Forecasting Apple Stock Prices Using LSTM, GRU, and Transformer Models: A Deep Learning Approach to Digital Market Prediction

Bentar Priyopradono[1,*], Dian Oktari[2]

[1]Teknik, Informatika, Univeritas Prof. Dr. Hazairin, SH, Bengkulu, Indonesia
[2]Faculty of Engineering, Universitas Prof. Dr. Hazairin, S.H., Indonesia

ABSTRACT

The rapid growth and volatility of digital financial markets have increased the demand for accurate, data-driven forecasting methods capable of handling complex temporal and nonlinear relationships. This study investigates the predictive performance of three deep learning architectures Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer Encoder in forecasting Apple Inc.'s stock prices. Using historical data spanning 1980 to 2025, each model was trained on 60-day sequences to predict the next trading day's closing price. The dataset was normalized using the MinMaxScaler and divided into 80 percent for training and 20 percent for testing. All models were trained for 20 epochs with a batch size of 32, employing the Mean Squared Error (MSE) loss function and the Adam optimizer. The results show that the GRU model achieved the best predictive accuracy, with an RMSE of 4.8281 and an MAE of 3.5578, followed by the LSTM model with an RMSE of 8.8865 and an MAE of 6.0122. The Transformer model performed less effectively, indicating its reliance on larger and more complex datasets to optimize its self-attention mechanism. Overall, the findings demonstrate that recurrent neural network-based models remain highly effective for stock price forecasting in digital financial environments. The study concludes that deep learning, particularly the GRU architecture, provides a powerful and computationally efficient framework for real-time prediction and decision-making in digital market analysis and algorithmic trading applications.

Keywords Deep Learning, Stock Price Forecasting, Gated Recurrent Unit (GRU), Digital Financial Markets, Transformer Model

INTRODUCTION

The rapid evolution of digital technologies has profoundly transformed global financial systems, giving rise to complex and interconnected digital markets. In this environment, stock price movements are influenced not only by traditional economic factors but also by high-frequency trading, social sentiment, and algorithmic decision-making. These elements introduce nonlinearities and uncertainties that make accurate forecasting a challenging yet essential task for investors and policymakers. Traditional statistical methods such as Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) have long been used for time-series forecasting. However, their underlying linear assumptions and limited ability to capture nonlinear temporal dependencies restrict their effectiveness in modeling the intricate behaviors of modern financial markets [1]. As digital markets continue to expand, there is an increasing need for more advanced analytical frameworks capable of learning from large, dynamic, and

Submitted 15 April 2026
Accepted 20 June 2026
Published 27 October 2026

Corresponding author
Bentar Priyopradono,
bentarpriyopradono@gmail.com

Additional Information and Declarations can be found on page 277

DOI: 10.63913/ftij.v2i4.39

Copyright
2026 Priyopradono and Oktari

Distributed under
Creative Commons CC-BY 4.0

How to cite this article: B. Priyopradono and D. Oktari, "Forecasting Apple Stock Prices Using LSTM, GRU, and Transformer Models: A Deep Learning Approach to Digital Market Prediction," *FinTech Innov. J.*, vol. 2, no. 4, pp. 263-279, 2026.

non-stationary data.

In recent years, artificial intelligence (AI) and deep learning (DL) have emerged as powerful alternatives to conventional econometric models in financial forecasting. Deep learning algorithms are particularly effective at uncovering complex patterns in sequential data, enabling them to learn and predict nonlinear relationships that traditional models fail to identify. Among these, recurrent neural networks (RNNs) and their advanced variants Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have gained significant attention for their capability to capture long-term dependencies within time-series data [2]. The LSTM model, introduced by Hochreiter and Schmidhuber, incorporates memory cells and gating mechanisms that allow it to retain relevant information over long sequences. The GRU model, proposed by Cho et al., simplifies this structure while maintaining comparable performance, making it computationally more efficient. These models have been successfully applied to various financial prediction tasks, including stock price forecasting, volatility estimation, and cryptocurrency analysis, demonstrating their adaptability to dynamic digital market environments [3].

More recently, Transformer-based architectures have attracted considerable interest in the financial forecasting domain due to their use of self-attention mechanisms, which allow the model to learn long-range dependencies without the need for sequential processing. Originally developed for natural language processing, Transformers have achieved state-of-the-art results in numerous sequential learning applications. In finance, researchers have begun exploring their potential for time-series prediction, sentiment-informed forecasting, and multi-asset market modeling [4]. Some studies report that attention-based models can outperform recurrent structures when sufficient data are available. However, other research suggests that Transformers require larger and more complex datasets to achieve optimal performance, as their attention mechanisms are computationally demanding and data-intensive. These mixed findings highlight the need for further empirical evaluation of Transformers in comparison to established RNN-based models such as LSTM and GRU, particularly in real-world financial contexts.

Despite the rapid development of deep learning methods in financial analytics, a clear research gap remains in systematically comparing different deep learning architectures using consistent experimental frameworks and long-term financial datasets. Most existing studies either focus on short-term predictions or apply a single model without examining the comparative advantages and limitations of multiple architectures. Moreover, there is limited research applying these methods to extended historical datasets that capture both pre-digital and digital market conditions [5]. Apple Inc., as one of the world's most influential technology companies, provides an ideal case for such an investigation. Its long-term stock data reflect shifts in market behavior, investor sentiment, and technological innovation, making it a comprehensive representation of both traditional and digital-era financial dynamics. Examining how various deep learning models perform on this dataset offers valuable insights into their adaptability and robustness across different temporal and market conditions.

