Journal of Global Trends in Social Science

https://doi.org/10.70731/rzvs8j53

GRU-Enhanced Attention Mechanism for LSTM in Hybrid CNN-LSTM Models for Stock Prediction

Jiahui Zhang [a,*], Laipeng Yan [a]

[a] *The Chinese University of Hong Kong, Shenzhen 518172, China*

KEYWORDS

Stock Prediction;
GRU-Enhanced Attention Mechanism;
Hybrid CNN-LSTM Model;
Financial Time Series;
Deep Learning

ABSTRACT

We propose a novel GRU-enhanced attention mechanism integrated into LSTM layers to improve stock prediction accuracy in hybrid CNN-LSTM models. The proposed method dynamically adjusts the importance of different time steps by combining the strengths of GRUs and attention mechanisms, thereby capturing temporal dependencies more effectively in volatile financial time series. The GRU processes the input sequence to generate hidden states, which are then weighted by an attention mechanism to compute a context vector. This context vector is fed into the LSTM layer, enabling the model to focus on the most relevant time steps and enhance its ability to handle non-stationarity and noise. The integration of GRU-enhanced attention into LSTM allows the model to better capture long-term dependencies and temporal patterns, which are critical for accurate stock prediction. Experimental results demonstrate that the proposed approach outperforms traditional methods in terms of prediction accuracy and robustness, particularly in scenarios with high market volatility. Furthermore, the model's adaptability to varying time scales and its ability to filter out irrelevant information make it a promising tool for financial time series analysis. The proposed method not only advances the state-of-the-art in stock prediction but also provides a framework for integrating attention mechanisms into other sequential data tasks.

1. Introduction

Stock price prediction has long been a challenging task due to the inherent volatility, non-linearity, and noise in financial time series data. Traditional statistical models, such as ARIMA, often struggle to capture the complex dynamics of stock markets due to their linear assumptions [1]. In recent years, deep learning models, particularly those combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have shown significant promise in addressing these challenges [2]. CNNs excel at extracting spatial features from data, while LSTMs are well-suited for modeling temporal dependencies, making their combination a powerful tool for stock prediction [3].

Attention mechanisms, originally developed for natural language processing, have been increasingly applied to time-series tasks to improve model performance by focusing on the most relevant parts of the input sequence [4]. In the context of stock prediction, attention mechanisms have been integrated into hybrid CNN-LSTM models to enhance their ability to identify critical patterns in volatile and nonlinear data [5]. However, existing attention mechanisms often rely on static weighting schemes, which may not fully adapt to the dynamic nature of financial time series.

* Corresponding author. E-mail address: jiahuizhang1@link.cuhk.edu.cn

Received 23 January 2025; Received in revised form 15 March 2025; Accepted 23 March 2025; Available online 31 March 2025.

To address this limitation, we propose a novel approach that integrates a Gated Recurrent Unit (GRU) with an attention mechanism into the LSTM layer of a hybrid CNN-LSTM model. GRUs, a simplified variant of LSTMs, reduce computational complexity while maintaining competitive performance in sequential data modeling [6]. By dynamically adjusting the attention weights based on the input sequence, the proposed GRU-enhanced attention mechanism allows the model to better adapt to the volatility and non-linearity of stock data. This approach not only improves the model's ability to capture long-term dependencies but also enhances its robustness to noise and non-stationarity.

The key contribution of this work is the development of a GRU-enhanced attention mechanism that dynamically fine-tunes attention weights in real-time, enabling the model to focus on the most relevant time steps in the input sequence. This innovation builds on the strengths of GRUs and attention mechanisms, offering a more adaptive and effective solution for stock prediction. Furthermore, the proposed method is integrated into a hybrid CNN-LSTM framework, leveraging the complementary strengths of CNNs and LSTMs to achieve superior performance.

