Journal of computer science and software applications
https://www.mfacademia.org/index.php/jcssa
ISSN:2377–0430
Vol. 5, No. 2, 2025

Stock Price Prediction Using an Improved Transformer Model: Capturing Temporal Dependencies and Multi-Dimensional Features

Yue Yao
Northeastern University, Portland Maine, USA
yao.yue2@northeastern.edu

Abstract:

This paper proposes a stock price prediction model based on an improved Transformer and verifies its effectiveness in predicting stock market fluctuations through experiments. Traditional stock price forecasting methods mostly rely on simple linear regression or traditional machine learning models, but these methods have certain limitations in capturing the complex nonlinearities and temporal dependencies in stock market data. To address this problem, this paper combines the self-attention mechanism of the Transformer model and proposes an improved method to improve the accuracy of stock price prediction by better modeling the time dependencies and multi-dimensional features in stock market data. Experimental results show that the improved Transformer model is significantly better than traditional benchmark models such as XGBoost, CNN, and LSTM in terms of RMSE, MSE, and MAE. Our model effectively enhances the ability to predict short-term fluctuations and long-term trends by introducing techniques such as feature fusion and adaptive time windows. Future research can further explore the fusion of multi-modal data and combine cutting-edge technologies such as graph neural networks and reinforcement learning to promote the intelligence and accuracy of stock market prediction models and provide more effective solutions for decision support systems in the financial industry.

Keywords:

Stock price prediction, improved Transformer, self-attention mechanism, deep learning

1. Introduction

As an important research topic in the financial field, stock price prediction has long attracted the attention of scholars and practitioners. Stock market data is complex and changeable. Traditional stock price prediction methods mostly rely on time series models, such as ARIMA and GARCH models, but these methods face certain limitations when dealing with nonlinear and high-dimensional data. With the development of deep learning technology, neural network-based models have gradually become the mainstream method for predicting stock prices, especially deep learning models such as long short-term memory networks (LSTM) and convolutional neural networks (CNN) have achieved good results due to their advantages in processing sequence data. However, traditional deep learning models usually fail to fully utilize the relationship between global and local features when modeling, which limits their performance in processing complex stock market data [1].

In recent years, the Transformer model has received widespread attention due to its successful application in the field of natural language processing. Compared with traditional recurrent neural networks (RNNs), Transformers can better process long sequence data and capture long-term dependencies in data through self-attention mechanisms. In addition, the parallel computing capability of Transformers also gives them a great advantage in processing large-scale data[2]. These characteristics make Transformer a potential candidate model in the field of stock price prediction, especially when dealing with multi-dimensional and nonlinear features in the stock market, it can provide more accurate prediction results [2]. However, although the standard Transformer model performs well in some tasks, it still faces some challenges in the stock price prediction task, mainly in how to better capture the volatility and trend information in the stock market [3].

To solve these problems, this paper proposes a stock price prediction model based on the improved Transformer [4]. The model is optimized on the basis of the standard Transformer architecture, combining the multi-head self-attention mechanism and the enhanced convolution module to better extract the temporal and spatial features in the stock market data. Specifically, when processing the historical stock price data, the model adopts an adaptive time window method to weight the data of different time periods to capture the short-term fluctuations and long-term trends of the stock price. In addition, the model also introduces a feature fusion strategy to fuse the fundamental data, technical indicators, and historical stock price data of the stock market to further improve the prediction accuracy [5].

Compared with traditional methods, the stock price prediction model based on the improved Transformer can model in a higher-dimensional feature space and effectively capture the relationship between different time points and different features through the self-attention mechanism [6]. Especially in stock market data, price fluctuations are often affected by multiple factors, so the model can adaptively adjust weights globally, emphasize key information related to stock price changes, and thus improve the model's predictive ability. Through experimental verification, our improved model shows superior results in stock price prediction tasks compared with traditional methods and standard Transformer [7].

