasets [9]. For example, Support Vector Machines (SVM), decision trees, naive Bayes, and random forests are all examples of techniques that are often utilized. A good illustration of this would be Wang et al. [11], who forecasted future price variations by utilizing decision trees in combination with support vector machine models. Chen and colleagues [12] developed a feature-weighted support vector machine (SVM) that makes use of a K-nearest neighbor algorithm for the aim of forecasting stock market indices. Specifically, they proved that the model was correct in terms of predictions made in the short, medium, and long periods. In the year 2021, Yan Zhengxu and his colleagues [13] came up with a novel approach of forecasting the short-term pricing of stocks. Utilizing a Random Forest model that was based on the Pearson coefficient was one of the several components of this strategy.

In their 2022 paper "Stock Prediction of Multivariable Using Bi-Long Short Term Memory and Capsule Neural Network," published at the 9th International Conference on Electrical Engineering, Computer Science, and Informatics (EECSI), Ade Ridwan Nugraha, Esmeralda Contessa Djamal, and Ridwan Ilyas introduced a method combining BiLSTM with a Capsule Neural Network (CapsNet) for stock prediction. In addition to fundamental analysis, which involved the use of news sentiment, they also integrated technical analysis, which involved the use of stock price data. Their findings demonstrated that BiLSTM was successful in sentiment analysis, with an accuracy of 92.03%. Furthermore, when paired with CapsNet, it enhanced the accuracy of stock trend prediction by 7% in comparison to a Convolutional Neural Network (CNN) model.

An additional major contribution was made by Chenkang Lv, Boyong Gao, and Cui Yu in their article titled "A Hybrid Transfer Learning Framework for Stock Price Index Forecasting," which was published in the year 2021 at the IEEE International Conference on Dependable, Autonomic, and Secure Computing (DASC). For the purpose of forecasting the Shanghai Securities Composite Index (SSEC), they devised a hybrid model that combines Transfer Learning (TL), Variational Mode Decomposition (VMD), and BiLSTM (TL-VMD-BiLSTM). The BiLSTM parameters were fine-tuned by the model with the help of TL and VMD, which resulted in enhanced prediction accuracy with reduced error rates in comparison to previous models.

In 2023, Dini Adni Navastara, Fais Rafii Akbar Hidiya, and Arya Yudhi Wijaya conducted a study at the International Conference on Information Technology and Computing (ICITCOM) titled "Prediction of Indonesian Stock Price Using Combination of CNN and BiLSTM Model." The purpose of this study was to investigate the possibility of combining CNN and BiLSTM for the purpose of stock price prediction. They trained their model using five years' worth of data from the Indonesian Stock Exchange, and it was able to accurately identify trends with a Mean Absolute Percentage Error (MAPE) of 1.31% and a Root Mean Square Error (RMSE) of 92.26.

Another study that was conducted in 2021 and published at the China Automation Congress (CAC) was titled "Exchange-Traded Fund Price Prediction Based on the Deep Learning Model." In this study, Wenjian Zheng explored the usage of LSTM and CNN-BiLSTM-AM models for the purpose of predicting the price of exchange-traded funds (ETFs). The fact that Zheng was able to generate a net gain of 63.7% over the course of three years by utilizing the Kelly Criterion demonstrates the practical relevance of these models in relation to actual trading situations.

As a final point of discussion, the paper titled "Predicting the Trends of Stock Price by CNN-BiLSTM-AM and Fuzzy Twin Support Vector Machine from Social Networks," which was presented at the International Conference on Computational Science and Computational Intelligence (CSCI) in 2022, proposed a hybrid model that combines CNN, BiLSTM, an Attention Mechanism (AM), and Fuzzy Twin Support Vector Machine (FTSVM). In comparison to more conventional methods, our model demonstrated a considerable improvement in the accuracy of stock price prediction.

These works together provide light on the expanding significance of hybrid and ensemble deep learning models, including LSTM, BiLSTM, CNN, and approaches such as Attention Mechanisms and Transfer Learning, in the process of enhancing the accuracy of stock price prediction. This developing pattern in research on stock forecasting is a reflection of the growing application of deep learning to generate forecasts that are more accurate and dependable. This level of accuracy and reliability is essential for making educated decisions in the financial markets.