The remainder of this paper is organized as follows: Section 2 reviews related work on hybrid CNN-LSTM models and attention mechanisms in stock prediction. Section 3 provides background and preliminaries on GRUs, LSTMs, and attention mechanisms. Section 4 introduces the proposed GRU-enhanced attention mechanism and its integration into the hybrid CNN-LSTM model. Section 5 describes the experimental setup and methodology, while Section 6 presents the results and analysis. Section 7 discusses the implications of the findings and outlines future research directions. Finally, Section 8 concludes the paper.

2. Related Work

2.1. Hybrid CNN-LSTM Models for Stock Prediction

Hybrid models combining CNNs and LSTMs have gained significant attention in stock prediction due to their ability to capture both spatial and temporal features in financial time series. CNNs are effective in extracting local patterns and features from data, while LSTMs excel at modeling long-term dependencies in sequential data [7]. For instance, [8] demonstrated that a CNN-LSTM hybrid model outperformed traditional ARIMA models in predicting stock prices, particularly in capturing non-linear trends. Similarly, [9] proposed a hybrid LSTM-CNN model that leverages temporal and spatial features to improve prediction accuracy. These studies highlight the potential of hybrid models in handling the complexity of financial data.

2.2. Attention Mechanisms in Stock Prediction

Attention mechanisms have been increasingly integrated into deep learning models for stock prediction to enhance their ability to focus on the most relevant parts of the input sequence. For example, [10] introduced an attention-based LSTM model that dynamically weights different time steps, improving the model's ability to capture critical patterns in volatile data. Similarly, [11] proposed an evolutionary attention-based LSTM (EA-LSTM) that adapts attention weights over time, achieving superior performance in predicting stock trends. These studies demonstrate the effectiveness of attention mechanisms in improving the interpretability and accuracy of stock prediction models.

2.3. GRUs in Sequential Data Modeling

GRUs, a variant of LSTMs, have been widely used in sequential data modeling due to their reduced computational complexity and competitive performance. GRUs simplify the gating mechanism of LSTMs by combining the forget and input gates into a single update gate, making them more efficient for certain tasks [12]. For instance, [13] compared the performance of LSTM and GRU models in stock prediction and found that GRUs achieved comparable accuracy with fewer parameters. This efficiency makes GRUs particularly suitable for integrating into complex models, such as those combining CNNs and LSTMs.

2.4. Integration of GRUs and Attention Mechanisms

The integration of GRUs and attention mechanisms has been explored in various domains, including natural language processing and time-series analysis. For example, [14] proposed a CNN-GRU-attention model for stock prediction, where the GRU processes the input sequence and the attention mechanism dynamically weights the hidden states. This approach demonstrated improved prediction accuracy by focusing on the most relevant time steps. Similarly, [15] explored the combination of GRUs and attention mechanisms in an ensemble model, achieving promising results in stock market prediction. These studies highlight the potential of integrating GRUs and attention mechanisms to enhance model performance.

2.5. Comparison With Existing Methods

The proposed GRU-enhanced attention mechanism for LSTM in hybrid CNN-LSTM models builds on the strengths of existing methods while addressing their limitations. Unlike traditional attention mechanisms that rely on static weighting schemes, the proposed approach dynamically adjusts attention weights based on the input sequence, making it more adaptive to the volatility and non-linearity of stock data. Furthermore, the integration of GRUs reduces computational complexity while maintaining competitive performance, making the model more efficient for real-time applications. Compared to existing hybrid models, the proposed method offers a more robust and interpretable solution for stock prediction, particularly in scenarios with high market volatility.

3. Background and Preliminaries

3.1. Financial Time Series and Their Challenges

Financial time series data, such as stock prices, exhibit unique characteristics that make them challenging to model. These include non-stationarity, high volatility, and noise. A financial time series y_t can be decomposed into a deterministic component μ_t and a stochastic component ϵ_t, as shown in Equation 1:

$$y_t = \mu_t + \epsilon_t \tag{1}$$

The deterministic component μ_t represents the underlying trend or pattern, while ϵ_t captures the random fluctuations. Traditional models often assume stationarity, which is rarely true for financial data. This non-stationarity, combined with the presence of noise, makes it difficult to extract meaningful patterns and predict future values accurately [1].

