Research and Analysis of Stock Prediction Based on Deep Learning

Hanyu Zhang

St. Francis Methodist School, Singapore, Singapore
hanyuzhang149@gmail.com

Abstract. In recent years, with all the relentless strides in artificial intelligence, deep learning could now be applied to almost all fields, including but not limited to health care, scientific research, and financial analytics. Among these applications, predicting and assessing stock market trends using deep learning is among the most fervent and promising as far as research is concerned. This method overcomes the shortcomings manifest in traditional statistical models, such as linear regression or ARIMA, for providing better prediction accuracy through complicated nonlinear patterns. Incorporating time-series modeling, natural language processing, and deep multimodal learning will allow an investor better to assess risk and opportunities in trading, try to understand the dynamics of the stock market. The paper at hand represents a holistic study of significant studies and updates in the field over the past late five years. Further, stock prediction using deep learning will likely achieve higher reliability, transparency, and interpretability, with mutual reinforcement of further technological innovation and the practicalities of financial management.

Keywords: Stock Prediction, Deep Learning, Convolutional Neural Network, Transformer Neural Network, Bidirectional Encoder Representations from Transformers

1 Introduction

Traditionally, predicting stock prices relied heavily on fundamental statistical methods. But these statistical methods have inflexible underpinnings: the assumption that market changes are linear, and fixed patterns repeat over time. Yet the very essence of price movements in the stock market is dynamic, influenced by incredibly complex phenomena that often transcend the imagination of traditional modeling methods.

However, just a couple of years ago, the field of artificial intelligence finally began to provide deep learning-what can be described as a very brilliant brain that can learn complex patterns and regularities from huge databases autonomously without reliance on previously established human rules. This fact alone drew great attention to this feature from all over the world, and the scientific community began testing different deep learning models concerning stock prediction.

Initially, the combination CNN-LSTM was very successful. Indeed, one can think about it in this way: it has two skills. First, it applies CNN to rapidly detect short-term

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_81

changes in prices, such as small trend patterns. Then, it uses an LSTM to connect those short-term patterns and analyze long-term trends in the data. In general, it performs better quite regularly compared to one of the two only.

However, later on, a better and stronger version was developed, the Transformer [1], along with its improved versions, such as the Informer [2]. This model's main strength lies in its holistic view, which enables it to view relationships among all data points over a very long period and, more importantly, adequately capture the periodic behavior of long-term patterns. Compared with older methods, it often outperforms them when longer-term perspectives are required.

Beyond the numbers, such as prices and trading volumes, researchers have recently realized that the market sentiment plays an important role. Therefore, they structured new approaches using an AI model to analyze financial news such as FinBERT [3], which classifies the news as positive and negative. The output of this model is then converted into a numerical measure of "market sentiment" and combined with historical price data as inputs to a deep learning model. The model, thus, captures "public opinion trends" in making judgments and shows significant prediction efficiency, especially in critical periods around these events.

First, though these deep learning methods have reached good experimental results, such as forecasting price direction with 65%-70% accuracy rates, the MAE is much lower than the traditional models- they are far from being considered "money-printing machines." The models are complicated and can "learn the wrong things," and they require vast high-quality data and fine-tuning. Moreover, an ideal lab setup is one thing, but running in the real-life market- which is subjected to countless surprises and transaction costs- is a whole different story. Thus, deep learning today is a very powerful help in human decision-making and not in replacing humans.

This area, indeed, moving at a fast pace, filled with challenges, pertinently warrants a review and summarization to cover all major scholarly work over the last five years. Thus, the review seeks to clarify the following facets:

The area has witnessed rapid development over the past few years, with various breakthroughs and challenges experienced along the journey. It will do justice to the past five years of significant work in this realm. Thus, this review aims at the following: how to select appropriate models under different circumstances, analyze their advantages and disadvantages, and offer perspectives on the future.

2 Overview of Mainstream Methods

2.1 CNN--LSTM Model

The hybrid CNN--LSTM model combines the Convolutional Neural Network's ability to extract local features with the Long Short-Term Memory network's ability to model temporal dependencies. The idea is simple: first, extract local fluctuation features from data, such as prices and trading volumes, via convolutional layers; then feed these feature sequences into an LSTM network to learn long-term temporal relationships; lastly, the price changes or probabilities of rise/fall are output through fully connected layers. Studies have shown that this hybrid architecture demonstrates significantly better short- and medium-term prediction performance than either CNN or LSTM alone [4, 5].