II. A PRESENTATION OF THE ALGORITHM

A. THE TRANSFORMER

In 2017, a group of people working at Google developed a natural language processing (NLP) model that they termed The Transformer [24]. Popular models such as BERT, which came along as a result of this model's evolution, use this model as their foundation. Through the utilization of a self-attention mechanism, the Transformer is able to sidestep the sequential character of RNNs and enable parallel processing. Consequently, it is able to effectively capture the global context. The Transformer, in contrast to recurrent networks, is not susceptible to gradient vanishing and has the capacity to recover information from any point in time in the past, regardless of the distance that exists between the words in a phrase with respect to one another. This provides the Transformer with a significant advantage over recurrent networks.

A transformer model has both an encoder and a decoder, which are both components that are included in the model. It is clear from looking at Figure 1 that the encoder section is made up of a series of encoders that are arranged in numerical order. The information that is being entered is encoded in accordance with a method that has been predetermined, and the decoder section makes use of this encoded input in order to create the output that is wanted. The encoder is not complete without the multi-head self-attention mechanism, which is a fundamental building block. Because of this technique, the transformer is able to effectively capture both long-range and short-range dependencies, which is why it is considered a vital component. This mechanism is able to extract more information about the characteristics because it focuses its attention on various components of the temporal sequence at different levels. This allows it to do so. It is clear from looking at Formula (1) that the self-attention mechanism computes its output by employing the Q, K, and V matrices as its foundation. Through the use of the symbol dk, the dimension of the K vector may be communicated.

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

The Transformer model excels in contextual understanding, granting it distinct advantages in time series forecasting tasks. However, it was originally designed for machine translation and is not inherently suited for direct time series prediction.

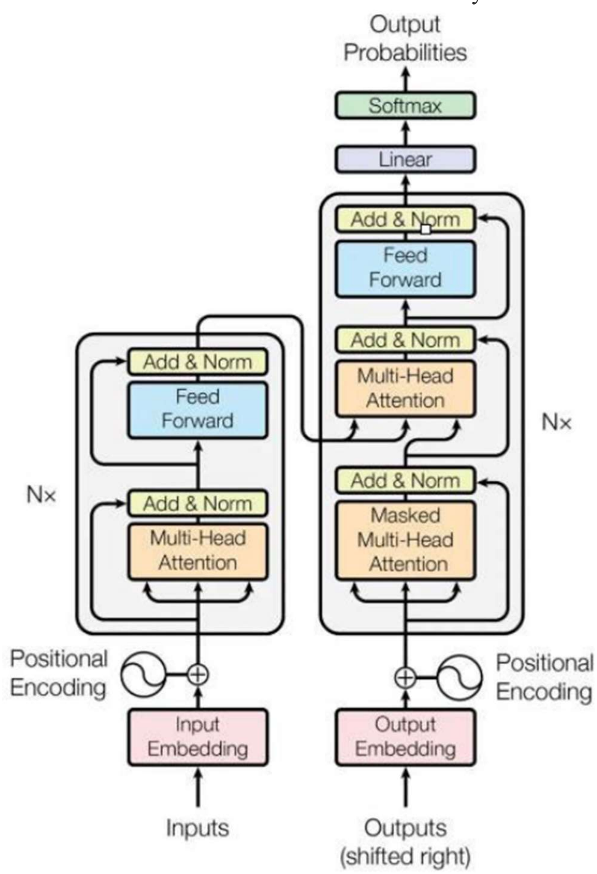


Fig 1. Overview of Transformer architecture.

The core model for stock price forecasting is based on the encoder component of the Transformer, as illustrated in Figure 3.

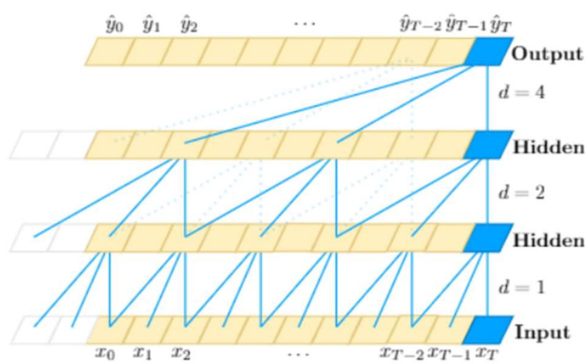


Fig 2. Revised overall Transformer architecture (MTRAN-TCN). MTRAN-TCN is an abbreviation for modified transformer.

