DOI: 10.54254/2754-1169/2026.BJ30243

# *A Feature-Attention Neural Framework for Multi-Factor Stock Return Forecasting*

**Shengzhe Xu**

*Department of Computer Science, University of Nottingham Ningbo China, Ningbo, China*
*scysx5@nottingham.edu.cn*

***Abstract.*** Stock return prediction is a fundamental issue in quantitative investment research. However, the non-linear and highly volatile nature of financial markets makes precise modeling extremely challenging. Although traditional machine learning models can depict some nonlinear relationships, they have difficulty capturing time-dependent and cross-factor interactions. Moreover, deep learning models often have complex structures, redundant parameters, and limited interpretability. This paper proposes a lightweight attention neural Network - LIFA-Net (Lag-InterFactor Attention Network), introducing the Lag-Wise Attention and Multi-Head Attention mechanism among factors under the multi-layer perceptron (MLP) backbone to achieve the joint modeling of time-dependent and factor relationships. The experimental results show that the LIFA-Net is significantly superior to the traditional baseline model in terms of prediction accuracy and robustness, and effectively improves the feature interpretability and generalization ability of the model. This study verified the effectiveness and practical value of the double-layer attention structure in the modeling of financial time series.

***Keywords:*** Machine Learning, Attention Mechanism, Stock Prediction, Financial Time Series, Multi-Factor Model

## 1. Introduction

Stock return prediction is a fundamental task in portfolio management and quantitative finance. Because the financial markets are nonlinear and irregular, stock return prediction is commonly formulated as a multivariate time series forecasting problem. However, since the stock market is highly chaotic and volatile, reliable forecasting remains an open challenge. Therefore, developing robust forecasting models is important to both academic research and practical financial applications.

Numerous efforts have been devoted to improving stock return prediction accuracy through a variety of machine learning and deep learning techniques. Sonkavde et al. compared several machine learning algorithms, including Random Forest, XGBoost, and LSTM, by constructing an ensemble model for stock price prediction and achieved a Root Mean Square Error (RMSE) of 2.14, demonstrating solid performance in medium and short-term predictions [1]. Yang et al. developed a hybrid model combining eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) and applied it to predict price trends of multiple stocks, achieving a prediction

DOI: 10.54254/2754-1169/2026.BJ30243

accuracy as high as 94.5% and verifying the advantages of gradient boosting algorithms in feature learning [2]. However, even though these algorithms are able to model nonlinear relationships, dynamic temporal dependencies, and cross-factor interactions are hard to capture, which limits interpretability and generalization ability. Nelson et al. employed Long Short-Term Memory (LSTM) to predict the short-term price fluctuations of the Brazilian stock market and achieved an average accuracy of 55.9%, which was higher than other traditional models' performance [3]. However, the sensitivity to input characteristics leads to high variance and a limited generalization effect. Mozaffari et al. introduced a Transformer-based architecture with Multi-Head Attention to predict the closing price of stocks, achieving an RMSE of 10.14, and the prediction accuracy is higher than that of LSTM (12.23) and Prophet (15.87) models [4]. However, the model required a complex training process and is too sensitive to hyperparameters, limiting its robustness under different market conditions.

This paper proposes a simple and powerful attention-based stock return forecast model. Specifically, the model combines Lag-Wise Attention and Inter-Factor Attention mechanisms to realize time series dependence and inter-factor relationship modeling. This can more effectively capture the interaction between dynamic patterns and factors in the financial time series, and improve the prediction accuracy and interpretability of the functional level under the lightweight architecture.

## 2. Methodology

### 2.1. Dataset

The data set used in this study is from Kaggle's S&P 500 index stock dataset, which covers the daily trading data of the U.S. stock market from January 2013 to December 2018 [5]. Each record contains the market information of stocks on the trading day, including the opening price, highest price, lowest price, closing price, and trading volume. The dataset contains a total of 505 listed companies with a total number of transaction records of 619,040, covering multiple industries, with good market representativeness and time continuity.

To extract the market features related to stock return forecast, this study constructs five technical factors of momentum, volatility, relative strength index (RSI), simple moving average ratio (SMA ratio), and volume changes. Concretely, these factors can reflect short-term market trends (Momentum), risk levels (Volatility), investor sentiment (RSI), price trend smoothness (SMA ratio), and trading activities (volume changes), respectively. In existing research, these technical factors have been widely validated as important signals for stock return prediction [6-10].