The main objective of this study is to evaluate and compare the predictive performance of three deep learning architectures LSTM, GRU, and Transformer Encoder in forecasting Apple's stock prices within the context of digital financial

markets. By employing a standardized data preprocessing pipeline, consistent training procedures, and uniform evaluation metrics, this research aims to provide a fair and empirical comparison of model performance. The study contributes to the state of the art in financial forecasting by addressing a significant empirical gap and offering evidence-based guidance on selecting appropriate architectures for different data conditions. Furthermore, it provides practical implications for the design of AI-driven decision support systems in algorithmic trading, investment management, and digital financial analytics. The findings are expected to deepen the understanding of how deep learning models can be optimized for real-time forecasting and enhance strategic decision-making in the increasingly complex landscape of digital financial markets.

Literature Review and Related Works

Accurate stock price forecasting remains one of the most challenging problems in the field of financial analytics due to the nonlinear and highly volatile nature of market data. Traditional time-series models such as ARIMA and GARCH have been widely used to analyze financial data, yet their linear assumptions limit their ability to capture the dynamic relationships inherent in complex digital markets. With the rise of artificial intelligence (AI) and data-driven trading, researchers have increasingly adopted deep learning (DL) architectures for financial forecasting tasks, demonstrating their superiority in identifying temporal dependencies and nonlinear interactions within large datasets [6]. Deep learning techniques, particularly Recurrent Neural Networks (RNNs) and their advanced variants, have shown great promise in modeling financial time series characterized by sequential dependencies and noise [7].

Among these architectures, the Long Short-Term Memory (LSTM) network has emerged as a robust approach for sequential data modeling. It addresses the vanishing gradient problem by incorporating memory cells and gating mechanisms that enable it to retain long-term dependencies [8]. Multiple studies have shown that LSTM models outperform conventional machine learning and statistical methods in predicting both short- and long-term market movements [9], [10]. Variations such as stacked LSTM and bidirectional LSTM have further improved prediction accuracy by enhancing the network's ability to capture both past and future information within financial sequences [11]. The Gated Recurrent Unit (GRU), a simplified derivative of LSTM, has gained attention for offering comparable performance with reduced computational complexity. Research comparing GRU and LSTM has found GRU to be more efficient in training while maintaining high forecasting precision, particularly for volatile assets and high-frequency market data [12], [13].

Recent advancements have also focused on hybrid deep learning frameworks that combine CNN, LSTM, and GRU models to integrate spatial and temporal feature extraction. These hybrid architectures have demonstrated superior predictive accuracy compared to single-network models, particularly when applied to complex financial datasets such as stock indices and technology-sector equities [14], [15]. Studies have also highlighted the importance of data preprocessing techniques, such as clustering and normalization, which can significantly enhance the stability and convergence of deep learning models [16]. Beyond conventional architectures, ensemble methods and multimodal approaches that combine numerical and textual data have shown promise in

improving prediction reliability in digital financial environments [17].

The emergence of Transformer-based architectures has marked a significant shift in time-series forecasting research. Unlike RNN-based models that process data sequentially, Transformers employ a self-attention mechanism that captures long-range dependencies across entire sequences simultaneously, improving scalability and parallelism [18]. Recent studies have applied Transformers and hybrid LSTM–Transformer networks to financial forecasting, demonstrating strong potential for improved accuracy when sufficient data and computational resources are available. However, empirical evidence also suggests that Transformers are highly data-dependent and may underperform on small or univariate datasets where sequential dependencies are best captured through recurrent architectures [19]. Hybrid models integrating Transformers with LSTM or MLP layers have emerged as promising alternatives, balancing the strengths of both recurrence and attention mechanisms [20].

Overall, existing studies demonstrate that deep learning models, especially recurrent and hybrid architectures, outperform traditional econometric approaches in stock price prediction. Nevertheless, gaps remain in the literature regarding standardized model comparison across architectures under consistent experimental conditions and on large, long-term datasets. Addressing this limitation is essential to advancing the understanding of how deep learning can be effectively deployed for digital market forecasting, particularly for assets with long historical records such as Apple Inc. stock. The present study builds upon this foundation by conducting a comparative analysis of LSTM, GRU, and Transformer models, thereby contributing empirical insights into the strengths and trade-offs of different deep learning architectures for financial time-series forecasting.

Methodology

This study adopts a quantitative experimental approach to evaluate and compare three deep learning architectures Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer Encoder for forecasting Apple Inc.'s stock prices in the digital market context. The methodological process consists of five sequential stages: data acquisition, data preprocessing, model architecture design, model training and evaluation, and experimental validation. These stages are illustrated in figure 1, which presents the overall research workflow, including the flow from raw data collection to model evaluation and interpretation. The figure summarizes the systematic procedure used to ensure methodological consistency, reproducibility, and fairness across all models. Each model was trained and evaluated under identical experimental conditions to ensure objective comparison and to minimize bias arising from differences in preprocessing, hyperparameters, or data segmentation.

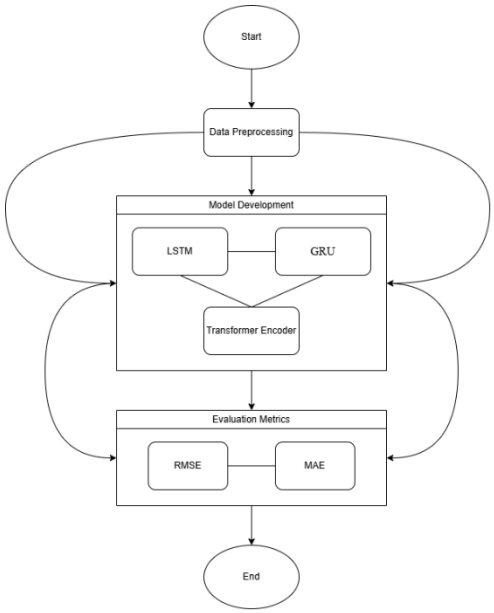


Figure 1 Research Steps

The dataset used in this research consists of daily historical data of Apple Inc. (AAPL) stock prices from December 12, 1980, to January 2025, obtained from public financial databases. Each entry includes the opening, highest, lowest, closing, and adjusted closing prices, as well as trading volume. The closing price was selected as the target variable because it is a widely accepted indicator of market performance and investor sentiment. The dataset covers more than four decades of trading activity, capturing significant economic and technological transitions that influence digital market behavior. This long temporal coverage enables the evaluation of the models' ability to learn from both long-term trends and short-term fluctuations while capturing the nonlinear dynamics of financial markets.

