Proceedings of ICMRED 2026 symposium: The Future of Work: Strategy, Workforce Transformation, and Organizational Renewal
DOI: 10.54254/2754-1169/2026.LD32987

Stock Market Forecasting Technology Driven by Artificial Intelligence

Junxin Deng

Shenzhen Audencia FinTech Institute, Shenzhen University, Shenzhen, China
2024362057@mails.szu.edu.cn

Abstract. Stock market prediction has become a difficult problem in the field of financial engineering due to high data noise, non-stationarity and nonlinear characteristics, which traditional methods can hardly handle effectively. This paper systematically combs the research achievements of artificial intelligence technology in the field of stock forecasting from 2015 to 2026. Starting from traditional machine learning to deep learning, from single modality to multimodal fusion, and from general architectures to financial-specific pre-training, it adopts a five-dimensional evaluation framework to quantify model performance. Studies have found that convolutional neural networks enable the automatic recognition of financial patterns, recurrent neural networks address the challenge of modeling temporal dependencies, Transformer and state-space models break through the efficiency bottleneck of long sequences, and multimodal fusion and financial pre-trained models significantly improve domain adaptability. Current research still faces challenges such as insufficient out-of-distribution generalization and lack of interpretability. Lightweight architectures and causal inference represent the main future directions. This study provides systematic technical references for the design of stock prediction models.

Keywords: Stock Market Forecasting, Artificial Intelligence, Deep Learning, Time Series Modeling Multimodal Fusion

1. Introduction

The stock market has long been a research hotspot in the financial sector. Influenced by multiple factors such as fundamentals, policies and investor sentiment, its price movements exhibit nonlinear, non-stationary and high-noise characteristics, making accurate prediction quite difficult.

Faced with these challenges, traditional econometric methods appear somewhat inadequate. Classic approaches such as ARIMA and GARCH, which are based on linear assumptions and stationarity premises, exhibit limited performance in scenarios involving extreme market conditions or high-frequency data. Early machine learning methods mainly relied on feature engineering, and their dependence on manual features made it difficult for models to mine deep patterns. Research attempted to improve performance through model fusion, and experiments showed that ensemble learning can indeed enhance prediction performance [1].

The emergence of the Transformer architecture in 2017 revolutionized sequence modeling technology. Transformer architecture achieved parallel capture of global dependencies through the

DOI: 10.54254/2754-1169/2026.LD32987

self-attention mechanism, which has become a theoretical cornerstone for temporal modeling [2]. Subsequent technological evolution has been quite rapid: PatchTCN addresses the efficiency issue of long sequences through segmented embedding [3], Auto former adopts a hierarchical autoregressive structure for further optimization [4], and ss-Mamba realizes long-distance dependency modeling with linear complexity [5], becoming an important alternative to Transformer. At the data fusion level, multimodal fusion frameworks have begun to integrate multi-source information such as price data, news texts, and K-line images. Research has demonstrated the complementary value of such heterogeneous data [6]. In terms of the pre-training paradigm, the Chronos model demonstrates the power of universal pre-training, yet its performance degrades significantly when applied to financial data. Subsequently, the finance-specific Kronos model was proposed [7], while the application of masked autoencoding in financial time series was explored [8].

Existing research has formed an evolutionary path from single models to integrated frameworks, from single modal to multi-source data fusion, and from general modeling to domain adaptation. However, there is a lack of systematic integrated analysis of the advantages and limitations of different technical paths. This paper will evaluate model performance from five dimensions: expressive ability, long-dependency modeling capability, data utilization capability, interpretability, and computational cost, and sort out the technological evolution context from three perspectives: model architecture, data fusion, and training paradigm.

As shown in figure 1, this study sorts out the technological evolution path from traditional machine learning to deep learning, from single modality to multimodal fusion, and from general modeling to financial-specific pre-training.