Moreover, in order to capture dynamic dependencies in financial time series, 15 lag features were generated for each factor to reflect the market memory effect across various time windows. The target variable is defined as the average return over the next five trading days by calculating the five-day rolling average of the daily rate of return.

In data preprocessing, the data was first sorted by stock name and date; for the missing values, grouped interpolation was used to impute them, and all continuous features were standardized using the StandardScaler method to ensure consistent feature scales and enhance the stability of model training.

DOI: 10.54254/2754-1169/2026.BJ30243

## 2.2. Model architecture

This study proposed a lightweight attention-based neural architecture named LIFA-Net (Lag-InterFactor Attention Network) for multi-factor stock return prediction. This model introduces two complementary attention mechanisms based on a multilayer perceptron (MLP) to model the interaction relationship between the lag features and factors of the time series [11]. This attention structure improves the accuracy of the stock return prediction and maintains the explainability of the feature level. The framework of LIFA-Net includes three main components: a Lag-Wise attention module, an Inter-Factor Multi-Head Attention Module, and a backbone MLP predictor.

Each information in the input layer contains multiple financial factors and historical lag features. Assuming that there are $N_f = 5$ factors and each factor contains an L-lag value, the input dimension in total is $N_f$ ×L. The Lag-Wise Attention module aims to capture the dynamic change pattern of these $N_f$ financial factors in the time series. In addition, the model structure calculates the attention weight of each lagging financial factor through the feedforward network, which is activated by Tanh, and normalizes them by using the Softmax. The process assigns adaptive weights to different historical points, so that the model can focus on the time window with the largest amount of information, instead of sliding windows that rely on a fixed length.

The Inter-Factor Multi-Head Attention module models the correlation between different financial factors. Each factor is represented as a compressed embedded vector after delayed attention aggregation, which is linearly projected into the low-dimensional representation space through linear mapping and input into the Multi-Head Attention layer [12]. This module captures the relationships among nonlinear factors, such as the synergy between volatility and momentum during market fluctuations. Through this module, LIFA-Net realizes the feature integration at the factor level and enhances the ability to capture market structure information.

In the prediction stage, the model connects the original input features, lag aggregation features, and factor-level context features in series and inputs them into the MLP predictor. The model is composed of two fully connected layers, using ReLU activation function, Batch Normalization, and Dropout regularization to enhance the generalization ability of the model and prevent overfitting. This model's output is the predicted average return rate for the next five trading days.

Eventually, the final attention weights are visualized to provide explanations for the predicted results. The visualization shows how LIFA-Net allocates attention across lag periods and financial factors, providing intuitive and interpretable support for investment decision-making.

## 2.3. Experimental design

This study conducted a series of systematic experiments using the preprocessed S&P 500 dataset to evaluate the effectiveness and robustness of the LIFA-Net framework in predicting multi-factor stock returns. The design of experiments not only aims to verify the prediction accuracy and interpretability of the model, but also presents the analysis of comparison among several benchmark methods.

The dataset is divided into a training subset (80%) and a test subset (20%) in chronological order to avoid data leakage. In addition, to ensure stable convergence during model training, all input variables and target values are standardized through the StandardScaler mechanism to maintain consistency in feature scaling.

Four typical traditional models were selected, including linear regression, random forest, XGBoost, and standard MLP, for comprehensive comparison in this experiment. These models

DOI: 10.54254/2754-1169/2026.BJ30243

contain different approaches of statistical learning, set tree methods, and neural architecture. Each baseline is finely tuned through grid search to ensure the fairness of the performance evaluation.

Regarding the evaluation indicators, this study adopts root mean square error (RMSE) and determination coefficient ($R^2$) as the main indicators. RMSE quantifies the average size of the predicted deviation, and $R^2$ measures the degree of interpretation of the change of dependent variables by the model. Better predictive performance means lower RMSE and higher $R^2$.

In terms of training configuration, the model used the Adam optimizer (learning rate 0.0005) and the mean square error (MSE) loss function. In order to prevent overfitting, an Early Stopping mechanism is introduced when the loss does not improve within 30 consecutive cycles. The model adopts Dropout (0.2) and Batch Normalization to enhance the generalization ability. In addition, the temperature parameter $\tau$ in the attention module is set to 2.0, and the number of heads for Multi-Head Attention is set to 4.

