Comparative Analysis of Mainstream Stock Prediction Models

Xiaolin Wei *

College of Optoelectronics and Information Engineering, Fujian Normal University, Fuzhou, China

* Corresponding Author Email: bzesetqe@krae.edu.kg

Abstract. In response to the challenges in stock price prediction caused by the high noise and non-stationary characteristics of the financial market, this paper systematically studies the technical evolution path of deep learning hybrid models. By deconstructing the collaborative mechanisms of four architectures: Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN)-LSTM, Gate Recurrent Unit (GRU), and Temporal Convolutional Network (TCN)-Transformer, and using various evaluation conditions such as Mean Absolute Error and Root Mean Square Error to analyze the advantages of the models in stock prediction, the research further reveals the insufficiency of cross-market adaptability and the bottleneck of extreme event prediction. It proposes a phased technical path - in the near term, build a federated learning data framework to improve the utilization rate of small samples, and in the medium and long term, integrate causal inference and quantum acceleration mechanisms to drive the prediction model towards an industrial-level decision-making system. The results show that the Mean absolute error (MAE) index of TRAformer is 66.8% lower than that of the ordinary LSTM model, and the composite model supported by the ARO algorithm also performs better than the ordinary LSTM model.

Keywords: Stock Price Prediction, Deep Learning Hybrid Models, Hyperparameter Optimization, Time Series.

1. Introduction

Among various financial assets, stocks have become the focus of market attention due to their high liquidity, wide participation and sensitivity to economic fundamentals. Stock prices, as a direct reflection of the market value of listed companies, their dynamic changes not only reflect the operating conditions, profitability and development prospects of enterprises, but also contain comprehensive information such as macroeconomic trends, industry rise and fall, market sentiment and investors' expectations. Therefore, accurately understanding and predicting stock price fluctuations is of extremely significant importance for investors, enterprises, regulatory authorities, and even the healthy operation of the entire macro-economy. Stock price prediction is a central challenge in financial quantitative analysis, complicated by the high-noise, non-linear, and dynamically evolving nature of markets [1]. Traditional methods such as time series models and fundamental analysis are constrained by linear assumptions and lagging effects, struggling to capture sudden market fluctuations. With advancements in artificial intelligence, machine learning (SVM, Random Forest) and deep learning (LSTM, Transformer) have emerged as dominant approaches. These techniques enhance prediction accuracy by integrating multi-source heterogeneous data (market trends, news, options/futures as leading indicators). Nevertheless, market efficiency and the unpredictability of black swan events remain fundamental obstacles. The simple and single Long Short-Term Memory (LSTM) model cannot meet the requirements of larger data and more diverse and complex situations. Recent studies focus on multi-modal data fusion and hybrid model architectures, they significantly advancing predictive frontiers. Some researchers have attempted to combine other suitable models with the LSTM model, giving rise to new composite models. For example, the stock sequence array convolutional LSTM (SACLSTM) is a new model generated by combining the Convolution Neural Network (CNN) and LSTM models [2], this new model needs to integrate historical data, leading indicators and other factors affecting stock prices into a sequence array, and input this array as the input image into the SACLSTM framework for prediction. The prediction quality is superior to that of frameworks such as SVM, CNN-corr, and CNNpred. In addition, some researchers have upgraded the performance of LSTM by using meta-heuristic algorithms, such as LSTM-ARO. The optimized model is divided into two steps: detour foraging (exploration) and random hiding (exploitation), and its prediction performance performs well under the four scoring criteria of MSE, MAE, MAPE and R^2. For instance, Time-series Recurrent Neural Network (TRNN) builds upon the principles of RNN models. Its core approach is to establish the trading volume of a certain stock in the stock market and expand the influence of the trading volume data on the stock value [3]. This significantly improves the accuracy and efficiency of the calculations. In the field of deep algorithms, apart from LSTM models, Transformer models are also popular. Therefore, some researchers have updated them on this basis, such as TCN-Transformer [4] which is based on the TRAformer stock price prediction system and integrates various functions such as data processing, refined analysis, and price prediction.

In addition to the aforementioned deep learning models, there are many other machine learning methods that can be used in stock prediction systems, such as traditional machine learning techniques, neural networks, time series analysis and graph-based approach [5]. Different types of methods naturally have their own advantages and disadvantages. The integrated composite system can play to its strengths and avoid its weaknesses, integrating and bringing out the advantages of different machine learning. To a certain extent, this enables the new stock price prediction system to be free from the troubles of certain disadvantages and ultimately achieve the goal of improving the prediction efficiency of the system.

