Comparative Study of LSTM and Transformer for A-Share Stock Price Prediction

Zhuoran Lin(✉)

Finance, Southern University of Science and Technology, Shenzhen, China
12010614@mail.sustech.edu.cn

Abstract. Forecasting stock prices in capital markets holds great significance for the economic development and social stability of a country, necessitating a robust predictive model. This study presents a comprehensive comparison between Long Short-Term Memory (LSTM) and Transformer neural network architectures for predicting A-share stock prices in the Chinese capital market. Although both LSTM and Transformer have demonstrated success in various applications, research comparing their performance in stock price prediction remains limited. In this paper, we employ both LSTM and Transformer models on daily and minute frequency data to predict the closing prices of the Shanghai Composite Index. Our findings indicate that although LSTM outperforms Transformer in terms of Mean Absolute Error and Mean Squared Error, it tends to simplify the problem by falling into the trap of autocorrelation. Conversely, Transformer learns unique dependencies, demonstrating potential for capturing the internal relationships of securities price changes. Our results suggest that Transformer is a more promising model for stock price prediction, as its self-attention mechanism may provide valuable insights to investors, financial practitioners, and fund managers. This study not only contributes to the existing literature on stock price prediction but also serves as a foundation for future research in the Chinese capital market, benefiting researchers and practitioners in finance, economics, and machine learning.

Keywords: Transformer · LSTM · A-share · stock prediction

1 Introduction

Stock price prediction is a crucial task for investors, traders, and fund managers as it holds significant implications for personal financial decision-making. Since the establishment of the securities market, various methods have been utilized to predict stock prices, including technical analysis, fundamental analysis, and machine learning algorithms. In recent years, the popularity of deep learning models such as Long Short-Term Memory (LSTM) and Transformer, along with their outstanding performance in various domains, has made them increasingly popular in stock price prediction.

LSTM, introduced by Hochreiter and Schmidhuber [1] in the 1990s, is a RNN with gate units particularly suited for processing sequential data. They have been widely applied in speech recognition, natural language processing, and time series prediction [2]. Based on the gate units, LSTMs can capture long dependencies in data and make

K. Hemachandran et al. (Eds.): ICAID 2023, AHIS 9, pp. 72–82, 2023.
https://doi.org/10.2991/978-94-6463-222-4_7

predictions based on patterns learned from historical data, which is particularly useful for exploring the patterns of stock price movements.

On the other hand, the Transformer, a neural network architecture introduced by Vaswani et al. in 2017 [3], is particularly suited for processing time-series data in a parallel and scalable manner. Unlike LSTMs that process input data sequentially, Transformers can process input data in parallel, thus, improving computational efficiency. Transformer has shown better performance than LSTM in natural language processing tasks such as machine translation and sentiment analysis [4]. Through multi-headed self-attention, Transformer can capture deeper patterns in time-series data.

Despite the recent success of LSTM and Transformer in various applications, researches comparing their performance in stock price prediction are still limited. Furthermore, research on this topic in the context of the Chinese capital market is relatively scarce compared to the rich research on price movements in Western markets. Given the unique characteristics and challenges of the Chinese stock market, including market volatility, information asymmetry, and regulatory policies, research on price movements in the Chinese securities market is particularly important [5].

The results of this study will provide valuable insights into the performance of LSTM and Transformer in predicting stock prices within the Chinese capital market. This will help us understand the advantages and disadvantages of these models in forecasting stock price fluctuations and offer a comparative analysis of their performance in the context of the Chinese capital market. The findings of this research will be of interest to scholars and practitioners in the fields of finance, economics, and machine learning. Furthermore, this study will lay the foundation for future research on stock price prediction in the Chinese capital market.