Before model training, data preprocessing was conducted to ensure the quality, consistency, and interpretability of the dataset. The data were chronologically sorted and examined for missing or duplicate entries, although no significant anomalies were found. The closing price series was normalized to a range between 0 and 1 using the MinMaxScaler transformation, improving numerical stability and accelerating the convergence of deep learning models. The normalization formula is given by:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Xdenotes the original data value, $X_{min}$and $X_{max}$are the minimum and maximum observed values, respectively, and X_{norm}represents the normalized output. This transformation ensures that all features contribute proportionally to model learning. A sliding time window of 60 trading days was used to create supervised input-output pairs, where the past 60 closing prices predict the next day's price. The dataset was then divided into 80 percent for training and 20 percent for testing, ensuring that the testing data represent unseen future trends and prevent data leakage.

The three deep learning architectures were designed with comparable depth and complexity to ensure fair comparison. The LSTM model employs memory cells and gating mechanisms to manage information flow through time. Its mathematical formulation is expressed as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \widetilde{C_t} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2)$$
$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t}, o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), h_t = o_t * \tanh(C_t)$$

$x_t$is the input vector, $h_t$represents the hidden state, $C_t$is the cell state, and $W$and $b$are the weight and bias parameters.

The GRU model simplifies the LSTM's internal structure by merging the input and forget gates into a single update gate, expressed as:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]), r_t = \sigma(W_r \cdot [h_{t-1}, x_t]), \widetilde{h_t} = \tanh(W_h \cdot [r_t * h_{t-1}, x_t]), h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h_t} \quad (3)$$

This formulation allows GRU to balance short- and long-term dependencies effectively while reducing computational cost.

The Transformer Encoder model applies a self-attention mechanism that learns contextual relationships among all time steps simultaneously. The self-attention operation is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

Q, K, and Vdenote the query, key, and value matrices, and $d_k$represents the dimension of the key vector. This mechanism enables the model to assign varying importance to different time steps, capturing long-range dependencies within the financial sequence.

The models were implemented using TensorFlow and Keras in Python 3.10, trained for 20 epochs with a batch size of 32. The Adam optimizer was employed due to its adaptive learning rate mechanism, ensuring stable and efficient convergence. The Mean Squared Error (MSE) was chosen as the loss function, mathematically expressed as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \quad (5)$$

$y_i$and $\hat{y_i}$are the actual and predicted stock prices, respectively, and $n$is the total number of samples. The model performance was quantitatively evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}, \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}| \quad (6)$$

RMSE emphasizes large deviations by squaring the errors, while MAE provides an intuitive measure of average error magnitude in actual price units. Together, these metrics offer a balanced evaluation of model precision and robustness.

All experiments were conducted on a workstation equipped with an NVIDIA GPU to accelerate matrix computations and reduce training time. Early stopping was used to prevent overfitting by monitoring validation loss, and model checkpoints were saved to preserve the best-performing model weights. Visual analyses of predicted versus actual prices were also generated to evaluate how closely the models followed market trends and volatility patterns. The consistent application of preprocessing steps, parameter tuning, and performance metrics ensures a fair and reproducible comparison across all deep learning architectures evaluated in this study.

Algorithm 1 Deep Learning-Based Stock Price Forecasting

Input: Historical stock price data $D = \{x_1, x_2, \ldots, x_T\}$, window size $w = 60$, learning rate η, epochs $E = 20$, batch size $B = 32$

Output: Predicted next-day closing price $\hat{y}_{t+1}$and performance metrics (MSE, RMSE, MAE)

Process:

Start

Data Normalization:

For each data point $x_t \in D$:

$$x_t^{norm} = \frac{x_t - X_{min}}{X_{max} - X_{min}}$$

Sliding Window Generation:

Construct input-output pairs:

$$X = [x_i^{norm}, x_{i+1}^{norm}, \ldots, x_{i+w-1}^{norm}]$$
$$Y = x_{i+w}^{norm}$$

where $i = 1,2,\ldots,T-w$.

Data Splitting:

Divide data into training (80%) and testing (20%) subsets:

$$(X_{train}, Y_{train}), (X_{test}, Y_{test}) = \text{split}(X, Y)$$

Model Initialization:

Select model $M \in$ {LSTM,GRU,Transformer} and initialize trainable parameters:

$$\Theta_M = \{W, b\}$$

Model Training:

For each epoch $e \in [1, E]$:

$$\hat{Y}_{train} = f_M(X_{train}; \Theta_M)$$
$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$
$$\Theta_M = \Theta_M - \eta \nabla_{\Theta_M} L$$

Model Evaluation:

Predict test data:

$$\hat{Y}_{test} = f_M(X_{test}; \Theta_M^*)$$

Compute performance metrics:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i |$

Model Selection:
Select best-performing model based on minimal RMSE:

$$M^* = \arg\min_M (\text{RMSE}_M)$$

Visualization:
Plot actual vs. predicted pricesand display comparative RMSE and MAE results.
End

Results

The preliminary analysis aimed to examine the overall behavior and historical evolution of Apple's stock prices to identify major patterns that could influence model selection and forecasting performance. Figure 2 presents the trend of Apple's closing stock price from 1980 to 2025, showing a persistent upward trajectory over time. The data reveal an extended phase of relatively moderate price movement between 1980 and the early 2000s, followed by a sharp and sustained increase beginning around 2005. This inflection point aligns with Apple's major innovations and product launches, including the introduction of the iPhone and the expansion of its digital ecosystem. The steady growth in stock price reflects the company's increasing global market presence and technological dominance, which have made Apple one of the most influential players in the digital economy.



