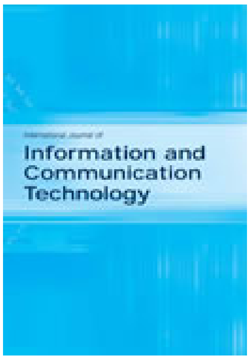


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Analysing trends in trading behaviour in financial markets using deep learning algorithms

Yiming Tan

Article History:

Received:	12 October 2024
Last revised:	28 October 2024
Accepted:	28 October 2024
Published online:	13 December 2024

Analysing trends in trading behaviour in financial markets using deep learning algorithms

Yiming Tan

Shenyang City University,
Shenyang, 110112, China
Email: 13390105079@163.com

Abstract: The complexity of financial markets demands analytical methods that capture nonlinear and time-varying data characteristics. Traditional methods often fall short, prompting the use of deep learning, particularly LSTM, for its time series processing prowess. However, LSTMs struggle with long sequences due to vanishing or exploding gradients, leading to the loss of early data significance. To address this, we introduce the LSTM-AT model, which integrates LSTM with an attention mechanism to enhance its focus on key data aspects. Our model, trained on historical financial data, outperforms traditional methods in predicting market trends. Despite its high accuracy, the model faces challenges like overfitting and data labelling requirements. Future work will focus on improving interpretability and exploring its application across various financial markets.

Keywords: LSTM-AT; financial markets; trading behaviour; trend prediction.

Reference to this paper should be made as follows: Tan, Y. (2024) 'Analysing trends in trading behaviour in financial markets using deep learning algorithms', *Int. J. Information and Communication Technology*, Vol. 25, No. 10, pp.34–46.

Biographical notes: Yiming Tan received her Master's degree from the University of New South Wales (Australia) in 2012. She is currently a Lecturer at the Shenyang City University and her current research interest is in financial market and investment.

1 Introduction

Financial markets, as the nerve centre of the global economy, not only carry the flow and distribution of capital, but their stability and efficiency are critical to the health of the economy (Pagano, 1993). The trading behaviours of the market, covering the purchase and sale of stocks, bonds, foreign exchange and derivatives, not only reflect the decision-making process of investors, but also serve as a barometer of macroeconomic indicators. However, the dynamics of trading are shaped by various influences, encompassing macroeconomic figures, policy modifications, and the prevailing sentiment in the market, etc., which together create the complexity and unpredictability of markets. Understanding and predicting trading behaviour in financial markets is therefore an important and difficult task for investors, regulators and policymakers (Al-Nasseri et al., 2015).

Traditional methods of financial market analysis, such as technical and fundamental analysis, rely heavily on historical data and statistical models, and these methods have had some success in dealing with linear relationships and short-term patterns in market data. However, as the volume and complexity of market data have proliferated, these traditional methods have become inadequate in capturing market dynamics, dealing with nonlinear patterns, and predicting long-term trends (Papanagnou and Matthews-Amune, 2018). In addition, traditional methods often ignore the deeper factors behind market behaviour, such as investor psychology and behavioural patterns, which are playing an increasingly important role in financial markets.

The progression of deep learning (DL) has been particularly swift in the last couple of years, especially the emergence of the LSTM architecture, which stands for long-short-term memory models, new perspectives have been provided for financial market analysis (Ting et al., 2019). LSTM models, capitalising on their proficiency with chronological data streams, can effectively capture the market trading behaviour of the long-term dependencies and complex patterns, which makes it show great potential in financial market forecasting.

In this field, researchers have achieved a number of results. Pham et al. (2019) applied LSTM neural networks to predict stock market prices. This study not only demonstrated the potential of LSTM in capturing market trends, but also laid the foundation for subsequent research. In the same year, Mamun et al. further extended the application of LSTM in financial market prediction, and their study improved the predictive ability of the model by combining macroeconomic factors and micro-trading data, providing a new perspective on financial market prediction.

