

Multi-Agent Stock Prediction Systems: Machine Learning Models, Simulations, and Real-Time Trading Strategies

Daksh Dave, Vikhyat Chauhan, Gauransh Sawhney,
Rishabh Karthik Ramesh

Abstract

This paper presents a comprehensive study on stock price prediction, leveraging advanced machine learning (ML) and deep learning (DL) techniques to improve financial forecasting accuracy. The research evaluates the performance of various recurrent neural network (RNN) architectures, including Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), and attention-based models. These models are assessed for their ability to capture complex temporal dependencies inherent in stock market data. Our findings show that attention-based models outperform other architectures, achieving the highest accuracy by effectively capturing both short- and long-term dependencies. This study contributes valuable insights into AI-driven financial forecasting, offering practical guidance for developing more accurate and efficient trading systems.

Keywords: Stock Prediction, Financial Forecasting, LSTM, AI-driven Trading Systems

1 Introduction

The stock market, characterized by its inherent volatility and complexity, poses a considerable challenge for accurate price forecasting and informed decision-making. In recent years, the rapid advancement of machine learning (ML) and deep learning (DL) techniques has brought transformative changes to the financial industry, offering more sophisticated tools for predicting market trends and developing trading strategies [1]. Traditional statistical methods, such as ARIMA [2] and linear regression [3], while foundational, have increasingly been replaced by neural network-based models that excel in capturing complex temporal patterns in financial data. Among these, architectures like Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and attention-based models have demonstrated superior performance in time-series prediction tasks, making them invaluable for stock price forecasting.

In this research, we systematically explore a broad array of ML and DL models, applying them to the task of stock price forecasting. Our study encompasses both univariate and multivariate time series data, providing a robust evaluation of several cutting-edge deep learning architectures. These include LSTM (and its variations such as Bidirectional LSTM and 2-Path LSTM), GRU (and its bidirectional and 2-path versions), as well as Seq2seq models and their extensions, such as Seq2seq VAE and hybrid approaches that integrate auto-encoders with boosting algorithms. Attention-based models, including the "Attention is All You Need" architecture, are also examined to assess their ability to enhance predictive accuracy.

Through case studies, particularly involving TESLA stock, we demonstrate the practical application of these models in real-world scenarios. The evaluation covers both predictive accuracy and their implications for automated trading strategies, providing a holistic view of the potential for AI-driven systems in the financial domain. The insights gained from this research contribute to the growing body of knowledge on AI-driven stock prediction systems and their capacity to revolutionize quantitative finance.

The major contributions of this paper are as follows:

1. A comprehensive evaluation of the effectiveness of various ML and DL models in forecasting stock prices.
2. A practical demonstration of the models through case studies on real-world stock data, specifically focusing on TESLA.
3. Insights into the potential of AI-driven systems to improve trading performance in stock markets.

The remainder of this paper is organized as follows: Section 2 presents the literature review, discussing prior work in the field. Section 3 describes the proposed methodology in detail. The experimental setup and datasets are outlined in Section 4. Finally, Section 5 discusses the results and their implications for future research.

2 Literature Review

Predicting the stock market is challenging yet crucial for investors, traders, and researchers. Various methods, including mathematical, statistical, and Artificial Intelligence (AI) techniques, have been proposed to forecast stock prices. Parmar et al. [4] discovered that machine learning and deep learning prediction methods significantly outperform traditional stock market forecasting techniques in terms of both speed and accuracy. W. Khan [5] noted that social media significantly influences stock predictions, with Random Forest consistently performing well in all scenarios. Additionally, utilizing larger datasets and incorporating sentiment analysis into ML and DL models can enhance the precision of stock price predictions. These findings lead to development of many machine learning models, one being the Support Vector Machines (SVM) which belongs to the class of supervised learning algorithms. Mankar et al. [6] found that SVM proved to be a more effective and practical machine learning model for predicting stock price movements based on the sentiment expressed in tweets. Implementation of other deep learning techniques like artificial neural networks (ANNs)

shows significant improvement over the previous machine learning solutions. Rao P. S. et al. [7] improved upon the existing model by proposing an artificial neural networks (ANNs), utilizing the backpropagation algorithm. The model outperformed the traditional regression methods and exhibited lower prediction errors. Naik N. and Mohan B. R. [8] found that deep learning techniques outperformed machine learning techniques in terms of results. Further development into deep learning techniques suggested usage of RNN which can capture temporal dependencies and can further improve upon the prediction accuracy.

Selvamuthu D., Kumar V., and Mishra A. [9] discovered that tick data provided more accurate predictions compared to 15-minute data. They reported that the Levenberg-Marquardt, Scaled Conjugate Gradient, and Bayesian Regularization methods achieved an impressive 99.9% accuracy with tick data. Additionally, they suggested that using LSTM RNN would be advantageous. W. Fang et al. [10] proposed the LSTM predictive model, which includes an embedded layer and an automatic encoder. They note that stock news is not fully leveraged and that the approach has only been applied to the Chinese stock market. Their findings demonstrate that shallow machine learning algorithms, such as SVM and backpropagation, deliver lower accuracy compared to the LSTM and embedded LSTM methods. There have been several variations of LSTM RNN applications over stock prediction which proved it as a robust method over previous implementations. S. Mohan, S. Mullapudi, et al. [11] demonstrated that RNN models incorporating LSTM outperformed the ARIMA model and showed a slight advantage over the Facebook Prophet algorithm. According to the research by Vignesh CK [12], the LSTM method achieved a mean accuracy of 66.83 for Yahoo Finance, while the SVM method only attained an accuracy of 65.20. Training with smaller datasets while increasing the number of epochs can enhance testing outcomes across different datasets. M. Nikou, G. Mansourfar, and J. Bagherzadeh [8] demonstrated that the LSTM-RNN block provides superior predictions of the closing price for the company's dataset compared to other methods. They recommend using combined models, such as the SVR model paired with a genetic algorithm, along with other hybrid approaches from machine learning algorithms. The findings of N. Sirimevan et al. [13] indicates that the LSTM-RNN model, when combined with weighted average and differential evolution techniques, effectively predicted stock prices.