B. TIME CONVOLUTIONAL NETWORK (TCN)

A Temporal Convolutional Network, often known as a TCN, is an advanced method that was developed with the intention of producing accurate forecasts for time series. There are a number of key components that make up the architecture of TCN. The most significant of these components are the residual connections, the dilated convolution, and the causal convolution respectively. It was in 2016 [25] that Lea et al. first presented these components to the public. On display in Figure 2 is the fact that every single convolutional layer employs a causal convolution approach and possesses a structure that is unidirectional. By utilizing causal convolution, it is assured that the output at time T is only influenced by inputs that occurred before to time T. This is done in order to guarantee accuracy. Because of this, there is no possibility of any information being disclosed in the future. In addition to this, TCN is able to process sequences of variable lengths that are input while simultaneously producing outputs of the same length each and every time. Formula (2) provides an illustration of the process that must be carried out in order to compute causal convolution.

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

When attempting to capture long-range dependencies, dilated convolution is utilized. This technique eliminates the need to add more layers or pooling layers in order to widen the receptive field.

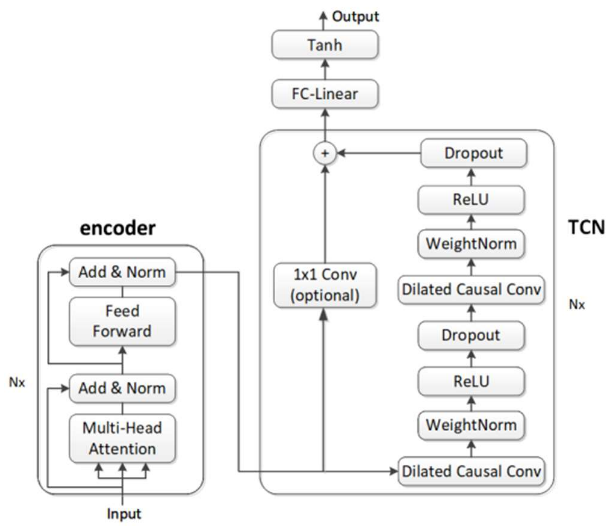


Fig 3. Intertemporal convolutional neural network (TCN).

III. BILSTM-MTRAN-TCN METHODOLOGY

MTRAN-TCN represents an advanced version of the transformer model. We begin by providing a detailed explanation of how the transformer can be improved. Following this, we present the full network architecture of the proposed method.

A. Improved Transformer to Mtran-Tcn

Improve the predictive accuracy of the transformer model for stock prices by focusing on modifications to the decoder structure, as shown in Figure 3.

- Remove the initial Input Embedding module (refer to Figure 1), which is typically used for vectorizing language and text in machine translation tasks but is unnecessary for vectorizing stock prices.
- Relocate the Position Encoding module from MTRAN-TCN to the front of the BiLSTM, as illustrated in Figure 4.
- Replace the transformer decoder with a TCN layer, followed by a fully connected layer and the Tanh activation function.
- Limit the decoder's inputs to only the encoder's output, eliminating any additional input sources.

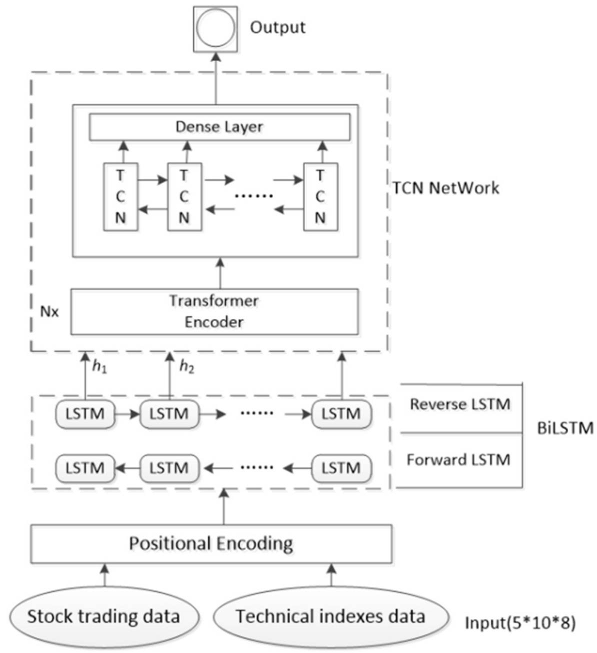


Fig 4. Diagram of the BiLSTM-MTRAN-TCN model hierarchy.