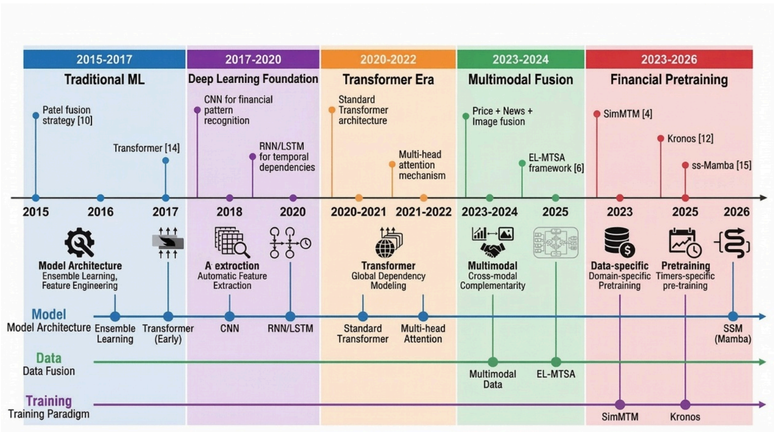

Figure 1. Timeline of the evolution of artificial intelligence stock forecasting technology (2015-2026)

2. Base model: traditional machine learning and deep learning methods

2.1. Fusion strategies for traditional machine learning

Traditional machine learning methods achieve pattern recognition of financial data through feature engineering and model integration. Algorithms such as support vector machines and random forests have become common tools in this field due to their nonlinear fitting capabilities. Ensemble learning improves generalization ability and stability by aggregating the results of multiple base learners.

This study selected 10 years of trading data from India's CNX Nifty and S&P BSE Sensex indices as experimental samples [1], used 10 technical indicators as input features, and compared the

DOI: 10.54254/2754-1169/2026.LD32987

prediction performance of single models (ANN, RF, SVR) with two-stage fusion models (SVR-ANN, SVR-RF, SVR-SVR). In this study, SVR was employed to denoise preprocessed features, and a rolling window was used to partition the dataset to avoid look-ahead bias. The experimental results show that for short-term predictions (1–10 days), the RMSE of the fusion model is reduced by 10%–15% compared with single models, while for long-term predictions (15–30 days), it is reduced by 20%–25%. Among them, the SVR-ANN combination is optimal, achieving an 18%–22% improvement over the standalone ANN. This study constructs a reusable machine learning fusion paradigm, providing a benchmark reference for subsequent deep learning integration frameworks.

2.2. Convolutional neural networks and financial pattern recognition

Convolutional neural networks achieve automatic feature extraction through local weight sharing and spatial convolution, a characteristic that gives them unique advantages in stock pattern recognition tasks From the application perspective, CNN in the financial field is mainly divided into two categories: K-line image pattern detection and one-dimensional time series local feature extraction.

This study took the K-line data of 500 A-share stocks from 2018 to 2023 as the research object [9], converted the K-line data into 28×28 grayscale images, and proposed a two-layer clustering + CNN framework for pattern recognition and prediction. This study adopts similarity matching to cluster K-line patterns, extracts features using CNN and combines them with LSTM for prediction, and employs five-fold cross-validation to ensure the robustness of the results. The experimental results show that the accuracy of morphological recognition reaches 89.2%, which is 12.5% higher than that of traditional template matching. The annualized rate of return of the strategy based on this morphology is 18.7%, with an excess return of 9.3%.

As shown in figure 2, this method first converts K-line data into 28×28 grayscale images, identifies pattern categories through two-layer clustering, and then uses CNN to extract features and make predictions. Experiments show that the morphological recognition accuracy of this method reaches 89.2%, which is an increase of 12.5% compared with traditional template matching.