A. Moghar and M. Hamiche [14] demonstrated that the accuracy of stock price forecasting using LSTM models improves as the number of training epochs increases, specifically for GOOGL and NKE assets. J. Eapen, D. Bein, and A. Verma [15] found that integrating Convolutional Neural Networks (CNN) with BiLSTM models enhanced stock market prediction accuracy by 9 percentage compared to using a single deep learning pipeline. Similarly, Adil Moghar and Mhamed Hamiche [16] concluded that BiLSTM models produce a lower Root Mean Squared Error (RMSE) compared to standard LSTM models, making them a more favorable option for stock prediction tasks. This hybrid approach also outperformed traditional Support Vector Machine (SVM) regressors in forecasting temporal sequences. Further optimizations have been proposed by incorporating external features into deep learning models to improve predictive accuracy. For instance, Khan, W., Malik, et al. [17] discovered that adding sentiment analysis as an attribute had a minimal impact on stock price predictions

but still improved the accuracy of machine learning algorithms by approximately 2 percentage. Moreover, M. Nabipour et al. [18] found that using binary data instead of continuous data resulted in a significant enhancement in stock price predictions, highlighting the importance of data representation in predictive modeling. The reviewed literature highlights the growing effectiveness of deep learning models, such as LSTM and BiLSTM, in improving stock price forecasting accuracy through model integration and external optimizations. The reviewed literature highlights the growing effectiveness of deep learning models in improving stock price forecasting accuracy through model integration and external optimizations. However the field is evolving with time abd further research needs to be done to deploy machine learning models that can mimic humans and ensure that profit is generated with minimal human intervention.

3 Proposed Methodology

In this section, we outline the methodology for our experiment, which involves utilizing three distinct architectures: Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Transformers. The following sections will provide a detailed explanation of the implementation and processes for each architecture.

3.1 LSTM

LSTM (Long Short-Term Memory) [19] is a type of Recurrent Neural Network (RNN) designed to capture long-term dependencies in sequential data and is suitable for processing and predicting with long intervals and delays in time series. LSTM can be used to make Stock Price Predictions where the network learns the temporal dependencies in historical stock price data, capturing different patterns to deduce future stock prices. The key feature of LSTM is its use of gating mechanisms to regulate the flow of information through the network, addressing the vanishing gradient problem that can occur in standard RNNs. The figure 1 denotes a LSTM cell.

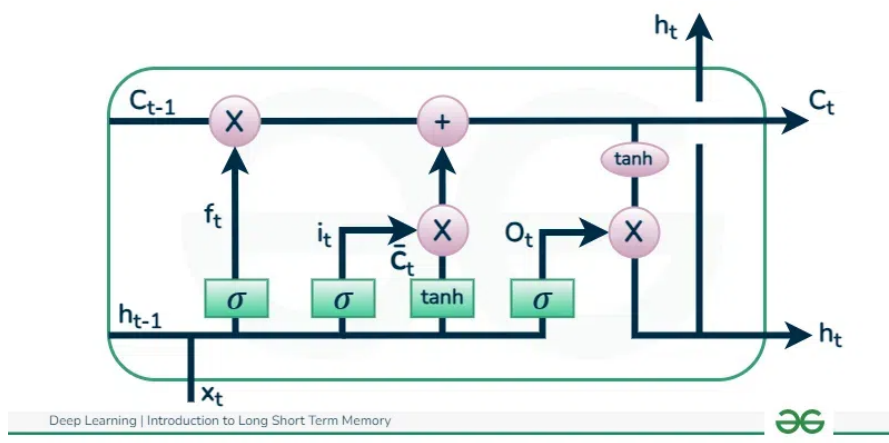


Fig. 1: LSTM cell [20]

An LSTM cell consists of three main gates:

1. **Forget Gate:** Decides which part of the previous cell state should be forgotten.
2. **Input Gate:** Determines which new information should be added to the cell state.
3. **Output Gate:** Controls the output of the current cell.

Additionally, an LSTM cell maintains a **cell state** C_t , which acts as a memory, and a **hidden state** h_t , which is used for the output at each time step.

At each time step t , the LSTM performs the following operations:

1. **Forget Gate:** The forget gate determines how much of the previous cell state C_{t-1} should be retained for the current step:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Where: - x_t is the input at time step t , - h_{t-1} is the hidden state from the previous time step, - W_f and b_f are the forget gate weights and biases, - σ is the sigmoid activation function, which outputs values between 0 and 1.

2. **Input Gate:** The input gate controls how much of the new information \tilde{C}_t should be added to the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

Next, a candidate cell state \tilde{C}_t is computed:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Where: - W_i , b_i , W_C , and b_C are the input gate and candidate cell state weights and biases, - \tanh is the hyperbolic tangent activation function.

3. **Cell State Update:** The new cell state C_t is a combination of the old cell state C_{t-1} (modulated by the forget gate) and the new candidate cell state \tilde{C}_t (modulated by the input gate):

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

Where \odot denotes element-wise multiplication.

4. **Output Gate:** The output gate determines what part of the cell state should be output as the hidden state h_t :

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

Finally, the hidden state h_t is computed by applying the \tanh function to the cell state and modulating it with the output gate:

$$h_t = o_t \odot \tanh(C_t)$$

- The **forget gate** f_t determines how much of the previous memory C_{t-1} to retain.
- The **input gate** i_t modulates how much new information \tilde{C}_t to add. - The **cell state** C_t is updated based on these two gates. - The **output gate** o_t controls what is output as the hidden state h_t .

By combining these gates, the LSTM cell can retain long-term dependencies over many time steps, which makes it effective for tasks involving sequential data, such as language modeling, time-series forecasting, and more.