To further evaluate the application potential of the LIFA-Net model in quantitative investment, the experiment designed a quantile-based long-short strategy based on predicted return signals. When the predicted rate of return is in the upper percentile (>70%), establish a long position; and when it is in the lower percentile (<30%), establish a short position. Each transaction is held for 5 trading days, and a single rollover cost of 0.1% is considered. Ultimately, the results were evaluated by calculating cumulative returns, annualized rate of return, volatility, and Sharpe ratio. Formatting your paper

## 3. Results

### 3.1. Predictive performance comparison

Table 1 shows the predictive performance of each model. LIFA-Net achieved RMSE = 0.003046 and R = 0.8048 indicators in the experiment, which is better than all baseline models. Although traditional tree models, including Random Forest, XGBoost, can describe nonlinear relationships, they cannot handle time series. The standard MLP can capture nonlinear features, but it cannot consider time series and factor interaction. To address these problems, LIFA-Net's Lag-wise Attention and Inter-Factor Attention can effectively capture time series and cross-factor features, which improve prediction accuracy and fitting performance.

Table 1. Comparison of predictive performance across models

| Model | RMSE | $R^2$ |
|---|---|---|
| Linear Regression | 0.003152 | 0.7910 |
| XGBoost | 0.003194 | 0.7854 |
| Random Forest | 0.004142 | 0.6391 |
| MLP | 0.003168 | 0.7889 |
| LIFA-Net (Proposed) | 0.003046 | 0.8048 |

### 3.2. Backtesting performance analysis

The backtest results presented in Table 2 show that LIFA-Net has achieved the highest performance among all evaluation models, including an annualized rate of return of 11.97%, a Sharpe rate of 2.999, and a volatility of 3.99%. These results mean that its prediction accuracy has been improved

DOI: 10.54254/2754-1169/2026.BJ30243

without a corresponding increase in risk exposure. Compared with traditional models, LIFA-Net adaptively focuses on key lag dependence and inter-factor correlation through its attention architecture. Therefore, LIFA-Net is capable of generating more meaningful economic transaction signals and showing greater robustness and practical applicability.

Table 2. Comparison of quantile-based backtesting results (5-day holding strategy)

| Model | Annualized Return | Annualized Volatility | Sharpe Ratio |
|---|---|---|---|
| Linear Regression | 11.74% | 4.00% | 2.933 |
| XGBoost | 11.87% | 3.99% | 2.975 |
| Random Forest | 11.81% | 3.97% | 2.977 |
| MLP | 11.80% | 4.00% | 2.952 |
| LIFA-Net (Proposed) | 11.97% | 3.99% | 2.999 |

In conclusion, the experimental results demonstrate that LIFA-Net outperforms the traditional machine learning and deep learning models in terms of prediction accuracy and investment performance in this experiment. This result further supports the effectiveness of integrating the attention mechanism into the multi-factor framework and proves the effect that LIFA-Net can reveal the relationship between time patterns and factors with higher efficiency and depth.

## 4. Discussion

The results of experiments state that the remarkable performance of LIFA-Net comes from its double-layer attention framework. In this architecture, the Lag-wise Attention mechanism adaptively redistributes the importance of historical data over different time spans, and the Multi-Head Attention module captures complex nonlinear interactions among financial factors. By integrating these two mechanisms, LIFA-Net can take the time dependence and the factor correlations into account simultaneously before making decisions. Eventually, this framework is able to improve prediction accuracy and maintain a high level of feature interpretability in this stock return prediction task.

Although the attention mechanism has been proven to be effective in improving the predictive accuracy and interpretability of financial models [13-15], this study still has some limitations that deserve further investigation. First, the current model only relies on technical indicators and does not include macroeconomic variables. Some research has shown that macroeconomic factors such as GDP growth and inflation have a significant impact on stock returns [16]. Therefore, the lack of some variables may limit the ability of models to capture basic economic dynamics. Second, the lack of investor sentiment signals will limit the ability of predictive models, because the inclusion of behavioral factors can improve the sensitivity of the model to market changes. Engelberg et al. found that sentiment indicators extracted from news and social media have significant predictive ability for asset prices [17]. Third, the S&P 500 stock dataset does not include extreme market conditions such as the COVID-19 pandemic. Hence, the models are not able to deal with some unprecedented market shocks and structural changes [18]. Finally, the verification was limited to the S&P 500 index market, which may restrict the applicability of LIFA-Net to other markets. Rai et al. have emphasized the importance of cross-market robustness testing during crisis periods [19].