This paper systematically reviews the application of various machine learning methods in stock price prediction systems, as well as the technological evolution and cutting-edge trends of stock price prediction systems. By critically analyzing the applicable scenarios of statistical models, machine learning, and deep learning frameworks, and combining the effect of hyperparameter optimization (ARO/GA) to evaluate each stock price prediction system, Analyze the adaptability and performance advantages and disadvantages of various machine learning methods in the stock price prediction system. This article aims to provide better reference opinions and solutions for those who want to participate in stock market trading and need stock market prediction tools, and can also offer new ideas for developers of stock market prediction systems. And in response to the existing problems, directions such as considering the sentiment index were proposed, and the development path of the future stock price prediction model was initially planned [6].

2. Key Technology Classification Analysis

2.1. LSTM Model

LSTM is a special type of recurrent neural network (RNN) proposed by Hochreiter and Schmidhuber in 1997. It is designed to address the problem of gradient vanishing/exploding encountered by standard RNNs when dealing with long sequence data, thereby enabling the learning of long-term dependencies within the sequence.

2.1.1. Model Characteristics:

The core advantage of LSTM lies in its ability to handle long sequence dependencies, enabling it to capture relevant information that is separated by a considerable distance within the sequence. Secondly, it has a crucial gating mechanism and memory unit. Through the forget gate, input gate, and output gate, it finely controls the flow of information (performing four steps: remembering, forgetting, updating, and outputting information in sequence), and the memory unit provides a relatively stable information transmission channel to maintain long-term states. Additionally, as mentioned earlier, alleviating gradient vanishing or explosion is more stable in long sequences compared to standard RNNs.

2.1.2. Advantages in the stock price prediction system:

Stock price data is essentially a strong time series. LSTM can effectively model the short-term fluctuation patterns and potential medium-long-term trends between historical data points of price and trading volume. The influencing factors in the financial market are extremely complex and interact highly nonlinearly. LSTM, as a neural network, has a strong ability to approximate nonlinear functions and can learn these complex patterns, while traditional linear models (such as ARIMA) are difficult to handle.

2.2. CNN-LSTM Model

The hybrid neural network framework of CNN and LSTM was integrated into a new model, named Stock Sequence Array Convolutional LSTM (SACLSTM). The core objective is to enhance the accuracy of stock price prediction through multi-dimensional financial data. Its principle consists of three steps: The first step is to input data representation, convert multi-dimensional time series data such as historical stock prices, options, and futures into a two-dimensional matrix (similar to an image), where the matrix rows represent time windows (default 30 days) and columns represent different indicators (opening price, closing price, option settlement price, etc.). The second step is to feature extraction, it is divided into two parts, CNN and LSTM. The former uses convolutional layers (3x3 filters) to capture local patterns (such as short-term price fluctuations and technical indicator correlations), and compresses redundant information through pooling layers. The latter uses the feature vectors extracted by CNN as the time series input and learns long-term dependencies (such as periodic trends, market sentiment inertia). The third step is the prediction output.

2.2.1. Model Characteristics:

Convert financial time series data into a matrix-like image, leveraging the spatial feature extraction capability of CNN (breaking through the limitation of traditional RNN that only relies on time series). It introduces leading indicators such as options and futures as inputs for the first time, capturing market expectation signals (such as implied volatility of options indicating short-term fluctuations). Automatically extract high-order features, avoiding subjective biases in manual design of technical indicators (such as MACD, RSI).

2.2.2. Advantages in the stock price prediction system:

This model demonstrates significant advantages in financial time series prediction. Firstly, its high robustness is attributed to the application of convolutional neural networks (CNN). The inherent local perception feature of CNN combined with pooling operations can effectively identify local patterns in the data and suppress the interference of market noise, thereby enhancing the model's ability to capture the core signals. Secondly, the model adapts to complex market dynamics through a multi-scale time series modeling strategy. Specifically, the CNN layer focuses on extracting and capturing local patterns in short-term price fluctuations, while the long short-term memory network (LSTM) is good at learning the long-term dependencies and trend evolution in the sequence. The collaborative work of CNN and LSTM enables the model to simultaneously understand and utilize market information at different time scales. Finally, the improvement in model performance is significantly dependent on the fusion of information from multiple data sources. Empirical studies have shown that when the model not only utilizes historical price data but also integrates relevant information from options and futures markets, its prediction accuracy has been significantly improved. This multi-source information fusion provides the model with a more comprehensive and rich representation of market states, which is a key element in building more accurate prediction models. The structural framework of SACLSTM is shown in Figure 1.

