A Study of Stock Price Forecasting Using Improved LSTM-Based Models

Zhuoyue Wu*

College of Economics and Management, Nanjing Forestry University, Nanjing, China

*Corresponding author: wu2310509121@njfu.edu.cn

Abstract. With the advancement of big data technology, machine learning has found widespread application in financial market forecasting due to its exceptional data processing and predictive capabilities. Upon reviewing recent research on stock prediction based on machine learning, it is found that under different market environments, Long Short-Term Memory significantly reduces prediction errors compared to other single-model methods. Therefore, to further explore the potential of LSTM models, this paper reviews recent research on stock price prediction based on improved LSTM models, focusing on the basic principles and characteristics of five models: BiGRU-LSTM、VMD-CSSA-LSTM、DMD-LSTM、SF-GET-LSTM、Doc-W-LSTM. Comparative analysis is conducted with other advanced models. Meanwhile, through comparative analysis of the literature, it is found that current research faces issues such as insufficient interpretability, limited research markets, and inadequate verification methods. It is proposed that introducing methods such as SHapley Additive exPlanations, causal inference, and time-series cross-validation may address these shortcomings. Finally, future research may endeavour to integrate the five models, constructing a multimodal adaptive and explainable intelligent financial forecasting system. This would provide theoretical and technical support for intelligent financial decision-making, driving further optimisation and innovation in deep learning within the forecasting domain.

Keywords: Stock price forecasting; Machine learning; Long Short-Term Memory.

1. Introduction

Stock price prediction has long been a difficult problem, stock market prices exhibit strong non-linearity, high complexity, and significant volatility, and are highly correlated with macroeconomics, market sentiment, business operations, and other factors, making them susceptible to external influences. With the advent of the big data era, the stock market has generated a massive amount of data. Machine learning, which can extract effective features from these nonlinear and noisy data, automatically learn patterns, and make predictions, is widely used in tasks such as stock price prediction, trend judgment, risk assessment, and investment strategy optimization. Simultaneously, machine learning can enhance stock prediction through feature engineering and model optimization, providing investors with more scientific decision-making basis, enabling funds to flow more efficiently to potential enterprises, thereby promoting the optimization of industrial structure and further enhancing market stability.

Currently, stock price forecasting can be categorised into four approaches: traditional machine learning methods, deep learning and neural network methods, time series analysis methods, and graph neural networks and graph theory methods. Payal Soni concluded through comparison that traditional machine learning is suitable for structured, small-sample data, while deep learning excels at handling complex nonlinear features and large-scale data; time series analysis performs well for short-term predictions; and graph neural networks can reveal potential connections between stocks [1]. Mehar Vijh utilized artificial neural networks and random forest models to forecast the stock prices of five large enterprises across various industries. She compared the accuracy of the two models using RMSE and MAPE as evaluation metrics, concluding that ANNs exhibited smaller prediction errors [2]. Zuriani Mustaffa employed the Salp Swarm Algorithm to optimise the weights and biases of the ANN, thereby enhancing the accuracy of predictions [3]. Chuyi Jia took BYD shares as the research subject, employing Backpropagation Neural Network, Extreme Learning Machine and Long Short-Term

Memory for forecasting and comparing the performance differences among the three models [4]; Yufi Jia compared the accuracy of three models, namely AutoRegressive Integrated Moving Average, Support Vector Machine and LSTM, in predicting the stock price of Bank of China, ARIMA struggles to capture nonlinear time series characteristics, while SVM lacks sufficient temporal memory capacity. Among the five models, LSTM performs the best, demonstrating strong time series modeling capabilities, stable model convergence, and no overfitting [5]. Due to its excellent performance, the LSTM model is widely favored. Sugam Agrawal used an LSTM model to predict the next day's closing price of Tata Steel and incorporated technical indicators such as Relative Strength Index, Moving Average Convergence Divergence, and Simple Moving Average as features. The resulting prediction curve closely matched the actual stock price trend [6]. In addition, this study also introduces many hybrid models based on LSTM with excellent fitting effects. This study describes a plethora of machine learning models for stock prediction, not only providing investors with technical analysis tools but also offering new research directions and theoretical references for future researchers in financial prediction, thereby promoting the development of the intelligent finance field.

2. Overview of Mainstream Technologies

LSTM is a special recurrent neural network structure, and its basic unit is called a memory cell, which consists of a forget gate, an input gate, and an output gate. Through the gating mechanism, the model is able to capture and retain long-term dependencies, perform controllable memory updates, and alleviate the vanishing gradient problem. Meanwhile, LSTM is applicable to a wide range of tasks, and thus has been widely used in the field of stock prediction, giving rise to many hybrid models based on this model.