De-Souza et al. (2021) proposed an LSTM-based approach to predict financial signals. Their work not only demonstrated the potential of LSTM in dealing with complex financial time series data, but also provided a publicly available dataset and benchmarking, providing a valuable resource for subsequent research. The distinctive innovation of this research is its approach to combining LSTM's extensive memory capacity with the nonlinear patterns observed in financial markets. As a result, it boosts the reliability and accuracy of projection systems. Guo et al. (2024) reviewed the application of DL to stock market forecasting. They not only summarised the advantages of LSTM models in stock market forecasting, but also explored their potential in dealing with market volatility and uncertainty. Guo et al.'s study provided researchers in the field of finance with a comprehensive view of DL, highlighting the importance of LSTM in financial data analysis. The cumulative results of these studies not only demonstrate the application of LSTM in financial market analysis, but also provide new analytical tools for researchers and practitioners in the field of finance.

Although LSTM performs well in dealing with short-term time series, the model may suffer from data forgetfulness leading to a decrease in transaction prediction accuracy when confronted with long time series data (Khan et al., 2023). To tackle this issue effectively and further improve the model's analytical accuracy, this paper proposes a novel method that combines the LSTM architecture augmented by AT. The constructed LSTM-AT model aims to heighten the model's discernment of pivotal data points with the aid of AT, which improves the processing effect on large-scale temporal data and optimises the correctness of anticipating trading patterns.

The intent behind this paper is to equip individuals in finance with valuable information on how to effectively use LSTM-AT models for market analysis and decision making. By utilising LSTM's strengths in long-term sequence retention and the capacity of AT to recognise key points, the LSTM-AT model is able to better understand and predict market behaviours, and provide investors and financial institutions with more accurate decision support. Through the research, we expect to promote the practical application of DL techniques in financial market analysis and prediction, and provide new ideas and methods for future research and development.

The main innovations and contributions of this work include:

1. Constructing the LSTM-AT model: this study proposes to construct the LSTM-AT model by combining the traditional LSTM model with AT. The amalgamation targets at refining the model's capability to pinpoint significant information in financial time series, with an emphasis on extended temporal relationships.
2. Enhancement of model interpretability: the introduction of AT enhances the model's predictive precision and elucidates its decision-making processes. This enables financial analysts to comprehend the model's reasoning behind its predictions and pinpoint the key determinants essential for anticipating market trends.
3. Advance the utilisation of DL techniques within the domain of financial market trading: through the research in this paper, it is expected to promote the practical application of DL techniques, especially the LSTM-AT model, in financial market analysis and prediction, and to bring new research directions and practical methods to the field of financial technology.

2 Relevant technologies

2.1 Long-short-term memory

When exploring the complexity and dynamics of financial markets, traditional forecasting models often have difficulty in detecting extended patterns within sequential data. To address this challenge, researchers have turned to recurrent neural network (RNN) (Sherstinsky, 2020). RNNs are designed to effectively manage and predict sequential data trends. Unlike traditional feed-forward neural networks, RNNs feature self-loops that facilitate the transmission of data across different time frames through the Hidden State, thereby enabling the recognition of patterns across time in sequential datasets.

The core property of RNNs is their ability to pass information from the preceding temporal stage to the present one through hidden states, thus creating temporal dependencies in sequential data. This ability allows RNNs to excel in areas including the interpretation of natural language, the identification of spoken words, and the scrutiny of financial market dynamics (Khurana et al., 2023).

The fundamental configuration of an RNN includes layers for input, hidden processing, and output. For every temporal interva t, the RNN takes in the input x_t and the preceding hidden state h_{t-1}, processes the current the hidden state ht using a nonlinear activation function, and subsequently generates the outputs o_t. The hidden state is passed recursively through time with the following formula:

$$h_t = f\left(W_h x_t + U_h h_{t-1} + b_h\right) \tag{1}$$

$$o_t = g\left(W_o h_t + b_o\right) \tag{2}$$

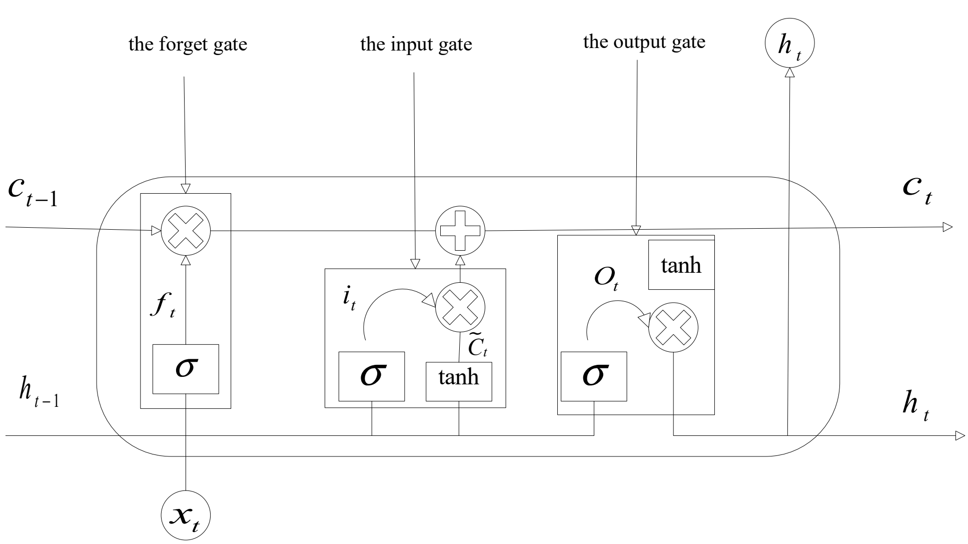
Here, x_t denotes the input at time step t; h_t represents the current hidden state, encapsulating data from the prior time step; W_h and U_h denote the weight matrices associated with the input-to-hidden layer transitions and the influence of the preceding hidden state on the current one. Respectively, b_h is the bias term; f is usually a nonlinear activation function (e.g., tanh or ReLU), and g represents the output layer's activation function, such as the softmax function.

Through this recursive structure, RNNs can model the contextual information before and after in a sequence. For example, the significance of a term is often contingent upon the context provided by preceding terms within the phrase, and RNNs are able to convey this contextual information through the recursive structure. Consequently, this offers more nuanced interpretations for applications like the interpretation of natural language and the identification of spoken language.

While RNNs are designed to handle long-term dependencies, they often face the challenge of vanishing gradients increase. The BPTT algorithm back propagates the gradient propagates forward from the current time step to earlier time steps, and when the sequence is long, the gradient values may grow or decay exponentially.

LSTM networks were devised to counteract the inherent limitations within RNN architectures (Yu et al., 2019). The LSTM, an advanced variant of RNNs, was initially introduced by Hochreiter and Schmidhube in their 1997 publication. LSTMs are engineered to regulate informational flow by employing forget gates, input gates, and output gates, which facilitate the effective capture of long-term dependencies. A conventional LSTM unit incorporates three regulatory gates – namely, the forget gate, the input gate, and the output gate – as well as a cell state. Memory cells in LSTM are capable of retaining or discarding information from past timesteps, mitigating the gradient vanishing problem. The configuration of an LSTM cell is depicted in Figure 1.

Figure 1 Structure of LSTM



Oblivion gate, input gate, output gate, candidate cell information generated within the input gate, new cell information, and the computations for the outputs and associated parameters are detailed below.

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + h_f\right) \tag{3}$$

$$i_t = \sigma\left(W_t\left[h_{t-1}, x_t\right] + b_i\right) \tag{4}$$

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{5}$$

$$\tilde{c}_t = \tanh\left(W_c\left[h_{t-1}, x_t\right] + b_c\right) \tag{6}$$

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \tag{7}$$

$$h_t = o_t \tanh\left(c_t\right) \tag{8}$$

In addition, LSTM significantly improves the gradient disappearance problem encountered by traditional RNN in processing long-sequence data through its exquisite gating mechanism. Forgetting gates control the process of forgetting information, ensuring that the model does not retain irrelevant historical information. The input gate determines the storage of new information, while the output gate generates output based on the current state. The synergy of these gates enables LSTM to capture long-term dependencies in financial time series data effectively.

These gating mechanisms of LSTM enable it to efficiently identify patterns across extended timeframes within sequential datasets while avoiding the challenge associated with vanishing or exploding gradients, and thus show great potential in financial market analysis.