Input historical price sequences go through a one-dimensional convolutional layer to extract local features, an LSTM layer to learn temporal dependencies, and finally, regression or classification prediction through a fully connected layer. Common input features are prices, trading volumes, and various technical indicators. From a practical point of view, using this hybrid architecture would enable the simultaneous capture of short-term fluctuations and long-term trends; hence, it enjoys the complementarity of both networks and thus draws on the advantage of possessing an excellent nonlinear relationship modeling capability for higher prediction accuracy. On the downside, the critical drawbacks of this hybrid architecture are its complex model structure, huge parameters, and high demands on computation and data volume; it is probably overfitted when the samples are scarce or noisy; its interpretability is poor; it is relatively inconvenient to use this model directly in practice decision making.

2.2 Methods Based on Transformer (Including Informer and Autoformer)

The Transformer model, which is based on a self-attention mechanism [1], can directly model dependencies between any two positions in a time series. Hence, it is well-suited for capturing long-time-span data. Unlike recurrent networks, the Transformers allow for parallel computations during training, which substantially boosts efficiency.

Inference and Autoformer are two variants of the Transformer architecture that have been developed specifically to address time series forecasting [2, 6]. Acquiring only those temporal dependencies that are necessary for achieving high-quality predictions, Informer employs a sparse self-attention mechanism, thereby greatly reducing its computational burden. The Autoformer applies trend and seasonal decomposition within the model architecture, thus simplifying capturing periodic change in financial data.

These types of models are generally developed on encoder-decoder architecture. The input is a price sequence, and the output denotes value predictions for future time periods. Many other features could be added, such as technical indicators and sentiment data, to make the model increasingly aware of market changes.

The Transformers and their variants have stellar performance in long-sequence predictions, capturing global and local dependencies. Compared to classical LSTM, it has relevant advantages: parallelized training, efficient modeling, and stronger generalization ability.

Complexity in modeling, vast parameters, and high demands in computation and data quality. With such a small amount of samples or high noise, prediction performance will suffer drastically. Besides, since financial time series do not typically exhibit stable patterns, the performance of such a model in real trading environments remains questionable.

2.3 Fusion Methods of Text Sentiment and Price Series

Multi-modal fusion methods combining text sentiment and price series improve the sensitivity of stock prediction by integrating textual information from financial news, announcements, social media, etc., with market data. The core idea is: using pre-trained language models in the financial domain (e.g., FinBERT [3], which is built upon the BERT architecture [7]) to extract text sentiment, then fusing these sentiment features with data such as prices and technical indicators to predict market trends.

FinBERT [3], trained on large corpora of financial text, can identify positive, neutral, or negative sentiment in financial texts. By aggregating daily news sentiment statistics, a sentiment time series can be generated and aligned temporally with price data.

Models typically input both sentiment features and price series into LSTM or Transformer models to learn the relationship between the two types of information. Research finds that after incorporating sentiment features, short-term price prediction accuracy usually improves significantly, especially during major news or earnings report periods [8].

3 Comparative Analysis of Experimental Results on Model Performance

3.1 General Description

To verify the performance differences of deep learning models in stock prediction tasks, this section synthesizes the results from multiple representative studies over the past five years [9-11], systematically comparing three mainstream model types: hybrid CNN--LSTM models, models based on Transformer and its variants (Informer, Autoformer, etc.), and multi-modal fusion models of text sentiment and financial time series (FinBERT + LSTM/Transformer). The analysis focuses on prediction accuracy (accuracy, MAE, RMSE), model training efficiency and stability, sensitivity to data scale, and applicable scenarios. Results indicate that CNN--LSTM performs stably in short-term prediction; Transformer models have significant advantages in modeling long-sequence dependencies; and the FinBERT fusion model performs best in event-driven markets.

3.2 Method One: Hybrid CNN--LSTM

Typical experiments use daily data from the CSI 300 Index and S&P 500 Index, spanning 2015 through 2023. Features include closing price, trading volume, and technical indicators such as MACD, RSI, and KDJ. The data are split into training, validation, and test sets in a 7:2:1 ratio, with a sliding window input of 30 days predicting the 31st-day price. The model structure comprises a two-layer CNN and a single-layer LSTM, the Adam optimizer, a learning rate of 0.001, and Mean Squared Error (MSE) as the loss function.