2.1. BiGRU-LSTM（SMP-DL）

Shaban W. M. has pioneered a novel stock market prediction system based on deep learning (SMP-DL), which comprises two stages: data preprocessing and stock price prediction [7]. Firstly, the author utilized web scraping technology to obtain stock data, removed missing values, deleted redundant features to prevent overfitting, and employed Min-Max normalization to enhance the convergence speed of gradient descent. Next comes the establishment of the model, the core of the system. The new model is a combined structure of Long Short-Term Memory and Bidirectional Gated Recurrent Unit. Initially, BiGRU is employed to extract bidirectional time series features, and then the output is passed to a multi-layer LSTM to capture long-term dependencies. Finally, a Dense layer is added for regression prediction. Adam is selected to automatically adjust the learning rate, and the data is traversed 40 times, with 64 samples used for each parameter update. While overcoming the problems of vanishing gradients and exploding gradients, the SMP-DL model captures both forward and backward dependency information in time series. Additionally, it boasts short training time and strong generalization ability, performing well across multiple datasets and effectively enhancing prediction accuracy. Meanwhile, the author also designed an intelligent trading platform, into which the designed model was integrated, enabling users to enjoy stock analysis and prediction services on the mobile terminal in a convenient, safe, and low-cost manner at any time.

2.2. VMD-CSSA-LSTM

To better predict stock prices, the author proposes a new hybrid model (VMD-CSSA-LSTM) [8]. Taking the prediction of the Shanghai Composite Index as an example, in order to reduce data complexity and extract key features, the author performs variational mode decomposition on the original closing price series, obtaining multiple intrinsic mode functions. This process eliminates noise components while minimizing decomposition loss. To further enhance the global search capability, the authors applied the Circle chaotic map to initialize the Sparrow Search Algorithm, enabling a more uniform population distribution and avoiding local optima. Consequently, the optimal number of neurons (272), learning rate (0.0091), and number of iterations (275) were determined and then fed into the LSTM model. The optimized LSTM was used to model and forecast each Intrinsic Mode Function, and the final closing price prediction was obtained by aggregating the individual IMF forecasts. The model integrates signal decomposition, chaos optimization, and deep learning, significantly enhancing the robustness and accuracy of predictions.

2.3. DMD-LSTM

Shi Jiannan proposed a prediction method based on dynamic modal decomposition and long short-term memory network [9]. When market trends exhibit a certain degree of persistence, the DMD-LSTM model effectively decomposes industry stock data influenced by market sector interplay using DMD, extracts modal features that truly reflect market trends and individual stock dynamics, and then inputs the extracted modal features and fundamental data into an LSTM network to perform modeling and prediction for different markets. In addition, Shi Jiannan has constructed a fusion model that performs weighted averaging between the prediction results of DMD-LSTM and the prediction values of a single LSTM model based on fundamental features. This model leverages strong feature extraction capabilities and takes into account both price fluctuation patterns and economic driving factors, resulting in more comprehensive and accurate prediction results. However, when the market continues to fluctuate, leading to a weaker linkage effect among industry sectors, the DMD algorithm becomes ineffective.

2.4. SF-GET-LSTM

The rapid development of social media has made market sentiment an increasingly influential factor on stock prices, and Guanghui Chen's new hybrid forecasting model, SF-GET-LSTM, fully takes this into account [10]. First, the paper uses BERT (Bidirectional Encoder Representations from Transformers), which has undergone fine-tuning, to analyze stock reviews and categorize them into positive and negative sentiments. Then, the paper quantifies the "public sentiment" into a specific numerical value, the Sentiment Factor (SF), for daily use in model prediction. Next, GARCH、EGARCH and TGARCH are employed to fit the return series, extracting conditional volatility and residuals. These volatility parameters will serve as exogenous input features for subsequent models. Finally, a two-layer LSTM structure is employed to process time series, with price features, sentiment factors, and volatility parameters jointly inputted for prediction. This model takes both price and sentiment into account, and it integrates volatility information in the most comprehensive manner. It can simultaneously capture clustering, asymmetry, and threshold effects. Even during periods of intense volatility or market reversals, it can effectively perceive market sentiment, price fluctuations, and structural characteristics from multiple dimensions, enabling accurate predictions.