2 Related Work

2.1 Stock Price in China

The topic of price research in securities markets has been popular for several decades. In modern finance, Fama's Efficient Market Hypothesis (EMH) [6] provides the theoretical foundation for all current methods of modeling security prices. Fama's weak-form EMH, semi-strong form EMH, and strong-form EMH allow for more machine learning and neural network algorithms to be used in predicting stock prices. In comparison to Fama's research on the US securities market, Hung verified the market efficiency of the Shenzhen and Shanghai Stock Exchange AB shares through multiple variance ratio tests [7]. The results indicate that the Shanghai Stock Exchange A shares are weakly efficient, whereas the weak efficiency of the Shenzhen Stock Exchange AB shares and Shanghai Stock Exchange B shares does not hold. With scholars shifting their focus to behavioral finance, Lo's Adaptive Market Hypothesis (AMH) [8] combines behavioral finance with the EMH and studies the coexistence of rationality and irrationality in the market, as well as the unfair pricing of securities in the US securities market, providing a better explanation for the reasons behind abnormal fluctuations in security prices. On the basis of AMH, Lim examined the effectiveness of the Chinese securities market by testing the problem of false return caused by the autocorrelation of returns in the market [9]. These theoretical studies in finance lay the foundation for artificial intelligence methods to model security market prices, and provide assistance in data processing and selection in this paper.

2.2 Deep Learning in Stock Prediction

Due to the presence of activation functions, deep learning algorithms have strong modeling capabilities with nonlinearity. LSTM and Transformer have demonstrated outstanding performance in fields such as text recognition. These two models have powerful modeling effects on data with sequential information, and can learn deep information from sequential data. This is also the main reason for their application in complex time series modeling problems such as stock price prediction. Lu et al. modeled and predicted the Shanghai Composite Index using a CNN-LSTM model based on LSTM. Sunny et al. also used bi-directional LSTM for Google stock price modeling [10]. Regarding the application of these two models in stock price, Althelaya et al. compared the short-term and long-term predictions of these two models on the Standard & Poor 500 Index (S&P500) dataset, and found that BLSTM performs better than LSTM in both short-term and long-term predictions with the same number of neurons [11]. Transformer has become popular in direct modeling of time series in recent years. In order to solve the problem of poor feature extraction ability of traditional RNN in high-noise stocks, Li & Qian improved Transformer inspired by frequency decomposition, and achieved good results in modeling and predicting the Shanghai-Shenzhen 300 Index [12]. Similarly, Ding et al. improved the multi-head self-attention mechanism, enabling Transformer to perform better in modeling the Nasdaq Index and A-shares [13].

Nowadays, many models derived from LSTM and Transformer have emerged, but there is a lack of horizontal comparison of their basic models, which hinders the selection of appropriate models for specific scenarios by relevant field workers. This paper will provide a comparison of the two models on different dimensions to provide references for relevant workers.

3 Methodology

In this study, we compared the performance of two deep learning models, LSTM and Transformer, in modeling and predicting the time series data of A-share stock prices.

3.1 LSTM

The LSTM model is a type of recurrent neural network (RNN) designed for modeling sequential data [1]. It utilizes gate units to control the flow of information, addressing the issues of gradient explosion and gradient vanishing encountered in traditional RNNs. Figure 1 demonstrates the basic structure of the LSTM. The LSTM architecture is composed of input gates, forget gates, and output gates, which selectively allow or discard information flow through the network. Through these gate units, the LSTM model can better handle long sequence information and has a better memory effect for recurring patterns. This is a crucial factor in exploring long-term dependencies in stock price modeling using the LSTM model.

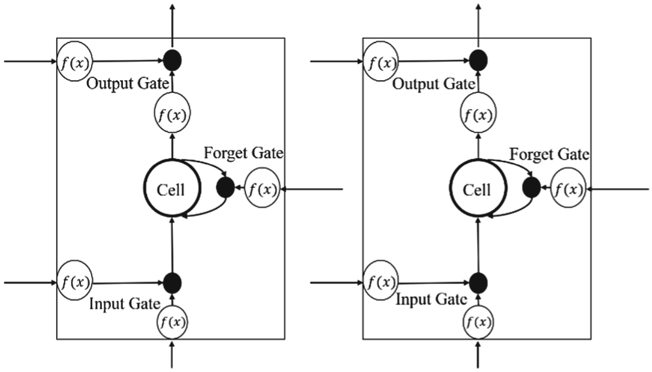

Fig. 1. Framework of LSTM [1]

LSTM with deeper architectures can learn more complex relationships in the data [14]. It is also essential to find the right balance between model complexity and performance. So we adjust the number of stacked LSTM layers in our model to fit the datasets with different size.