3.2 GRU

GRU (Gated Recurrent Unit) [21] is a simpler variant of the LSTM (Long Short-Term Memory) network. It also addresses the vanishing gradient problem by using gating mechanisms but with a simplified structure. GRU can capture temporal dependencies in sequential data and is effective in predicting stock prices by learning from historical patterns and trends. A GRU has two gates: the **update gate** and the **reset gate**, which control the flow of information through the unit. The figure 2 denotes a GRU cell

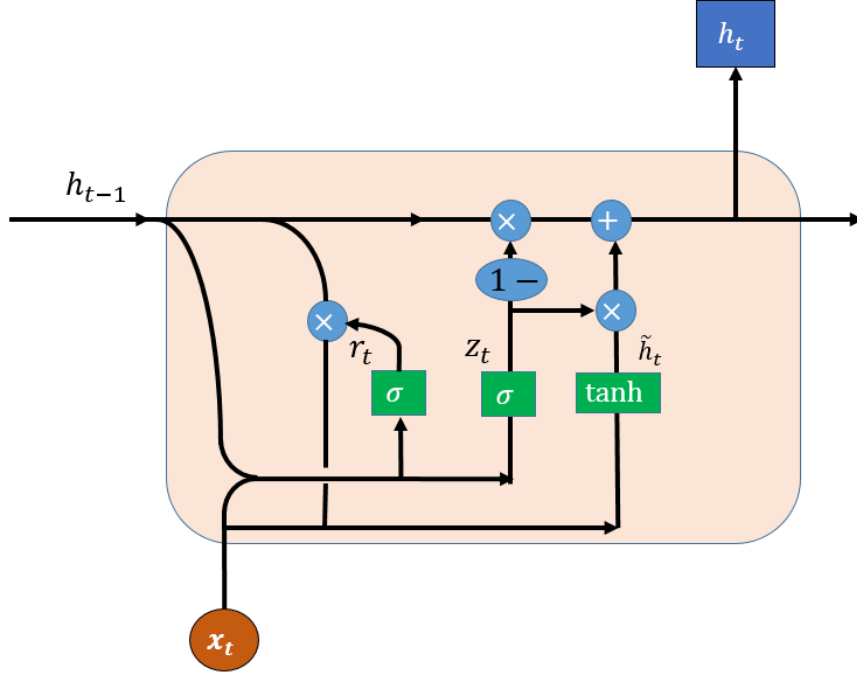


Fig. 2: GRU cell [22]

At each time step t , the GRU performs the following operations:

1. **Update Gate:** The update gate z_t determines how much of the previous hidden state h_{t-1} should be retained in the current hidden state:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z)$$

Where: - x_t is the input at time step t , - h_{t-1} is the hidden state from the previous time step, - W_z and b_z are the weights and biases for the update gate, - σ is the sigmoid activation function.

2. Reset Gate: The reset gate r_t controls how much of the previous hidden state h_{t-1} should be forgotten when updating the current hidden state:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r)$$

3. Candidate Hidden State: The candidate hidden state \tilde{h}_t is calculated using the reset gate r_t , which modulates the contribution of the previous hidden state:

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h)$$

Where: - W_h and b_h are the weights and biases for the candidate hidden state, - \odot represents element-wise multiplication.

4. New Hidden State: The new hidden state h_t is a combination of the previous hidden state h_{t-1} and the candidate hidden state \tilde{h}_t , controlled by the update gate:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t$$

- The **update gate** z_t controls how much of the past information is retained.
- The **reset gate** r_t determines how much of the previous hidden state should influence the current state.
- The **candidate hidden state** \tilde{h}_t is modulated by the reset gate.
- The final **hidden state** h_t is a blend of the previous hidden state and the candidate state, based on the update gate.

GRUs have fewer parameters than LSTMs because they do not have a separate memory cell, which often makes GRUs faster to train and perform comparably well in many tasks involving sequential data.

3.3 Transformer

The paper *Attention is All You Need* [23] introduced the Transformer model, which relies solely on attention mechanisms to handle sequential data, without the need for recurrent or convolutional neural networks. The Transformer architecture has since become the foundation for many state-of-the-art models in natural language processing, such as BERT and GPT, and has also been applied to other domains, including time-series analysis and stock price prediction.

Unlike RNNs and GRUs, which process data sequentially, the Transformer processes the entire input sequence at once, allowing for better parallelization and faster training. This architecture is particularly useful for capturing long-range dependencies in the data. Figure 3 depicts the Transformer architecture.

The Transformer consists of an encoder and a decoder. For stock price prediction, you typically use only the encoder portion, which applies multi-head self-attention and feed-forward layers to learn the temporal patterns in the data. The attention mechanism allows the model to focus on different parts of the input sequence when making predictions. Since the Transformer lacks the inherent sequential structure of

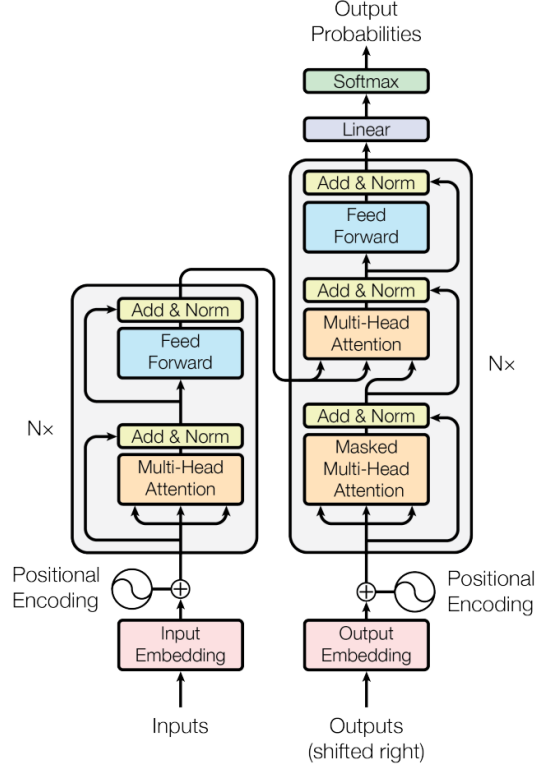


Fig. 3: Transformer architecture

RNNs, positional encodings are added to the input data to provide information about the order of the time steps.