2.2 Attention mechanism

Attention is a model in DL based on the human visual nervous system that emulates the human capacity to concentrate on the most pertinent aspects of the information being processed, enabling the model to discern and concentrate on the most relevant aspects of the input (Niu et al., 2021). This mechanism plays a crucial role in applications including the identification of visual data, the interpretation of spoken words, and the understanding of textual content, where it can help models improve efficiency and focus on key information, enhancing performance. Attention models include the ability to selectively focus, process complex data, and capture long-distance dependencies, enhancing the accuracy and interpretability of the model (Soydaner, 2022). The structure of the Attention mechanism is shown in Figure 2.

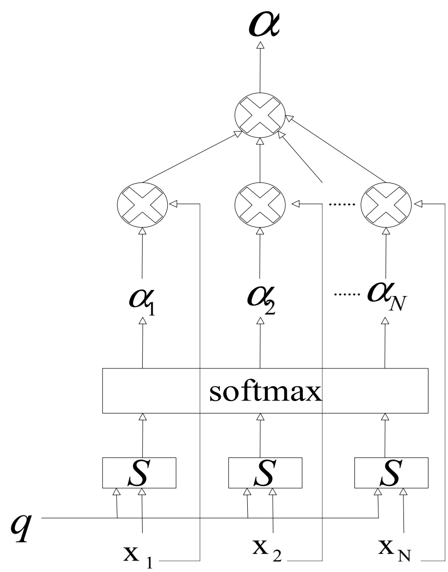
Attention specific expressions are as follows:

$$\begin{aligned}\alpha_n &= p(z = n \mid X, q) = Softmax\left(s\left(x_n, q\right)\right)\\ &= \frac{\exp\left(s\left(x_n, q\right)\right)}{\sum_{j=1}^{N}\left(s\left(x_j, q\right)\right)}\end{aligned} \tag{9}$$

$$attention(X, q) = \sum_{n=1}^{N} \alpha_n x_n \tag{10}$$

In this formula, X is a sequence of variables containing N elements, q is a query variable, α_n is the weight of the n^{th} element, $s(x_n, q)$ is an attention scoring function to measure the similarity between x_n and the query variable q, and $attention(X, q)$ is an attention variable.

Figure 2 Structure of AT



In the LSTM-AT model, the attention layer is added after the LSTM layer to enhance the model's ability to focus on key time steps (Song et al., 2016). This mechanism is achieved by calculating a weight for each time step, the size of which reflects the importance of that time step to the current forecast. By assigning higher weights to this critical information, the model can more accurately predict the future movement of the market.

3 Building the LSTM-AT model for financial market trend prediction

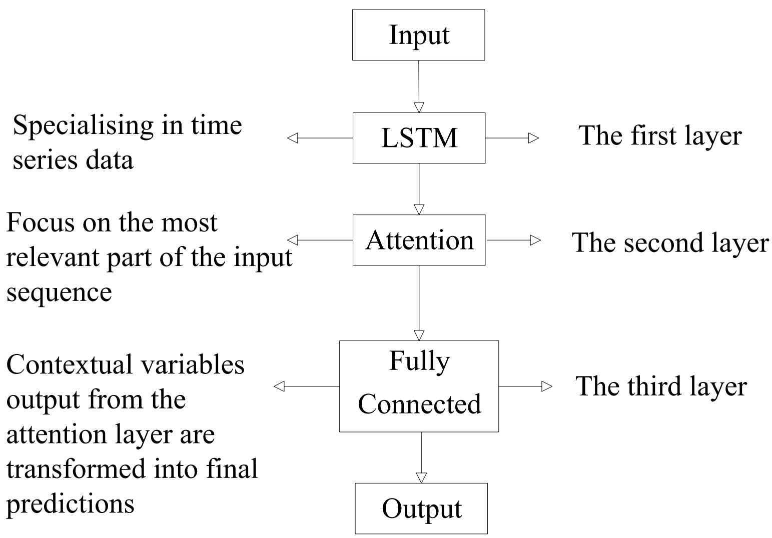
3.1 The LSTM-AT model

This study proposes a model that combines LSTM and attention mechanism, i.e., LSTM-AT model to analyse the trend of trading behaviour in financial markets (Wang et al., 2020). In the process of building LSTM-AT model, we use convolution layer to enhance the model's ability to capture local features. Although we are dealing with financial time series data rather than text, the convolution layer still plays to its advantage in capturing local patterns and trends. By using convolution cores of different sizes, the model can extract multi-granularity information from the data, which is similar to identifying keywords and phrases in text analysis. This method enables us to quickly extract key information from a large amount of financial data, providing a rich feature representation for the subsequent LSTM layer.