3.2 Transformer

The Transformer is a type of deep learning model designed specifically for handling sequential data, such as that found in natural language processing tasks. Predicated on a multi-head self-attention mechanism, the Transformer model allows each element in the input sequence to interact with all other elements, thereby understanding their relationships more effectively and capturing richer temporal sequence information [3]. This mechanism is visually represented in Fig. 2. The versatility and efficiency of Transformer models have led to their extensive use in a variety of tasks, including but not limited to machine translation, text classification, and language generation.

Unlike LSTM that processes data sequentially, the Transformer performs better when processing long sequence data. It can parallel process all data, thereby better capturing long-range dependencies. Therefore, it has become a mainstream choice in natural language processing tasks. In this study, we explore whether the Transformer can continue to maintain its advantages in this type of modeling task by using stock price sequence modeling.

3.3 Evaluation Metrics

There are many evaluation metrics to choose from when assessing the accuracy of prediction results. We have chosen two common and concise evaluation metrics, MAE and MSE. Through these two metrics, we can directly compare the accuracy of the prediction results of two models.

MAE, as a measure of absolute error, pays more attention to the size of the difference between the predicted value and the true value. It measures the average magnitude of errors between the predicted values and the actual observed values, without considering

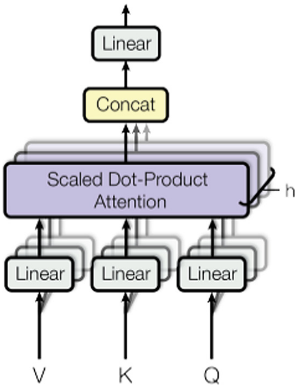

Fig. 2. Multi-Head Attention [3]

the direction of the errors. In other words, MAE provides an estimate of how far the predicted close prices are from the actual outcomes, on average. We can directly observe the size of the difference between the actual stock price and the predicted value through MAE. The formula to compute MAE is:

$$MAE = \frac{\sum_{i=1}^{n} \left| y_i - \widehat{y_i} \right|}{n} \tag{1}$$

MSE, as a measure of squared error, it measures the average squared difference between the predicted values and the actual observed values. By squaring the errors, MSE emphasizes larger errors, making it sensitive to outliers or significant deviations in predictions. The formula to compute MSE is:

$$MSE = \frac{\sum_{i=1}^{n} \left(y_i - \widehat{y_i} \right)^2}{n} \tag{2}$$

For MAE, it does not emphasize large errors compared to other metrics like MSE. Since MSE squares the errors, the metric can be more sensitive to outliers than other metrics like Mean Absolute Error (MAE). Additionally, MSE is not as straightforward to interpret as MAE, as its unit is squared compared to the original data. So we use these two complementary metrics to evaluate the prediction results of LSTM and Transformer.

4 Experiments

4.1 Data Selection

According to the Adaptive Market Hypothesis (AMH), modeling market efficiency as a variable is more suitable for A-share market conditions. However, currently there are no accurate variables that can describe market efficiency, so we directly use A-share volume-price data: opening price, closing price, highest price, lowest price, trading volume, and turnover rate for fitting, hoping that LSTM and Transformer can find the relationship of efficiency changes from volume-price indicators. We chose the Shanghai Stock Index as our dataset. The Shanghai Stock Index is a stock index composed of a batch of stocks selected by the Shanghai Stock Exchange based on their large market size, good liquidity, and strong representativeness. The Shanghai Stock Index is one of the main reference indicators for the Chinese stock market, so we use it to measure the overall performance and market trend of the Chinese stock market [15].

In order to improve the predictive capabilities of our models, the proper scale of dataset is of necessity. In financial markets, technical analysis is an essential component of traditional stock price prediction. Traditional technical analysis indicators based on daily trading data, such as MACD and EMA, have been applied to stock price prediction problems using machine learning algorithms like SVM [16]. However, M. Wu and X. Diao [17] find that have found that technical analysis based on these indicators does not yield significant predictive results in the Chinese capital market. To allow LSTM and Transformer models to identify more accurate predictive signals compared to these indicators, we first selected the 1-day frequency data of the Shanghai Stock Index as our dataset.