The attention mechanism in Transformers is defined by the following equation:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

where:

- Q (Query): The set of queries.
- K (Key): The set of keys.
- V (Value): The set of values.
- d_k : The dimension of the key vectors, used to scale the dot product for numerical stability.

This equation computes a weighted sum of the values V , where the weights are determined by the dot product between the queries Q and the keys K , scaled by $\sqrt{d_k}$, and passed through a softmax function to produce attention scores.

The attention mechanism in Transformers allows the model to capture both short- and long-term dependencies efficiently, making it highly suitable for tasks like stock price prediction where relationships between data points may span over long intervals.

4 Experimental Design

For our experiments, we plan to use the **Tesla Stock dataset**. A Tesla stock dataset typically contains historical data on the stock performance of Tesla Inc. (TSLA) over a given period of time. This dataset is often used for financial analysis, stock market prediction, machine learning models, and trend analysis. Here’s a general description of the common columns we found to be useful in the Tesla stock dataset:

- **Date:** The date when the stock data was recorded.
- **Open:** The stock price at the opening of the trading day.
- **High:** The highest price reached during the trading day.
- **Low:** The lowest price reached during the trading day.
- **Close:** The stock price at the close of the trading day. This is often the most used column for trend analysis.
- **Adj Close:** The adjusted closing price, which accounts for corporate actions like stock splits, dividends, etc.
- **Volume:** The total number of shares traded during the day.

Date	Open	High	Low	Close	Adj Close	Volume
2024-01-02	720.00	725.00	710.00	715.00	715.00	15,200,000
2024-01-03	715.50	730.00	705.50	725.00	725.00	17,500,000

Table 1: Sample rows from the Tesla Dataset

The Table 1 depicts sample rows from the Tesla dataset. For training and testing, we will partition the dataset into training and testing subsets. The training set comprises all data points from the initial timestamp up to 30 days prior to the final recorded timestamp. The test set includes the data from the last 30 days of the dataset. This approach ensures that the model is trained on historical data and is evaluated by forecasting based on the most recent 30-day period.

For each model in the experiments, the number of layers was set to 1. The size of the hidden layer was fixed at 128 units. Training was performed over 10 epochs and the learning rate was set to 0.01.

For the implementation, we use Python in a Jupyter notebook environment hosted on Google Colab and use L4 GPUs for training. The TensorFlow framework is employed to build and train the forecasting models. This setup provides an efficient, scalable environment for model training and experimentation, with access to GPU acceleration where necessary.

In addition to the baseline recurrent neural network models (LSTM and GRU), we explore multiple architectural variants designed to enhance predictive capabilities. One such variant is the **2Path** model, which processes the input sequence through two

separate recurrent pathways before merging their outputs. This dual-branch design can capture multiple scales or perspectives of the time series. For instance, one path might focus on short-term volatility while the other attends to longer-term trends, and their combined embedding provides a richer context for the prediction task.

We also incorporate the **Seq2Seq** (Sequence-to-Sequence) framework, an encoder-decoder architecture that compresses an input sequence into a fixed-length latent representation and then generates the output sequence step-by-step from this representation. Originally developed for machine translation, Seq2Seq models are well-suited here for forecasting future stock prices, as they can handle variable-length input histories and produce multi-step forecasts that adapt to the complexity of financial time series.

To further enhance the model’s ability to understand temporal dependencies, we leverage **Bidirectional** configurations. Bidirectional architectures process the input sequence in both forward and reverse directions, providing the model with future context when analyzing historical data. This approach mitigates the inherent limitation of unidirectional models, which only have access to past data points. By incorporating information from subsequent time steps, the model can develop a more nuanced and context-rich representation of the time series dynamics.

Lastly, we combine the advantages of both bidirectionality and the encoder-decoder paradigm in **Bidirectional-Seq2Seq** models. In this setup, the encoder is bidirectional, thereby extracting features informed by both past and future elements in the input sequence. The resulting latent representation is then utilized by the decoder, which generates forecasts step-by-step. The synergy of bidirectionality and Seq2Seq allows the model to fully leverage the temporal structure of the data, capturing intricate long-range dependencies and subtle market shifts that simpler architectures might overlook.

5 Results and Discussion

5.1 Model Performance Overview

The evaluation of various models is summarized in Table 2. The table compares the accuracy of LSTM, GRU, their variants, and the Attention mechanism. The results highlight the differences in performance based on the model architecture.

Model Name 1	Accuracy 1	Model Name 2	Accuracy 2
LSTM	89.2522	LSTM-2Path	94.1620
LSTM-Seq2Seq	88.3902	LSTM-Bidirectional	94.1620
LSTM-Bidirectional-Seq2Seq	95.0921	GRU	84.1694
GRU-2Path	83.7746	GRU-Seq2Seq	90.8855
GRU-Bidirectional	87.0331	GRU-Bidirectional-Seq2Seq	87.3672
Attention	95.1467		

Table 2: Model Accuracy Comparison

Additional figures that support the results presented in this section are provided in the Appendix for further reference and analysis.

5.2 Analysis of LSTM Variants

Figures 4 to 8 depict the performance trends of various LSTM-based models:

- **LSTM** (Figure 4) achieved an accuracy of **89.2522%**. Although it captures baseline temporal dependencies, it struggles with nuanced or abrupt changes in the input sequences.
- **LSTM-2Path** (Figure 5) improved to **94.1620%**. The inclusion of a two-path mechanism allows simultaneous processing of parallel data streams, enabling the model to capture more complex patterns that a single-path model might miss.
- **LSTM-Seq2Seq** (Figure 6) achieved **88.3902%**. Contrary to expectations, integrating a seq2seq framework did not substantially enhance performance. One possible reason is that while seq2seq architectures shine in tasks like machine translation, this particular dataset may not benefit as strongly from the encoding-decoding paradigm.
- **LSTM-Bidirectional** (Figure 7) reached **94.1620%**. This notable improvement underlines the importance of bidirectional processing, enabling the model to leverage information from both past and future temporal contexts to form a richer representation.
- **LSTM-Bidirectional-Seq2Seq** (Figure 8) outperformed all other LSTM models with **95.0921%** accuracy. This configuration harnesses the strengths of both bidirectionality and a seq2seq architecture, indicating that when combined, these elements can more effectively model complex temporal dependencies.