Specifically, the convolution layer moves over the time series through a sliding window mechanism, with each convolution kernel responsible for capturing specific patterns, such as price fluctuations or local peaks in trading volume. These local features are then fed into the LSTM layer, which handles the long-term dependencies between these features. In this way, our model can more effectively understand and predict the financial market trading behaviour.

The LSTM-AT model's design includes layers for input, LSTM processing, attention mechanism, dense connections, and output generation. Figure 3 depicts the architecture of the LSTM-AT model.

Figure 3 Architecture of the LSTM-AT model



3.1.1 Input layer

The initial layer of the model is designed to input raw financial market data. These data may include historical prices, trading volumes, opening and closing prices, etc. At the input layer, the data undergoes pre-processing, such as normalisation, to ensure that the input features received by the model are on the same scale, thus facilitating model learning. Enhancing the model's convergence rate and predictive precision is vital, with x representing the raw data and x' denoting the normalised version.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{11}$$

3.1.2 LSTM layer

The LSTM layer, central to the model, excels at handling sequential data. This layer consists of multiple LSTM units, each unit is equipped with three regulatory mechanisms: a gate for memory erasure, a gate for new information admission, and a gate for data release (Md et al., 2023). These gating mechanisms enable the LSTM to capture long-term data patterns, circumventing the typical gradient diminution issues faced by standard RNNs. This capability is particularly important in financial market analysis, as this capability empowers the model to discern persistent factors shaping market dynamics.

3.1.3 Attention layer

After the LSTM layer extracts features, the attention mechanism bolsters the model's ability to forecast accurately. This layer determines which time steps are most critical to the current prediction by calculating the weights of each time step. This approach enables the model to adaptively concentrate on the input sequence's key elements, thus improving the accuracy of the prediction. The attention layer generates a contextually weighted variable as its output that combines information from the entire sequence, providing a rich context for the final prediction (Hao et al., 2019).

The attention score formula, the attention weight formula, and the context variable formula are as follows, where v is the learnable weight variable, and W_a and b_a are the weights and biases of the attention mechanism.

$$e_t = v^T \tanh\left(W_a h_t + b_a\right) \tag{12}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{T} \exp(e_i)} \tag{13}$$

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{14}$$

3.1.4 Fully connected layer

After the LSTM layer extracts features, the attention mechanism strengthens the model's ability to make accurate predictions. This layer determines which time steps are most critical to the current prediction by calculating the weights of each time step. The system grants the model the capability to purposefully highlight significant elements within the input data, thus improving the accuracy of the prediction. The attention layer produces a context vector that is weighted according to relevance that combines information from the entire sequence, providing a rich context for the final prediction.

Following the attention mechanism is a layer of densely interconnected neurons, and its task is to transform the contextual variables of the output of the attention layer into the final prediction. This layer consists of densely connected neurons, each connected to all the outputs of the previous layer. The fully connected layer maps the context variables to the prediction target by learning the weights of these connections. This layer is designed to allow the model to learn the complex mapping relationships from features to predicted outcomes.

The fully-connected formula is as follows, where W_z and b_z denote the parameters, including weights and biases, utilised in the fully connected layer, and c is the context variable obtained from the attention layer.

$$z = W_z c + b_z \tag{15}$$

3.1.5 Output layer

The model culminates with the output layer, which is the last in the sequence, which is responsible for generating the final forecast. In financial market forecasting, this may include future movements of stock prices, market volatility, etc. The output layer typically uses activation functions appropriate to the particular task, encompassing linear activation functions for predicting continuous values and softmax functions for discrete classification. The design of this layer ensures that the model is able to output results that match the prediction task.

