



# MALETE JOURNAL OF ACCOUNTING AND FINANCE

A Publication of the Department of Accounting and Finance, Faculty of Management and Social Sciences, Kwara State University, Malete

Vol. 6 No. 1, June, 2025

ISSN: 2735-9603

## FORECASTING GLOBAL STOCK PRICES USING GATED RECURRENT UNIT, LONG SHORT-TERM MEMORY WEIGHTED-LSTM AND LSTM WITH ATTENTION

**Emmanuel Imuede Oyasor**

Department of Accounting Science, Walter Sisulu University, Mthatha, South Africa

[emmanueloyasor247@gmail.com](mailto:emmanueloyasor247@gmail.com)

### **Abstract**

*This research assesses the forecasting potential of four deep learning models: Gated Recurrent Unit (GRU), Simple Long Short-Term Memory (LSTM), LSTM with Attention, and Weighted LSTM (W-LSTM), specifically for predicting stock prices in five African countries - Tanzania, South Africa, Nigeria, Kenya, and Morocco. The evaluation utilized historical price data, and model performance was measured using four metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination ( $R^2$ ). The GRU model consistently produced the lowest error rates and highest accuracy in South Africa (MSE: 0.87, MAE: 0.89, RMSE: 0.87,  $R^2$ : 0.94) and Nigeria (MSE: 0.90, MAE: 0.89, RMSE: 0.92,  $R^2$ : 0.93), demonstrating a strong predictive power in these regions. Meanwhile, in Kenya and Morocco, the Simple LSTM model excelled, achieving  $R^2$  scores of 0.92 and 0.91, respectively, along with commendable MSE and RMSE figures (Kenya RMSE: 0.84, Morocco RMSE: 0.84). Although both the LSTM with Attention and W-LSTM models produced competitive results, they did not consistently achieve lower error metrics or higher  $R^2$  values compared to the simpler models. These results indicate that streamlined architectures like GRU and LSTM are effective for stock price forecasting in emerging markets. The findings offer valuable insights into the predictive effectiveness of sophisticated recurrent neural network models. The study holds practical significance for financial organizations, investors, and policymakers in utilizing deep learning for informed decision-making and risk management, while pinpointing areas for future improvements in model development and data integration.*

### **INTRODUCTION**

The performance of a nation's financial market is a significant factor that influences its overall economic state, allowing economists financial analysts to assess country's current economic-being. Among the different types of financial markets, the stock market is particularly important. A nation's economic circumstances affect various sectors, including finance, agriculture, metals, and investment banking, either directly or indirectly. The development of these sectors is contingent upon their volatility, which adheres to the basic principle of supply and demand. The demand for a specific sector has a direct effect on the stock market, with rising supply encouraging traders and financial institutions to invest in that sector or stock, which leads to an increase in prices. Moreover, the regular distribution of dividends plays a role in generating profits and returns on invested capital. It is essential for investors to recognize the right time to sell their shares in order to achieve their targeted returns. Financial markets include several types, such as stock markets, derivatives markets, bond markets, and commodity markets (Obthong et al. 2020). The stock market acts as a venue where investors can purchase and hold a stake or share of a company. As companies expand, they frequently need additional funding to finance their future projects. With the consent of existing shareholders—who may experience diluted ownership due to the issuance of additional shares—companies are able to offer these shares to investors to raise funds. Successful ventures lead to an augmented stock market value for these

shares. Stocks available on the stock market can be acquired for both short-term and long-term investment strategies. Long-term investments involve retaining shares for an extended duration, while short-

Term trading involves the rapid buying and selling of shares, with investors aiming to generate profits within days or. Various trading strategies are employed by traders, such as swing trading, day trading, position trading, and scalping (Mann and Kutz 2016). However, the stock market is a complex field influenced by numerous factors like political events, economic indicators, and public sentiment. Predicting stock prices accurately is a challenging task that requires comprehensive research covering many years. The emergence of advanced machine learning and deep learning techniques has made it possible to create models that can effectively forecast stock values. Historically, stock market predictions were based on two primary approaches. The first, technical analysis, relies on historical stock prices – including opening and closing prices, trading volume, and closely related closing values – to estimate future stock prices. The second approach, qualitative analysis, considers external factors such as a company's profile, market conditions, political and economic influences, and textual information from financial news articles, social media, and expert blogs. Despite these analyses, predictions often lack precision, making it hard for investors to achieve returns on their investments. Fundamental, technical, and primary analyses are widely regarded methods for analyzing market trends (Manish and Thenmozhi 2014), but they have inherent drawbacks related to lagging indicators and prediction errors. This limitation has led researchers to develop more sophisticated techniques using machine learning and deep learning models for real-time market scenarios.