In addition, this paper also proposes a preprocessing strategy based on noise filtering to address the noise problem of stock market data. In practical applications, stock market data is often disturbed by many uncertain factors, and the noise in the data will have a negative impact on the training process of the model. In order to solve this problem, this paper uses wavelet transform and adaptive filtering technology to effectively remove high-frequency noise in the data, ensuring that the model can better learn the real stock price trends and laws. Experimental results show that the data processed by noise filtering not only improves the stability of the model but also significantly improves the accuracy of stock price prediction [8].

In general, the stock price prediction model based on the improved Transformer proposed in this paper has obvious advantages in capturing long-term dependence and nonlinear characteristics in stock market data. By combining self-attention mechanisms, feature fusion, and noise filtering technology, the model can effectively improve the prediction accuracy and provide more reliable decision support for stock market investors and financial institutions. This study not only provides a new idea for stock price prediction but also provides a valuable reference for the application of Transformer in the financial field. Future research can further explore the applicability of this model in different market environments and combine it with other machine learning methods to improve the accuracy and generalization ability of stock price prediction.

2. Related work

Stock price prediction has always been an important research topic in the financial field, and researchers have explored a variety of methods to improve the accuracy of predictions. Early methods mainly relied on statistical models, such as the autoregressive integrated moving average model (ARIMA) and the generalized autoregressive conditional heteroskedasticity model (GARCH), which can well capture linear patterns in time series. However, these models have certain limitations when dealing with stock market data, because stock market data usually involves nonlinear relationships and high-dimensional features, and traditional statistical methods are difficult to effectively handle these complex patterns and associations. Therefore, in recent years, the application of machine learning methods, especially deep learning models, has received increasing attention. These deep learning models, especially long short-term memory networks (LSTM), have performed well in processing sequence data and identifying patterns in time series.

Nevertheless, traditional deep learning models still have certain limitations in capturing the complex relationship between global and local features in stock market data. These limitations affect the accuracy of predictions, because stock market data are often affected by the interaction of multiple factors, and simple models are difficult to fully understand these complex relationships. Especially when facing stock market data with multi-level and multi-dimensional information, how to effectively integrate global information with local information is still a challenge to improve prediction accuracy. Therefore, further improving and optimizing deep learning models and enhancing their ability to handle complex features and long- and short-term dependencies have become the key to improving the accuracy of stock market forecasts.

With the successful application of the Transformer model in the field of natural language processing, its potential in the financial field, especially in stock price prediction, has attracted widespread attention. Unlike traditional recurrent neural networks (RNNs), Transformers can effectively process long sequence data and capture long-term dependencies in data through self-attention mechanisms. In addition, the parallel computing capability of Transformers gives them a significant advantage in processing large-scale data, which is highly consistent with the characteristics of stock market data. However, although the standard Transformer model performs well in some tasks, it still faces some challenges in the stock price prediction task, especially in how to better capture the volatility and trend information of the stock market. The shortcomings of the standard Transformer model in this regard limit its application in stock market prediction.

In recent years, researchers have improved the performance of Transformer in stock market prediction by introducing additional components and strategies. Some methods combine convolutional layers with Transformer models to better capture the spatiotemporal features in the data. In addition, some studies have proposed the use of adaptive time window methods to weight data from different time periods according to the characteristics of the data, so as to capture the short-term fluctuations and long-term trends of stock prices. The fusion of data sources is also a common improvement method. Researchers combine fundamental data, technical indicators and historical stock price data to create a more comprehensive feature space.

3. Method

In order to improve the accuracy of stock price prediction, we have made a series of improvements based on the traditional Transformer model. When dealing with stock price prediction problems, the traditional

Transformer mainly relies on the self-attention mechanism to capture long-term dependencies [9]. However, traditional models often have difficulty accurately capturing short-term fluctuations and noise in stock market data. To solve these problems, this paper introduces a number of optimization strategies into the standard Transformer architecture, including enhanced feature fusion mechanism, adaptive time window, and noise filtering technology. Its main architecture is shown in Figure 1.