The predicted output formula is as follows, where σ is the activation function, W_y and b_y are the parameters, namely the weights and biases, associated with the output layer, and $\hat{y}$ constitutes the model's forecasted results.

$$\hat{y} = \sigma\left(W_y z + b_y\right) \tag{16}$$

Through the synergy of these layers, the LSTM-AT model is able to efficiently process complex time series data from financial markets and provide accurate forecasting results. Every layer is engineered to refine the model's capacity for feature extraction and predictive accuracy, ensuring its efficacy in the analysis of financial markets.

3.2 Evaluation indicators

For a complete analysis of the LSTM-AT model's performance, we used an array of evaluation metrics, including mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) (Tian et al., 2018). Here, N denotes the number of samples, $\hat{y}_i$ represents the initial forecasted outcome, and y_i signifies the corresponding observed value.

These indicators can not only measure the prediction accuracy of the model, but also reflect the performance of the model in different aspects. The accuracy-focused model predicts the correct proportion, the recall measure model captures the proportion of positive cases to all actual positive cases, and the F1-score is the harmonic average of accuracy and recall, provides a comprehensive measure of performance. Together, these indicators help us assess the validity and reliability of the model in predicting mental toughness.

$$C_{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{17}$$

$$C_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{18}$$

$$C_{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{19}$$

4 Experimental results and analyses

4.1 Ablation experiments

In order to gain a deeper understanding of the contribution of individual components of the LSTM-AT model to the prediction performance, we designed a series of ablation experiments (Luo et al., 2020). These experiments assess the impact of key components of the model on the final results by gradually removing them. In this way, we can more accurately identify which components are critical for improving model performance.

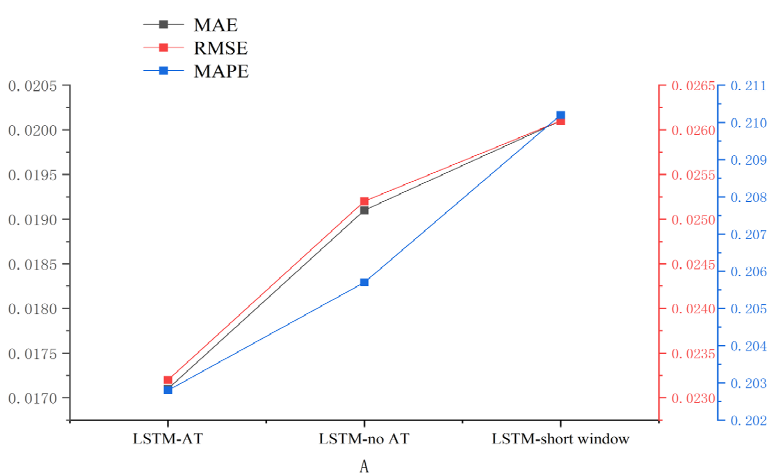
These experiments systematically removed the attention mechanism to evaluate its impact on model performance. The experimental framework is based on our main model settings, using the same dataset and hyperparameter configuration to ensure the comparability of results.

In our ablation experiments, we first removed the attentional layer while preserving the LSTM layer and other model components. This change allows us to directly observe the contribution of the attention mechanism to the performance of the model. We used the same performance evaluation criteria, including mean square error (MSE), MAE, and R-squared, to compare the performance of the complete model with that of the ablation model.

The experimental results show that when the attention mechanism is removed, the performance of the model decreases on all the evaluation indexes. Specifically, MAE increased from 0.0171 to 0.0191, RMSE increased from 0.0232 to 0.0252, and R-squared also decreased. These results highlight the importance of the attention mechanism in identifying and focusing key information in time-series data, thereby improving the predictive accuracy of the model.

To assess the effect of time window length on model performance, we shortened the original long sequence data to simulate the model's performance when dealing with shorter time sequences, and recorded the change in performance as shown in Figure 4.

Figure 4 Results of ablation experiments (see online version for colours)



After removing the attention mechanism, we found that the performance of the model decreased. Specifically, the MAE of the model increased from 0.0171 to 0.0191, the

RMSE increased from 0.0232 to 0.0252, and the MAPE increased from 20.28% to 20.57%. This indicates that the attention mechanism plays an important role in capturing the key information in the input data and thus improving the prediction accuracy.

