Vol. 16 (2026): Proceedings of the 2026 International Conference on AI and Technological Innovation Applications (AITIA 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/aitia.2026441447

The Application of Artificial Intelligence in A-Share Market Prediction: A Case Study of the CSI 300 Index

Yao Tong*

Institute of Business, Hunan University of Science and Technology, Xiangtan, China

**Corresponding author: Yao Tong.*

Abstract

As a core component of the modern financial system, the stock market's price fluctuations not only reflect the business conditions of enterprises, the development trends of industries, and macroeconomic dynamics but also are closely related to investors' asset allocation, market risk management, and national finance. With the development of the times, the application of artificial intelligence with strong data processing capabilities, nonlinear fitting capabilities, and autonomous learning capabilities in the stock market can assist researchers in making more accurate predictions of stock price trends. This review summarizes and analyzes the application scenarios and technical paths of artificial intelligence in predicting the stock market. Through different models, hidden patterns are mined from massive data to provide more scientific decision-making bases for individual investors and institutions, reducing blind trading and irrational operations. At the same time, the combination of artificial intelligence and stock prediction promotes the upgrading of quantitative investment strategies and provides new technical methods for the industry.

Keywords

SI 300 index, A-share market, artificial intelligence, stock prediction, sentiment analysis

1. Introduction

In recent years, the rapid advancement of artificial intelligence has profoundly reshaped financial research and practice, particularly in the field of stock market prediction [1]. Traditional stock prediction models primarily rely on statistical methods such as time series analysis, regression analysis, and financially theory-driven models derived from the Efficient Market Hypothesis. For example, time series models utilize the autocorrelation of historical prices or returns for forecasting [2]; econometric linear regression methods predict returns based on different explanatory variables [3]; and multi-dynamic-factor models analyze and forecast stock market trends from multiple dimensions [4]. These approaches typically assume idealized conditions, such as linear relationships and normally distributed market data, which makes it difficult to capture the nonlinear relationships and extreme fluctuations caused by factors including investor sentiment, unexpected events, and policy changes [5]. Consequently, the predictive accuracy of traditional models tends to be limited under the influence of various exogenous factors.

The emergence of artificial intelligence models has provided new approaches to addressing these challenges. Machine learning and deep learning techniques have demonstrated remarkable capabilities in identifying nonlinear dependencies and hidden patterns within high-dimensional financial data. These models enable researchers to extract key information from massive datasets, including technical indicators, macroeconomic variables, and structured information such as news sentiment and social media activity, thereby improving the accuracy and robustness of stock market predictions.

Although artificial intelligence has been increasingly applied to stock market prediction in China, existing studies on the A-share market remain fragmented. Most research focuses on individual forecasting models, such as LSTM, CNN, Transformer, or hybrid approaches, with limited efforts to systematically compare their applications, advantages, and limitations within a unified framework. In particular, despite the CSI 300 Index being one of the most widely adopted benchmarks for the Chinese A-share market, there is still a lack of comprehensive reviews that synthesize AI-based prediction methods using the index as a representative case.

This study systematically reviews the major categories of artificial intelligence approaches, ranging from traditional machine learning algorithms to advanced deep learning architectures, and evaluates their performance in capturing complex market dynamics. In addition, it discusses hybrid and ensemble models that integrate statistical methods, AI techniques, and sentiment-driven factors to further enhance prediction accuracy. Using the CSI 300 Index as a case study, this research comprehensively examines the applicability, strengths, limitations, and potential optimization directions of various AI models for forecasting the Chinese A-share market. By synthesizing the existing methodologies and empirical findings on AI-based stock market prediction in China, this study aims to provide a comprehensive overview of current research and contribute to the deeper integration of artificial intelligence technologies into stock market forecasting.