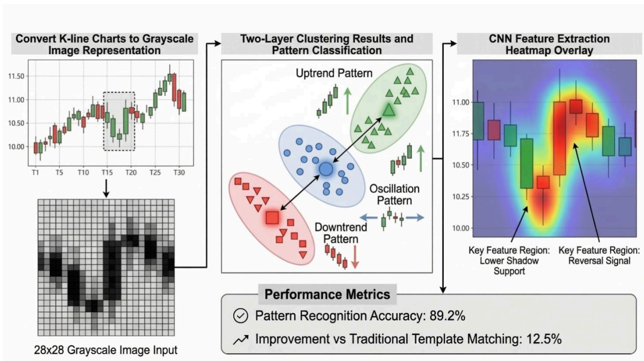

Figure 2. CNN financial pattern recognition flowchart

2.3. Recurrent neural networks and temporal dependence modeling

Recurrent Neural Networks dynamically memorize sequence information through the recursive structure of hidden states, reducing reliance on manual features. However, standard RNNs suffer from the vanishing gradient problem, which restricts their ability to process long sequences. Long Short-Term Memory networks and Gated Recurrent Units alleviate this problem through gating mechanisms and have been widely applied in the prediction of stock price trends and volatility.

DOI: 10.54254/2754-1169/2026.LD32987

As an important variant of RNN, LSTM demonstrates unique advantages in forecasting stock price trends and volatility Studies show that the multivariate LSTM directional prediction accuracy reaches 63.8%, an increase of 7.2% compared with the univariate model, and the RMSE is reduced by 11.5%. This case has established the fundamental paradigm of temporal dependency modeling, providing an important reference for the subsequent design of hybrid architectures.

2.4. Summary of this section

Traditional machine learning methods have verified the effectiveness of ensemble learning in high-noise data, CNN has realized the automatic recognition of financial patterns, and RNN has solved the difficult problem of temporal dependency modeling. However, these basic models still have obvious limitations when dealing with ultra-long sequences and extreme market conditions, laying the foundation for the subsequent introduction of deep learning methods.

3. Modern time series modeling models: transformer and state space models

3.1. Core mechanisms of the standard transformer and financial adaptation

With the self-attention mechanism as its core, Transformer breaks through the limitations of RNN sequential computing and enables the parallel capture of global dependencies The multi-head attention mechanism can simultaneously capture the synergistic relationships among multiple variables, and positional encoding addresses the issue of sequence order sensitivity. These features provide a new technical path for financial time series processing using the Transformer.

Transformer architecture was proposed [2], which achieves parallel capture of global dependencies through the self-attention mechanism and has become the theoretical cornerstone of temporal modeling. Further analysis analyzed the adaptation challenges of Transformer in the financial sector [10], pointing out that financial data exhibits unique characteristics such as high noise, irregularity, and extreme volatility. Direct application of the standard Transformer fails to achieve desirable performance, requiring specialized optimizations in noise processing, irregular sampling, positional encoding, and other aspects.

As shown in figure 3, the standard Transformer adopts an encoder-decoder architecture, which realizes the parallel capture of global dependencies through the multi-head self-attention mechanism, breaking through the serial computing limitations of traditional recurrent neural networks.

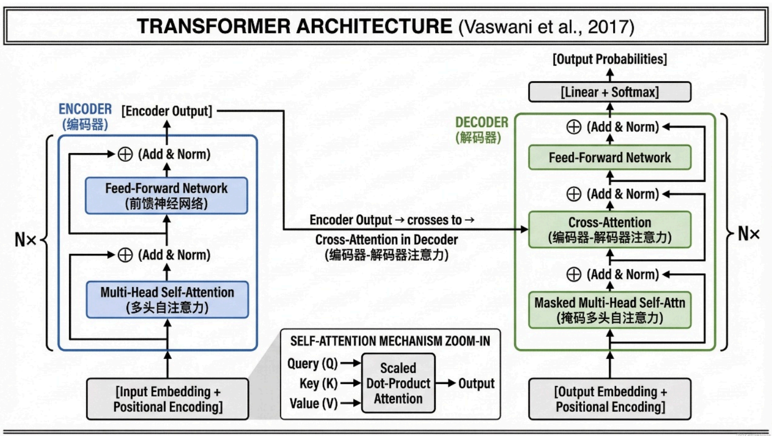

Figure 3. Standard transformer architecture diagram

DOI: 10.54254/2754-1169/2026.LD32987

3.2. Technical improvements of the second-generation temporal transformer

The second-generation temporal Transformer addresses the efficiency bottleneck of long-sequence modeling through technologies such as sparse attention and sequence decomposition, demonstrating outstanding performance in general time-series tasks.