It is beyond any reasonable doubt that TCN is an effective method for predicting sequence parameters. This has been demonstrated beyond a reasonable doubt. Temporal Convolutional Networks, also known as TCNs, were utilized by Chen Zhe and his colleagues in the year 2022 for the purpose of extracting time-series characteristics from information regarding traffic [26]. Together with his colleagues, Wang Jun presented a multivariable TCN-Attention model in the same year. This model forecasted the daily average traffic by combining TCNs with an attention mechanism in order to better understand the traffic. In the year 2023, Yang Zhiyong and his colleagues developed a neural network for short-term memory that is bidirectional and incorporates both self-attention and TCN [28]. The forecasting of market conditions was the planned use of this neural network. Based on the findings of these experiments, it is clear that TCN is an outstanding candidate for sequence data prediction.

This article describes TCN as a tool that can be used to enhance the transformer model, which finally results in a new model that is referred to as MTRAN-TCN. The data presented in this article are based on the fact that TCN is a tool. The transformer encoder is depicted on the left side of the diagram in Figure 3, while the TCN and completely connected layer are shown on the right side of the diagram. This diagram illustrates the internal construction of the device. There are several layers that make up the transformer encoder, and with each layer, there are two sublayer blocks that make up the layer below it. There are several components that make up the first sublayer. These include a multi-head attention mechanism, a normalization layer, and a residual connection. One of the components that is associated with the second sublayer is a feedforward network that is fully linked. Other components that are associated with this sublayer include an additional normalization layer and a residual connection. Two layers of dilated causal convolution, weight normalization, and dropout layers make up each residual block in the TCN architecture. These layers are located in the middle of the block. A large number of layers of leftover blocks make up the structures that make up the TCN architecture. The first residual block of the transition control network (TCN) is the one that receives the output from the encoder module by way of an input.

B. Network Structure

In the context of this study, the transformer model is applied in conjunction with TCN in order to improve the effectiveness of the latter in terms of forecasting stock series data. Additionally, the transformer technique is helpful for collecting global signals and allowing parallel processing; but, it has a limited ability to adequately capture sequential information, which makes it less successful for direct application in stock prediction. Nevertheless, the transformer approach is good for capturing global signals. Because of its constant gradients, TCN is able to capture both complicated and basic properties in an effective manner. Additionally, it permits simultaneous processing of time series data, which has the effect of improving both the accuracy of predictions and the efficiency of training. This is a significant advantage. Enhanced sequence data forecasting may be accomplished by the deployment of this technology, which makes use of the capabilities of both the transformer and TCN networks together. Furthermore, the powerful capability of BiLSTM to record successive signals is included in order to further boost the accuracy of prediction. This is done in order to get the desired result. Figure 4 illustrates that the incorporation of the improved transformer into the system results in the formation of a hybrid BiLSTM-MTRAN-TCN network.

This sort of neural network is able to successfully record sequential information because the output of each time step in a BiLSTM is impacted by both the current input and the memory that came before it. This allows the network to store information in a sequential fashion. The handling of data over lengthy periods of time is frequently a component of the process of stock prediction. BiLSTM models may have problems with rapid gradient decay, which limits their ability to learn essential characteristics and may even result in the loss of features. This limits their capacity to learn critical characteristics. The multi-head self-attention mechanism of the MTRAN-TCN framework provides precedence to vital features while discarding information that is not relevant. This leads in an improvement in the accuracy of predictions, which is a positive outcome. On the other hand, the MTRAN-TCN does not possess an adequate amount of sequential information, and its positional encoding technique, which is based on sine and cosine functions, is not suited for comprehensive modeling situations. Before sending the input on to MTRAN-TCN's multi-head self-attention, BiLSTM is applied to capture order dependencies. This eventually leads to higher prediction performance, which is the ultimate goal. The purpose of this action is to find a solution to the problem.