5.3 Analysis of GRU Variants

Figures 9 to 13 show that GRU-based models generally trail behind their LSTM counterparts. The simpler gating mechanism of GRUs, while computationally appealing, may not be as adept at handling more intricate long-term dependencies present in the data.

- **GRU** (Figure 9) obtained **84.1694%** accuracy. While GRUs can capture basic temporal correlations, the baseline performance indicates limitations in handling more complex patterns.
- **GRU-2Path** (Figure 10) marginally decreased to **83.7746%**, suggesting that the two-path strategy did not yield the same gains observed in the LSTM variant. The simpler GRU gating may not fully leverage parallel information streams as effectively as LSTM cells.
- **GRU-Seq2Seq** (Figure 11) improved significantly to **90.8855%**. The seq2seq structure appears beneficial here, likely because it provides a more flexible encoding-decoding dynamic that helps the GRU model better capture temporal dependencies.
- **GRU-Bidirectional** (Figure 12) achieved **87.0331%**, demonstrating that bidirectionality does provide a performance lift, albeit a more modest one than in LSTM models.

- **GRU-Bidirectional-Seq2Seq** (Figure 13) yielded **87.3672%**, a slight improvement over the bidirectional-only variant. While beneficial, the combination of bidirectionality and seq2seq for GRUs does not approach the accuracy gains seen with LSTMs, indicating that the intrinsic memory mechanisms of LSTMs may be better suited to exploiting these architectural enhancements.

5.4 Attention Mechanism

The **Attention** model (Figure 14) attained the highest overall accuracy of **95.1467%**. Unlike RNN-based models that rely on iterative hidden-state updates, attention mechanisms enable the model to directly “focus” on the most relevant portions of the input sequence. This ability to weigh and highlight critical time steps dynamically results in a stronger handling of long-range dependencies and subtle variations that may be difficult for standard recurrent architectures to capture adequately.

5.5 Comparative Analysis and Reasoning

- **LSTM vs. GRU:** Across all configurations, LSTM-based models outperform their GRU counterparts. The richer gating mechanism of LSTMs seems better suited to extracting and preserving critical information over long sequences. As tasks become more complex, the LSTM’s capacity to handle intricate dependencies provides a tangible advantage.
- **Bidirectionality:** Introducing bidirectionality consistently enhances performance for both LSTM and GRU models. By processing input sequences in both forward and reverse temporal directions, the model obtains a more holistic view of the data. This expanded context is particularly valuable in tasks where future observations can inform the interpretation of earlier events.
- **Sequence-to-Sequence Architectures:** The seq2seq configurations yield mixed results. While LSTM-Bidirectional-Seq2Seq shows a notable improvement, the LSTM-Seq2Seq alone does not surpass the baseline. For GRUs, seq2seq integration markedly improves accuracy. These mixed outcomes suggest that seq2seq frameworks are beneficial primarily when combined with other architectural enhancements (e.g., bidirectionality) or when the base model can exploit the encoder-decoder paradigm effectively.
- **Attention:** The attention-based model emerges as the top performer. Its ability to directly attend to relevant parts of the input, rather than relying solely on hidden-state propagation, empowers it to capture complex, non-linear relationships. This proves especially valuable in datasets where essential cues are scattered over long time spans, and maintaining a single vector representation of the entire past may lead to information loss.

6 Conclusion

The results highlight the intricate nature of sequence modeling and the importance of selecting appropriate architectures to effectively capture temporal dynamics. While recurrent networks such as LSTMs and GRUs demonstrate a solid baseline

for modeling sequential data, their performance can be significantly elevated through enhancements like bidirectionality, sequence-to-sequence (Seq2Seq) frameworks, and attention mechanisms. Bidirectional processing enriches the model’s contextual understanding by incorporating information from both past and future time steps. Seq2Seq architectures introduce flexibility by enabling the model to handle variable-length input and output sequences effectively. However, the most notable improvement arises from the incorporation of attention mechanisms, which allow the model to dynamically focus on the most relevant parts of the input sequence. This capability enables attention-based models to efficiently capture both short- and long-term dependencies, leading to state-of-the-art performance. Overall, these findings underscore that augmenting recurrent networks with advanced architectural features is essential for decoding complex temporal patterns. The attention mechanism, in particular, stands out as a transformative element, offering a powerful means to overcome the limitations of traditional RNNs and unlocking superior predictive capabilities for tasks like stock price prediction. Future work could explore hybrid models that integrate attention mechanisms with graph neural networks (GNNs) or transformers with domain-specific embeddings. Additionally, deploying these models in real-time trading environments and applying them to diverse datasets, including multi-asset portfolios, could further validate their effectiveness and generalizability.

References

- [1] Lin, C.Y., Lobo Marques, J.A.: Stock market prediction using artificial intelligence: A systematic review of systematic reviews. *Social Sciences Humanities Open* **9**, 100864 (2024) <https://doi.org/10.1016/j.ssaho.2024.100864>
- [2] Ariyo, A.A., Adewumi, A.O., Ayo, C.K.: Stock price prediction using the arima model. In: 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp. 106–112 (2014). <https://doi.org/10.1109/UKSim.2014.67>
- [3] Bhuriya, D., Kaushal, G., Sharma, A., Singh, U.: Stock market predication using a linear regression. In: 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA), vol. 2, pp. 510–513 (2017). <https://doi.org/10.1109/ICECA.2017.8212716>
- [4] Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., Chouhan, L.: Stock market prediction using machine learning. In: 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), pp. 574–576 (2018). IEEE
- [5] Khan, W., Ghazanfar, M.A., Azam, M.A., Karami, A., Alyoubi, K.H., Alfakeeh, A.S.: Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 1–24 (2022)
- [6] Mankar, T., Hotchandani, T., Madhwani, M., Chidrawar, A., Lifna, C.: Stock