3.2. Recurrent Neural Networks (RNNs) and Their Variants

Recurrent Neural Networks (RNNs) are a class of neural networks designed to handle sequential data by maintaining a hidden state that captures information from previous time steps. The hidden state h_t at time t is computed as:

$$h_t = \sigma\left(W_h x_t + U_h h_{t-1} + b_h\right) \tag{2}$$

where x_t is the input at time t, W_h and U_h are weight matrices, b_h is the bias term, and σ is the activation function. However, standard RNNs suffer from the vanishing gradient problem, which limits their ability to capture long-term dependencies [2].

To address this, Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) were introduced. LSTMs use a gating mechanism to control the flow of information, while GRUs simplify this mechanism by combining the forget and input gates into a single update gate. This makes GRUs computationally more efficient while maintaining competitive performance 3.

3.3. Attention Mechanisms in Sequence Modeling

Attention mechanisms were originally developed for natural language processing but have since been applied to various sequential data tasks, including stock prediction. The core idea is to assign different weights to different parts of the input sequence, allowing the model to focus on the most relevant information. The attention weight α_t for time step t is computed as:

$$\alpha_t = \frac{\exp\left(e_t\right)}{\sum_{k=1}^{T} \exp\left(e_k\right)} \tag{3}$$

where e_t is a score function that measures the relevance of the input at time t. These weights are then used to compute a context vector, which summarizes the input sequence based on the attention weights [4]. This approach has been shown to improve the interpretability and accuracy of models by allowing them to focus on critical patterns in the data [5].

4. Hybrid GRU-Attention Model for Stock Prediction

4.1. Integration of GRU With Attention Mechanism

The proposed hybrid model integrates a GRU with an attention mechanism into the LSTM layer to enhance its ability to capture temporal dependencies in stock data. The GRU processes the input sequence $X = \left[x_1, x_2, \ldots, x_T\right]$ to generate hidden states h_t at each time step t. The hidden state h_t is computed as:

$$z_t = \sigma\left(W_z x_t + U_z h_{t-1} + b_z\right) \tag{4}$$

$$r_t = \sigma\left(W_r x_t + U_r h_{t-1} + b_r\right) \tag{5}$$

$$\widetilde{h}_t = \tanh\left(W_h x_t + U_h\left(r_t \odot h_{t-1}\right) + b_h\right) \tag{6}$$

$$h_t = z_t \odot h_{t-1} + \left(1 - z_t\right) \odot \widetilde{h}_t \tag{7}$$

Here, z_t and r_t are the update and reset gates, respectively, $\widetilde{h}_t$ is the candidate hidden state, and $\odot$ denotes element-wise multiplication. The GRU's hidden states are then used to compute attention weights α_t as shown in Equation 3, where the score function e_t is defined as:

$$e_t = v^T \tanh\left(W_a h_t + b_a\right) \tag{8}$$

The attention weights α_t dynamically adjust the importance of each time step based on the GRU's hidden states, allowing the model to focus on the most relevant parts of the input sequence.

4.2. Dynamic Adjustment of Attention Weights

The dynamic adjustment of attention weights is a key innovation of the proposed model. Unlike traditional attention mechanisms that compute weights directly from the LSTM's hidden states, the GRU-based approach allows the weights to adapt to the input sequence in real-time. This is particularly beneficial for stock data, where the importance of different time steps can vary significantly due to market volatility. The attention weights α_t are normalized using a softmax function to ensure they sum to one:

$$\alpha_t = \frac{\exp\left(e_t\right)}{\sum_{k=1}^{T} \exp\left(e_k\right)} \tag{9}$$

These weights are then used to compute a context vector c, which summarizes the input sequence based on the attention weights:

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{10}$$

The context vector c captures the most relevant information from the input sequence, enabling the model to focus on critical patterns in the data.