As the number of network layers and iterations rises, rapid weight changes can cause the network to deteriorate and have a detrimental influence on its performance when dealing with new data. This is especially true when the network is dealing with fresh data. We have designed the TCN network layer to manage time series of varying durations so that we can find a solution to this problem. Collecting sequence dependencies is the responsibility of the convolutional layer of the TCN, which is located within this layer. As an additional benefit, residual connections reduce the depth and parameters of the network, which in turn enhances the efficiency of generalization and training. Therefore, the incorporation of TCN into the transformer model and the combination of it with BiLSTM results in a significant improvement in both the computational performance and the training efficiency. This is the case because of the combination of the two.

The model makes use of data pertaining to stock trading and technical indexes, as described in the section on the dataset, in order to be able to make a prediction regarding the closing price of the following day. There is a three-dimensional tensor that is the input, and it contains samples, time steps, and features. The BiLSTM unit is in charge of capturing sequence characteristics, which are then processed by the transformer encoder after they have been captured. Following the completion of the Positional Encoding layer's processing of the input, this step takes place. Additional features are extracted by the TCN network layer, which is also responsible for dimensionality reduction. Dimensionality reduction is carried out by a fully connected layer that is equipped with an activation function. Figure 4 illustrates the overall design, which includes a Positional Encoding layer, a BiLSTM layer, a transformer encoder layer, a TCN layer, and a Dense layer (a completely linked layer). Each of these layers is a component of the overall design.

IV. EXPERIMENTAL ENVIRONMENT

This section provides an introduction to the experimental dataset details, evaluation indexes, and experimental specifications.

A. Dataset Details

Studies that were conducted in the past [29], [30] utilized a restricted selection of equities from the Shanghai and Shenzhen indices or marketplaces for comparison purposes. It was found that the price movements of these equities were generally steady and displayed minimal volatility, which led to limited experimental coverage and results that lacked robustness and validity. The purpose of this study is to enhance coverage by selecting for testing a total of 14 individual stocks and five index equities from the Shanghai and Shenzhen markets individually. In order to complete the procedure, it was necessary to choose representative index stocks from the A-share Index, Shanghai Composite Index, Shenzhen Component Index, CSI 300, and Growth Enterprise Board Index. The selected index stocks are presented in Table 1 for your perusal.

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

TABLE 1. Certain index stocks.

The experiment utilizes a bidirectional stock selection method to choose equities from the Shanghai and Shenzhen markets, broadening the study's scope. Horizontally, companies are selected based on market capitalization, classified as large-cap or small-cap stocks; vertically, companies are selected according to stock categories, covering seven broad sectors including finance, real estate, coal, steel, non-ferrous metals, petrochemical, and automotive. The selected equities are presented in Table 2.

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

TABLE 2. Securities from the Shenzhen and Shanghai markets.

This information was gathered from JoinQuant and can be seen at https://www.joinquant.com/research. The index stocks were also included in this data. The information was obtained from Tushare (https://tushare.pro/) for a total of fourteen equities that are listed in Shanghai and Shenzhen. In this study, data on stocks were collected from the most recent 2,700 trading days for each stock. The time period covered by this study began in February 2012 and ended in May 2023. The information on the index stock comprises historical trading data with metrics such as the closing price, the highest price, the lowest price, the opening price, price fluctuations, changes, turnover, and volume. Two technical indicators, namely the 5-day moving average and the 10-day moving average, are included in the stock dataset for the Shanghai and Shenzhen equities. Additionally, the first six of these parameters are included in the dataset. In order to account for the large differences that existed between the various components of the stock dataset, it was required to normalize (33). The Z-score approach, which is stated in formula (3), was utilized for the purpose of standardization in this investigation.

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

The standardized value is denoted by yi, the input data is denoted by xi, the average of the input data is represented by x2, and the standard deviation of the input data is represented by s.

B. Evaluation of the Performance

The criteria for assessing procedures include the mean square error (MSE), mean absolute error (MAE), root mean square error (RMSE), and R-squared (R^2). The formulas for calculating these error metrics are shown in Equation (4).

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

yi is the value that is really being measured, yi^ is the value that is being expected, and y2 is the value that is being measured as the mean. When the MSE, MAE, and RMSE values are less, it indicates that the performance is higher. The value of R2 might be anything between 0 and 1. Its performance improves in proportion to the degree to which it approaches 1.