- market prediction based on social sentiments using machine learning. In: 2018 International Conference on Smart City and Emerging Technology (ICSCET), pp. 1–3 (2018). IEEE
- [7] Rao, P.S., Srinivas, K., Mohan, A.K.: A survey on stock market prediction using machine learning techniques. In: ICDSMLA 2019: Proceedings of the 1st International Conference on Data Science, Machine Learning and Applications, pp. 923–931 (2020). Springer
 - [8] Nikou, M., Mansourfar, G., Bagherzadeh, J.: Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management* **26**(4), 164–174 (2019)
 - [9] Selvamuthu, D., Kumar, V., Mishra, A.: Indian stock market prediction using artificial neural networks on tick data. *Financial Innovation* **5**(1), 1–12 (2019)
 - [10] Fang, W.-X., Lan, P.-C., Lin, W.-R., Chang, H.-C., Chang, H.-Y., Wang, Y.-H.: Combine facebook prophet and lstm with bpnn forecasting financial markets: the morgan taiwan index. In: 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pp. 1–2 (2019). IEEE
 - [11] Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., Anastasiu, D.C.: Stock price prediction using news sentiment analysis. In: 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), pp. 205–208 (2019). IEEE
 - [12] Vignesh, C.: Applying machine learning models in stock market prediction. *IRJET, Journal* (2018)
 - [13] Sirimevan, N., Mamalgaha, I., Jayasekara, C., Mayuran, Y., Jayawardena, C.: Stock market prediction using machine learning techniques. In: 2019 International Conference on Advancements in Computing (ICAC), pp. 192–197 (2019). IEEE
 - [14] Moghar, A., Hamiche, M.: Stock market prediction using lstm recurrent neural network. *Procedia computer science* **170**, 1168–1173 (2020)
 - [15] Eapen, J., Bein, D., Verma, A.: Novel deep learning model with cnn and bi-directional lstm for improved stock market index prediction. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0264–0270 (2019). IEEE
 - [16] Aguirre, H.: Proceedings of the Genetic and Evolutionary Computation Conference. ACM, ??? (2018)
 - [17] Khan, W., Malik, U., Ghazanfar, M.A., Azam, M.A., Alyoubi, K.H., Alfakeeh, A.S.: Predicting stock market trends using machine learning algorithms via public

- sentiment and political situation analysis. *Soft Computing* **24**(15), 11019–11043 (2020)
- [18] Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S., Mosavi, A.: Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis. *Ieee Access* **8**, 150199–150212 (2020)
 - [19] Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997) <https://doi.org/10.1162/neco.1997.9.8.1735>
 - [20] GeeksforGeeks: What is LSTM - long short term memory? GeeksforGeeks (2024). <https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/>
 - [21] Cho, K.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014)
 - [22] Huang, Z., Yang, F., Xu, F., Song, X., Tsui, K.-L.: Convolutional gated recurrent unit-recurrent neural network for state-of-charge estimation of lithium-ion batteries. *Ieee Access* **7**, 93139–93149 (2019)
 - [23] Vaswani, A.: Attention is all you need. *Advances in Neural Information Processing Systems* (2017)

