Proceedings of the 10th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/2026.37196

Study of Mainstream Methods in Stock Price Prediction: Performance Characteristics, Comparative Analysis and Fusion Innovation Directions

Yanlin Huang

Beijing Royal School, Beijing, China
hyl20091007@gmail.com

Abstract. Stock price prediction is a typical but difficult problem in quantitative finance that can help investors make decisions and manage risks. However, due to the high volatility of the financial market and other reasons, it cannot be known in advance. The three main methodological paradigms in the current studies are introduced systematically here. The first is feature selection and extraction to improve the quality of the model's input data from raw data. The second type is a neural network-based approach; RNNs, LSTMs and Transformers are used to learn complex, non-linear temporal dependencies in the historical data. The third kind is a graph-structure-based method that builds a model of the connections among all stocks to study cross-asset spillover effects. The review shows that each method has distinct strengths and is suited to different market conditions, with no single approach dominating all scenarios. A primary insight is that future research should move towards integrating these complementary methodologies and leveraging multi-modal data sources, which holds the potential for developing more robust and economically interpretable predictive frameworks.

Keywords: Stock price prediction, Feature selection, Extraction, Neural Network, Graph Structure

1. Introduction

Stock price prediction is an old and basic problem in modern finance that has been studied for more than a century and is closely related to the development of the world's capital markets. Based on the practical demands that have arisen from a hundred years of development of global financial markets and the continuous progress of financial theory, after Bachelier's 1900 paper "The Theory of Speculation" first applied mathematical methods to describe stock price fluctuations, this subject has become one of the main research areas in modern finance. After the spread of the Efficient Market Hypothesis, scholars have been studying whether stock prices are predictable for many years. Empirical studies have found many market anomalies, such as momentum and reversal effects, and have continuously expanded this area. With the development of the world's capital markets, institutional investors have sought higher returns; facing the risks of market fluctuations, they also need to meet stricter risk management requirements, and recently, big data and artificial intelligence

DOI: 10.54254/2754-1169/2026.37196

technology have emerged; thus, stock price prediction has evolved from an early application of trading experience into an interdisciplinary field integrating finance, computer science and statistics.

The demands for research in stock price prediction can be divided into three types.

A large number of historical data can be used to train and learn predictive models to discover hidden statistical patterns and non-linear relationships in market fluctuations, provide deeper insights into the process of price formation, market efficiency, and behavioral finance biases at the cognitive level for researchers and investors.

A good predictive model can offer objective and timely buy-or-sell signals for institutional and individual investors at the decision-making level, help them adjust asset allocation and timing under risk control, etc. Algorithms can reduce the effect of human cognitive bias and emotion in an information-abundant market.

Stress testing and scenario analysis based on predictive models can be carried out at the level of risk management to identify extreme market fluctuations and systemic risks in advance, provide macroprudential tools for regulators, and help investors build more resilient portfolios to control downside risks better. Nowadays, this area is closely linked to artificial intelligence, big data and high-performance computing, and new frontiers in investment science are being continuously opened for practical application.

The main research problem in stock price prediction is that stock price changes are influenced by many different sources simultaneously, such as changes in economic policies, individual company operations, breaking news, and various investor sentiments; these factors have nonlinear and non-stationary fluctuations and cannot be fully accounted for by a single model. Research data commonly used in this field at present include OHLCV prices and trading volume in the price-volume dimension, derived technical indicators, corporate fundamental data, and public news text information.

There has been an obvious trend in the development of stock price prediction methods over time: from the early days of fundamental and technical analysis, to traditional econometric models such as ARIMA, then to older machine-learning techniques including support vector machines and random forests, and finally to current deep-learning technologies such as LSTMs and graph neural networks; they have gradually moved from linear fitting to nonlinear complex feature mining. Existing academic research in this field generally divides mainstream stock prediction methods into three categories: Feature selection and feature extraction methods, neural network methods, and graph structure methods [1]. This paper will also carry out systematic sorting and comparative analysis from these three categories of methods.