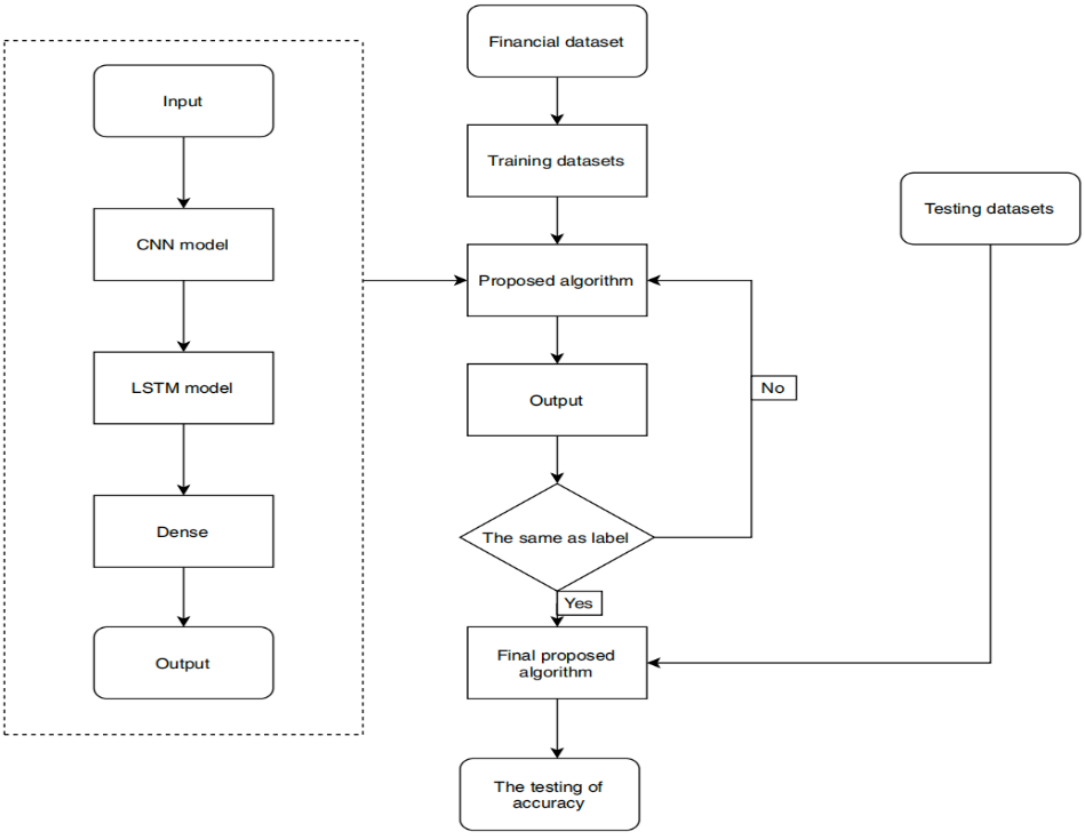

Fig 1. Flowchart of proposed SACLSTM [2].

2.3. GRU Model

Gate Recurrent Unit (GRU), also known as the gate-based recurrent unit, combines the forget gate and input gate of LSTM into an update gate, thereby reducing the total number of parameters (experimental results show that it has 24,000 fewer parameters than LSTM) and lowering the complexity of the model [7].

2.3.1. Model Characteristics:

The core strength of this model lies in its high efficiency and powerful feature processing capabilities. By reducing the number of parameters, the model significantly shortens the training period, enhances training and operational efficiency, making it particularly suitable for online real-time prediction scenarios that require rapid responses. Additionally, the model demonstrates outstanding multi-dimensional time series data processing capabilities, effectively integrating heterogeneous features such as prices, trading volumes, and various technical indicators, thereby significantly improving the accuracy of prediction results.

At the model mechanism level, the key lies in achieving a precise balance between long-term and short-term memory. This is mainly attributed to its built-in update gate and reset gate mechanism, which can dynamically and adaptively adjust the weight ratio between historical state information and current input data. Thus, it can capture long-term dependencies while also sensitively responding to short-term dynamic changes.