Today, machine learning and deep learning algorithms present considerable advantages over traditional methods like technical and fundamental analysis. These algorithms utilize machine learning and artificial intelligence to enhance stock price and index forecasting. Machine learning acts as an additional tool alongside technical and fundamental analysis, creating a robust trading platform. These models can solve various challenges, including stock price predictions, classification, portfolio management, algorithmic trading, market sentiment analysis, and risk assessment. This article aims to examine the various methods used for stock price forecasting and classification.

Several prominent machine learning and deep learning algorithms discussed in this article include linear regression, random forest, logistic regression, k-nearest neighbors, support vector machines, naïve Bayes, ARIMA (autoregressive integrated moving average), FB Prophet (Facebook Prophet), LSTM (long short-term memory), GRU (gated recurrent units), along with ensemble algorithms like random forest and XG-Boost (extreme gradient boosting) (Zhong and Enke 2019; Sidra and Sen 2020; Parray et al. 2020). For assessing the accuracy of classification models, metrics such as recall, precision, and F-score are commonly utilized, while root mean square error (RMSE) and mean absolute percentage error (MAPE) are employed for regression or price forecasting models (Jose et al., 2019). Numerous review articles on stock price prediction and forecasting have been published (Polamuri et al., 2019; Kumar et al., 2021; Soni et al., 2022). With the rise of artificial intelligence and machine learning, the volume of publications has significantly increased. This article presents the latest cutting-edge machine learning and deep learning methods for predicting stock prices. Additionally, this study will concentrate on forecasting the South African stock market index using artificial intelligence techniques such as

GRU, LSTM, WLSTM, and LSTM with Attention, which fall under the category of technical analysis for stock market prices.

Different levels of evaluation metrics such as Mean Absolute Percentage Error (MAE), Mean Squared Error (MSE), Directional Accuracy (DA), Dynamic Time Wrapping (DTW), Symmetric Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). With these, we can be able to suggest the best models for the prediction of stock market price in South Africa based on their levels of performances or accuracy measures. Lastly, this work is segmented into four different parts; section one is the introduction, section two is related work, section three focuses on the methodology while section four is finding and analysis and section five is conclusion and recommendation.

## LITERATURE REVIEW

### *Conceptual Review*

Machine learning (ML) has become a transformative tool across various industries due to its ability to improve predictive accuracy and reduce errors in complex tasks. The concept of applying ML to financial markets stems from the need to better forecast price movements, manage risks, and optimize trading strategies in highly volatile environments. The rapid advancement in information technology has expanded ML applications beyond finance, influencing domains such as healthcare, agriculture, and cybersecurity. For example, ML techniques have been employed in breast cancer detection, crop disease monitoring, and malware detection, all of which rely on accurate pattern recognition to inform timely decisions.

In financial markets, the core conceptual challenge is to extract meaningful signals from noisy and often imbalanced datasets, such as those encountered in stock price or foreign exchange forecasting. Various ML algorithms have been explored, including ensemble methods like Random Forest and AdaBoost, single classifiers such as Support Vector Machines (SVM), Neural Networks, Logistic Regression, and Gradient Boosting Machines (GBM). The use of feature selection to identify relevant variables and reduce dimensionality further enhances the performance of these models, emphasizing the importance of data preprocessing in financial predictions.

A key conceptual insight is that model performance varies across different financial contexts, time frames, and datasets. For instance, while ensemble methods often provide superior accuracy by combining multiple models, single classifiers can perform competitively when carefully tuned and paired with effective feature selection. Moreover, dealing with imbalanced data, common in fraud detection or anomaly identification, requires specific techniques to reduce false positives and improve classification precision, highlighting the interplay between data characteristics and model design.

### *Theoretical Review*

The theoretical foundation for applying machine learning in financial forecasting is rooted in pattern recognition, statistical learning theory, and computational intelligence. Researchers have approached the problem of stock price prediction by leveraging classification and regression frameworks, where historical financial indicators serve as input features to predict future price movements or returns.

Oyasor, E. I. (2025). Forecasting Global Stock Prices Using Gated Recurrent Unit, Long Short-Term Memory Weighted-LSTM and LSTM with Attention. *Malete Journal of Accounting and Finance*, 6 (1), 147 – 161

Studies such as Gu, Kelly, and Xiu (2018) extended ML applications to less-explored asset classes like foreign exchange markets, addressing theoretical challenges like class imbalance and model generalization. Ensemble learning theories suggest that aggregating multiple weak learners, such as in Random Forests or AdaBoost, can produce stronger, more robust predictors by reducing variance and bias.