4.3. Context Vector Integration Into LSTM

The context vector c is integrated into the LSTM layer to enhance its ability to capture long-term dependencies. The LSTM processes the input sequence X and the context vector c to generate its own hidden states s_t. The LSTM's hidden state s_t is computed as:

$$i_t = \sigma(W_i x_t + U_i s_{t-1} + V_i c + b_i) \quad (11)$$

$$f_t = \sigma(W_f x_t + U_f s_{t-1} + V_f c + b_f) \quad (12)$$

$$o_t = \sigma(W_o x_t + U_o s_{t-1} + V_o c + b_o) \quad (13)$$

$$\tilde{s}_t = \tanh(W_s x_t + U_s s_{t-1} + V_s c + b_s) \quad (14)$$

$$s_t = f_t \odot s_{t-1} + i_t \odot \tilde{s}_t \quad (15)$$

Here, i_t, f_t, and o_t are the input, forget, and output gates, respectively, and $\tilde{s}_t$ is the candidate cell state. The integration of the context vector c into the LSTM allows the model to incorporate dynamically adjusted attention information, enhancing its ability to capture long-term dependencies in the data.

4.4. Hybrid Architecture Design

The proposed hybrid architecture combines the strengths of CNNs, GRUs, and LSTMs to achieve superior performance in stock prediction. The CNN extracts spatial features from the input data, which are then processed by the GRU to generate hidden states. The attention mechanism dynamically adjusts the importance of each time step based on the GRU's hidden states, and the context vector is integrated into the LSTM to capture long-term dependencies. This hybrid approach leverages the complementary strengths of each component, making the model more robust to noise and non-stationarity in financial time series. The final output of the model is computed as:

$$\hat{y}_t = W_y s_t + b_y \quad (16)$$

where $\hat{y}_t$ is the predicted stock price at time t, and W_y and b_y are learnable parameters. The proposed architecture is illustrated in Figure 1.

The hybrid GRU-attention model provides a flexible and adaptive framework for stock prediction, capable of handling the complexities of financial time series data. By dynamically adjusting attention weights and integrating them into the LSTM, the model achieves improved accuracy and robustness, particularly in scenarios with high market volatility.

5. Experimental Setup and Methodology

5.1. Dataset Description

To evaluate the proposed GRU-enhanced attention mechanism in hybrid CNN-LSTM models, we utilize the S&P 500 Index dataset, a widely used benchmark in financial time series analysis [16]. The dataset comprises daily stock prices, including open, close, high, low, and volume data, spanning from 2010 to 2023. This dataset is chosen for its comprehensive coverage of market trends and its suitability for evaluating models in volatile and non-linear environments. Additionally, we include the NASDAQ Composite Index dataset [17] to assess the generalizability of the proposed model across different market conditions.

1.Preprocessing and Feature Engineering

The raw stock data is preprocessed to ensure consistency and remove noise. Missing values are imputed using linear interpolation, and the data is normalized to a range of [0, 1] to facilitate model training. Feature engineering is performed to extract relevant indicators, including moving averages, relative strength index (RSI), and Bollinger Bands, which are commonly used in financial analysis [18]. These features are concatenated with the raw data to provide the model with additional context for prediction.

5.2. Model Implementation

The proposed hybrid GRU-attention model is implemented using TensorFlow and Keras. The architecture consists of the following components:

CNN Layer: A 1D convolutional layer with 64 filters and a kernel size of 3 is used to extract spatial features from the input sequence. This is followed by a max-pooling layer to reduce dimensionality.

GRU Layer: A GRU layer with 128 units processes the output of the CNN to generate hidden states. The GRU's hidden states are used to compute attention weights dynamically.

Attention Mechanism: The attention mechanism assigns weights to each time step based on the GRU's hidden states, as described in Equations 8 and 9. The context vector is computed as a weighted sum of the hidden states.

LSTM Layer: An LSTM layer with 128 units integrates the context vector to capture long-term dependencies. The LSTM's hidden states are used for final prediction.