While daily trading data provides long-term trends of stock price changes and lower noise, high-frequency trading data has played an increasingly important role in stock trading with the advancement of big data technology. In addition to displaying real-time volume and price information, high-frequency data performs well in reflecting investor sentiment and conducting risk assessment [18]. Consequently, we also employed minute-level trading data of the Shanghai Stock Index as our dataset. Apart from providing more detailed and accurate information, high-frequency data substantially increases the dataset's capacity, which is highly advantageous when training LSTM and Transformer models. A larger and more diverse dataset ensures that the LSTM model learns from various market conditions and scenarios, making it more robust and accurate in its predictions [19]. In the case of the Transformer architecture, its attention mechanism and multi-head self-attention allow it to capture long-range dependencies and intricate correlations between features, thus making it well-suited for large-scale financial data analysis [20]. Simultaneously, large-scale datasets can alleviate the overfitting problem for LSTM and Transformer models [21]. Therefore, we also chose the 1-min frequency data of the Shanghai Stock Index as another dataset for our study.

The dataset is divided into two categories: low-frequency daily trading data and high-frequency minute-level data. The daily trading data spans from 2002/12/17 to 2022/12/17, while the minute-level data spans from 9:30 on 2019/12/23 to 15:00 on 2022/12/23.

4.2 Data Pre-processing

During the daily trading process, there is a period of collective bidding before the continuous bidding phase, in which buy and sell orders are executed immediately if a matching counterparty order is available at the specified price. During collective bidding, traders submit their buy and sell orders, and the prices are not yet determined. At the end of the collective bidding phase, the opening price of each stock is determined based on the accumulated buy and sell orders. This opening price aims to maximize the trading volume and minimize the price difference between buy and sell orders. In the Chinese

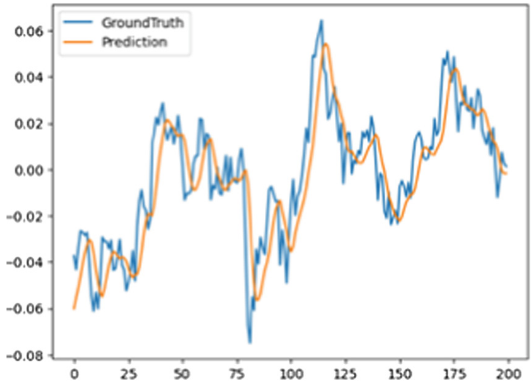

Fig. 3. LSTM Prediction Under Daily Frequency Data

stock market, the collective bidding phase takes place before the market opens in the morning. Although there is no transactions in non-trading period, the expectation of the stocks may still changed. Indeed, the announcement of industry-related policies and the broadcasting of crucial news generally occur during non-trading hours [22]. This implies that all information from the end of the previous trading day to the opening of the next trading day is reflected through the collective bidding process. The information contained in this process is exceedingly diverse and complex, making it difficult for deep learning models to capture. Therefore, we choose to remove the trading data within the collective bidding period to prevent it from influencing the pattern learning during the continuous bidding stage.

And then we standardized all data using the StandardScaler() in sklearn. Standardization makes the data easier to train because the data range is limited to a relatively small interval, which helps the optimizer find the global optimal solution more quickly. This is helpful for fitting stock prices that change over time and ensuring high computational efficiency [23].

Furthermore, scaling can reduce the influence of outliers on the model. Although irrational market behavior can cause prices to enter abnormal ranges, which is a constant occurrence in the real financial world, we still want to obtain a universally applicable model. Scaling can map outliers to a range closer to normal values, thus reducing their impact on the model and improving its robustness.

Lastly, scaled data have better comparability because they are within the same numerical range. This helps the model better capture the relationship between different features, thus improving the model's accuracy.