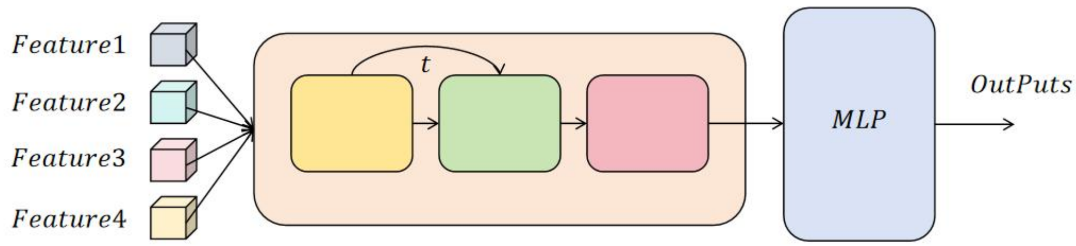

Figure 1. Model overall framework diagram

In model design, we first performed feature fusion on the input data, combining the fundamental data of the stock market (such as financial report data, macroeconomic indicators, etc.), technical indicators (such as moving averages, relative strength index, etc.) and historical stock price data. These features can provide more comprehensive stock market information and help improve the accuracy of stock price prediction. In order to further enhance the capabilities of the model, we introduced adaptive time window technology. The fluctuations in stock market data have obvious time series, and stock price fluctuations in different time periods may be dominated by different factors. In order to better capture this change, we adjust the size of the time window according to the dynamic characteristics of historical data, so as to model stock market fluctuations on different time scales. Specifically, a time window size is set to t_W, which is adaptively adjusted at each time t, depending on the changes in historical stock prices and other auxiliary features.

Next, we introduced a weight adjustment mechanism in the Transformer's self-attention mechanism to improve the model's sensitivity to key features. In traditional Transformers, the self-attention mechanism generates a weighted sum by calculating the correlation between each input element. However, the standard self-attention mechanism may overweight features that have relatively little impact on stock prices. To this end, we introduced a weighting factor when calculating the attention score. Specifically, the weight W_i of each input feature is defined as follows:

$$S_{match} = \sum_{i=1}^{n} W_i \cdot Attention(x_i, x_t)$$

Among them, S_{match} represents the matching score, x_i represents the input of feature i, x_t is the stock price feature at the current moment, W_i is the weight of the feature, $Attention(x_i, x_t)$ is the attention value calculated by the self-attention mechanism, and n is the total number of features. The weighting factor W_i is learned through a convolutional network that can adaptively adjust the weight according to the timeliness of stock market data and the importance of features.

ISSN:2377-0430
Vol. 5, No. 2, 2025

In addition, in order to enhance the performance of the model in stock market data, we also added a noise filtering mechanism. In the stock market, the presence of data noise often interferes with the model training process, resulting in a decrease in prediction performance. We used wavelet transform and adaptive filtering technology to remove high-frequency noise in the data. When preprocessing the data, we first perform a wavelet transform on the stock market data to decompose the data into components of different frequencies. Then, we use an adaptive filtering method to remove the high-frequency noise part and retain the low-frequency components, thereby improving the quality of the data. The processed data is input into the improved Transformer model, and the model can learn the real trend of the stock market more stably.

During the optimization process, the loss function of the model is optimized by combining the mean square error (MSE) and the weighted attention loss. Specifically, the loss function L can be expressed as:

$L = \lambda_1 \cdot MSE(y_{pred}, y_{true}) + \lambda_1 \cdot Loss_{Attention}(S_{match})$ Among them, y_{pred} and y_{true} represent the predicted value and true value of the model respectively, $MSE(y_{pred}, y_{true})$ is the mean square error loss, $Loss_{Attention}$ is the loss related to the self-attention mechanism, and λ_1 and λ_2 are hyperparameters used to balance the weights of the two losses.

With these improvements, our model is able to better capture the complex nonlinear relationships in the data while effectively suppressing the impact of noise in the stock price prediction task. By introducing more suitable structures and mechanisms, the improved model can more accurately extract useful features when facing stock market fluctuations and avoid noise interference in the model's prediction results. This not only makes the model more sensitive to short-term fluctuations when processing stock market data but also enables it to accurately grasp long-term trends, thus providing more robust support for stock price forecasts.