Figure 2 Apple Stock Closing Price Over Time

In addition to the overall growth pattern, the analysis reveals a notable increase in price volatility in the more recent period, especially after 2015. This heightened fluctuation indicates stronger market reactions to economic events, investor sentiment, and digital financial activities, suggesting that Apple's stock price has become more sensitive to rapid changes in global market conditions. The presence of such volatility also implies nonlinear and time-dependent relationships within the data that are not easily captured by traditional econometric models. Therefore, applying deep learning architectures becomes a logical and necessary approach. These models are designed to handle complex temporal dependencies and nonlinear dynamics, making them more effective for understanding and predicting stock price movements in the fast-evolving landscape of digital financial markets.

To assess the predictive capability of deep learning models in forecasting stock prices, three architectures were implemented: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and a Transformer Encoder. Each model was trained using Apple's historical closing price data, where a 60-day sliding time window was applied to predict the closing price on the following trading day. This temporal segmentation allowed the models to learn sequential dependencies within the data. Before training, the dataset was randomly divided into two subsets, consisting of 80 percent for training and 20 percent for testing, ensuring that the models could generalize beyond the training samples. The input values were normalized between zero and one using the MinMaxScaler to reduce numerical instability and improve convergence speed. All models were trained for 20 epochs with a batch size of 32 to balance computational efficiency and learning accuracy. The optimization process employed the Adam optimizer, which adaptively adjusted learning rates to enhance convergence, while the Mean Squared Error (MSE) function was used as the loss metric to minimize the average squared difference between predicted and actual prices. This methodological framework ensured that each deep learning model was trained under consistent and reproducible conditions, enabling a fair and reliable comparison of their forecasting performance.

The predictive performance of the models was evaluated using two widely recognized statistical error metrics, namely Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). These metrics were chosen because they provide complementary perspectives on model accuracy. RMSE measures the square root of the average squared differences between predicted and actual values, making it highly sensitive to large prediction errors. MAE, on the other hand, calculates the average of the absolute differences, providing a more interpretable measure of the average deviation in actual price units. Together, these metrics allow for a balanced evaluation of each model's forecasting precision and robustness. Table 1 summarizes the results obtained from the testing dataset for all three deep learning architectures under identical experimental conditions.

Table 1 Performance Comparison of LSTM, GRU, and Transformer Models

Model	RMSE	MAE
LSTM	8.8865	6.0122
GRU	4.8281	3.5578
Transformer	31.6977	23.4890

The results indicate that the GRU model achieved the highest predictive accuracy, with an RMSE of 4.8281 and an MAE of 3.5578, outperforming both the LSTM and Transformer models. The LSTM model produced an RMSE of 8.8865 and an MAE of 6.0122, demonstrating moderate accuracy but slightly higher sensitivity to short-term price fluctuations. The Transformer model showed the weakest performance, with an RMSE of 31.6977 and an MAE of 23.4890, indicating that it struggled to capture the temporal relationships within this relatively small and univariate dataset. These results suggest that models based on recurrent neural networks, particularly GRU, are more effective in learning the nonlinear and sequential patterns present in stock price movements compared to attention-based architectures when limited data are available.

The LSTM model demonstrated a strong ability to capture the dominant

temporal patterns in Apple's stock price data, accurately reflecting both the overall upward trajectory and the cyclical fluctuations present in the time series. As shown in figure 3, the predicted values generated by the LSTM model closely followed the actual market trends over the testing period, indicating that the model effectively learned long-term dependencies across multiple sequential observations. This capability stems from the LSTM architecture's memory cell mechanism, which allows the model to retain relevant information from earlier time steps while filtering out less significant signals. The relatively smooth prediction curve produced by the LSTM suggests that the model performs well in recognizing sustained market momentum and broader economic trends, characteristics that are essential for long-horizon financial forecasting.

Despite these strengths, the LSTM model showed some limitations in capturing short-term price volatility, particularly during periods of sharp upward or downward movements. The lagging behavior observed during these rapid fluctuations indicates that while the LSTM can generalize well over long temporal intervals, it tends to respond slowly to sudden shifts in market sentiment or unexpected financial events. The computed RMSE of 8.89 reflects a moderate prediction error, signifying satisfactory overall accuracy, yet there remains room for improvement. Enhancing the model's responsiveness could be achieved through more extensive hyperparameter tuning, such as adjusting the number of hidden units, learning rate, or sequence length. Additionally, incorporating more complex architectures or external market indicators, including trading volume or sentiment-based variables, may help improve short-term predictive precision and further strengthen the LSTM model's adaptability in dynamic financial environments.