Experimental verification shows that this model has excellent adaptability across different time granularities. Whether it is low-frequency data at the daily level or high-frequency data at the 5-minute level, the model can maintain stable performance. Notably, on high-frequency trading data (such as 5-minute intervals), the model's prediction accuracy is particularly outstanding, with the maximum deviation between the predicted and actual values controlled within approximately 2%, fully demonstrating its effectiveness in fine-grained prediction tasks.

2.3.2. Advantages in the stock price prediction system:

Gated Recurrent Unit (GRU) is an improved version of RNN. Through update gates and reset gates, it dynamically adjusts the information flow, effectively capturing the time dependence of stock prices and avoiding the gradient disappearance problem of traditional RNN. The total number of parameters of the GRU model is significantly reduced compared to the LSTM model (24,000 fewer [4]), improving the time efficiency of the stock price prediction system.

2.4. TCN-Transformer Model

The model fusion architecture TRAformer combines the time-domain convolutional network (TCN) and the improved Transformer. The TCN module: expands the receptive field through dilated convolution to capture short-term local dependencies (such as intraday fluctuations) in the stock price sequence. Its residual connection structure alleviates the problem of gradient disappearance. The Transformer module: introduces probabilistic sparse multi-head self-attention (PS-MSA), reducing complexity through sparse computation. The improvement lies in the residual accumulation of attention scores, enhancing the ability to extract high-dimensional features.

2.4.1. Model Characteristics:

The core of this model architecture lies in its multi-scale feature extraction capability and efficient computational design. Specifically, the time convolutional network (TCN) is used to precisely capture short-term local patterns in the input sequence, such as subtle fluctuations in technical indicators; while the Transformer structure is adept at modeling long-term dependencies, such as the potential impact of macroeconomic factors on the market. These two structures do not operate independently but are integrated through complementarity to form a hybrid model that can comprehensively understand the dynamics of time series.

To address the computational challenges posed by long-term sequential data, the model adopted a dual-efficiency optimization strategy. On one hand, the Partial Sequence Multi-head Self-Attention (PS-MSA) mechanism was introduced, which significantly reduced the complexity of attention calculations by limiting the range of sequences involved in the attention computation. On the other hand, the Distillation Layers were employed to gradually reduce the dimension of the sequences, effectively alleviating the memory and computational burdens of processing long sequences, ensuring the model's feasibility in handling large-scale historical data.

In terms of enhancing the expression ability and stability of the model, an attention enhancement mechanism based on residuals was designed in the architecture. The core of this mechanism lies in the residual accumulation of cross-layer attention scores. This approach not only deepens the fusion depth between different hierarchical features, making the information transmission and integration more comprehensive, but also significantly improves the stability of gradient flow, reduces the risk of information attenuation during the training of deep networks, and ultimately enhances the overall performance and robustness of the model.

2.4.2. Advantages in the stock price prediction system:

It can handle 20-dimensional features (technical indicators + sentiment scores) simultaneously, overcoming the limitation of traditional models that rely on a single historical price, and making further upgrades on the basis of LSTM, breaking through the limitations of a single system. Supports real-time data synchronization and scheduled updates to ensure the timeliness of the prediction input. Can obtain stock price prediction results as quickly as possible for reference.

3. Literature Analysis Application

3.1. Dataset

In the research of the TCN-Transformer model, the data sources are: data comes from third-party financial data interfaces (such as Eastmoney website) and stock discussion platforms. The types and ranges of stocks are: select representative stocks from four industries in the Chinese A-share market from January 30, 2021 to January 30, 2024. Guizhou Moutai (600519, liquor industry), ZTE Communication (000063, communication industry), CITIC Securities (600030, financial industry), and Hengrui Medicine (600276, pharmaceutical industry). Each stock has approximately 728 data entries. In the research of LSTM network with artificial rabbit optimization algorithm, the data comes from Yahoo Finance API, covering the constituent stocks of the Dow Jones Industrial Average (DJIA), including 30 large companies from January 1, 2018 to January 1, 2023, such as Apple and Boeing, but specific stock codes are not listed. In the stock price prediction based on the GRU model, the data source of the dataset is not clearly specified, but the data includes historical trading data and technical indicators (such as MACD), mainly selecting two stocks from the Chinese A-share market, namely Baofengchuang (002371) and SMIC (688981), but the details of the number of data entries or time range are not mentioned. The main difference lies in the geographical and market differences of the stocks, which are the same as the language region used in the researchers' published papers, reflecting the differences in stock prices under different market structures and national conditions. Additionally, the time span of the selected stock price data is different. A longer time span of the selected data may improve the adaptability of the research model to the economic cycle, but the time span is not a criterion for testing the model's predictive ability. The datasets selected in these three studies reflect different research focuses. In the TCN study, the emphasis is on multi-source data fusion (technical indicators + sentiment analysis) to improve the prediction accuracy in complex market environments, suitable for the policy-sensitive Chinese A-share market. The combined algorithm research of artificial rabbit and LSTM uses a simple price sequence and long-term data to validate the algorithm (ARO optimized LSTM) [8], highlighting computational efficiency and generalization ability, suitable for mature markets such as the US stock market. The research on stock price prediction based on the GRU model processes two Chinese A-share stocks with high-frequency data and basic technical indicators (MACD), optimizing short-term predictions, suitable for real-time trading systems.