PatchTCN adopts piecewise embedding and channel-independent design to reduce the computational cost of attention, reducing the RMSE by 28% and improving the inference speed by 40% compared with the standard Transformer [3]. Autohformer adopts an efficient hierarchical autoregressive structure based on PatchTCN [4], decomposes sequences into trend and seasonal components, improves efficiency by 15%, and reduces the 30-step prediction MAE by 9%. These two studies provide architectural references for lightweight modeling of financial time series.

3.3. State space models and efficient long sequence modeling

State-space models implement long-sequence dependency modeling with linear time complexity, making them an important alternative to Transformer. Mamba architecture captures long-range dependencies through state-space dynamic systems, avoiding the quadratic complexity problem and demonstrating unique advantages in processing ultra-long sequences.

The ss-Mamba model was proposed, which integrates semantic embeddings with the Mamba module to enhance the representation capability of non-stationary patterns [5]. This study compared the performance of ss-Mamba with that of Transformer and LSTM on general time-series datasets and synthetic financial noise data. Experimental results show that in the 1000-step prediction task, ss-Mamba reduces the RMSE by 19% and shortens the training time by 52% compared with Transformer and reduces the RMSE by 25% compared with LSTM. This study provides a lightweight solution for ultra-long financial time series processing.

4. Data-driven expansion: multimodal fusion

4.1. Motivations and hierarchies of multimodal fusion

Single-modal data is difficult to comprehensively characterize market dynamics, and multimodal fusion improves representation capabilities through cross-modal information complementarity. From the perspective of fusion levels, multimodal fusion can be divided into three levels: feature-level fusion, decision-level fusion, and deep fusion, with different levels applicable to different data scenarios and task requirements.

Research on multimodal fusion can be traced back to explorations of vision-language models such as VideoBERT, and these studies have provided a paradigm reference for the fusion of K-line images and textual sentiment. In recent years, the development of multimodal large models has further advanced the progress of multimodal fusion technology in the financial sector.

As shown in figure 4, the three main paradigms of multimodal fusion include feature-level fusion, decision-level fusion, and deep fusion. The three are parallel paradigms and are suitable for different data scenarios and task requirements.