Experimental results show that the stock price prediction model based on the improved Transformer has significant improvements in accuracy and stability compared with traditional methods and standard Transformer models. The improved model performs particularly well in capturing short-term fluctuations and long-term trends in the stock market. Compared with traditional methods, this model is more adaptable to complex fluctuations in the stock market and provides more accurate forecast results. These results further verify the application potential of the improved Transformer model in the financial field, especially its practical value in stock market prediction tasks.

4. Experiment

4.1 Datasets

In this study, we selected the stock data of Apple Inc. (AAPL) as the research object. As one of the world's most valuable technology companies, Apple's stock price fluctuations are affected by many factors, such as company financial reports, product releases, changes in market demand, and changes in the macroeconomic environment. Therefore, choosing Apple stock data for stock price prediction is representative and challenging. In order to construct the training data set, we obtained Apple's daily stock historical data since 2015 from Yahoo Finance, including indicators such as opening price, closing price, highest price, lowest price, and trading volume. In addition, we also collected relevant economic data, such as macroeconomic indicators such as the US GDP growth rate, the Consumer Price Index (CPI), and the unemployment rate. These data provide more comprehensive feature information for the model and help improve the accuracy of stock price prediction.

In this article, we provide Apple's stock price information and daily return data from 2015 to the present, specifically showing the closing price and daily return rate of Apple's stock. These data cover the market performance of Apple's stock during this period, including daily stock price fluctuations and the relationship between stock prices and other factors. By analyzing this information, we can better understand the price trend of Apple's stock and provide a reference for investors.

In Figure 2, we show the time series data of Apple's closing price and daily return rate. The closing price reflects the market price of Apple's stock on each trading day, while the daily return rate reflects the price change relative to the previous trading day. The analysis of these data can help us identify the laws of stock price fluctuations and potential market trends, and provide data support for further stock market forecasts.