2. Overview of the A-share Market and the CSI 300 Index

The Chinese A-share market is one of the largest and most distinctive stock markets in the world. A-shares are stocks denominated in Renminbi (RMB) and primarily traded on the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE). These listed companies are incorporated in mainland China and mainly target domestic investors. However, access for foreign institutional investors has gradually been liberalized through programs such as the Qualified Foreign Institutional Investor (QFII) scheme and the Stock Connect mechanisms linking the mainland with Hong Kong (the Shanghai–Hong Kong Stock Connect and the Shenzhen–Hong Kong Stock Connect). This dual-access framework renders the A-share market both dynamic and complex.

A prominent feature of the A-share market is the high proportion and activity of retail investors [6]. Unlike institutional investors, who typically pursue long-term, strategy-driven investments, Chinese retail investors often exhibit short-term speculation, herd behavior, and emotion-driven trading patterns. This investor composition results in pronounced market volatility and low price efficiency, as market movements are frequently influenced by collective sentiment reactions rather than by fundamental value assessments. Consequently, traditional models that assume rational expectations and stable return distributions often fail to capture the behavioral complexity of the A-share market.

Another defining characteristic of the A-share market is the significant influence of government policies and regulatory interventions. For instance, daily price limits are imposed to constrain stock price fluctuations, providing a certain degree of protection to the domestic market [7]. These policy-driven mechanisms form a semi-administrative market structure distinct from the self-regulating frameworks commonly found in developed markets. While such interventions may mitigate systemic risks, they also introduce regime shifts and nonlinear dependencies, thereby increasing the complexity of modeling and prediction.

In addition, the A-share market is prone to frequent short-term fluctuations, reflecting both speculative trading and strong sensitivity to policy changes. Such volatility is often exacerbated by investors’ overreactions to news, online discussions, and macroeconomic signals, which can lead to temporary bubbles and abrupt corrections. As a result, the market exhibits non-stationary and nonlinear characteristics. These properties pose challenges to traditional econometric forecasting techniques but simultaneously create opportunities for artificial intelligence models capable of capturing complex patterns.

Furthermore, the structural segmentation among the Main Board, the ChiNext Board, and the STAR Market increases market heterogeneity. The Main Board primarily accommodates large, well-established corporations, while the ChiNext and STAR boards focus on innovation-driven, small and medium-sized technology enterprises. Overall, the unique structural features of the A-share market distinguish it from mature Western markets. These characteristics not only complicate traditional quantitative modeling but also open avenues for artificial intelligence approaches that can better handle behavioral complexity, nonlinear dynamics, and multifactor dependencies in financial data. Therefore, understanding these structural attributes is essential for assessing the effectiveness and limitations of AI-based techniques in predicting the CSI 300 Index.

The CSI 300 Index is composed of the top 300 listed companies by market capitalization from the Shanghai and Shenzhen stock exchanges. It is the most influential and important index in China's stock market, comprehensively reflecting the overall performance of A-share listed stock prices.

3. Traditional Approaches to Stock Market Prediction

The traditional methods for predicting the stock market mainly rely on statistics, econometrics and financial theory, emphasizing the analysis of historical prices and fundamental data. These methods usually assume that the market conforms to a certain extent to linear relationships, normal distribution and the efficient market hypothesis.

Time series analysis methods forecast stock prices by examining historical price and return data from the perspective of temporal dependence [8]. By utilizing ordered sequences or collections of data points observed at equal time intervals, this approach can effectively construct models capable of predicting future stock price movements.

Econometric models can also provide a certain degree of predictive power in the stock market. For example, a regression model may be established in which principal component analysis is employed to reduce dimensionality and assign appropriate weights to variables, thereby enabling the model to better accommodate the time-varying characteristics of financial time series [9].

Fundamental analysis, as a traditional forecasting approach, is likewise an effective predictive method. Given the complexity and diversity of factors influencing stock market fluctuations, fundamental analysis evaluates a wide range of determinants affecting stock prices [10]. By examining both microeconomic and macroeconomic factors, it allows for informed inferences regarding a company’s future economic performance and business prospects.