The performance of the model similarly decreased after shortening the length of the time window for LSTM layer processing (Kwon et al., 2020). In this case, the MAE increased from 0.0171 to 0.0201, the RMSE increased from 0.0232 to 0.0261, and the MAPE increased from 20.28% to 21.02%. This result suggests that a longer time window helps the model to capture more time-dependent information, which improves the accuracy of the prediction.

4.2 Integrated experiments

Through ablation experiments, we validate the importance of the LSTM layer and the attention mechanism in the model (Samek et al., 2016). These findings provide us with insights into the impact of model components on prediction performance. However, in order to fully evaluate the performance of the LSTM-AT model, we also need to compare it with other existing models. This will help us understand the relative strengths and potential limitations of LSTM-AT models for real-world financial market prediction tasks.

For this purpose, we have selected common financial forecasting models: TCN (temporal convolutional neural network), ARIMA (autoregressive moving average model), SVR (supported variables regression model), RNN model, and LSTNet are selected to achieve the same task as the LSTM-AT model outlined in this research. The objective of forecasting shifts in trading behaviour within financial markets, along with the corresponding experimental outcomes, is presented in Table 1.

Table 1 Findings from comprehensive experiment

Model	*MAE*	*RMSE*	*MAPE*
TCN	0.0276	0.0342	22.77
ARIMA	0.0503	0.0595	30.95
SVR	0.0691	0.0762	37.77
RNN	0.0373	0.0430	27.01
LSTNet	0.0285	0.0357	22.83
LSTM-AT	0.0171	0.0232	20.28

The experimental data indicates that the LSTM-AT model surpasses alternative models in forecasting shifts in trading behaviour within the financial market, proving its reliability and validity in its application to financial data, and that the model has the smallest error in each of the evaluation metrics, demonstrating its enhanced effectiveness.

5 Conclusions

Within the present study, we provide a thorough analysis and prediction of the changing trends of trading behaviour in the financial market by constructing an LSTM-AT model based on LSTM and attention mechanism. Through the training and validation of financial market data, our model demonstrates superior performance in several evaluation indexes and shows significant advantages over existing methods.

The LSTM-AT model performs well in the task of predicting trading behaviours in financial markets, effectively capturing market dynamics and nonlinear features. Experimental results show that LSTM-AT has a significant advantage in prediction accuracy over other baseline models, due to the LSTM's capacity for recognising extended patterns and the attention mechanism's focus on critical data points. Although no specific accuracy figures are available, the model's efficacy is confirmed through its evaluation against alternative models.

Even though the LSTM-AT model demonstrated favourable outcomes in the experimental trials, we also recognise some limitations of the model. Firstly, the model may be sensitive to overfitting, especially in the face of the complexity and variability of financial market data. Second, the model may have a significantly higher demand on computational resources when dealing with very long time series. In addition, the model's reliance on large amounts of labelled data may limit its application in data-scarce market environments.

To address the limitations of the model, future research can explore the following directions: firstly, the generalisation ability of the model can be enhanced by introducing regularisation techniques or adopting a more complex network structure. Second, investigate how to optimise the model structure and training process to reduce the demand for computational resources. In addition, unsupervised learning or semi-supervised learning methods are explored to decrease the dependency on extensive labelled datasets. Finally, the LSTM-AT model's versatility and flexibility are evaluated by its application across various financial markets and products.

The LSTM-AT model provides a new tool for financial market analysis that helps investors and financial institutions to better understand and predict market behaviour. We expect this research to promote the utilisation of DL methodologies for financial market analytics and forecasting, and to bring new research directions and practical approaches to the field of financial technology.

In conclusion, this study demonstrates the potential of LSTM-AT models in the prediction of trading behaviours in financial markets and provides new ideas for future research. With the continuous development of DL technology, we believe that the LSTM-AT model will play a greater role in financial market analysis.

References

Al-Nasseri, A., Tucker, A. and De-Cesare, S. (2015) 'Quantifying StockTwits semantic terms' trading behavior in financial markets: an effective application of decision tree algorithms', *Expert Systems with Applications*, Vol. 42, No. 23, pp.9192–9210.