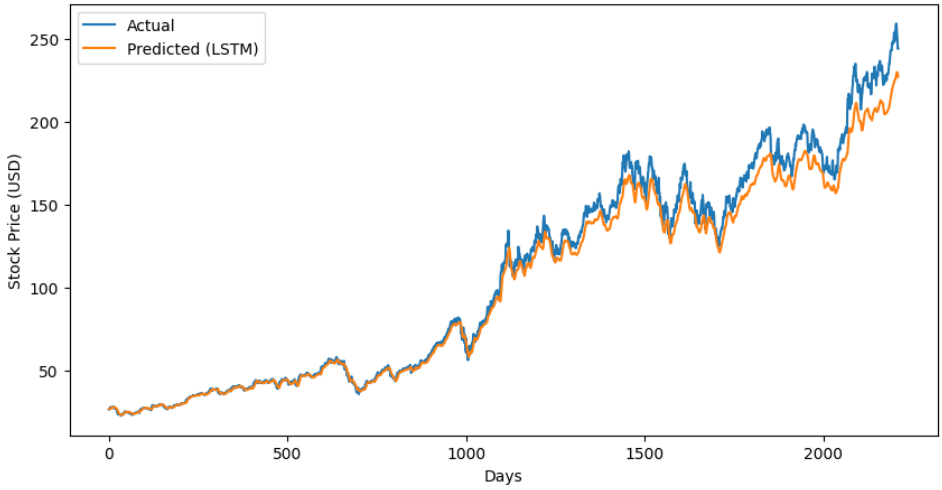


Figure 3 LSTM Model Prediction vs Actual Stock Prices

The GRU model exhibited the strongest predictive capability among all models evaluated in this study, achieving both higher accuracy and greater stability across different market conditions. As illustrated in figure 4, the GRU's predicted stock prices closely aligned with the actual historical prices throughout the testing period, including intervals of heightened market volatility. This close alignment indicates that the GRU effectively captured the nonlinear temporal dependencies inherent in financial time series data. The architecture's simplified design, which integrates update and reset gates, enables efficient information

retention over multiple time steps without the risk of overfitting or gradient vanishing. As a result, the GRU successfully balanced model complexity with computational efficiency, allowing it to generalize well across varying market dynamics while maintaining high predictive precision.

In addition to its accuracy, the GRU demonstrated superior convergence behavior during the training process, reaching an optimal solution faster than both the LSTM and Transformer models. The faster convergence can be attributed to the GRU's compact structure, which reduces the number of trainable parameters while preserving the ability to model long-term dependencies. This efficiency made the model more robust against noise and sudden fluctuations in the data, characteristics that are common in stock market movements. The GRU's lower RMSE and MAE values further validate its effectiveness in managing both gradual and abrupt price changes. These findings highlight the GRU model as a particularly well-suited architecture for digital financial forecasting applications, where market data are often nonlinear, volatile, and influenced by multiple interacting economic factors.

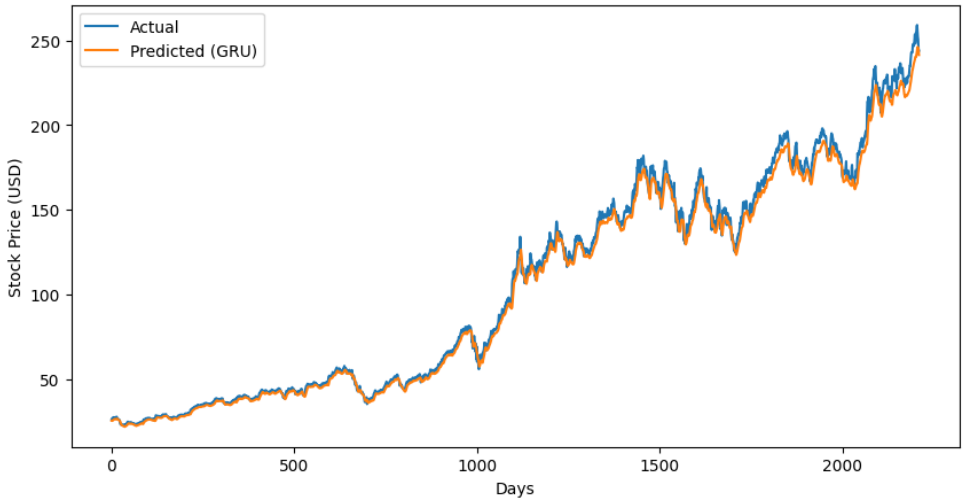


Figure 4 GRU Model Prediction vs Actual Stock Prices

In contrast to the recurrent neural network-based models, the Transformer architecture produced less accurate predictions when applied to Apple's stock price data. As shown in figure 5, the Transformer model's predicted values diverged considerably from the actual prices, particularly during periods of sharp upward trends and sudden fluctuations in market activity. The model struggled to follow rapid transitions in the time series, resulting in an overall weaker fit. This performance is reflected in its evaluation metrics, with an RMSE of 31.70 and an MAE of 23.49, both of which are substantially higher than those achieved by the LSTM and GRU models. These results indicate that the Transformer was unable to generalize effectively from the available data, likely due to the limited size and univariate nature of the dataset. The self-attention mechanism, which is the core component of the Transformer, requires large amounts of training data to identify and learn meaningful relationships across time steps. When data are limited or lack diversity, the model tends to overfit to specific patterns and loses its ability to make reliable predictions on unseen data.

Despite its relatively weak performance in this experiment, the Transformer

model should not be dismissed as ineffective for financial forecasting tasks. Its architectural design offers significant advantages for capturing long-range dependencies and modeling interactions between multiple variables, especially when applied to complex and high-dimensional datasets. Future research could explore the potential of the Transformer by incorporating additional input features such as trading volume, volatility indices, macroeconomic indicators, or sentiment data derived from digital news and social media sources. Such enhancements could allow the self-attention mechanism to exploit a richer contextual understanding of market behavior, improving predictive accuracy and robustness. Moreover, applying more advanced optimization techniques, including fine-tuned learning rates and larger batch sizes, may enhance the model's performance on financial time series data. Overall, while the Transformer underperformed in this study, its flexibility and scalability suggest considerable promise for future digital market forecasting applications.