4. The Application Methods of Artificial Intelligence in Stock Prediction

With the rapid development of artificial intelligence, stock market forecasting has gradually evolved from traditional machine learning algorithms to deep learning, graph-based learning, and multimodal information fusion. Compared with conventional statistical models, AI methods provide stronger capabilities for modeling nonlinear relationships and extracting high-dimensional features from financial data. Existing studies on the CSI 300 Index indicate that different AI models exhibit distinct strengths depending on forecasting objectives, prediction horizons, and input data characteristics. Early AI-based studies mainly relied on artificial neural networks (ANN) and support vector machines (SVM). These models primarily utilize historical prices, trading volume, and technical indicators to establish nonlinear relationships between market information and future index movements [11]. ANN demonstrate strong nonlinear approximation capability and achieve relatively high prediction accuracy for continuous price forecasting, whereas SVM generally perform better in directional classification problems and remain effective when training samples are limited. Studies investigated the effectiveness of ANN for forecasting the CSI 300 Index by comparing three different input-output mapping structures. The empirical results demonstrated that ANN successfully captured the nonlinear characteristics of the CSI 300 Index and consistently achieved prediction hit rates above 60%, outperforming random forecasting. Moreover, incorporating multiple input variables, including historical highest, lowest, and closing prices, significantly improved prediction accuracy compared with single-input structures. The study also found that an input window length of approximately 14-20 trading days produced the most stable forecasting performance, indicating that appropriate feature selection and input structure design are critical for ANN-based stock market prediction [12]. However, both methods depend heavily on manual feature engineering and have limited ability to capture long-term temporal dependencies, restricting their performance in highly volatile financial markets.

To address these limitations, researchers increasingly adopted deep learning models, including deep neural networks (DNN), recurrent neural networks (RNN), and particularly long short-term memory (LSTM) networks. Unlike traditional machine learning approaches, deep learning models automatically learn hierarchical representations from raw financial data and are more effective in modeling nonlinear and sequential market dynamics [13]. Studies based on the Chinese A-share market have shown that LSTM consistently outperforms ANN and SVM in predicting CSI 300 price movements because its memory mechanism effectively captures long-term temporal dependencies and reduces information loss during sequence learning. The hybrid information extractor-based long short-term memory (H-LSTM) model achieved strong forecasting performance for the CSI 300 Index, with the best mean absolute percentage error (MAPE), directional symmetry (DS), and Theil's U values of 0.009%, 52.19%, and 0.006, respectively. Furthermore, the model recorded a worst-case mean absolute relative percentage error of only 0.371%, indicating that the deviation between the predicted and actual values of the CSI 300 Index remained within 0.4%. Overall, the predicted price series closely followed the actual movements of the CSI 300 Index, demonstrating the effectiveness of the proposed H-LSTM model in capturing market dynamics and generating highly accurate forecasts [14]. Consequently, LSTM has become one of the most widely adopted benchmark models for medium- and short-term CSI 300 forecasting.

More recently, AI research has shifted from single time-series modeling toward relationship-aware and multimodal learning. Graph neural networks (GNN) explicitly model structural relationships among listed companies, such as industry affiliations, supply chain linkages, and market interactions, thereby capturing information beyond individual price series. Compared with LSTM, GNN demonstrates greater effectiveness when stock correlations play a significant role in driving index fluctuations [15]. In addition, multimodal models integrate heterogeneous information from financial news, macroeconomic indicators, technical indicators, and investor sentiment through Transformer-based architectures and cross-modal fusion mechanisms [16]. The study employed a multilayer perceptron (MLP) model to evaluate the stock price forecasting performance for the CSI 300 Index under two settings: with and without an attention mechanism. The results showed that the attention-enhanced MLP model achieved superior predictive performance. Using the CSI 300 Index and its constituent stocks as the scope of analysis, the model attained a test accuracy of 97.77% and an F1-score of 96.99%, demonstrating the effectiveness of incorporating an attention mechanism into MLP for improving stock market prediction [17]. Multimodal frameworks achieve superior forecasting accuracy and robustness by jointly exploiting structured numerical data and unstructured textual information, especially during periods of heightened market uncertainty.