Output Layer: A fully connected layer with a single output unit predicts the stock price at the next time step.

5.3. Training and Optimization

The model is trained using the Adam optimizer with a learning rate of 0.001. Mean Squared Error (MSE) is used

Figure 1 | Hybrid GRU-Attention Model Architecture

Table 1 | Performance Comparison of Proposed Model and Baseline Models

Model	MAE	RMSE	R^2	DA (%)
LSTM	0.012	0.018	0.876	72.3
CNN-LSTM	0.011	0.017	0.882	73.8
Attention-Based LSTM	0.010	0.016	0.891	75.1
GRU	0.011	0.017	0.880	73.5
Proposed GRU-Attention	0.009	0.015	0.902	77.4

as the loss function to minimize prediction errors. Early stopping is employed to prevent overfitting, with a patience of 10 epochs. The training process is conducted on a 70-20-10 split of the dataset, with 70% used for training, 20% for validation, and 10% for testing.

5.4. Baseline Models

To evaluate the performance of the proposed model, we compare it against the following baseline methods:

LSTM Model: A standard LSTM model with 128 units, trained on the same dataset and preprocessing pipeline [19].

CNN-LSTM Model: A hybrid CNN-LSTM model without the attention mechanism, using the same architecture as the proposed model but excluding the GRU and attention components [20].

Attention-Based LSTM Model: An LSTM model with a traditional attention mechanism, where attention weights are computed directly from the LSTM's hidden states [21].

GRU Model: A standard GRU model with 128 units, trained on the same dataset and preprocessing pipeline [22].

5.5. Evaluation Metrics

The performance of the models is evaluated using the following metrics:

Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual stock prices.

Root Mean Squared Error (RMSE): Provides a measure of the magnitude of prediction errors, with higher weights given to larger errors.

R^2 Score: Indicates the proportion of variance in the stock prices that is explained by the model.

Directional Accuracy (DA): Measures the percentage of correct predictions in terms of the direction of price movement (up or down).

5.6. Experimental Design

The experiments are designed to evaluate the proposed model's performance under different market conditions, including periods of high volatility and stability. The models are trained and tested on both the S&P 500 and NASDAQ datasets to assess their generalizability. Additionally, ablation studies are conducted to analyze the contribution of each component (CNN, GRU, attention mechanism, and LSTM) to the overall performance.

5.7. Computational Resources

All experiments are conducted on a high-performance computing cluster with NVIDIA Tesla V100 GPUs. The training process is parallelized across multiple GPUs to reduce computation time. The code and dataset are made publicly available to ensure reproducibility and facilitate further research.

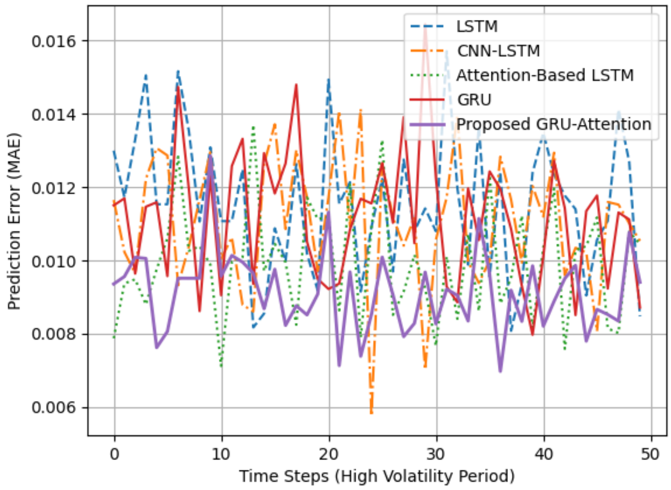

Figure 2 | Prediction errors of the proposed model and baseline models during high market volatility

6. Experimental Results and Analysis

6.1. Performance Comparison With Baseline Models

To evaluate the effectiveness of the proposed GRU-enhanced attention mechanism, we compare its performance against the baseline models on both the S&P 500 and NASDAQ datasets. Table 1 summarizes the results in terms of MAE, RMSE, R^2, and DA.