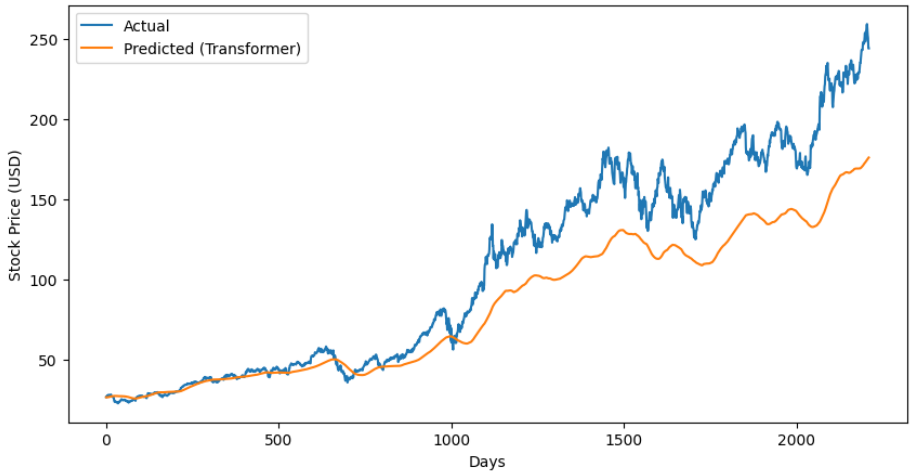


Figure 5 Transformer Model Prediction vs Actual Stock Prices

Figure 6 provides a comparative visualization of the Root Mean Squared Error (RMSE) values obtained from the three deep learning architectures evaluated in this study. The comparison clearly demonstrates that the GRU model produced the lowest RMSE, indicating superior predictive precision relative to both LSTM and Transformer models. The LSTM model achieved a moderately higher RMSE, reflecting its solid ability to capture long-term temporal patterns but with slightly reduced responsiveness to abrupt market changes. In contrast, the Transformer recorded the highest RMSE, confirming its limited effectiveness on this specific dataset. This comparative result underscores the importance of selecting architectures that align with the characteristics of the dataset, particularly its size, dimensionality, and temporal complexity. For univariate financial data such as daily stock prices, models with recurrent memory components like GRU and LSTM appear to be better suited for capturing sequential dependencies and mitigating overfitting risks.

Overall, the findings reaffirm that Recurrent Neural Network (RNN)-based architectures continue to offer strong advantages for time-series forecasting in digital market environments. The GRU model, in particular, demonstrated a balanced combination of accuracy, training efficiency, and generalization capability, making it the most effective approach for predicting Apple's stock price movements. The LSTM model remains a competitive alternative,

especially for long-term market trend analysis, due to its capacity to retain information over extended temporal sequences. Although the Transformer model underperformed in this experiment, it still presents valuable opportunities for future research when applied to larger and more complex datasets that include multiple financial indicators or sentiment-driven variables. Incorporating such contextual features could allow attention-based mechanisms to capture richer relationships across time and improve predictive accuracy. Consequently, these comparative insights highlight that while recurrent-based models currently dominate univariate stock forecasting, transformer-based architectures hold long-term potential for advancing digital financial prediction when supported by sufficiently comprehensive datasets.

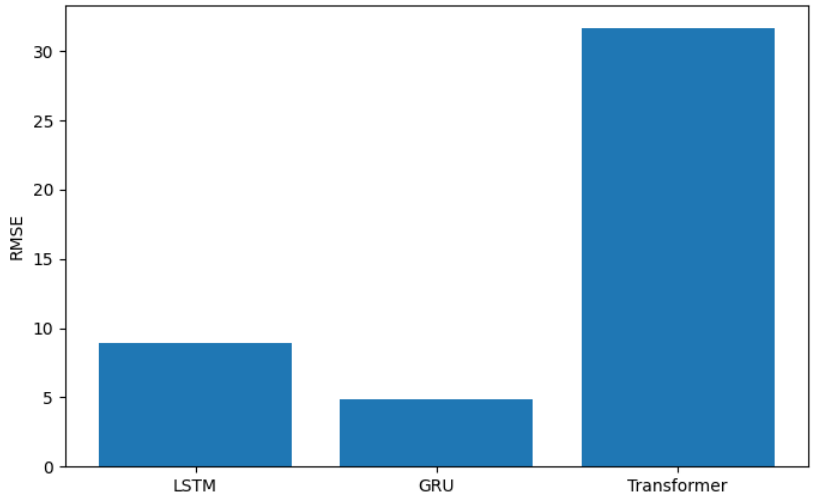


Figure 6 Comparison of RMSE Values for LSTM, GRU, and Transformer Models

Discussion

The results of this study indicate that the GRU model achieved the highest predictive accuracy and demonstrated stable learning performance compared with the LSTM and Transformer models. This finding is consistent with previous research showing that GRU-based architectures can perform effectively in stock price forecasting, although the relative performance of GRU and LSTM depends on the dataset and experimental settings [21], [23]. Furthermore, research combining GARCH with recurrent neural networks suggests that GRU-based models can provide competitive forecasting accuracy while offering computational efficiency compared with LSTM-based alternatives [22].

The LSTM model also demonstrated good predictive performance, but it was less responsive to sudden market fluctuations in the present experiment. This behavior may be related to differences in model architecture, hyperparameter configuration, and the temporal characteristics of the dataset. Previous studies indicate that the effectiveness of recurrent and attention-based architectures varies across forecasting tasks, meaning that no single architecture is guaranteed to perform best under all conditions [21], [25].