In an empirical study, Wang et al. determined an average RMSE of 0.041 is achieved in the daily prediction task using CNN—LSTM [5], and the directional prediction accuracy was about 70.2%. This represents a significant improvement from standalone LSTM (64.7%) and a standalone CNN (61.5%). The model captures local price fluctuation patterns through convolutional layers and learns about trend changes via LSTM, making its performance much more reliable during "oscillating markets" and "trend reversal" phases. In minute-level high-frequency data experiments, CNN–LSTM showed an approximately 18% decrease in RMSE in 3-minute window predictions, but its advantage diminished in low-volatility intervals, indicating the model is more suitable for capturing short-term nonlinear features [12]. The training time is about 30% longer than for traditional LSTM; however, convergence on the validation set is much more stable. They observed that if the amount of data was sufficient and if the Dropout and regularization techniques were applied, the model's risk of overfitting could be effectively reduced.

CNN--LSTM excels in capturing short-term market fluctuations and pattern features, making it a robust baseline model for stock prediction. In comparative experiments across multiple papers (e.g., "Deep Convolutional Transformer Network for Stock Movement Prediction" [13]), CNN--LSTM is often used as a performance benchmark. Its advantages lie in its clear structure and ease of implementation, but it relies heavily on data scale and struggles to cover long-term dependencies.

CNN--LSTM is a good baseline model in stock prediction, given its strength in capturing short-term market fluctuations and features in the pattern. In most comparative works, such as those for the paper "Deep Convolutional Transformer Network for Stock Movement Prediction" above, CNN-LSTM serves as a performance benchmark. Its clarity in structure and ease of implementation is provided, and it highly depends on the scale of data and poor functioning over long-term dependencies.

3.3 Method Two: Based on Transformer and Its Variants

Most Transformer-based models consider long-cycle data (60-240 day windows) to show their effectiveness at modeling long-term dependencies. The experimental data set consists of Chinese A-shares and US stock market data, with features including opening price, high price, trading volume, market sentiment index, etc. The Informer model draws on a sparse attention mechanism and hierarchical decoding strategy; Autoformer further builds on this by adding trend and seasonal decomposition modules, thus improving its handling of the non-stationary nature of financial time series.

According to the Informer paper by Zhou et al. [2], Informer reduced RMSE on benchmark tasks like ETT and NASDAQ100 by 15% to 25% compared to LSTM; thus, training speed was increased by about 1.7 times. It also obtained 71.5% directional prediction accuracy in stock market prediction experiments [6], while Autoformer achieved 70.8% directional prediction accuracy. The two outperform RNN and CNN-LSTM in long-cycle fluctuation prediction. As for these long-term window predictions, the Informer remains quite stable; the computation is redundant, showing good performance in a 120-day window prediction. The Autoformer also has the potential to explore long-term trend signals in cyclical industries like energy and banking. However, Transformer models are computationally expensive; they require 2 to 3 times more GPU memory than LSTM, and they are sensitive to data quality and hyperparameters.

Transformer and its variants have straightforward strengths in modeling long-sequence dependencies and representing complex financial structures, which naturally guide their application in macro-trend forecasting and investment signal mining over a medium-to-long horizon. Compared to CNN--LSTM, it strikes with periodicity better but could lose ground when data scarcity comes with high noise.

3.4 Method Three: Multi-modal Fusion of Text Sentiment + Financial Time Series

This approach synthesizes financial text and times series data. Pre-trained language models—particularly, FinBERT [3]—extracts sentiment signals from financial news, corporate announcements, and social media postings. The daily aggregated positive, neutral, and negative sentiment scores provided by FinBERT are then aligned with stock prices. Common experiments feed the sentiment features obtained via FinBERT into an LSTM or a transformer network together with price indicators. In the paper by Liu et al. on "Financial Sentiment Analysis Using FinBERT," such a modeling approach is compared—in terms of performance with and without sentiment features—by utilizing the S&P 500 and Dow Jones stock indices.