2.5. Doc-W-LSTM

To enhance the predictive performance of the Long Short-Term Memory (LSTM) model, Xuan Ji transforms social media texts which carry market sentiment into new input features. Xuan Ji employs the Doc2Vec model, which includes both the Distributed Memory (DM) and Distributed Bag of Words (DBOW) architectures, to train the textual data. On the basis of understanding and preserving semantic information, this approach generates more representative textual feature vectors. To address the mismatch between text vectors and financial feature dimensions, the author employs a stacked autoencoder to compress the text dimensions to 21 dimensions, balancing the dimensions of the two while avoiding information loss. Due to the presence of a large amount of random noise in stock price data, this paper also utilizes wavelet transform to decompose the time series, eliminating noise and non-stationary components, and achieving stationarity processing for non-stationary sequences. Finally, the financial features and social media text features are inputted into LSTM for stock price prediction. As shown in fig. 1, this is the new model Doc-W-LSTM constructed by Xuan Ji [11].

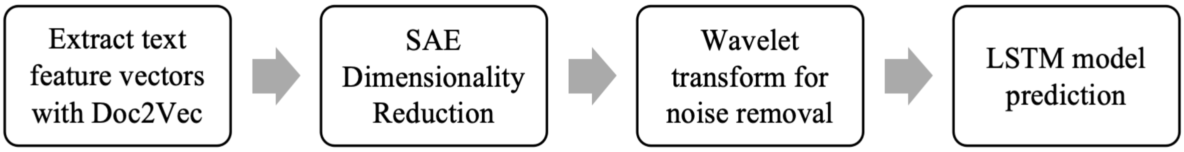

Fig. 1 Schematic diagram of Doc-W-LSTM model principle

3. Application and Results Analysis

3.1. BiGRU-LSTM（SMP-DL）

Shaban W. M. primarily uses RMSE, R^2, and training time as evaluation metrics [7], as shown in Table 1, by comparing the model with ABC-ANFIS-SVM, CNN-BiLSTM-AM and PSO-LSTM, it is concluded that although ABC-ANFIS-SVM improves computational accuracy, the training time is excessively long (6.5s). CNN-BiLSTM-AM faces difficulties in training and "exposure bias" issues. PSO-LSTM has completed parameter optimization, but the model tends to "remember" the training data, leading to overfitting, resulting in a high RMSE (0.3987) on the test data. However, the SMP-DL system effectively mitigates this issue by introducing Dropout and Gated Recurrent Unit. Currently, BiGRU-LSTM has been applied to predict the closing prices of IBM, AAPL, and GOOG for the next trading day. By combining the bidirectional feature extraction capability of BiGRU with the long-term dependency modeling capability of LSTM, it leverages its time series modeling ability, leading other algorithms in both accuracy and speed.

Table 1. Differences in RMSE, R^2 and training time among different models

Model	RMSE	R^2	Training duration	shortcoming
ABC-ANFIS-SVM	0.3638	0.9458	6.5	Training is time-consuming
CNN-BiLSTM-AM	0.3487	0.9590	8.6	The training of the attention mechanism is unstable
PSO-LSTM	0.3987	0.9400	9	Prone to overfitting
BiGRU-LSTM(best)	0.2883	0.9948	2.5	Requires a large amount of training data

3.2. VMD-CSSA-LSTM

Huang H. J. compares LSTM, VMD-LSTM, VMD-SSA-LSTM and VMD-CSSA-LSTM [8]. The single LSTM model shows the worst fitting performance with the target curve, yielding the largest RMSE, which exceeds 95. The introduction of Variational Mode Decomposition (VMD) reduces data complexity and partially suppresses noise interference, lowering the RMSE by 30.7204 and reducing the MAE to 53.6392 compared with the standalone model. Meanwhile, relative to the standard Sparrow Search Algorithm (SSA) with an RMSE of 36.8762, the Chaotic Sparrow Search Algorithm (CSSA) generates a population with broader and more uniformly distributed initial positions. During the iterative optimization, it achieves a better balance between local search and global search, thereby avoiding local optima, accelerating convergence, and enhancing the stability and robustness of the prediction results. The VMD-CSSA-LSTM model using this algorithm achieved an RMSE of only 6.4549 and an MAE of only 4.9177, while the other three models had values for both evaluation metrics greater than 28. Therefore, the VMD-CSSA-LSTM model performed the best. The author also proposed that when the stock closing price reaches its highest point, this model has the best predictive ability.