In contrast, the Transformer model showed comparatively weaker generalization performance on the univariate dataset used in this study. However, this result should not be interpreted as evidence that Transformers are inherently unsuitable for financial forecasting. Their performance can

depend on factors such as model configuration, learning-rate scheduling, data augmentation, and the characteristics of the input time series [24]. Research on forecasting architectures also suggests that the benefits of attention mechanisms and recurrent components depend on the specific task and model design [25].

Recent studies have explored hybrid architectures that combine the sequence-learning capabilities of recurrent networks with the attention mechanisms of Transformers. For example, Transformer–GRU and LSTM–Transformer models have been investigated for financial forecasting, with reported improvements under their respective experimental settings [26], [27]. These findings suggest that hybrid models may be a promising direction for future research. Nevertheless, their effectiveness should be evaluated using the same dataset, data preprocessing procedures, evaluation metrics, and validation strategy to ensure a fair comparison with standalone GRU, LSTM, and Transformer models.

Overall, the findings of this study suggest that the GRU model was the most suitable among the three architectures evaluated for the dataset and experimental conditions considered. Future research could investigate hybrid architectures, additional financial indicators, and sentiment information to determine whether these approaches improve forecasting accuracy and robustness across different market conditions [22], [23], [26], [27].

Conclusion

This research examined the application of three advanced deep learning architectures, namely Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Transformer Encoder, for forecasting Apple's stock prices within the framework of digital financial market prediction. The experimental design utilized a 60-day sliding time window to predict the subsequent trading day's closing price, providing a consistent temporal basis for evaluation. The results indicated that the GRU model achieved the highest accuracy and stability among the three models, with superior performance reflected in its lower RMSE and MAE values. This outcome highlights the GRU's ability to efficiently capture both short-term and long-term dependencies in sequential financial data through its simplified gating mechanism. The LSTM model also demonstrated strong predictive capabilities, particularly in modeling long-term temporal relationships, although it exhibited a moderate lag in responding to sudden market fluctuations. In contrast, the Transformer model underperformed, suggesting that its attention-based architecture is less effective when applied to small, univariate datasets with limited contextual diversity. These findings collectively reinforce the importance of aligning model architecture with the nature and complexity of the data used in digital market forecasting.

Beyond the comparative performance of the models, the findings of this study offer broader insights into the role of artificial intelligence in financial analytics. The superior performance of recurrent neural network-based models, such as GRU and LSTM, underscores their continued relevance for capturing the nonlinear and volatile behavior of stock prices in digital markets. From a practical perspective, the GRU model presents an efficient and scalable solution for developing real-time trading support systems, portfolio management tools, and automated forecasting platforms. The results also reveal promising directions

for future research, including the integration of multidimensional data sources such as trading volume, economic indicators, and sentiment metrics from digital media. These enhancements could provide richer contextual inputs for attention-based models like the Transformer, allowing them to achieve higher accuracy and interpretability in complex financial environments. Overall, this study affirms the value of deep learning as a transformative analytical framework for digital financial forecasting and highlights the potential of combining recurrent and attention-based architectures to advance the predictive capabilities of artificial intelligence in dynamic market conditions.

Declarations

Author Contributions

Conceptualization: B.P., D.O.; Methodology: B.P., D.O.; Software: B.P., D.O.; Validation: B.P., D.O.; Formal Analysis: B.P.; Investigation: B.P., D.O.; Resources: B.P., D.O.; Data Curation: D.O.; Writing – Original Draft Preparation: B.P.; Writing – Review and Editing: B.P., D.O.; Visualization: D.O.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Xing, C. Yan, and C. Xie, “Predicting NVIDIA's next-day stock price: A comparative analysis of LSTM, MLP, ARIMA, and ARIMA-GARCH models,” *arXiv*, vol. 2024, no. May, pp. 1–15, May 2024, doi: 10.48550/arXiv.2405.08284.

[2] A. Sendi, F. Hoque, and M. T. Hoque, “Decoding stock trends: A comparative study of GRU, LSTM, and Transformer models in tech sector prediction,” *Trans. Eng. Comput. Sci.*, vol. 13, no. 3, pp. 1–15, 2025, doi: 10.14738/tmlai.1303.18843.

[3] H. Mehtarizadeh, N. Mansouri, B. M. Hasani Zade, and M. M. Hosseini, “Stock price prediction with SCA-LSTM network and statistical model ARIMA-GARCH,” *J. Supercomput.*, vol. 81, no. 1, pp. 1–25, 2025, doi: 10.1007/s11227-024-06775-6.

[4] L. Feng and A. Sinchai, “Deep context-attentive transformer transfer learning for financial forecasting,” *PeerJ Comput. Sci.*, vol. 11, no. 1, pp. 1–20, 2025, doi: 10.7717/peerj-cs.2983.

[5] C. Fan and X. Zhang, “Stock price nowcasting and forecasting with deep learning,” *J. Intell. Inf. Syst.*, vol. 63, no. 1, pp. 639–656, 2024, doi: 10.1007/s10844-024-00908-2.

[6] C. Zhang, N. N. A. Sjarif, and R. Ibrahim, “Deep learning techniques for financial time series forecasting: A review of recent advancements: 2020–2022,” *arXiv*, vol. 2023, no. May, pp. 1–20, May 2023, doi: 10.48550/arXiv.2305.04811.

[7] V. Duong, D.-T. Nguyen, T. Pham, V.-H. Nguyen, and V. Q. A. Le, “Comparative study of deep learning models for predicting stock prices,” *Proc. Fed. Conf. Comput. Sci. Inf. Syst.*, vol. 2022, no. Sep., pp. 103–108, Sep. 2022, doi: 10.15439/2022R02.

[8] I. Valova, N. Gueorguieva, A. Thakkar, P. Nikitha, and H. Mohamed, “Hybrid deep learning architectures for stock market prediction,” *Proc. World Congr. Electr. Eng. Comput. Syst. Sci.*, vol. 2023, no. 1, pp. 1–10, 2023, doi: 10.11159/cist23.121.