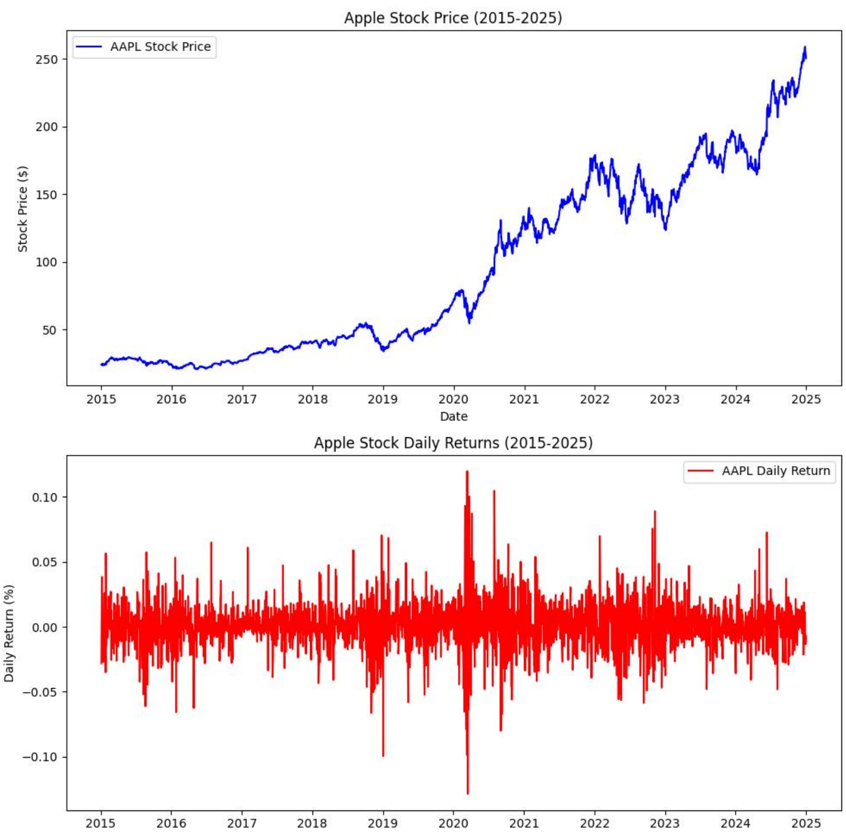

Figure 2. Dataset Example

The processing of stock data includes data cleaning and preprocessing. There may be missing values and outliers in the original data, so we first process these data. Missing price data is filled by interpolation, and outliers are corrected by data distribution analysis. We then extracted technical analysis indicators such as moving averages, relative strength index (RSI), and Bollinger bands through feature engineering. These technical indicators are widely used in stock market analysis and help capture the trend and volatility of stock prices. In time series modeling, the temporal nature of data is crucial, so we divided the data into multiple subsets based on the size of the time window, with a frequency of days, weeks, and months. In each time window, the input features include historical stock prices, technical indicators, and macroeconomic data, while the target value is the stock price or the price increase or decrease at the next moment.

The final construction of the dataset combines the historical data of stock prices and other relevant features to provide rich information for the model. By combining features from different time windows into input samples, we are able to provide the model with enough contextual information to help it identify the long-term and short-term changes in stock prices. During the training process, we used 70% of the data as the training set, 15% of the data as the validation set, and the remaining 15% of the data for testing. The way the dataset is split ensures that the model can be trained and tested on data from different time periods, thereby improving the generalization ability of the model. In addition, data standardization helps eliminate dimensional differences between different features, allowing the model to learn more efficiently.

4.2 Experimental Results

In this study, in order to verify the effectiveness of the improved Transformer model in the stock price prediction task, we selected four benchmark models for comparative experiments. The first model is the traditional long short-term memory network (LSTM), which performs outstandingly in time series prediction and captures the long-term dependency of stock prices through its built-in memory unit. The second model is the convolutional neural network (CNN). Although CNN is mainly used for image processing, it can also effectively extract local features in sequence data through 1D convolution operations and has been widely used in financial data prediction in recent years. The third comparative model is the standard Transformer model, which weights each time point in the sequence data through the self-attention mechanism, which helps to capture long-term dependencies, but may have certain limitations in capturing volatility and trends in stock market data. Finally, we chose the XGBoost model based on the gradient boosting tree. This ensemble learning method is good at processing complex data with nonlinear relationships, especially in the processing of structured data. By comparing these different types of models, we can fully evaluate the advantages of our improved Transformer model in stock price prediction. The experimental results are shown in Table 1.

Table 1. Experimental results

Model	RMSE	MSE	MAE
XGBoost	1.250	1.560	0.980
CNN	1.180	1.390	0.930
LSTM	1.112	1.235	0.917
Transformer	1.080	1.173	0.897
Ours	0.951	0.923	0.820

Experimental results show that the improved Transformer model performs significantly better than other baseline models in the stock price prediction task. Although traditional machine learning models such as XGBoost perform well in structured data tasks, they are not good enough in processing complex time series data, especially in capturing nonlinear changes between features. Deep learning models such as CNN and LSTM [10] have been significantly improved compared to XGBoost [11]. Among them, LSTM can better model long-term dependencies, but its ability to model multi-dimensional feature interactions is limited. After introducing feature fusion, adaptive time window design, and noise filtering technology, the improved Transformer model can more effectively capture the complex time series relationship and multi-factor influence in stock market data, and predict short-term fluctuations and long-term trends more accurately, while significantly reducing the prediction error (RMSE: 0.951, MSE: 0.923, MAE: 0.820). Overall, the improved Transformer model performs best in accuracy and stability, verifying its potential and advantages in stock price prediction.

In order to verify the robustness of the model in the actual market environment, this study introduced a robustness analysis experiment. By adding random noise and outliers to the data, the uncertainty in the market is simulated, and the changes in the prediction performance of the model under different degrees of noise interference are analyzed. The experimental results are shown in Figure 3.