7 Appendix

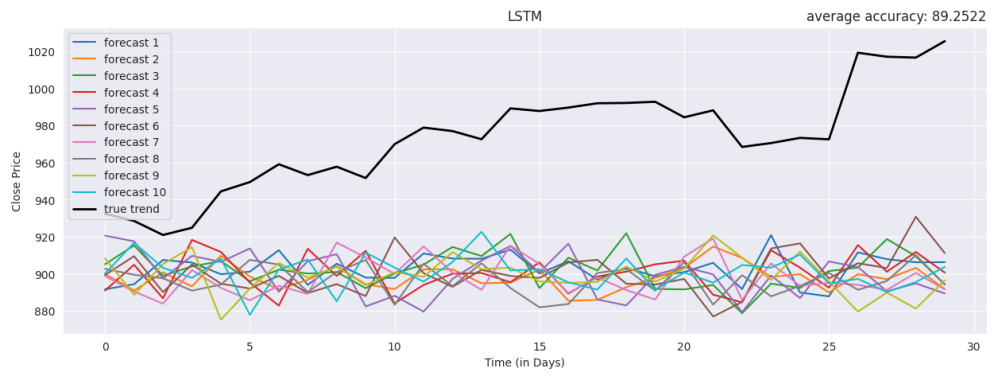


Fig. 4: LSTM

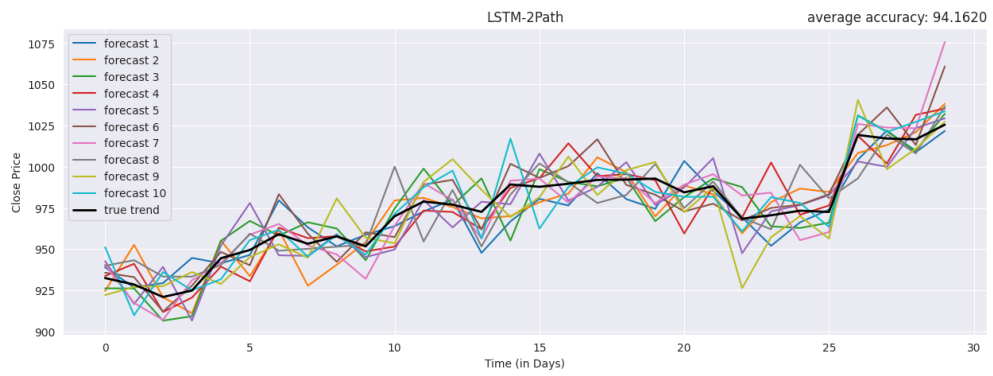


Fig. 5: LSTM-2 Path

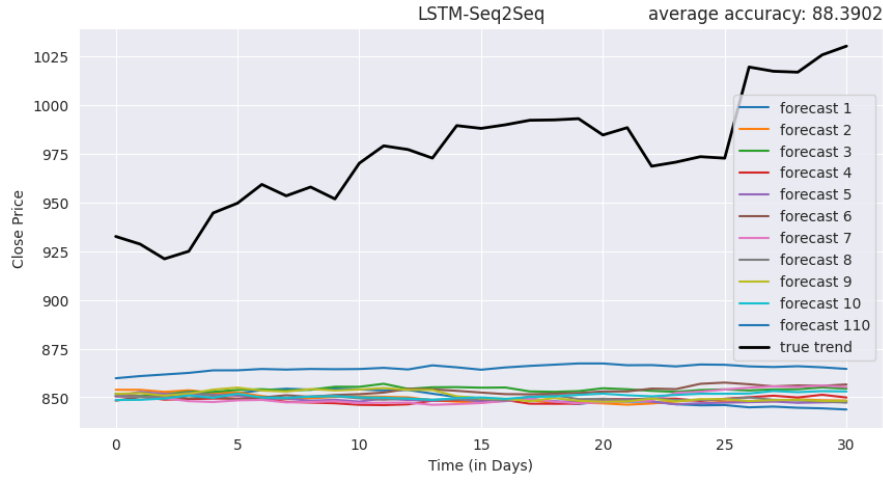


Fig. 6: LSTM-Seq2Seq

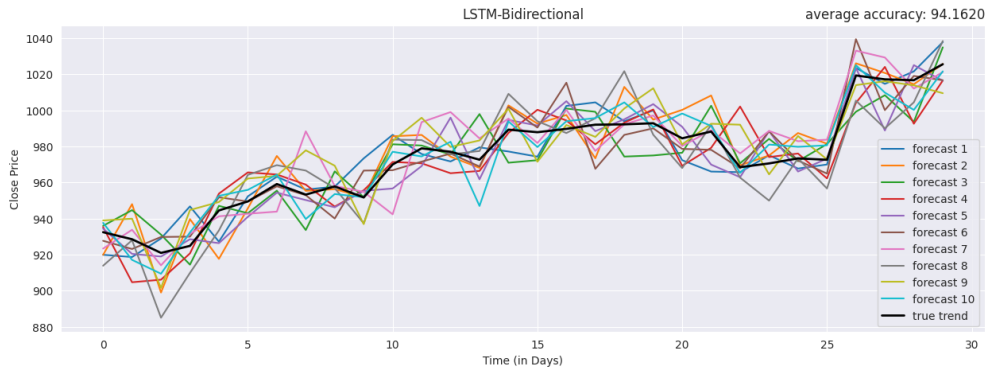


Fig. 7: LSTM-Bidirectional

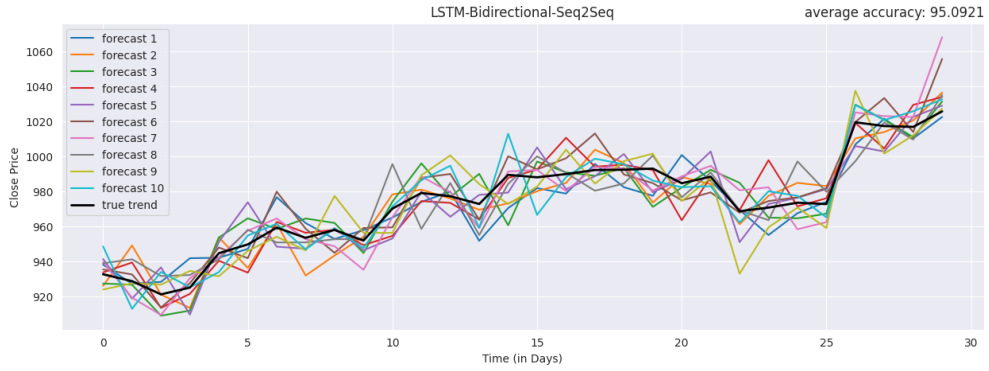


Fig. 8: LSTM-Bidirectional-Seq2Seq

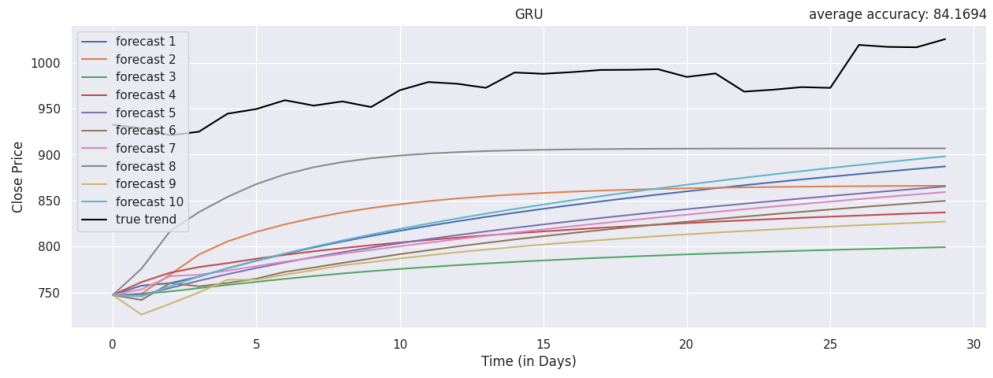