3.2. Evaluation Index

The main evaluation indicators include error metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) [5]. These metrics are used to quantify the deviation between the predicted values and the actual values. Fit indicators such as the Coefficient of Determination (R^2) [9] are used to measure the model's ability to explain variance. The higher the R^2 value, the better the fitting effect of the model. Mean Squared Error (MSE) has become a fundamental indicator for evaluating the accuracy of prediction models and the performance of estimators due to its clear mathematical definition, good differentiability, and significant penalty for large errors. It decomposes the model error through bias-variance decomposition, providing profound theoretical insights into the sources of model errors. However, when researchers and practitioners apply MSE, they must fully recognize its sensitivity to outliers and the issue of unit interpretability, and should carefully select or supplement it with other evaluation indicators (such as MAE, RMSE, R^2, MAPE, etc.) based on the specific data characteristics and requirements of the problem to obtain a more comprehensive and robust assessment of model performance [10].

Additionally, when evaluating a new model that combines two models, it will be specifically compared with the separate models before the fusion, for example, when evaluating the TRAformer model, it will be compared separately with the TCN model and the Transformer model before its combination.

3.3. Analysis of experimental results

First, the experimental results of the TRAformer model are analyzed. Taking the results on the ZTE communication dataset as an example (Table 1), MAE = 0.842 (better than 2.533 of LSTM and 2.947 of Transformer), MAPE = 3.035% (reduced by 6.35 percentage points compared to LSTM). This indicates that the TCN enhances local feature extraction and the residual attention mechanism effectively improves the prediction accuracy. The model's prediction effect is highly consistent with the actual stock price curve, verifying the effectiveness of multi-feature input (price, technical indicators, sentiment score). The experimental data are shown in Table 1 [4].

Table 1. Experimental results of the comparison model.

Model	ZTE Communication dataset		
	MAE	RMSE	MAPE (%)
TRAformer	0.842	0.988	3.035
Transformer	2.947	3.644	9.699
Informer	4.432	4.637	15.933
Reformer	2.008	2.300	7.258
LSTM	2.533	2.825	9.385

In the LSTM-ARO model, its performance outperforms the benchmark model for most stocks (Table 2). Taking Apple's stock as an example [7]:

Table 2. Comparison data of ARO-LSTM with other models.

AAPL	MAE	MAPE	R^2
LSTM-ARO	3.846	0.025	0.857
LSTM-GA	3.955	0.026	0.848
LSTM1D	5.851	0.037	0.675
ANN	7.673	0.050	0.458

MSE = 22.731 (lower than 24.109 of LSTM-GA and 51.696 of standard LSTM), MAE = 3.846 (50.2% lower than ANN), R^2 = 0.857 (significantly superior to 0.602 of LSTM3D), the prediction for high-frequency fluctuating stocks (such as UnitedHealth UNH) has the most significant improvement: MAE decreased from 40.169 of LSTM1D to 9.016, indicating that the ARO optimization effectively adapts to complex fluctuation patterns.

In the research on stock price prediction based on the GRU model, the experimental results were to test the influence of different input features and data frequencies. Taking North China Huacun (002371) as an example: the results of three-feature input were RMSE = 1.586 and MAE = 1.107, while the result of single-price input was RMSE = 2.942 and MAE = 2.305. These data prove that technical indicators (MACD) and trading volume can effectively supplement the information gap in the price sequence.