The results have indicated that, while sentiment features are included, a significant improvement in short-term prediction performance occurs: LSTM accuracy increases from 67.5% to 74.3%, and RMSE reduces from 0.042 to 0.034; accuracy exceeds 78% during major event windows, for instance, policy releases or earnings periods. The negative sentiment captured by FinBERT tends to be positively correlated with declines in the market; thus, it reflects a change in market psychology in advance. For this reason, during the impact of the 2020 pandemic, the FinBERT-LSTM model picked the trend reversal signals up two days earlier than the traditional LSTM. However, multi-modal fusion is computationally intensive—both FinBERT's encoding phase runs on a separate GPU, while any delays or biases in the textual data will propagate through to the predictions.

The FinBERT fusion model substantially enhances the short-term prediction capability of event-driven markets and adds interpretive value to the model. It primarily benefits from capturing non-numerical "sentiment information," provided that the underlying text data is of requisite high quality. Its output is most relevant when utilized in a composite prediction framework alongside price models.

3.5 Comprehensive Comparative Analysis

Synthesis of the experimental results thus far suggests that these three classes of models have different areas in which they can be applied. As shown in table 1, CNN--LSTM has a relatively simple structure and high stability, allowing it to be applied for short-cycle predictions. Transformer models can process long-sequence data, making them more effective in conducting macro-trend and periodic analyses. At the same time, multimodal fusion models can integrate external textual information, making them well suited for event-driven or news-driven markets.

Table 1. Performance comparison of different models

Model type	Average accuracy	RMSE (erroe)	Calculation efficiency	Requirement of data	Main advantages and disadvantages
CNN–LSTM	about 70%	0.041	medium	medium to high	Good at short-term forecasting and stable structure
Transformer（Informer/Autoformer）	71–72%	0.037–0.039	low	high	Strong long-term prediction, high computational cost
FinBERT	about 74%	0.034–0.035	medium to low	very high	Strong long-term prediction, high computational cost

Overall, CNN-LSTM balances performance and stability in short-term tasks; the Transformer shows significant advantages in long-sequence modeling. In event-driven scenarios, FinBERT multi-modal fusion results perform the best [14].

4 Challenges and Prospects

4.1 The Primary Challenges Currently

Although deep learning methods get considerable experimental improvements in stock predictions, they suffer from constraining issues for actual financial applications:

Markets' Nonstationarity and High Noise. Markets are affected by changes in macroeconomics and policy, unpredictable events, and many other factors, leading to highly nonstationary and noisy data. Generally, in such highly nonstationary conditions, the patterns learned by a model on the training set are unlikely to hold in perpetuity. Especially for the estimation of long-range dependencies, if an underlying sequence has a substantial random component, the model tends to mistake noise for the signal. The transformer and its variants thus retain an advantage in long sequence modeling but have to rely on sequence decomposition and sparse strategies to mitigate the noise.

Overfitting and Concept Drift. Deep networks, particularly large-parameter models like Transformers, are prone to overfitting either when data are insufficient or when the nature of sampled distribution changes over time; along with structural changes in the market (concept drift), it renders models passing offline backtests but failing in live trading. The common mitigants are cross-period rolling training, time series regularization, and online update strategies, but gaining robustness is pretty challenging.

Interpretability and Compliance Requirements. High demands of interpretability exist for financial decision-making scenarios; however, deep models are mostly treated as "black boxes," which generally prohibits their adoption in institutional trading, risk control, and compliance processes. Some progress was made in interpretability research—for instance, attention visualization and local surrogate models—but still, a gap is present.

Quality and Coverage of Data. Multi-modal data, such as high-frequency quotes, company announcements, news, and all data from social media, is highly expensive to ensure quality. Temporal alignment, denoising, and debiasing are also huge problems. While text sentiment can add signals during event-driven windows, For instance, FinBERT proved precious in extracting financial sentiment [3]; however, in the case of poor or delayed text coverage, the fusion effect would hardly result.

Computational Costs and Engineering Barriers. Large model training and real-time inference require powerful compute resources, limiting small research groups or small-to-medium institutions. Further model deployment entails engineering considerations around latency, memory, and stability.

4.2 Future Research Directions

These kinds of improvements can be divided into several dimensions with respect to their practical usability and robustness:

Adaptive and Online Learning: Development and research into online learning algorithms—incremental learning, meta-learning, etc.—that support rapid adaptation under concept drift to maintain live performance, hence being closer to the financial market's one-time training characteristic.