In terms of theoretical advances, the benchmarking of various classifiers against financial datasets, as seen in Ballings et al. (2015) and Gündüz et al. (2017), contributes to understanding the comparative strengths of models like Random Forest, SVM, Gradient Boosting Machines, and Logistic Regression. The findings emphasize the role of non-linear relationships and feature interactions that ensemble methods and tree-based models can capture effectively.

Additionally, the theoretical work around imbalanced datasets, highlighted in Singh et al. (2015) and Rajani and Padmavathamma (2012), addresses the challenge of skewed class distributions in financial fraud detection and anomaly classification. Techniques such as cost-sensitive learning, resampling, and rule-based methods are grounded in statistical theory to improve detection accuracy while managing the trade-off between false positives and false negatives.

Finally, the time series nature of financial data brings in theories related to sequential modeling and temporal dependencies, motivating the use of recurrent architectures like LSTM and GRU in later research. These models theoretically capture long-term dependencies and non-linear temporal patterns that traditional statistical models often fail to detect.

### *Empirical Review*

With differing degrees of success, machine learning techniques have been used in a number of earlier research to forecast changes in stock prices. AI (machine learning) applications have become crucial for forecasting or recognizing jobs, especially when predicting accuracy is crucial, as a result of the substantial changes brought about by the development of information technology in many industries, including finance. For instance, the application of machine learning techniques to the development of trading apps, the detection of breast cancer, weather forecasting, crop disease monitoring in the agricultural sector, etc., has not only improved our quality of life but also has an impact on human prediction accuracy, which in turn reduces forecasting errors. This section will examine a few research topics pertaining to the application of machine learning methods and how well they predict stock prices.

In a close study, the authors propose an SVM based technique to detect metamorphic malware. Gu, Kelly, and Xiu (2018) provide the first comprehensive approach to quantifying the effect of using machine learning (ML) to the prediction of monthly stock returns. Their goal is to apply machine learning methods in a relatively unexplored asset class: foreign exchange (FX). They examine the issue of imbalanced data sets, which have fewer malware samples than benign files, and how to successfully detect them with high precision and accuracy (Singh et al., 2015).

In order to determine whether the stock prices of 5767 European companies will increase by a predefined amount, Ballings et al. (2015) benchmarked three ensemble classification methods (Random Forest, AdaBoost, and Kernel Factory) against four single classifier models (Logistic Regression, Neural Networks, KNearest Neighbors, and Support Vector Machines). They used a variety of balance sheet and P&L statement variables as well as some general economic features as i

Oyasor, E. I. (2025). Forecasting Global Stock Prices Using Gated Recurrent Unit, Long Short-Term Memory Weighted-LSTM and LSTM with Attention. *Malete Journal of Accounting and Finance*, 6 (1), 147 – 161

inputs to see if ensemble classification methods could outperform single classifier models in terms of stock price prediction. The results show that Random Forest is the best predictor (with a median AUC value of 0.9037) out of all seven models and significantly outperforms all other classifiers except SVM.

Logistic Regression and Gradient Boosting Machine were used by Gündüz et al. (2017) to predict the daily closing returns of the three most traded stocks in ISE (GARAN, THYAO, and ISCTR). They also applied feature selection methods to determine the ideal number of features, using a variety of technical and fundamental variables as inputs. The results indicate that using feature selection methods improves the prediction performance for selected stocks; in terms of prediction accuracy, following the implementation of the feature selection process, GBM produced 59.90% for GARAN and 55.80% for THYAO, while LR produced 58.10% for ISCTR. Rajani and Padmavathamma (2012) suggest a rule-based technique applied to fraud detection problem. This paper discusses the problem of imbalanced data that results in a very high number of false positives and proposes techniques to alleviate this problem.

To forecast the direction of movement in the Daily Tehran Stock Exchange Index, Imandoust and Bolandraftar (2014) employed three classification techniques: Random Forest, Decision Tree, and Naive Bayesian Classifier.

They discovered that the Decision Tree technique outperforms Random Forest and Naive Bayesian Classifier in terms of hit rate (respectively 80.08%, 78.39%, and 72.15%) when ten technical and three fundamental factors are used as inputs.