Future research can focus on these limitations through several pathways. By extending the dataset, models can improve their robustness and generalizability. Exploring the combination of

DOI: 10.54254/2754-1169/2026.BJ30243

LIFA-Net with large language models (LLMs) to capture cross-market structural information and sentiment signals from multimodal data sources, such as financial news, social media, and analyst reports, presents a promising direction for enhancing both predictive accuracy and economic interpretability.

## 5. Conclusion

This paper proposes a lightweight attention neural network architecture, LIFA-Net, for multi-factor stock return prediction. By integrating Lag-Wise attention and Inter-Factor Multi-Head Attention, the model achieved the joint modeling of the interaction relationship between time-dependent features and factors. The experimental results show that LIFA-Net outperforms traditional machine learning and deep learning models in both RMSE and $R^2$ metrics, and achieves a higher risk-adjusted return in quantitative backtesting. Future work will combine LIFA-Net with large pre-trained models to build a multimodal financial forecasting framework.

## References


[1] Sonkavde, A., Adhikari, S., and Katti, S. (2023). Forecasting stock market prices using machine learning. International Journal of Financial Studies, 11(3), 94.

[2] Yang, J., Wang, J., and Li, X. (2021). Stock price prediction based on XGBoost and LightGBM. E3S Web of Conferences, 275, 01040.

[3] Nelson, D.M.Q., Pereira, A.C.M. and de Oliveira, R.A. (2017). Stock market's price movement prediction with LSTM neural networks. Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 1419-1426.

[4] Mozaffari, L. and Zhang, J. (2024). Predictive modeling of stock prices using Transformer model. Proceedings of the 9th International Conference on Machine Learning Technologies (ICMLT 2024), ACM, 41-48.

[5] Camenisch, J. (2017). S& P 500 stock data. Kaggle. https: //www.kaggle.com/datasets/camnugent/sandp500

[6] Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. Journal of Finance, 48(1), 65-91.

[7] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307-327.

[8] Chong, T.T.L., Ng, W.K., and Liew, V.K.S. (2014). Revisiting the performance of MACD and RSI oscillators. Journal of Risk and Financial Management, 7(1), 1-12.

[9] Brock, W., Lakonishok, J, . and LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, 47(5), 1731-1764.

[10] Karpoff, J.M. (1987). The relation between price changes and trading volume: A survey. Journal of Financial and Quantitative Analysis, 22(1), 109-126.

[11] Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning representations by back-propagating errors. Nature, 323, 533-536.

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems (NeurIPS), 30.

[13] Ma, T., Wang, W., and Chen, Y. (2023). Attention is all you need: An interpretable transformer-based asset allocation approach. International Review of Financial Analysis, 90, 102876.

[14] Nguyen, A.Q., Ha, S., and Nguyen, P. (2023). A lightweight multi-head attention transformer for stock price forecasting. SSRN Electronic Journal.

[15] Shabani, M., Tran, D.T., Magris, M., Kanniainen, J., and Iosifidis, A. (2022). Multi-head temporal attention-augmented bilinear network for financial time series prediction. arXiv preprint arXiv: 2201.05459.

[16] Pettenuzzo, D. and Timmermann, A. (2014). Forecasting macroeconomic variables under model instability. Journal of Financial Economics, 114(3), 527-545.

[17] Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all FEARS: Investor sentiment and asset prices. Review of Financial Studies, 28(1), 1-32.

[18] Baker, S.R., Bloom, N., Davis, S.J., Kost, K., Sammon, M., and Viratyosin, T. (2020). The unprecedented stock market reaction to COVID-19. Review of Asset Pricing Studies, 10(4), 742-758.

DOI: 10.54254/2754-1169/2026.BJ30243

[19] Rai, A., Mahata, A., Nurujjaman, Md., Majhi, S., and Debnath, K. (2021). A sentiment-based modeling and analysis of stock price during the COVID-19: U- and Swoosh-shaped recovery. Physica A: Statistical Mechanics and its Applications, 584, 126810.