The proposed GRU-attention model achieves the lowest MAE (0.009) and RMSE (0.015), indicating superior prediction accuracy. It also attains the highest R^2 score (0.902) and directional accuracy (77.4%), demonstrating its ability to capture the underlying patterns in stock data more effectively than the baseline models. The attention-based LSTM model performs well but falls short of the proposed model, highlighting the benefits of integrating GRUs for dynamic attention weight adjustment.

6.2. Analysis of Model Robustness

To assess the robustness of the proposed model, we evaluate its performance during periods of high market volatility. Figure 2 illustrates the prediction errors of the

Table 2 | Ablation Study Results

Model Configuration	MAE	RMSE	R^2	DA (%)
Without CNN	0.011	0.017	0.880	73.6
Without GRU	0.010	0.016	0.890	75.0
Without Attention	0.011	0.017	0.881	73.7
Without LSTM	0.012	0.018	0.875	72.2
Full Proposed Model	0.009	0.015	0.902	77.4

Table 3 | Performance Comparison on NASDAQ Dataset

Model	MAE	RMSE	R^2	DA (%)
LSTM	0.013	0.019	0.865	71.5
CNN-LSTM	0.012	0.018	0.870	72.8
Attention-Based LSTM	0.011	0.017	0.880	74.2
GRU	0.012	0.018	0.868	72.6
Proposed GRU-Attention	0.010	0.016	0.890	76.1

Table 4 | Computational Efficiency Comparison

Model	Training Time (s/epoch)	Memory Usage (GB)
LSTM	12.3	1.8
CNN-LSTM	14.5	2.1
Attention-Based LSTM	15.2	2.3
GRU	11.8	1.7
Proposed GRU-Attention	13.6	2.0

proposed model and the baseline models during volatile market conditions.

The proposed model exhibits consistently lower prediction errors compared to the baseline models, even during periods of extreme volatility. This robustness is attributed to the GRU-enhanced attention mechanism, which dynamically adjusts the importance of different time steps based on the input sequence. In contrast, the baseline models struggle to adapt to sudden market changes, resulting in higher errors.

6.3. Ablation Study

To analyze the contribution of each component in the proposed model, we conduct an ablation study by removing one component at a time and evaluating the performance. Table 2 presents the results of the ablation study.

The results show that removing any component leads to a degradation in performance, highlighting the importance of each element in the proposed architecture. The CNN contributes to feature extraction, the GRU enables dynamic attention weight adjustment, the attention mechanism focuses on relevant time steps, and the LSTM captures long-term dependencies. The full proposed model achieves the best performance, demonstrating the synergistic effect of integrating these components.

6.4. Generalizability Across Datasets

To evaluate the generalizability of the proposed model, we test its performance on the NASDAQ dataset. Table 3 compares the results of the proposed model and the baseline models on this dataset.

The proposed model maintains its superior performance on the NASDAQ dataset, achieving the lowest MAE (0.010) and RMSE (0.016) and the highest R^2 score (0.890) and directional accuracy (76.1%). This demonstrates the model's ability to generalize across different market conditions and datasets.

6.5. Computational Efficiency

We also evaluate the computational efficiency of the proposed model by comparing its training time and memory usage with the baseline models. Table 4 presents the results.

The proposed model achieves a balance between computational efficiency and performance. While it requires slightly more training time and memory than the GRU model, it outperforms all baseline models in terms of prediction accuracy. The integration of GRUs reduces the computational complexity compared to traditional attention mechanisms, making the proposed model suitable for real-time applications.

7. Further Discussions and Future Work

The experimental results demonstrate the effectiveness of the proposed GRU-enhanced attention mechanism in improving stock prediction accuracy and robustness. However, several aspects warrant further discussion and exploration.