Stronger Multi-modal Fusion Frameworks: Systematically combine price series, company fundamentals, macroeconomic indicators, and text (news, announcements, social media). Further consider temporal delays and causal structures between modalities to enhance the system's responsiveness to event-driven fluctuations. FinBERT [3] is an important starting point.

Graph Neural Networks and Cross-Asset Relationship Modeling: Represent the relationship between stocks, sectors, and industries using graphs and employ Graph Neural Networks to model group dynamics and contagion effects for improved prediction and risk estimation both at the level of individual stocks and portfolios [15].

Lightweight and Efficient Variants: Sparse attention, low-rank approximation, and model distillation will be applied for computation savings on Transformers and large language models in financial applications, thus making them deployable and operable in real time. Informer/Autoformer already gave directions and ideas in this regard [2, 6].

Interpretability and Robustness Tools: These would, hence, be required to cover the auditability requirement of the systems: causal inference, invariance model checking, and counterfactual analysis for understandable and input perturbation-robust financial prediction systems.

Standardization of Multi-institution Shared Datasets and Evaluation Systems: Intensive data sharing will standardize more extensive multi-modal datasets and unified backtesting standards across the industry/sector, assisting in lessening the reproducibility plight and allowing fair comparisons of methods.

5 Conclusion

This article has compared, against representative research and recent experimental results, the performance and applicable scenarios of three categories of deep learning methods concerning stock prediction. The hybrid CNN--LSTM is robust in capturing short-term and local patterns and thus should serve as a baseline solution for applied engineering. The Transformer-based variants, such as Informer and Autoformer, are better than others for modeling long sequences and periodic structures and thus are suited for medium-to-long-term and macro-trend analyses. Indeed, the multi-modal fusion of text sentiment and time series, as represented by FinBERT, can significantly improve the sensitivity and explanation power of the short-term prediction in an event-driven setting.

References

1. Vaswani, A. et al.: Attention Is All You Need (The Proposal of the Transformer Model) https://arxiv.org/abs/1706.03762 (2017)
2. Zhou, H. et al.: Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. https://arxiv.org/abs/2012.07436 (2021)
3. Araci, D.: FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. https://arxiv.org/abs/1908.10063 (2019)
4. Lu, W. et al.: Stock Price Prediction Model Based on CNN—LSTM. Complexity. (2020) https://doi.org/10.1155/2020/4717309
5. Wang, X. et al.: Research on the Application of Hybrid CNN--LSTM Model in Financial Time Series Forecasting. IEEE Access. https://doi.org/10.1109/ACCESS.2022.3142079 (2022)
6. Wu, H. et al.: Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. https://arxiv.org/abs/2106.13008 (2021)
7. Devlin, J. et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://arxiv.org/abs/1810.04805 (2018)
8. Liu, Y. et al.: Financial Sentiment Analysis Using FinBERT and Its Application in Stock Prediction. https://arxiv.org/abs/2306.02136 (2023)
9. Hu, Z. et al.: A Review of Deep Learning Applications in Forex and Stock Price Prediction https://arxiv.org/abs/2103.15781 (2021)
10. Olorunnimbe, K. et al.: A Survey of Deep Learning Research in the Stock Market. https://doi.org/10.1007/s10462-023-10607-2 Artificial Intelligence Review. (2023)

11. Bao, W. et al.: A Review of Data-Driven Stock Prediction Based on Neural Networks. Information Sciences. https://doi.org/10.1016/j.ins.2025.01.010 (2025)
12. Khodaee, P. et al.: A Hybrid CNN--LSTM Based Method for Predicting Stock Turning Points. Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2022.116111 (2022)
13. Xie, L. et al.: Deep Convolutional Transformer Network for Stock Movement Prediction. Electronics. https://www.mdpi.com/2079-9292/13/21/4225 (2024)
14. Zhang, S. et al.: Research on Multi-modal Financial Prediction Model Combining Transformer and Sentiment Features. Computational Economics. https://doi.org/10.1007/s10614-024-01365-7 (2024)
15. Dong, J. et al.: Application of Hybrid CNN--LSTM--GNN Network in Chinese A-Share Prediction. Entropy. https://www.mdpi.com/1099-4300/27/8/881 (2025)