Fig. 9: GRU

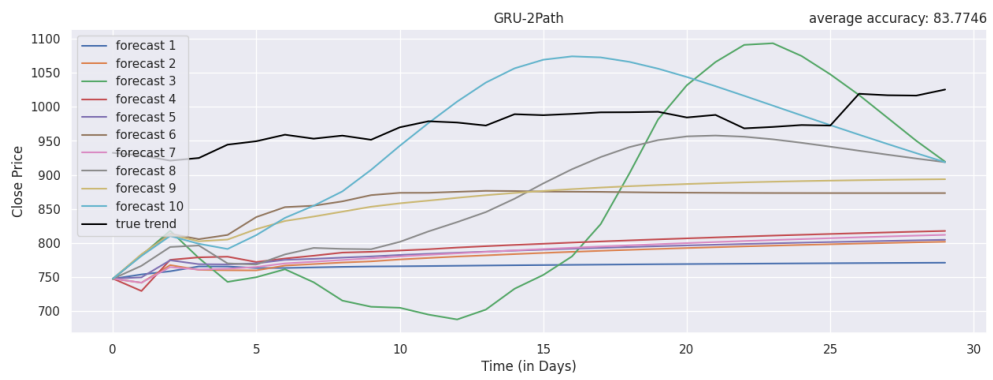


Fig. 10: GRU-2Path

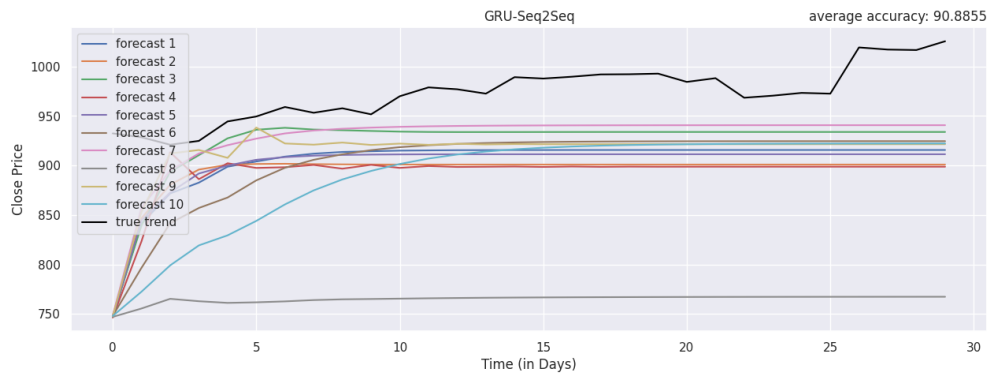


Fig. 11: GRU-Seq2Seq

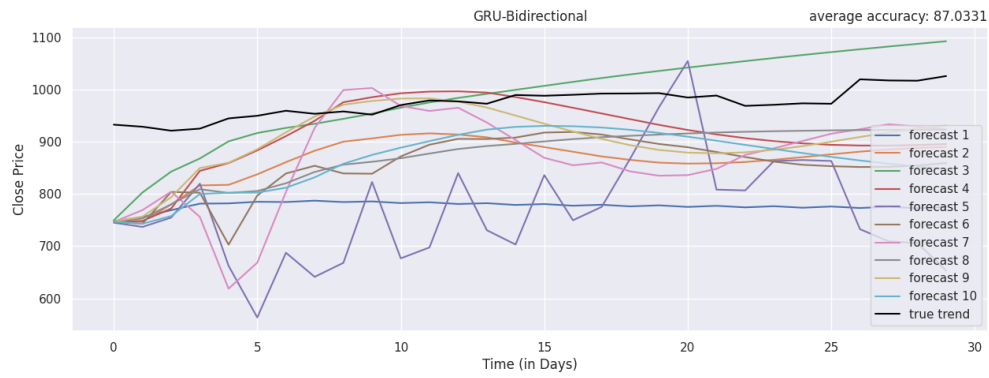


Fig. 12: GRU-Bidirectional

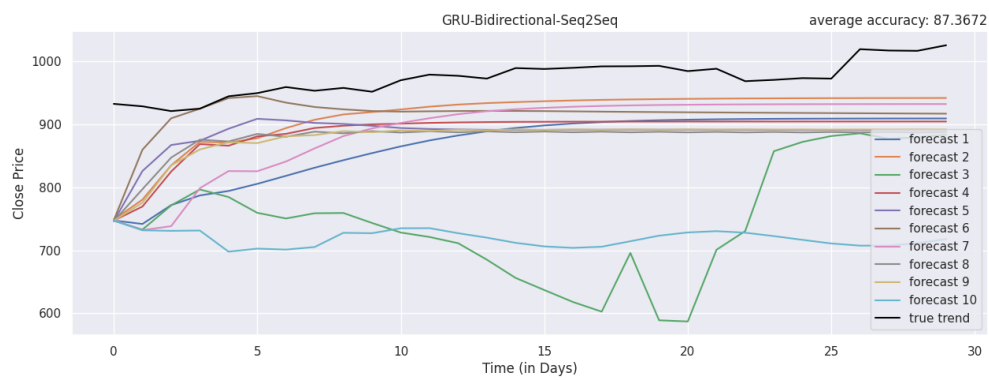


Fig. 13: GRU-Bidirectional-Seq2Seq

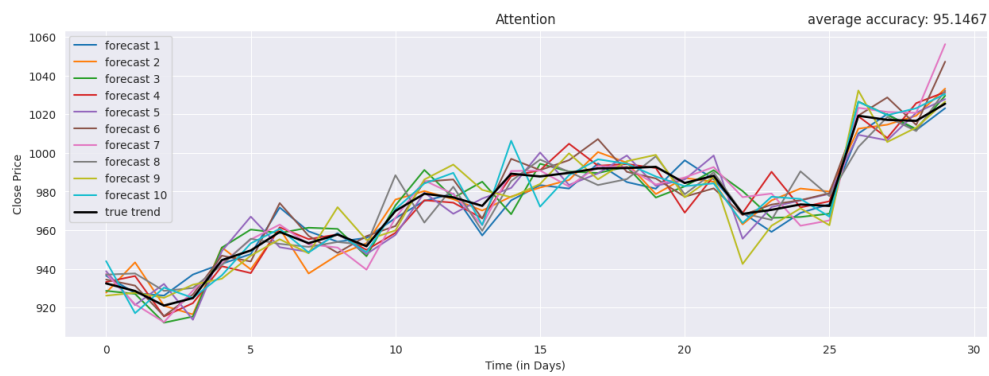


Fig. 14: Attention Mechanism Visualization