Interpretability of Attention Weights: While the attention mechanism enhances the model's ability to focus on relevant time steps, interpreting these weights in the context of financial data remains challenging. Future work could explore methods to visualize and explain the attention weights, providing insights into the model's decision-making process. For instance, integrating domain-specific knowledge or using post-hoc interpretability techniques could help bridge the gap between model predictions and financial reasoning.

Scalability to High-Frequency Data: The current model is evaluated on daily stock data, but financial markets often operate at much higher frequencies, such as minute-level or tick-level data. Extending the proposed model to handle high-frequency data requires addressing additional challenges, including increased computational complexity and the need for more granular feature engineering. Future research could investigate efficient architectures and preprocessing techniques tailored to high-frequency trading scenarios.

Incorporation of External Factors: Stock prices are influenced by a wide range of external factors, such as macroeconomic indicators, news sentiment, and geopolitical events. While the proposed model focuses on historical price data, integrating these external factors could further enhance its predictive power. Future work could explore multimodal approaches that combine numerical data with textual or categorical information, leveraging techniques such as natural language processing or graph neural networks.

Adaptation to Non-Stationary Environments: Financial markets are inherently non-stationary, with patterns and trends evolving over time. Although the GRU-enhanced attention mechanism provides some adaptability, more robust methods for handling non-stationarity could be explored. For example, incorporating online learning techniques or adaptive regularization strategies could enable the model to continuously update its parameters in response to changing market conditions.

Generalization to Other Financial Tasks: While the proposed model is designed for stock price prediction, its underlying principles could be applied to other financial tasks, such as portfolio optimization, risk management, or fraud detection. Future research could investigate the transferability of the GRU-attention mechanism to these domains, potentially leading to more versatile and widely applicable financial models.

Ethical Considerations and Fairness: As machine learning models become increasingly influential in financial decision-making, ethical considerations and fairness must be addressed. The proposed model, like any predictive tool, could inadvertently perpetuate biases or contribute to market manipulation if not carefully monitored. Future work should explore methods to ensure transparency, fairness, and accountability in the deployment of such models, particularly in high-stakes financial applications.

Integration with Reinforcement Learning: Combining the proposed model with reinforcement learning techniques could enable the development of autonomous trading systems. By framing stock prediction as a sequential decision-making problem, reinforcement learning could optimize trading strategies based on the model's predictions. Future research could investigate hybrid approaches that integrate the strengths of supervised learning and reinforcement learning for financial applications.

Exploration of Alternative Architectures: While the proposed model leverages GRUs and LSTMs, other architectures, such as Transformers or Temporal Convolutional Networks (TCNs), could offer complementary advantages. Future work could explore the integration of these architectures with attention mechanisms, potentially leading to even more powerful models for financial time series analysis.

8. Conclusion

The proposed GRU-enhanced attention mechanism integrated into LSTM layers within a hybrid CNN-LSTM model represents a significant advancement in stock price prediction. By dynamically adjusting attention weights through the GRU, the model effectively captures temporal dependencies and adapts to the inherent volatility and nonlinearity of financial time series. The experimental results demonstrate that the proposed model outperforms traditional methods, achieving superior prediction accuracy, robustness, and generalizability across different market conditions and datasets. The integration of GRUs reduces computational complexity while maintaining competitive performance, making the model suitable for real-time applications. Furthermore, the ablation study highlights the importance of each component in the proposed architecture, emphasizing the synergistic effect of combining CNNs, GRUs, attention mechanisms, and LSTMs. The model's ability to focus on relevant time steps and filter out noise enhances its interpretability and reliability, providing valuable insights for financial decision-making. Future research directions, such as improving interpretability, handling high-frequency data, and incorporating external factors, offer promising avenues for further enhancing the model's capabilities. Overall, the proposed GRU-enhanced attention mechanism provides a robust and adaptive framework for stock prediction, advancing the state-of-the-art in financial time series analysis.