[9] Y. Ma, “Comparative study on stock price forecasting based on deep learning,” *Adv. Econ. Manage. Res.*, vol. 12, no. 1, pp. 1–10, 2024, doi: 10.56028/aemr.12.1.917.2024.

[10] K. Al-Khatib, H. Khazaleh, H. Alkhazaleh, A. Alsoud, and L. Abualigah, “A new stock price forecasting method using active deep learning approach,” *J. Open Innov. Technol. Market Complex.*, vol. 8, no. 2, pp. 1–15, 2022, doi: 10.3390/joitmc8020096.

[11] A. John and T. Latha, “Stock market prediction based on deep hybrid RNN model and sentiment analysis,” *Automatika*, vol. 64, no. 4, pp. 981–995, 2023, doi: 10.1080/00051144.2023.2217602.

[12] C. Manjunath, Christ, B. Marimuthu, and B. Ghosh, “Deep learning for stock market index price movement forecasting using improved technical analysis,” *Int. J. Intell. Eng. Syst.*, vol. 14, no. 5, pp. 129–141, 2021, doi: 10.22266/ijies2021.1031.13.

[13] Y. S. Khodke and S. Deshpande, “Stock price prediction using LSTM and GRU,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 2023, no. 1, pp. 1–10, 2023, doi: 10.22214/ijraset.2023.53826.

[14] W. Chen, W. Hussain, F. Cauteruccio, and X. Zhang, “Deep learning for financial time series prediction: A state-of-the-art review of standalone and hybrid models,” *Comput. Model. Eng. Sci.*, vol. 2023, no. 1, pp. 1–25, 2023, doi: 10.32604/cmes.2023.031388.

[15] N. Nag, R. Murugesan, and M. George, “Multivariate input data for stock prediction using deep learning and hybrid models,” *J. Bus. Anal.*, vol. 9, no. 1, pp. 265–284, 2025, doi: 10.1080/2573234X.2025.2587153.

[16] H. Lyu, “Hybrid ARIMA-LSTM model for stock market prediction: A time series and deep learning integration approach,” *Informatica*, vol. 49, no. 1, pp. 1–15, 2025, doi: 10.31449/inf.v49i22.8510.

[17] J.-L. Kong, X. Zhao, W. He, X.-B. Yang, and X.-B. Jin, “EL-MTSA: Stock prediction model based on ensemble learning and multimodal time series analysis,” *Appl. Sci.*, vol. 15, no. 9, pp. 1–20, 2025, doi: 10.3390/app15094669.

[18] K. Bentaleb, M. Ben Houad, and M. Mestari, "A generative adaptive decomposition hierarchical Transformer for stock price forecasting," *IEEE Access*, vol. 14, no. 1, pp. 130991–131006, 2026, doi: 10.1109/ACCESS.2026.3727035.

[19] A. K. Kakde and M. P. Dale, "Multimodal stock price prediction using a hybrid GRU-multi-head attention model with ensemble sentiment and technical indicators," *Int. J. Electron. Commun. Eng.*, vol. 13, no. 5, pp. 1–15, 2026, doi: 10.14445/23488549/IJECE-V13I5P119.

[20] L. Zhou, Y. Zhang, J. Yu, G.-L. Wang, Z.-Z. Liu, S. Yongchareon, and N. Wang, "LLM-augmented linear Transformer–CNN for enhanced stock price prediction," *Mathematics*, vol. 13, no. 3, pp. 1–20, 2025, doi: 10.3390/math13030487.

[21] H. Abdullah, A. Tariq, I. Khan, R. Iqbal, F. Khan, and A. Iqbal, "Recurrent neural networks in time-series forecasting: A deep learning approach to stock market prediction," *Annu. Methodol. Arch. Res. Rev.*, vol. 2025, no. 1, pp. 1–15, 2025.

[22] J. Wei, S. Y. Yang, and Z. Cui, "Unified GARCH-recurrent neural network in financial volatility forecasting," *SSRN Electron. J.*, vol. 2025, no. 1, pp. 1–20, 2025.

[23] P. Mishra, "AI-driven stock price prediction: A case study on Agricultural Development Bank Limited (ADBL) using LSTM, GRU and TCN with and without news-headline sentiment," *Aadim J. Multidiscip. Res.*, vol. 2, no. 1, pp. 1–15, 2026, doi: 10.3126/ajmr.v2i1.97551.

[24] T. T. Thach, "Robust Transformer-based one-step stock index forecasting via shifted data augmentation," *arXiv*, vol. 2026, no. Jun., pp. 1–15, Jun. 2026, doi: 10.48550/arXiv.2606.15701.

[25] H. Heidenreich, P. R. Vlachas, and P. Koumoutsakos, "Deconstructing recurrence, attention, and gating: Investigating the transferability of Transformers and gated recurrent neural networks in forecasting of dynamical systems," *arXiv*, vol. 2024, no. Oct., pp. 1–20, Oct. 2024, doi: 10.48550/arXiv.2410.02654.

[26] H. Singh et al., "A hybrid Transformer GRU architecture for stock price prediction: Optimization and performance evaluation using the Artificial Rabbits Algorithm," *Int. J. Comput. Intell. Syst.*, vol. 2026, no. 1, pp. 1–20, 2026, doi: 10.1007/s44196-026-01446-1.

[27] M. R. Kabir, D. Bhadra, M. Ridoy, and M. Milanova, "LSTM–Transformer-based robust hybrid deep learning model for financial time series forecasting," *Sci.*, vol. 7, no. 1, pp. 1–15, 2025, doi: 10.3390/sci7010007.