3.3. DMD-LSTM

DMD-LSTM is primarily applied in stock prediction for the steel industry, metal products, and other related industries. To test the performance of this model, Jiannan Shi used Angang Steel as the data source and conducted multiple experiments under different market conditions. He also introduced Direction Accuracy (DA) as an evaluation metric to measure the accuracy of the predicted direction of price movement [9]. Using the results of the traditional SVR model as a benchmark, Table 2 shows that during the unilateral uptrend of stock prices, the RMSE of LSTM decreased by 0.0246 compared to the benchmark. However, its analysis of market trends remains incomplete. By utilizing Dynamic Mode Decomposition (DMD) to explore trend change information from market behavior to improve the single LSTM model, the RMSE of DMD-LSTM is only 0.0898. However, it has higher requirements for market trends. The establishment of the fusion model combines modal features and fundamental information, resulting in a price prediction accuracy 1.34% higher than that of the DMD-LSTM model. The prediction effect has reached its optimal, as shown in Table 3. In the case of a unilateral downtrend of stock prices, the prediction curve of the fusion model also closely matches the target curve, demonstrating excellent prediction accuracy.

Table 2. Comparison of prediction results of various models under unilateral increase

Model	Characteristic	RMSE	R^2	DA
SVR	Traditional regression with a linear kernel	0.1160	0.9307	0.4576
LSTM	Nonlinear memory networks	0.0914	0.9570	0.5198
DMD-LSTM	Modal + Temporal Fusion	0.0898	0.9592	0.4936
fusion model	Weighted ensemble	0.0886	0.9596	0.5127

Table 3. Comparison of prediction results of various models under unilateral decline

Model	RMSE	R^2	DA
SVR	0.1094	0.904	0.4746
LSTM	0.0984	0.9275	0.5424
DMD-LSTM	0.0989	0.9235	0.5953
fusion model	0.0970	0.9294	0.5953

3.4. SF-GET-LSTM

Guanghui Chen collected data from the CSI300 Index spanning eight years, as well as commentary data from Eastmoney, as the sources for testing the model [10]. Using the LSTM single model as the baseline, this model achieved an MAE of 0.0189 and a MAPE of 1.4044%. By incorporating sentiment factors reflecting investor psychological expectations, the MAPE was reduced by 0.1488%. To further enhance the model's stability, the authors introduced a volatility feature, resulting in a significant reduction in MAPE. The introduced volatility features can be categorized into three types. SF-GE-LSTM takes into account both overall volatility clustering and asymmetric volatility, with a decrease in MAE of 0.005. SF-GT-LSTM is extremely sensitive to negative sentiment shocks, with MAE reduced to 0.0136, making it the model with the best performance among the six except for the target model. If EGARCH and TGARCH are used to extract volatility parameters, they can depict multi-level volatility structures and achieve the strongest theoretical fitting ability. However, it is prone to overfitting, resulting in slightly lower prediction accuracy than SF-GT-LSTM. The SF-GET-LSTM model deeply integrates multi-source information, balancing accuracy and robustness, with the lowest MAE of 0.0119 and MAPE of 0.8849%, making it the best-performing structure among the six models.

3.5. Doc-W-LSTM

Xuan Ji primarily selected investor reviews from pharmaceutical companies as the data source and used the stock of Meinian Health as the experimental sample to train the model [11]. In the experiments, Xuan Ji tuned the parameters of Doc2Vec and found that the Distributed Memory model achieved the best performance at a dimensionality of 200, while the Distributed Bag of Words model performed best with a dimensionality of 100 and a window size of 5. Furthermore, based on the Recurrent Neural Network, Xuan Ji constructed a hierarchical comparison by sequentially introducing memory mechanisms, multi-dimensional financial features, and semantic features extracted from social media texts, as shown in Table 4. Traditional models have the largest error due to their inability to handle long-term dependencies and non-stationary sequences. Although LSTM has solved the long-term dependency problem and improved its nonlinear modeling ability, its simple structure ignores market characteristics, and the MAE has not significantly decreased. LSTM-F, which incorporates multidimensional financial features on this basis, to some extent compensates for the shortcomings of the LSTM model, and its predictive ability has been significantly improved (MAE=0.046). The new model proposed in this paper, Doc-W-LSTM, incorporates text features and specifically takes into account investor sentiment and market expectations, achieving multi-source information fusion and robust prediction. The prediction curve closely aligns with the actual trend.