De-Souza, F.S.H., Hojo-Souza, N.S., Dos Santos, E.B. et al. (2021) 'Predicting the disease outcome in COVID-19 positive patients through machine learning: a retrospective cohort study with Brazilian data', *Frontiers in Artificial Intelligence*, Vol. 4, p.579931.

Guo, R., Alves, M., Mehdikhani, M. et al. (2024) 'Synthesising realistic 2D microstructures of unidirectional fibre-reinforced composites with a generative adversarial network', *Composites Science and Technology*, Vol. 250, p.110539.

Hao, S., Lee, D-H. and Zhao, D. (2019) 'Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system', *Transportation Research Part C: Emerging Technologies*, Vol. 107, pp.287–300.

Khan, A., Fouda, M.M., Do, D-T. et al. (2023) 'Short-term traffic prediction using deep learning long short-term memory: taxonomy, applications, challenges, and future trends', *IEEE Access*, Vol. 11, pp.94371–94391.

Khurana, D., Koli, A., Khatter, K. et al. (2023) 'Natural language processing: state of the art, current trends and challenges', *Multimedia Tools and Applications*, Vol. 82, No. 3, pp.3713–3744.

Kwon, B-S., Park, R-J. and Song, K-B. (2020) 'Short-term load forecasting based on deep neural networks using LSTM layer', *Journal of Electrical Engineering & Technology*, Vol. 15, pp.1501–1509.

Luo, C., Li, X. and Ye, Y. (2020) 'PFST-LSTM: a spatiotemporal LSTM model with pseudoflow prediction for precipitation nowcasting', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, pp.843–857.

Md, A.Q., Kapoor, S., AV, C.J. et al. (2023) 'Novel optimization approach for stock price forecasting using multi-layered sequential LSTM', *Applied Soft Computing*, Vol. 134, p.109830.

Niu, Z., Zhong, G. and Yu, H. (2021) 'A review on the attention mechanism of deep learning', *Neurocomputing*, Vol. 452, pp.48–62.

Pagano, M. (1993) 'Financial markets and growth: an overview', *European Economic Review*, Vol. 37, Nos. 2–3, pp.613–622.

Papanagnou, C.I. and Matthews-Amune, O. (2018) 'Coping with demand volatility in retail pharmacies with the aid of big data exploration', *Computers & Operations Research*, Vol. 98, pp.343–354.

Pham, H.H.N., Futakuchi, M., Bychkov, A. et al. (2019) 'Detection of lung cancer lymph node metastases from whole-slide histopathologic images using a two-step deep learning approach', *The American Journal of Pathology*, Vol. 189, No. 12, pp.2428–2439.

Samek, W., Binder, A., Montavon, G. et al. (2016) 'Evaluating the visualization of what a deep neural network has learned', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 28, No. 11, pp.2660–2673.

Sherstinsky, A. (2020) 'Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network', *Physica D: Nonlinear Phenomena*, Vol. 404, p.132306.

Song, J., Tang, S., Xiao, J. et al. (2016) 'LSTM-in-LSTM for generating long descriptions of images', *Computational Visual Media*, Vol. 2, pp.379–388.

Soydaner, D. (2022) 'Attention mechanism in neural networks: where it comes and where it goes', *Neural Computing and Applications*, Vol. 34, No. 16, pp.13371–13385.

Tian, C., Ma, J., Zhang, C. et al. (2018) 'A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network', *Energies*, Vol. 11, No. 12, p.3493.

Ting, D.S., Peng, L., Varadarajan, A.V. et al. (2019) 'Deep learning in ophthalmology: the technical and clinical considerations', *Progress in Retinal and Eye Research*, Vol. 72, p.100759.

Wang, J.Q., Du, Y. and Wang, J. (2020) 'LSTM based long-term energy consumption prediction with periodicity', *Energy*, Vol. 197, p.117197.

Yu, Y., Si, X., Hu, C. et al. (2019) 'A review of recurrent neural networks: LSTM cells and network architectures', *Neural Computation*, Vol. 31, No. 7, pp.1235–1270.