The evolution of AI methods reflects a transition from single-source data learning to heterogeneous information fusion and from individual stock prediction to market-wide relationship modeling. For forecasting the CSI 300 Index, traditional machine learning models remain suitable for relatively small datasets and scenarios requiring higher interpretability, whereas deep learning methods provide stronger predictive performance for nonlinear time-series forecasting. GNN further improves forecasting accuracy by incorporating inter-stock relationships, while multimodal models demonstrate the greatest potential by integrating market, textual, and macroeconomic information into a unified prediction framework.

Existing evidence suggests that no single AI model performs optimally across all forecasting tasks. Traditional machine learning models, such as SVM and ANN, remain competitive for small-sample classification problems because of their relatively high interpretability and computational efficiency. However, for forecasting the CSI 300 Index, which exhibits strong nonlinear dynamics, temporal dependence, and sensitivity to external information, deep learning models generally provide superior predictive performance. Among them, LSTM has become the most widely adopted benchmark owing to its ability to capture long-term temporal dependencies. Recent studies further demonstrate that GNN-based models outperform conventional sequence models by incorporating inter-stock relationships, while multimodal Transformer frameworks achieve the highest forecasting accuracy through the integration of market data, financial text, macroeconomic indicators, and event information. Therefore, current evidence suggests that LSTM remains the most practical baseline model for CSI 300 forecasting, whereas GNN and multimodal Transformer architectures represent the most promising directions for improving forecasting accuracy under complex market environments.

5. Conclusion

With the continuous growth in the scale of financial market data and the increasing diversity of data types, artificial intelligence has gradually become an important research direction in stock market forecasting. Taking the Chinese A-share market as the research context and the CSI 300 Index as the case study, this paper systematically reviews the application and development of both traditional forecasting methods and AI-based approaches in stock market prediction. Traditional methods, including time-series analysis, econometric models, and fundamental analysis, have established an important foundation for stock forecasting research. However, because these approaches are typically built upon assumptions of linear relationships, stable distributions, and rational market behavior, they exhibit certain limitations when confronted with the complex nonlinear characteristics, behavioral biases of investors, and policy-driven factors that characterize the A-share market.

In contrast, AI-based approaches have demonstrated significant advantages in stock market forecasting owing to their powerful data-processing capabilities, nonlinear modeling capacity, and autonomous learning ability. From early artificial neural networks, support vector machines, and genetic algorithms to deep neural networks, recurrent neural networks, and long short-term memory networks, and more recently to graph neural networks, natural language processing, sentiment analysis, and multimodal fusion models, AI technologies have continuously advanced stock forecasting models toward higher predictive accuracy, stronger generalization capability, and more comprehensive information integration. Existing studies indicate that these approaches can effectively uncover hidden patterns from historical price data, market relationship networks, investor sentiment information, and macroeconomic variables, thereby outperforming traditional models in terms of forecasting accuracy, risk management, and investment decision support.

However, several challenges remain in the current literature. Financial markets are inherently characterized by high uncertainty and dynamic changes, making forecasting models susceptible to structural market shifts and unexpected events. In addition, deep learning models generally suffer from limited interpretability, while the mechanisms for integrating multi-source heterogeneous data still require further optimization. Moreover, model training often depends heavily on data quality and computational resources. Therefore, improving model robustness, interpretability, and real-time adaptability while maintaining high predictive accuracy remains an important direction for future research.