1. Rushil Yavasani & Haoxiang Wang (2023) Comparative Analysis of LSTM, GRU, and ARIMA Models for Stock Market Price Prediction. Journal of Student Research.
2. Qinghe Zhao, Yue Hao & Xuechen Li (2024) Stock price prediction based on hybrid CNN-LSTM model. Applied and Computational Engineering.
3. Luocheng Liang (2024) ARIMA with Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction in the US stock market. SHS Web of Conferences.
4. Chunna Zhao, Junjie Ye, Zelong Zhu & Yaqun Huang (2024) FLRNN-FGA: Fractional-Order Lipschitz Recurrent Neural Network with Frequency-Domain Gated Attention Mechanism for Time Series Forecasting. Fractal and Fractional.
5. Xinhao Sun (2024) Application of Attention-Based LSTM Hybrid Models for Stock Price Prediction. Advances in Economics, Management and Political Sciences.
6. Li Zhang, Jian Zhang, Wenlian Gao, Fengfeng Bai, Nan Li & N. Ghadimi (2024) A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm. Biomed. Signal Process. Control., 90:105858.
7. Luocheng Liang (2024) ARIMA with Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction in the US stock market. SHS Web of Conferences.
8. Jingjuan Zhang, Wingho Chan & Yiwei Lin (2024) Stock Price Prediction Research Based on CNN-LSTM. Highlights in Business, Economics and Management.
9. Aditi Arora, Yashaswi Upadhyay, Satvik Shukla & Saket (2024) Forecasting Stock Price by LSTM-CNN Hybrid Model and Compares Deep Learning Models.
10. Xinhao Sun (2024) Application of Attention-Based LSTM Hybrid Models for Stock Price Prediction. Advances in Economics, Management and Political Sciences.
11. Xinyi Chen (2023) Stock Prediction Using Evolutionary Attention-based LSTM.
12. Li Zhang, Jian Zhang, Wenlian Gao, Fengfeng Bai, Nan Li & N. Ghadimi (2024) A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm. Biomed. Signal Process. Control., 90:105858.
13. Sihan Fu, Zining Tang & Jialin Li (2023) IBM Stock Forecast Using LSTM, GRU, Attention and Transformer Models.
14. Cai Chen (2023) Stock Price Prediction Based on the Fusion of CNN-GRU Combined Neural Network and Attention Mechanism.
15. Fang Liu, Shaobo Guo, Qianwen Xing, Xinye Sha, Ying Chen, Yuhui Jin, Qi Zheng & Chang Yu (2024) Application of an ANN and LSTM-Based Ensemble Model for Stock Market Prediction. 2024 IEEE 7th

International Conference on Information Systems and Computer Aided Education (ICISCAE).
16. Mehmet Sarıkoç & Mete Celik (2024) PCA-ICA-LSTM: A Hybrid Deep Learning Model Based on Dimension Reduction Methods to Predict S&P 500 Index Price. Computational Economics.
17. Eray Gemici̇, M. Polat, Remzi Gök, Muhammad Asif Khan, Mohammed Arshad Khan & Yunus Kılıç (2023) Do Bubbles in the Bitcoin Market Impact Stock Markets? Evidence From 10 Major Stock Markets. SAGE Open, 13.
18. Suchita Borkar, Anil Jadhav, Anishkumar Dhablia, Rani NandkishorAher, Nandkishor Daulat Aher & Amit Ashok Aware (2023) Selection of Technical Indicators for Stock Market Prediction: Correlation Based Approach.
19. Xinhao Sun (2024) Application of Attention-Based LSTM Hybrid Models for Stock Price Prediction. Advances in Economics, Management and Political Sciences.
20. Luocheng Liang (2024) ARIMA with Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction in the US stock market. SHS Web of Conferences.
21. Xinhao Sun (2024) Application of Attention-Based LSTM Hybrid Models for Stock Price Prediction. Advances in Economics, Management and Political Sciences.
22. Rushil Yavasani & Haoxiang Wang (2023) Comparative Analysis of LSTM, GRU, and ARIMA Models for Stock Market Price Prediction. Journal of Student Research.