Table 4. Differences in MAE, RMSE and R^2 Among Models

Model	MAE	RMSE	R^2
RNN	0.435	0.301	0.882
LSTM	0.385	0.240	0.906
LSTM-F (Using only financial characteristics)	0.046	0.579	0.774
Doc-W-LSTM	0.019	0.110	0.957

4. Discussion

The five improved LSTM models introduced above each possess unique innovations and complement each other in terms of advantages. The DMD-LSTM and VMD-CSSA-LSTM models excel in trend decomposition and noise suppression, while the BiGRU-LSTM demonstrates remarkable effectiveness in time series and volatility modeling. The SF-GET-LSTM model addresses the shortcomings of the aforementioned models, which overlook market psychology and heteroskedasticity structures. The Doc-W-LSTM model possesses excellent sentiment mining and text feature fusion capabilities. Although the five models each have their own advantages, they still share some common issues. For example, the research on the model is limited in scope, as it has not been tested for generalization across different markets and industries. It does not fully consider abnormal situations such as structural breaks and black swan events, lacks interpretability, and lacks analysis of the contribution of various features as well as causal explanations at the economic mechanism level. Furthermore, the verification methods are inadequate.

Future researchers can incorporate transfer learning techniques to evaluate the model's generalizability and robustness across markets and multiple time horizons. Online learning strategies can also be introduced to cope with extreme market conditions, enabling models to evolve in real time and adjust to the current market state. In addition, combining SHAP with causal inference methods can enhance interpretability and provide causal explanations. Employing time-series cross-validation can further enrich the validation framework and assess the stability and generalizability of hyperparameter selection. Overall, the research should advance toward an intelligent and verifiable financial forecasting system.

5. Conclusion

This article primarily introduces five improved models of LSTM. The BiGRU-LSTM model is capable of learning from both past and future directions of the time series, and performs optimally in short-term price forecasting. DMD-LSTM can capture the linkage effects across different industries and sectors, and performs well in long-term trend prediction. VMD-CSSA-LSTM excels in mitigating the impact of non-stationarity and noise on output results, making it suitable for predicting short-term stock price fluctuations. The SF-GET-LSTM model is capable of simultaneously handling heteroskedastic volatility and market sentiment, predicting the impact of investor sentiment on price fluctuations. The Doc-W-LSTM model can make predictions based on market sentiment, public opinion, and relevant financial data reflected in social media texts, making it suitable for predicting stock prices that are significantly influenced by public opinion and news. In the future, researchers can attempt to integrate these five models to form a three-tier framework of "signal decomposition-emotion perception-fluctuation modeling", combining the advantages of the five models to construct a multimodal adaptive and interpretable financial prediction system, significantly improving prediction accuracy and robustness. Finally, it is hoped that the research presented in this paper can provide new insights for related studies on improved LSTM models, and promote further optimization and innovation in the field of deep learning for time series prediction.

References

[1] Soni P, Tewari Y, Krishnan D. Machine learning approaches in stock price prediction: A systematic review. Journal of Physics: Conference Series, 2022, 2161(1): 012065.

[2] Vijh M, Chandola D, Tikkiwal V A, Kumar A. Stock closing price prediction using machine learning techniques. Procedia Computer Science, 2020, 167: 599-606.

[3] Mustaffa Z, Sulaiman M H, Aziz A A. Artificial neural network-salp-swarm algorithm for stock price prediction. Iraqi Journal of Science, 2024, 65(12): 7207-7219.

[4] Jia C Y. Comparison of advantages of multiple machine learning methods in stock prediction. Finance, 2025, 15(1): 238-245.

[5] Jia Y F. Research on stock price prediction based on machine learning. E-Commerce Letters, 2024, 13(2): 2253-2258.

[6] Agrawal S, Bhadauriya S, Narayan V. LSTM price movement prediction for stock market. Journal of Neonatal Surgery, 2025, 14(27s): 495-501.

[7] Shaban W M, Ashraf E, Slama A E. SMP-DL: a novel stock market prediction approach based on deep learning for effective trend forecasting. Neural Computing and Applications, 2024, 36(4): 1849-1873.

[8] Huang H J, Li B. Stock price prediction based on VMD-CSSA-LSTM combination model. Journal of Nanjing University of Information Science & Technology, 2024, 16(3): 332-340.

[9] Shi J N, Zou J Z, Zhang J, Wang C M, Wei Z C. Research of stock price prediction based on DMD-LSTM model. Application Research of Computers, 2020, 37(3): 662-666.

[10] Chen G H, Wang L. GARCH-LSTM for stock price prediction using sentiment analysis. Engineering Letters, 2025, 33(7): 2261-2271.

[11] Ji X, Wang J C, Yan Z J. A stock price prediction method based on deep learning technology. International Journal of Crowd Science, 2021, 5(1): 55-72.