DOI: 10.54254/2754-1169/2026.LD32987

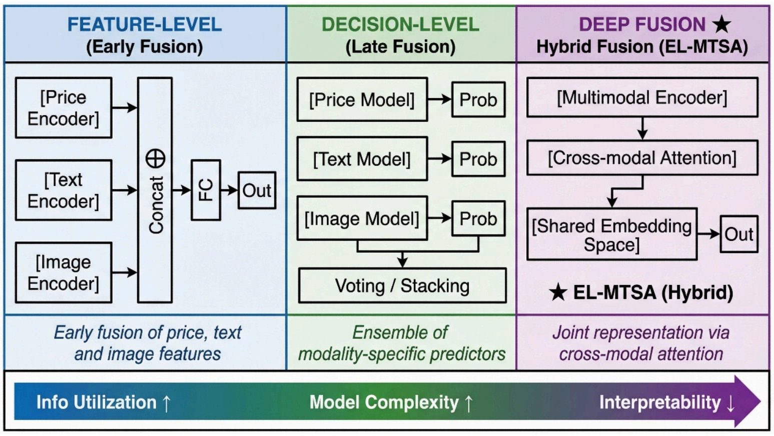

Figure 4. Hierarchical structure of financial multimodal fusion paradigm

4.2. Empirical progress in multimodal financial fusion

This study took data from the Hong Kong stock market from 2020 to 2024 as experimental subjects [6], constructed a multimodal dataset including prices, trading volumes, news sentiment, and macroeconomic indicators, and proposed the EL-MTSA framework that integrates LSTM, TextCNN, and Transformer sub-models. This study adopts feature-level fusion of multimodal encodings, dynamically assigns weights through the attention mechanism, and employs five-fold cross-validation to ensure robust results. The experimental results show that the accuracy of the 7-day rise and fall prediction reaches 76.5%, with an F1-score of 74.2%. Compared with the single-modal model, the accuracy is improved by 8.9%, and the Sharpe ratio rises from 1.2 to 1.8, which effectively proves the economic value of multimodal fusion. This study provides a complete multimodal fusion engineering solution and quantitatively demonstrates the complementary value of multimodal information.

5. Model paradigm upgrade: pre-training and foundation models

5.1. Limitations of general temporal pre-trained models

The pre-training-fine-tuning paradigm learns universal representations from large-scale data to adapt to downstream tasks, making it an important paradigm in current deep learning. However, general time-series models struggle to adapt to the unique characteristics of financial data.

The Chronos model reduces the zero-shot prediction MAE by 15% compared with LSTM on standard datasets, demonstrating the powerful capability of universal pre-training [11]. However, the study also reveals the fundamental challenge of domain adaptation: when applied to financial data, the performance of Chronos degrades by up to 22%, highlighting the severe limitations of general-purpose time-series pre-trained models in handling the characteristics of financial data such as extreme noise and non-stationarity.

5.2. Specialized financial time-series pre-trained model

In response to the domain adaptation problem of general pre-trained models, financial-specific temporal pre-trained models have emerged as the times require. The Kronos model was proposed [7], which uses a finance-specific tokenizer to process K-line data and conducts pre-training in an autoregressive manner on 20 years of historical data from 5,000 individual stocks across 10 global

DOI: 10.54254/2754-1169/2026.LD32987

markets. The results show that zero-shot predictions reduce the MSE by 22% compared with Chronos and by 35% compared with LSTM; the performance degradation in cross-market transfer is only 8%, which is significantly better than the 25% of general pre-trained models. This study has established a dedicated pre-training paradigm for finance, which effectively improves the efficiency of few-shot and cross-market prediction, and opens up a new path for financial time series pre-training

5.3. The pre-training paradigm of masked autoencoding

Masked autoencoding learns data representations by randomly masking parts of the input and reconstructing the original signal. It has achieved remarkable results in the fields of computer vision and natural language processing and has been gradually applied to time series data analysis in recent years.

The SimMTM framework was proposed [8], which adopts a multi-random masked sequence aggregation reconstruction strategy combined with manifold learning constraints, and conducted experiments on price sequence data of US stocks and A-shares from 2010 to 2022. The results show that the downstream stock price prediction RMSE is reduced by 14% compared with MAE pre-training, the rise-fall classification Accuracy reaches 72.1%, and a 60% masking ratio is the optimal setting. This study constructs a new paradigm for self-supervised pre-training of financial time series and verifies the effectiveness of the masked reconstruction strategy in the financial domain.

The causal robustness rating method proposes provides a systematic quantitative tool for evaluating the robustness of pre-trained models under distribution shift [12]. This method is based on the causal inference framework and can distinguish whether the improvement in model performance stems from genuine causal relationship learning or merely captures spurious correlations, providing a scientific basis for evaluating the generalization ability of pre-trained models such as Kronos and SimMTM.

6. Conclusion

This study systematically reviews the research achievements of artificial intelligence technologies in the field of stock market forecasting from 2015 to 2026. Based on a five-dimensional evaluation framework covering expressive power, long-range dependency modeling capability, data utilization capability, interpretability, and computational cost, it constructs a technical system ranging from traditional machine learning to deep learning, from unimodal modeling to multimodal fusion, and from general time-series architectures to financial-specific pre-trained models. Studies have found that CNN enables automatic recognition of financial patterns, RNN addresses the modeling challenges of temporal dependencies, Transformer and state-space models break through the efficiency bottlenecks of long sequences, multimodal fusion achieves complementarity among heterogeneous data, and dedicated pre-trained models for finance significantly improve domain adaptability.