2. Predication methods based on feature election and feature extraction

In the quantitative modeling framework for stock price prediction, "features" refer to all structured information input into the model to support trend judgment and price prediction. They not only cover raw trading data in the dimensions of open, high, low, close and trading volume, but also include various technical indicators derived from price-volume relationships such as moving average, relative strength index and MACD. Some studies further incorporate asset pricing dimension factors including the market risk premium in the CAPM model and industry beta coefficient into the feature system, to expand the economic interpretation boundary of features. The core logic of such methods is to screen out key variables with strong predictive power from high-dimensional raw datasets, or to map and compress redundant multi-dimensional information into a low-dimensional feature space, before generating more representative composite features for subsequent model analysis. This approach solves the "curse of dimensionality" that commonly exists

DOI: 10.54254/2754-1169/2026.37196

in traditional time series modeling, and it also serves as the core preprocessing step in the stock price prediction review research you are conducting.

Current mainstream implementations in this field can be clearly divided into two major branches: feature selection and feature extraction. Representative methods in the feature selection dimension include correlation analysis and random forest [2, 3]. Correlation analysis measures the linear or nonlinear correlation strength between each variable and future stock return through statistical indicators such as Pearson and Spearman correlation coefficients, so as to pre-eliminate noise variables with extremely low correlation with the prediction target. In the quantitative modeling scenarios related to the NEC economics competition, this step enables rapid filtering of meaningless redundant indicators. The random forest, relying on the ensemble learning mechanism of multiple decision trees, simultaneously counts the contribution of each feature to the prediction error during node splitting, and outputs a standardized feature importance ranking. It completes automatic feature screening while retaining the original physical meaning of variables, and is especially suitable for the collinearity scenario that commonly exists in financial data, avoiding result deviation caused by mutual interference of multiple variables [4].

The mainstream methods in the feature extraction dimension take Principal Component Analysis (PCA) and autoencoders as the core. PCA transforms original variables with significant collinearity into a set of uncorrelated principal components through orthogonal transformation, and extracts the first N principal components according to the descending order of variance contribution [5]. It realizes linear dimensionality reduction on the premise of retaining as much original data information as possible, and is the most widely used unsupervised dimensionality reduction tool in financial quantitative research. Many extended quantitative cases in AP Microeconomics will take it as a basic dimensionality reduction method for explanation. The autoencoder, relying on the nonlinear mapping capability of deep learning, compresses high-dimensional raw data into a low-dimensional latent space through the encoder, and then reconstructs the original input through the decoder. In the process of minimizing the reconstruction error, it automatically learns the most representative latent features, which can capture the nonlinear correlations that traditional linear methods cannot identify, and adapt to complex non-stationary financial time series data.

The core advantage of such methods lies in their ability to effectively filter market noise and irrelevant variables, significantly reduce the computational complexity of subsequent prediction models, shorten the time consumption of model training and inference, and reduce the overfitting risk brought by redundant features, helping the model focus more on the core driving factors of stock price fluctuations. However, their limitations are equally prominent. Whether it is feature screening based on statistical rules or dimensionality reduction operations based on mapping transformation, there is a potential risk of information loss. If marginal information with potential predictive value, such as niche technical indicators with low correlation but indicative effect under specific market conditions, is mistakenly deleted during the screening or transformation process, it will instead cause information loss in the model input, ultimately weakening the generalization ability of stock price prediction, and even leading to significant amplification of prediction deviation under extreme market conditions [6].

3. Prediction method based on neural networks

In the research context to deep learning for stock price prediction, neural network-based prediction methods represent the most widely applied technical branch at present. Their core feature is that they can reduce the dependence on manual feature engineering, and automatically capture the complex nonlinear variation rules hidden behind stock price fluctuations from massive historical trading data.

Researchers do not need to pre-define variable correlations through statistical rules, which greatly expands the adaptation boundary of quantitative models for non-stationary financial markets. The underlying logic of such methods is to build an end-to-end architecture with hierarchical feature extraction capability through the stacking of multi-layer neurons and the nesting of nonlinear activation functions, so that the model can independently fit the potential mapping relationship between multi-dimensional information such as price, trading volume and market sentiment in historical data and the future stock price trend during repeated iterative training. Representative models in this field have formed a clear technical evolution path. Starting from the early basic Artificial Neural Network, this fully connected architecture first realized the nonlinear fitting of simple price-volume relationships, laying a foundation for the application of deep learning in the financial quantitative field. Subsequently, the emergence of Recurrent Neural Network (RNN) introduced the time series memory mechanism for the first time, endowing the model with preliminary historical information retention capability. However, when processing long-cycle financial sequences, it is prone to the problem of gradient vanishing, making it difficult to capture the lagging impact of long-term market events on current stock prices. The Long Short-Term Memory (LSTM) network derived on this basis optimizes the time series memory unit in a targeted manner, and accurately controls the retention and forgetting of historical information through the gating mechanism. It can efficiently learn the dynamic rules of stock prices evolving over time, so it has become the most mainstream benchmark model in the field of stock price prediction in the past decade, and a large number of business quantitative studies take it as the core tool for time series prediction [7, 8].