5 Experiment Result

Through experiments, we employed LSTM and Vanilla Transformer to predict the closing price of the Shanghai Composite Index. The four figures presented in this study show the performance of LSTM on the training and validation sets. Figure 3 and Fig. 5 exhibit the prediction results of LSTM on daily and minute frequency data, respectively. From

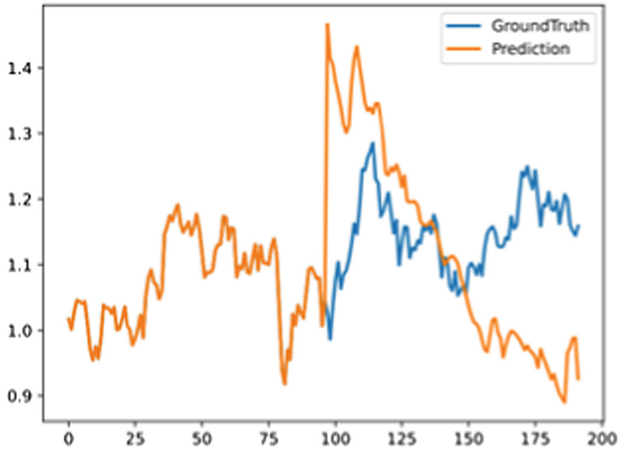

Fig. 4. Transformer Prediction Under Daily Frequency Data

Fig. 5. LSTM Prediction Under Minute Frequency Data

the graphs, it seems that LSTM captures long-term dependencies of stock prices very well.

However, upon closer inspection, we found that LSTM suffers from a lag problem, meaning that its prediction for the closing price at t + 1 is actually the closing price at t. This problem is a common issue in sequential modeling of stock prices due to their autocorrelation [24]. As LSTM aims to minimize the loss function, predicting the lagged result becomes the easiest and most common choice. Thus, it appears that LSTM has not learned the interdependencies among stock prices but rather took a shortcut. In this situation, the impressive performance of LSTM in terms of MSE and MAE shown in Table 1 seems to be of little significance.

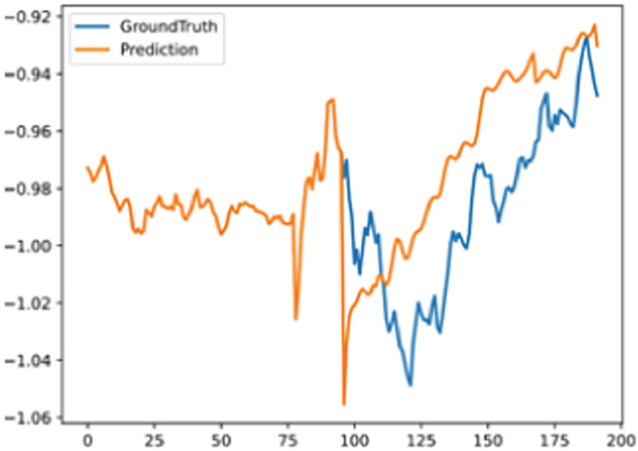

Fig. 6. Transformer Prediction Under Minute Frequency Data

Table 1. MAE and MSE of LSTM and Transformer in different frequency data

	Daily Frequency Data		**Minute Frequency Data**	
Model Types	**MAE**	**MAE**	**MSE**	**MSE**
LSTM	0.04434045	0.00453911	0.00961478	0.00022005
Transformer	0.27439132	0.11555387	0.04452187	0.00365713

On the other hand, Fig. 4 and Fig. 6 illustrate the prediction results of Transformer on the training and validation sets. Figure 4 and Fig. 6 present the prediction results of Transformer on daily and minute frequency data, respectively. Observing the graphs directly, we found that Transformer performs well on the training set, which is consistent with its characteristic of being less prone to overfitting compared to LSTM [25]. However, its prediction results on the validation set are not very satisfactory.

Furthermore, we found that predicting with daily frequency data yields unfavorable results as the trend of the index is opposite to the real movement trend. Nevertheless, using minute-level data successfully predicted the trend of stock price movement. From Table 1, we also observe that the Transformer model not performed better than LSTM in terms of MAE and MSE on the minute frequency data.

Overall, unlike LSTM, Transformer does not simply lag behind the real prices by one period for prediction. We could argue that Transformer has learned certain patterns in the stock prices, although the results are not ideal. From the perspective of learning dependencies, Transformer is a better choice.