This study provides a systematic technical reference for the design and empirical research of stock forecasting models. The five-dimensional evaluation framework offers a quantitative basis for comparing the performance of different technical approaches, helping researchers and practitioners select appropriate model architectures according to specific scenarios.

This study has the following limitations: First, no empirical comparison was conducted on various technical paths, and there is a lack of unified quantitative analysis of the performance characteristics of each model in different tasks; second, the research mainly relies on existing

DOI: 10.54254/2754-1169/2026.LD32987

literature reviews and lacks original empirical verification for specific markets; third, the study fails to fully cover the latest model architectures and training paradigms.

Future research should integrate lightweight architectures, causal inference, multimodal alignment, domain pre-training, and compliance risk control to construct prediction models with balanced performance, efficiency, interpretability, and compliance. Optimization of hybrid architecture based on Mamba and Transformer to adapt to high-frequency trading requirements; introduction of structural causal models to distinguish correlation from causality and improve decision-making transparency; development of a unified heterogeneous data representation space to capture event-driven cross-modal influence mechanisms; conducting instruction fine-tuning and reinforcement learning alignment for financial-specific pre-trained models; integration of economic theories and market microstructure knowledge to enhance robustness under extreme market conditions.

References

[1] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015) Predicting stock market index using fusion of machine learning techniques. Expert systems with applications, 42(4), 2162-2172.

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017) Attention is all you need. Advances in neural information processing systems, 30.

[3] Zhang, J., Guo, L., Song, L., Gao, S., Hao, C., & Li, X. (2024, November) PatchTCN: Patch-based transformer convolutional network for times series analysis. In Proceedings of the 2024 3rd international symposium on computing and artificial intelligence (pp. 1-9).

[4] Zhang, Q., Wen, H., Li, M., Huang, D., Yiu, S. M., Jensen, C. S., & Liò, P. (2025) Autohformer: Efficient hierarchical autoregressive transformer for time series prediction. arXiv preprint arXiv: 2506.16001.

[5] Ye, Z. (2025) ss-Mamba: Semantic-Spline Selective State-Space Model. arXiv preprint arXiv: 2506.14802.

[6] Kong, J., Zhao, X., He, W., Yang, X., & Jin, X. (2025) EL-MTSA: stock prediction model based on ensemble learning and multimodal time series analysis. Applied Sciences, 15(9), 4669.

[7] Shi, Y., Fu, Z., Chen, S., Zhao, B., Xu, W., Zhang, C., & Li, J. (2026, March) Kronos: A foundation model for the language of financial markets. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 30, pp. 25366-25373).

[8] Dong, J., Wu, H., Zhang, H., Zhang, L., Wang, J., & Long, M. (2023) Simmtm: A simple pre-training framework for masked time-series modeling. Advances in Neural Information Processing Systems, 36, 29996-30025.

[9] Li, X., Liu, Q., Hu, Y., & Liu, H. (2024) The Double-Layer Clustering Based on K-Line Pattern Recognition Based on Similarity Matching. Information, 15(12), 821.

[10] Saravana, M. K., Roopa, M. S., Arunalatha, J. S., & Venugopal, K. R. (2026) Transformers for Multivariate Time Series Forecasting: Comprehensive Analysis, Challenges, Research Opportunities and Future Prospects. IEEE Access.

[11] Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., ... & Wang, Y. (2024) Chronos: Learning the language of time series. arXiv preprint arXiv: 2403.07815.

[12] Lakkaraju, K., Kaur, R., Zehtabi, P., Patra, S., Valluru, S. L., Zeng, Z., ... & Valtorta, M. (2025) On creating a causally grounded usable rating method for assessing the robustness of foundation models supporting time series. arXiv preprint arXiv: 2502.12226.