With the continuous increase of time series data length, the academic community has introduced Convolutional Neural Network (CNN) in the field of computer vision into the stock price prediction scenario, using its local feature capture capability to extract the local price-volume correlation patterns in K-line charts. In recent years, the Transformer architecture in the field of natural language processing has also been migrated to financial time series tasks, realizing parallel correlation modeling of full-sequence information relying on the self-attention mechanism. The Informer model further developed on this basis improves the calculation efficiency of self-attention and the design of long-sequence position coding, which is specially adapted to ultra-long-span financial time series. It can cover daily-level trading data of several months or even years at one time, solving the pain point that traditional time series models cannot efficiently process long-cycle stock price sequences [9].

The core advantage of such methods lies in their extremely strong nonlinear learning capability, which can dig out complex market rules that cannot be identified by traditional statistical methods, and often achieve better prediction accuracy in scenarios with sufficient data volume. However, their limitations are also very prominent: the training of such models is highly dependent on large-scale, high-quality labeled datasets. If there are problems such as missing data or excessive outliers, the model is prone to overfitting. At the same time, the training process of large-scale neural networks requires a large amount of computing resources, and the training time is much longer than that of traditional statistical models. More critically, such deep models generally have the "black box" property, and the intermediate logic of the prediction process is difficult to be clearly explained, which forms an obvious conflict with the requirement of financial market for decision interpretability, and has become the core unsolved research gap in this direction.

DOI: 10.54254/2754-1169/2026.37196

4. Prediction method based on graph structure

In the cross-asset research system of stock price prediction, graph structure-based prediction methods are a rapidly emerging frontier branch in recent years. It breaks through the limitation of traditional time series models that only model a single stock independently, and incorporates the linkage relationships between different targets in the financial market into the prediction framework. The underlying logic of such methods can be illustrated with a very simple concrete example: abstract each independently traded stock in the market as a "node" in the graph structure, and abstract various correlations between different stocks as "edges" connecting the nodes, so as to construct a financial relationship graph that can completely map the interaction logic of targets across the whole market, completely breaking the information boundary of traditional univariate time series modeling. The correlation edges between nodes in this graph structure are not arbitrarily defined by humans, and their construction basis completely comes from the operation logic of the real financial market [10]. The mainstream correlation generation dimensions can be divided into four categories: the first category is the return correlation calculated based on historical price series, and the correlation edges are generated by counting the linkage degree of the long-term trends of different stocks; the second category is the affiliation relationship based on industry classification standards, which automatically establishes correlations for stocks in the same segmented track under industry classification systems such as SW and GICS; the third category is the physical business correlations such as publicly disclosed commercial cooperation and supply chain shareholding of listed companies, which can directly reflect the real interaction at the enterprise operation level; the fourth category is the temporary correlation generated based on common market events. When multiple stocks are simultaneously impacted by the same event such as policy release or industry positive news, temporary correlation edges will be generated in the graph structure. Relying on this pre-constructed financial graph structure, researchers can break through the information island limitation of a single stock, aggregate the price-volume and fundamental information of the target stock and other related targets into the same analysis framework, and capture the cross-target spillover effect that cannot be reflected in the time series data of a single stock itself. The most mainstream implementation tool in this direction at present is the Graph Convolutional Network. This type of model can complete convolution operations on the graph structure, automatically aggregate the feature information of adjacent nodes, and learn the transmission impact of different correlation relationships on the price fluctuation of the target stock during iterative training. Compared with traditional independent modeling methods, it can capture more hidden market linkage rules [11]. The core advantage of such methods is that they systematically incorporate the mutual influence mechanism between different stocks into the prediction process for the first time, solving the core defect of traditional single-target time series models that completely ignore the market linkage effect, and can significantly improve the rationality of prediction under complex market conditions such as sector rotation and risk transmission. However, its limitations are also very obvious: the correlation between stocks is not static and constant, and will continue to dynamically evolve with the switching of market styles, adjustment of industry policies and changes in enterprise operations. To continuously and accurately construct a graph structure that fully fits the operation state of the real market, it needs the support of high-frequency data covering multiple dimensions, and the difficulty of modeling and maintenance is much higher than that of traditional time series models, which has also become the core research difficulty in the current field of practical application [12].