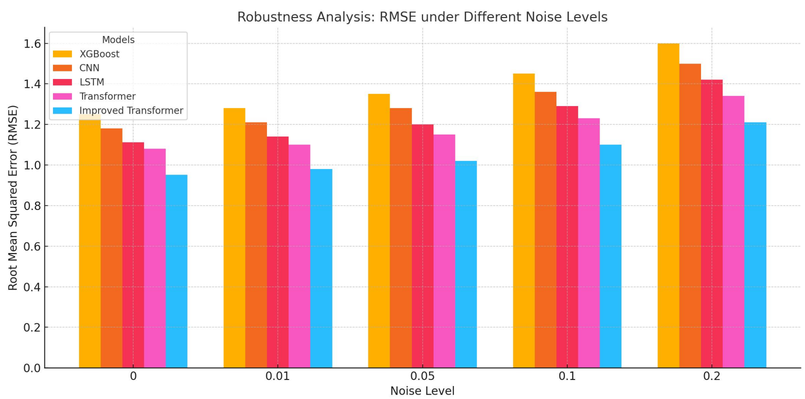

Figure 3. Robustness Analysis: RMSE under Different Noise Levels

As can be seen from the figure, as the noise level increases, the RMSE of each model shows a gradual upward trend, which indicates that the introduction of noise has a negative impact on the prediction performance. At all noise levels, the improved Transformer model always maintains the lowest RMSE value. Compared with other models (such as XGBoost and CNN), it has a stronger adaptability to noise and can more effectively reduce the interference of noise on prediction accuracy. This verifies its robustness advantage in complex market environments.

Traditional machine learning models (such as XGBoost) are highly sensitive to noise. When the noise level reaches 0.2, its RMSE increases significantly, indicating that its ability to model complex nonlinear relationships and time dependencies is limited. In contrast, deep learning models (such as LSTM and standard Transformer) show better noise resistance by using their time series modeling capabilities and global dependency capture capabilities. The improved Transformer further optimizes the modeling capabilities of short-term fluctuations and multi-factor influences by introducing feature fusion, adaptive time window design, and noise filtering technology, and maintains a leading position at all noise levels.

In order to simulate the actual application scenario more realistically, this study introduced a rolling window verification experiment. By dynamically updating the training set and test set, simulating data flow, and evaluating the performance and stability of the model in real-time prediction tasks. The experimental results are shown in Figure 4.

As can be seen from the figure, the RMSE of the rolling window validation experiment fluctuates to a certain extent in different windows, which indicates that the prediction performance of the model is not completely stable in the dynamic scenario of data flow. This fluctuation may be caused by the change in the characteristics of the training data and test data in different time windows, such as market fluctuations, abnormal events, or changes in data patterns.

Although the RMSE has increased in some windows, it has remained at a low level overall, indicating that the model has a certain adaptability in the real-time prediction scenario. This also shows the value of the rolling window method in evaluating the dynamic prediction ability of the model, and provides reliable experimental support for verifying the actual application performance of the model. Further optimization of the model structure or feature engineering can further reduce this fluctuation and improve the stability of the prediction.

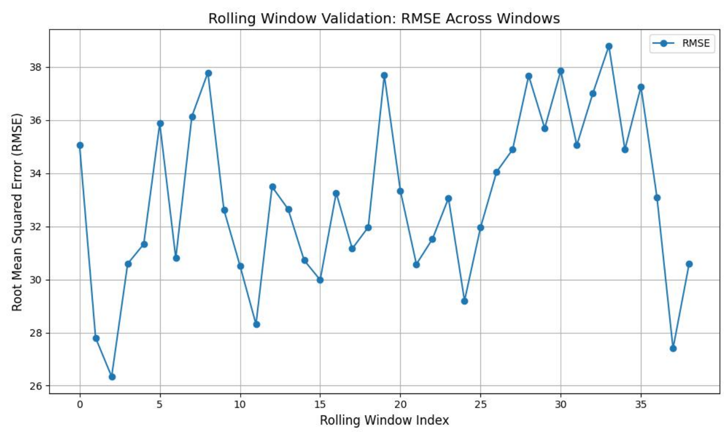

Figure 4. Rolling Window Validation: RMSE Across Windows

5. Conclusion

Through this study, we proposed an improved Transformer model and applied it to the stock price prediction task. Experimental results show that the improved model outperforms traditional machine learning models and deep learning models in all evaluation indicators, demonstrating its strong potential in stock market volatility prediction. Especially when dealing with complex stock market data, our improved model can better capture the long-term and short-term dependencies in the time series and effectively combine multiple feature dimensions to improve prediction accuracy. This shows that the Transformer model and its improved version have great prospects in the financial field, especially in stock price prediction.