6 Conclusions

In this article, we compared the performance of LSTM and Transformer in A-share stock price prediction. Based on intuitive performance evaluation, LSTM outperformed Transformer. However, judging from whether the long-term relationships within the sequence were learned, LSTM clearly simplified the problem and fell into the trap of autocorrelation. In contrast, Transformer did not fall into the pattern of autocorrelation, but instead learned a certain unique dependency. This indicates that Transformer has the potential to learn the internal dependencies of securities price changes. Therefore, we believe that Transformer is a more promising model, and based on its self-attention mechanism, it is highly likely to provide some help to investors, financial practitioners, and fund managers in studying stock prices.

References

1. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.
2. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," Neural computation, vol. 12, no. 10, pp. 2451-2471, 2000.
3. A. Vaswani et al., "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.
4. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.
5. J. Fernald and J. H. Rogers, "Puzzles in the Chinese stock market," Review of Economics and Statistics, vol. 84, no. 3, pp. 416-432, 2002.
6. E. F. Fama, "The behavior of stock-market prices," The journal of Business, vol. 38, no. 1, pp. 34-105, 1965.
7. J.-C. Hung, "Deregulation and liberalization of the Chinese stock market and the improvement of market efficiency," The Quarterly Review of Economics and Finance, vol. 49, no. 3, pp. 843-857, 2009.
8. A. W. Lo, "The adaptive markets hypothesis," The Journal of Portfolio Management, vol. 30, no. 5, pp. 15-29, 2004.
9. K.-P. Lim, M. S. Habibullah, and M. J. Hinich, "The weak-form efficiency of Chinese stock markets: thin trading, nonlinearity and episodic serial dependencies," Journal of Emerging Market Finance, vol. 8, no. 2, pp. 133-163, 2009.
10. M. A. I. Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using LSTM and bidirectional LSTM model," in 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020: IEEE, pp. 87–92.
11. K. A. Althelaya, E.-S. M. El-Alfy, and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," in 2018 9th international conference on information and communication systems (ICICS), 2018: IEEE, pp. 151–156.
12. C. Li and G. Qian, "Stock Price Prediction Using a Frequency Decomposition Based GRU Transformer Neural Network," Applied Sciences, vol. 13, no. 1, p. 222, 2022.
13. Q. Ding, S. Wu, H. Sun, J. Guo, and J. Guo, "Hierarchical Multi-Scale Gaussian Transformer for Stock Movement Prediction," in IJCAI, 2020, pp. 4640–4646.
14. M. Hermans and B. Schrauwen, "Training and analysing deep recurrent neural networks," Advances in neural information processing systems, vol. 26, 2013.

15. S. Yao, S. L. Morgan, and D. Luo, "Shanghai stock exchange composite index and bank stock prices in China: A causality analysis," University of Nottingham Research Paper, no. 2008/25, 2008.
16. I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," Artificial Intelligence Review, vol. 53, no. 4, pp. 3007-3057, 2020.
17. M. Wu and X. Diao, "Technical analysis of three stock oscillators testing MACD, RSI and KDJ rules in SH & SZ stock markets," in 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), 2015, vol. 1: IEEE, pp. 320–323.
18. X. Zhou, Z. Pan, G. Hu, S. Tang, and C. Zhao, "Stock market prediction on high-frequency data using generative adversarial nets," Mathematical Problems in Engineering, 2018.
19. E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," Expert Systems with Applications, vol. 83, pp. 187-205, 2017.
20. G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, and C. Eickhoff, "A transformer-based framework for multivariate time series representation learning," in Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 2114–2124.
21. C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 843–852
22. G. S. Oldfield and R. J. Rogalski, "A theory of common stock returns over trading and non-trading periods," The Journal of Finance, vol. 35, no. 3, pp. 729-751, 1980
23. V. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the influence of normalization/transformation process on the accuracy of supervised classification," in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020: IEEE, pp. 729–735.
24. I. Martin, "On the autocorrelation of the stock market," Journal of Financial Econometrics, vol. 19, no. 1, pp. 39-52, 2021.
25. A. Zeyer, P. Bahar, K. Irie, R. Schlüter, and H. Ney, "A comparison of transformer and lstm encoder decoder models for asr," in 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2019: IEEE, pp. 8–15.