5. Conclusion

The three main directions of change in the methodology of stock price prediction are now well-established, and they focus on different aspects of information processing. Feature selection and extraction aim to reduce the number of high-dimensional raw data or enhance the quality and density of input information for models by selecting necessary attributes. Neural network-based prediction methods aim to learn the hidden nonlinear temporal patterns in historical data and autonomously discover complex dynamic features of price series through end-to-end architectures. Graph structure-based prediction methods address the problem of single-target modelling by adding correlation information among several stocks to the analysis system and considering the effect of market linkage on the prices of individual stocks. Together, the three types of methods make up a full-featured system for information pre-processing, temporal pattern mining and cross-asset correlation modelling.

Overall, none of the methods can achieve top performance in all market cycles and for all stock targets. Feature processing methods are more sensitive to complex datasets with high information redundancy; neural network models have shown excellent performance in strong-regularity market conditions and when there is a large amount of data, and graph structure methods have demonstrated unique advantages in strong-correlation market phases, such as sector rotation and risk resonance. The performance of prediction models is always significantly constrained by the quality of data, market conditions and sudden events; therefore, the output of any prediction framework built on historical data should not be regarded as a certain guarantee of future stock price trends but rather as an approximate description of the probability distribution under specific constraints.

Looking ahead, the main direction of research in stock price prediction will gradually be multi-method fusion and multi-modal data integration. Researchers can try to complementarily combine feature engineering, neural networks and graph networks, for example, embedding feature selection modules in graph convolutional networks, or introducing prior information from correlation networks into LSTM temporal modeling. At the same time, future research should also expand from single historical price data to the integration of multi-source heterogeneous data such as company fundamentals, industry chain graphs, news sentiment texts, and macro policy signals, to build hybrid prediction frameworks with stronger explanatory power and robustness. Such fusion innovation may ultimately promote stock price prediction from technical indicator fitting to more economically meaningful systematic modeling.

References

[1] Mandal, R. C., & Kler, R. (2024). Enhancing stock price prediction with deep cross-modal information fusion network. *Fluctuation and Noise Letters, 23*(2).

[2] Bosancic, T., Nie, Y., & Mulvey, J. (2024). Regime-aware factor allocation with optimal feature selection. *The Journal of Financial Data Science, 6*(3), 10-37.

[3] Huang, H., & Li, B. (2024). Stock price prediction based on VMD-CSSA-LSTM combination model. *Journal of Nanjing University of Information Science & Technology, 16*(3), 332-3340.

[4] Wang, Y., & Guo, Y. (2019). Application of improved XGBoost model in stock forecasting. *Computer Engineering and Applications, 55*(20), 202-207.

[5] Zou, J., Zhao, Q., Jiao, Y., et al. (2023). Stock market prediction via deep learning techniques: A survey. *Neurocomputing, 532*, 121-148.

[6] Li, M., Huang, Z., & Xv, J. (2023). Stock price prediction based on LSTM-Wavelet model using deep learning and wavelet analysis. *Journal of Chongqing Technology and Business University (Natural Science Edition)*.

[7] Jain, S., & Kaur, K. (2026). AI in stock market forecasting: A systematic review of regional performance, crisis robustness, and transfer learning. *Research Square*.

DOI: 10.54254/2754-1169/2026.37196

[8] Zhen, K., Xie, D., & Hu, X. (2025). A multi-feature selection fused with investor sentiment for stock price prediction. *Expert Systems with Applications*.

[9] Zhu, P., Li, Y., Liu, Q., Cheng, D., & Jiang, C. (2025). Financial time series prediction with multi-granularity graph augmented learning. *IEEE Transactions on Knowledge and Data Engineering, 37*(11), 6426–6449.

[10] Petal, M., Jarimala, K., & Chattopadhyay, C. (2024). A systematic review on graph neural network-based methods for stock market forecasting. *ACM Computing Surveys*.

[11] Brown, J. D., & Davis, R. E. (2024). Return predictability with machine learning in 46 stock markets around the world. *Journal of Economic Dynamics and Control*.

[12] Bao, W., Cao, Y., Yang, Y., Che, H., Huang, J., & Wen, S. (2025). Data-driven stock forecasting models based on neural networks: A review. *Information Fusion*.