Overall, artificial intelligence has fundamentally transformed the research paradigm of stock market forecasting by significantly enhancing the ability to model nonlinear relationships, temporal dependencies, and complex market interactions. Compared with traditional forecasting methods, AI-based approaches have demonstrated superior predictive performance for the CSI 300 Index, particularly in capturing the dynamic characteristics of the Chinese A-share market. Nevertheless, no single AI model can fully address the unique challenges posed by China's policy-sensitive and retail investor-dominated market environment. Future research should focus on developing dynamic AI forecasting frameworks that explicitly incorporate policy events and macroeconomic shocks to improve model adaptability under rapidly changing regulatory and economic conditions. At the same time, multimodal learning frameworks should further integrate heterogeneous information sources, including market data, macroeconomic indicators, financial news, and retail investor sentiment from Chinese online investment platforms, to better capture the information transmission mechanisms driving fluctuations in the CSI 300 Index. In addition, improving the interpretability of financial AI models remains essential for enhancing model transparency, increasing decision reliability, and facilitating their practical application in quantitative investment and portfolio management. By combining high predictive accuracy with stronger interpretability and broader information integration, future AI-based forecasting systems are expected to provide more robust and practical support for investment decision-making in the Chinese A-share market.

References

[1] Chopra R, Sharma G D. Application of artificial intelligence in stock market forecasting: a critique, review, and research agenda[J]. Journal of risk and financial management, 2021, 14(11): 526.

[2] Box G E P, Jenkins G M, Reinsel G C, et al. Time series analysis: forecasting and control[M]. John Wiley & Sons, 2015.

[3] Altay E, Satman M H. Stock market forecasting: artificial neural network and linear regression comparison in an emerging market[J]. Journal of Financial Management & Analysis, 2005, 18(2): 18.

[4] Ng V, Engle R F, Rothschild M. A multi-dynamic-factor model for stock returns[J]. Journal of Econometrics, 1992, 52(1-2): 245-266.

[5] Cakra Y E, Trisedya B D. Stock price prediction using linear regression based on sentiment analysis[C]//2015 international conference on advanced computer science and information systems (ICACSIS). IEEE, 2015: 147-154.

[6] Los C A, Yu B. Persistence characteristics of the Chinese stock markets[J]. International Review of Financial Analysis, 2008, 17(1): 64-82.

[7] Su D, Fleisher B M. Risk, return and regulation in Chinese stock markets[J]. Journal of economics and business, 1998, 50(3): 239-256.

[8] Idrees S M, Alam M A, Agarwal P. A prediction approach for stock market volatility based on time series data[J]. Ieee Access, 2019, 7: 17287-17298.

[9] Ghorbani M, Chong E K P. Stock price prediction using principal components[J]. PloS one, 2020, 15(3): e0230124.

[10] Baresa S, Bogdan S, Ivanovic Z. Strategy of stock valuation by fundamental analysis[J]. UTMS Journal of Economics, 2013, 4(1): 45-51.

[11] Patel J, Shah S, Thakkar P, et al. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques[J]. Expert systems with applications, 2015, 42(1): 259-268.

[12] Chengzhao Z, Heiping P A N. Experimenting With Three Different Input-Output Mapping Structures of ANN Models for Predicting CSI 300 Index[J]. Management Science and Engineering, 2014, 8(1): 22.

[13] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions[J]. European journal of operational research, 2018, 270(2): 654-669.

[14] MA C, YANG J, REN Y, et al. The Prediction of Shanghai and Shenzhen 300 Index Based on H-LSTM Model[J]. China Journal of Econometrics, 2021, 1(2): 437-451.

[15] Feng F, He X, Wang X, et al. Temporal relational ranking for stock prediction[J]. ACM Transactions on Information Systems (TOIS), 2019, 37(2): 1-30.

[16] Liu Y, Lin Y, Xu C, et al. Multi-modal fusion approaches for stock market prediction: A comprehensive survey[J]. Pattern Recognition, 2026: 114384.

[17] Seo K, Dong Z, Song Y. Neural network-based CSI300 stock prediction: Feature importance and attention mechanism analysis[J]. Electronics, 2025, 14(9): 1729.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.