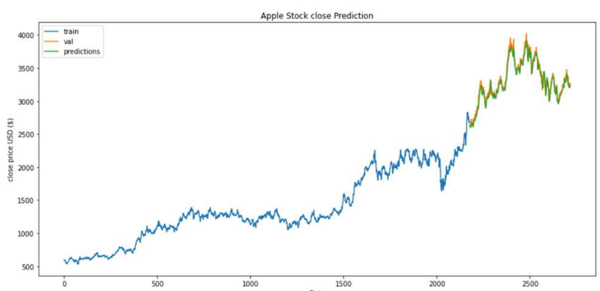
C. Parameters of the Network

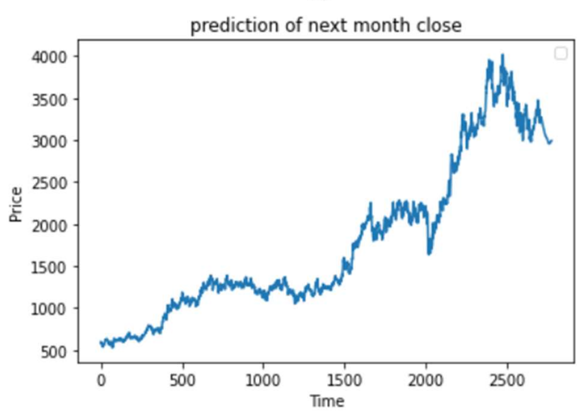
The experimental parametric settings of the BiLSTM-MTRAN-TCN model are outlined in Table 3. The Mean Squared Error (MSE) is used as the loss function, with Adam as the optimizer, and a learning rate of 0.00001. A window size of 10 is selected, allowing the prediction of the next day's stock closing price based on data from the previous week.

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 encoder layer	6
Number of TCN layer neurons	32
TCN later kernel size	1
Number of TCN hidden layer	4
Kernel size of TCN layer	7
Activation function of TCN layer	RELU

TABLE 3. The settings for the respective parameters of the BiLSTM-MTRAN-TCN technique.

V. RESULT





VI. CONCLUSION

This research paper presents the BiLSTM-MTRAN-TCN methodology for predicting the closing prices of investments in the stock market. Using this technique, the transformer model is modified by deleting the Input Embedding, replacing the original decoder with a TCN layer and a fully connected layer, and utilizing the encoder output as the only input for the decoder, with no additional inputs being used. Following the completion of the Position Encoding process, the data is initially processed by the BiLSTM in order to get sequence-dependent signals. After that, the data is sent to the modified transformer (MTRAN-TCN) for additional processing. In order to improve the accuracy of predictions, this hybrid model integrates different strategies in order to capitalize on their respective strengths and reduce their respective flaws.

The study investigates the efficiency of the BiLSTM algorithm, the influence of transformer adjustments, the precision of the approach, the generalization possibilities of the method, and the difficulties associated with the timeliness concept. The findings of empirical research indicate that the modified transformer (MTRAN-TCN) performs at its highest level when paired with BiLSTM, with BiLSTM having a substantial influence on the prediction of stock prices.

When it comes to the accuracy of the predictions, the BiLSTM-MTRAN-TCN technique does a better job than other models that are available in the published literature. LSTM, BiLSTM, CNN-BiLSTM, CNN-BiLSTM-AM, and BiLSTM-SA-TCN are some of the models that fall under this category. The scope of this study encompassed the investigation of fourteen stocks from the Shenzhen market as well as five benchmark shares from the Shanghai market. There were a total of fourteen equities that were selected using a bidirectional selection strategy. These stocks were selected from seven distinct broad groupings.

It has been demonstrated, on the basis of the results of the experiment that used index stocks, that the BiLSTM-MTRAN-TCN approach leads to an increase in R2 that ranges from 1.5% to 12.4% when compared to other models. During the stock tests that were carried out in Shanghai and Shenzhen, our technique achieved the highest R2 in 85.7% of the instances and the lowest RMSE in 78.6% of the cases. R2 increased by 0.3% to 17.6%, while the root mean square error (RMSE) decreased by 24.3%. Both of these changes occurred simultaneously.

Over the duration of the experiment on timeliness, which was conducted on five index stocks over the course of four unique time periods, the error index had a low degree of variation for the entirety of the trial. The results of the forecast were in agreement with one another. As a consequence of this, the BiLSTM-MTRAN-TCN method demonstrates a high degree of accuracy, the ability to generalize, and robustness in the process of handling new data without any problems pertaining to timeliness.

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 ARIMA-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. 1–6.

[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..