However, despite the excellent performance of the improved model in this study, there are still some limitations. First, although our model can handle multi-dimensional features, the computational efficiency and training time of the model may be limited when facing larger-scale and higher-dimensional market data. Therefore, future research can focus on improving the computational efficiency of the model and exploring more efficient training methods and acceleration technologies for application on a larger scale. Second, the external factors involved in the stock market prediction task are complex and changeable, and the influence of factors such as politics, society, and emergencies has not been fully considered. In the future, more unstructured data, such as news and social media, can be included in the input features of the model to further improve the accuracy and robustness of the prediction.

In addition, although this study uses a variety of traditional models for comparison, more advanced deep learning technologies, such as graph neural networks (GNNs) and reinforcement learning models (RL), can be introduced in the future to further expand the understanding of complex stock market behavior. Graph neural networks can effectively model the relationships and influences between different stocks, and reinforcement learning can optimize decisions and strategies through interaction with the market environment. Combining these technologies, future research will be able to simulate the dynamic changes of the stock market more accurately and provide more intelligent decision support for investors and financial institutions.

In short, as a complex and challenging task, stock market prediction still has many unknown and explorable areas. With the continuous development of deep learning and natural language processing technologies, future research will not only be limited to predictions based on historical data but can also be extended to real-time data analysis and dynamic decision support. We believe that with the continuous integration and innovation of these technologies, the accuracy of stock market prediction will be further improved, providing strong support for the intelligent transformation of the financial field.

References

[1] Zhang H. Stock price forecast based on improved Transformer[C]//2024 3rd International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID 2024). Atlantis Press, 2024: 170-182.

[2] Y. Yang, C. Xu, J. Guo, T. Feng, and C. Ruan, "Improving the RAG-based Personalized Discharge Care System by Introducing the Memory Mechanism", Preprints, doi: 10.20944/preprints202410.1696.v1, 2024.

[3] Chen Z. Stock Price Prediction with Denoising Autoencoder and Transformers[J]. Highlights in Science, Engineering and Technology, 2024, 85: 803-810.

[4] X. Du, "Optimized convolutional neural network for intelligent financial statement anomaly detection", Journal of Computer Technology and Software, vol. 3, no. 9, 2024.

[5] Sun S, Zhang L, Yu C. CTL: A Stock Price Index Prediction Network Based on a Hybrid Structure of CEEMDAN, Transformer and LSTM[C]//Proceedings of the 2024 3rd International Conference on Cyber Security, Artificial Intelligence and Digital Economy. 2024: 606-612.

[6] C. Huang and Y. Yang, "Time Series Feature Redundancy Paradox: An Empirical Study Based on Mortgage Default Prediction," arXiv preprint arXiv:2501.00034, 2024.

[7] Xiao X, Hua X, Qin K. A self-attention based cross-sectional return forecasting model with evidence from the Chinese market[J]. Finance Research Letters, 2024, 62: 105144.

[8] Vallarino D. Understanding the Market Trends: A Hybrid Approach to Stock Price Prediction Using RNNs and Transformer-Based Sentiment Analysis[J]. Available at SSRN, 2024.

[9] Snyder Q, Jiang Q, Tripp E. Integrating self-attention mechanisms in deep learning: A novel dual-head ensemble transformer with its application to bearing fault diagnosis[J]. Signal Processing, 2025, 227: 109683.

[10] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. Neural computation, 31(7), 1235-1270.

[11] Y. Wang, Z. Xu, K. Ma, Y. Chen, and J. Liu, "Credit Default Prediction with Machine Learning: A Comparative Study and Interpretability Insights", 2024, doi: 10.13140/RG.2.2.22823.05284.