4. Conclusion

This paper systematically investigates the technological evolution and application efficacy of machine learning in stock price prediction. The main contributions include a detailed dissection of the architectures of four core models: LSTM, CNN-LSTM (SACLSTM), GRU, and TCN-Transformer, and an elucidation of their technical advantages in handling the non-linear and non-stationary characteristics of financial time series data. For instance, the SACLSTM model overcomes the limitations of single time series modeling and the TCN-Transformer model's residual attention mechanism enhances multi-scale feature extraction capabilities, significantly improving the long-term and short-term dependency modeling effects. Additionally, through comparative experiments based on multi-market datasets (such as Chinese stocks, US stocks, etc.), using multi-dimensional evaluation criteria to quantitatively verify model performance, and discovering that the hybrid model shows significant advantages (such as TRAformer achieving an MAE of 0.842 on ZTE communication data, 66.8% lower than LSTM), the superparameter optimization strategy (such as the ARO algorithm) increases the R^2 of LSTM-ARO to 0.857, demonstrating the significant enhancement of the model's generalization ability by the meta-heuristic method. The GRU model performs exceptionally well in high-frequency scenarios, with high computational efficiency and excellent real-time feedback capabilities.

By mining the stock discussion and comments on these platforms, a more comprehensive calculation of the market sentiment index can be conducted, providing more diverse challenges and variables for new research models and increasing the training difficulty of the models. For the industrial application of stock price prediction technology, future research should achieve systematic breakthroughs along a phased path. Short-term goals (1-2 years) will focus on building a cross-institutional data collaboration framework based on federated learning, integrating private data from securities firms and multimodal information from social media through a secure sharing mechanism, aiming to increase the utilization rate of small sample market data by more than 40%, significantly alleviating the bottleneck of insufficient training in emerging markets. Mid-term evolution (3-5 years) requires deep integration of causal inference and generative adversarial denoising techniques, establishing a multi-body simulation system of "macro policy orientation - industry rotation patterns - market sentiment transmission", with a focus on breaking through the prediction blind spots of extreme events (such as sudden changes in volatility | > 5%), striving to reduce the prediction error of such events by 50%. Long-term vision (5 years and above) should explore quantum computing-enabled parallel acceleration mechanisms of spatiotemporal attention, achieving millisecond response of high-frequency predictions for thousands of stocks, and developing dynamic risk break-even modules (threshold adaptive calibration mechanism), ultimately promoting the transformation of academic models into industrial-level decision systems, providing intelligent guarantees for the stability of the financial market.

References

[1] Qi, T. W., Yu, W. N.: Multi-index Stock Prediction Based on LSTM.’ Computer and Digital Engineering, 2024, 52(2): 337-342.

[2] Wu, J. M. T., Li, Z., Herencsar, N., et al.: ‘A graph-based CNN-LSTM stock price prediction algorithm with leading indicators.’ Multimedia Systems, 2023, 29(3): 1751-1770.

[3] Lu, M., Xu, X.: TRNN: An efficient time-series recurrent neural network for stock price prediction.’ Information Sciences, 2024, 657: 119951.

[4] Hang, Y. X.: ‘Design and Practice of Stock Price Prediction System Based on TCN-Transformer.’ East China normal university, 2024. DOI: 10.27149 /, cnki. Ghdsu. 2024.003762.

[5] Soni, P., Tewari, Y., Krishnan, D.: Machine Learning approaches in stock price prediction: A systematic review’ Journal of Physics: Conference Series. IOP Publishing, 2022, 2161(1): 012065.

[6] Mu, G., Gao, N., Wang, Y., et al.: ‘A stock price prediction model based on investor sentiment and optimized deep learning.’ Ieee Access, 2023, 11: 51353-51367.

[7] Lu, X. Y., Lu, H. B.: ‘Stock Price Prediction Based on GRU Model.’ Shanxi Electronic Technology, 2024, (01):7-8+12.

[8] Gülmez, B.: ‘Stock price prediction with optimized deep LSTM network with artificial rabbit’s optimization algorithm.’ Expert Systems with Applications, 2023, 227: 120346.

[9] Biswas, A. K., Bhuiyan, M. S. A., Mir, M. N. H., et al.: ‘A Dual Output Temporal Convolutional Network with Attention Architecture for Stock Price Prediction and Risk Assessment.’ IEEE Access, 2025.

[10] Ma, D., Yuan, D., Huang, M., et al.: VGC-GAN: A multi-graph convolution adversarial network for stock price prediction.’ Expert Systems with Applications, 2024, 236: 121204.