Last but not least, Labiad et al. (2016) used 23 technical indicators to train three machine learning approaches (Random Forest, Gradient Boosted Trees, and Support Vector Machines) to forecast the Moroccan Stock Exchange for a relatively little period (10 minutes in advance). They have determined that, in terms of testing accuracy, Random Forest and Gradient Boosted Trees perform better than Support Vector Machines (95%, 94%, and 90%, respectively).

## METHODOLOGY

### *Machine Learning Models*

To predict of stock prices, the study considers four machine learning models - Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Weighted LSTM (W-LSTM) and LSTM with Attention Mechanism.

#### *Gated Recurrent Unit (GRU)*

Another recurrent neural network variation that attempts to overcome some of the drawbacks of LSTM is the Gated Recurrent Unit (GRU). GRU is made to capture long-term dependencies and model sequential data. But compared to LSTM, GRU is easier to use and more computationally efficient. Compared to LSTM, which has three gates, GRU only has two: update and reset. This lowers computational complexity without sacrificing the capacity to detect temporal relationships in time-series data. Figure 2 shows that  $x_t$  is the input at time step  $t$ ,  $h_{t-1}$  is the hidden state from the previous time step,  $z_t$  and  $r_t$  are the update gate and reset gate values, and  $\tilde{h}_t$  is the potential new hidden state.  $\odot$  represents element-wise multiplication,  $\sigma$  is the sigmoid function, and  $\tanh$  represents the hyperbolic tangent function.

One kind of recurrent neural network (RNN) used in deep learning is called a Gated Recurrent Unit (GRU). GRUs are especially good at processing data sequences for tasks like speech recognition, natural language processing, and time series prediction. They solve some of the drawback

s of conventional RNNs, especially those pertaining to long term dependencies in sequence data.

$$z_t = \sigma(Wzxt + Uzht-1 + bz) \tag{1}$$

$$r_t = \sigma(Wrxt + Urht-1 + br) \tag{2}$$

$$\tilde{h}^t = \tanh(W\tilde{h}xt + U\tilde{h}(rt \odot ht-1) + bh) \tag{3}$$

$$h_t = (1 - z_t) \odot ht-1 + z_t \odot \tilde{h}^t \tag{4}$$

where:  $x_t \in R^n$  is the input vector at time step  $t$ ,  $h_{t-1}, h_t \in R^m$  are the previous and current hidden states,  $z_t, r_t \in R^m$  are the update and reset gates,  $\tilde{h}_t \in R^m$  is the candidate hidden state,  $W_*, U_*$  are weight matrices,  $b_*$  are bias vectors,  $\sigma$  is the sigmoid function, and  $\odot$  denotes element-wise multiplication.

With encouraging outcomes, GRU has been used for stock price prediction tasks. For stock price prediction, Kyriakidis et al. (2019) showed that GRU outperforms LSTM while using less computing power. Numerous research have employed the GRU model to forecast stock price direction, volatility, and returns; the findings show that it is more successful at capturing long-term dependencies and is less likely to overfit than LSTM. Additionally, GRU has been shown to function well with sparse or noisy data, providing resilience to market swings (Chung et al., 2014).

#### Long Short-Term Memory (LSTM)

One kind of recurrent neural network (RNN) called Long Short-Term Memory (LSTM) is intended to capture long-range dependencies and handle sequential input. LSTM is appropriate for time-series forecasting tasks like stock price prediction because of its internal memory structure which allows it to retain information for extended periods of time. The input, output, and forget gates in the LSTM architecture regulate the information flow, enabling it to retain or forget data as required. The input gate  $i_t$ , forget gate  $f_t$ , and output gate  $o_t$  are the three gates added to LSTM, as seen in Figure 1. It introduces the cell state  $C_t$  as the long-term memory layer and the candidate state  $\tilde{C}_t$  as a temporary storage for information to be saved in the long-term memory. These three gates are functions of the previous time step's output  $h_{t-1}$  and the current time step's input  $x_t$ . LSTM has a well-constructed structure with gates named as forget gate, input gate and output gate.

$$f_t = \sigma(Wfxt + Ufht-1 + bf) \tag{5}$$

$$i_t = \sigma(Wixt + Uiht-1 + bi) \tag{6}$$

$$o_t = \sigma(Woxt + Uoht-1 + bo) \tag{7}$$

$$\tilde{c}_t = \tanh(W\tilde{c}x_t + U\tilde{c}h_{t-1} + b_c) \tag{8}$$

$$c_t = \tanh(Wcx_t + Uch_{t-1} + b_c) \tag{9}$$

where:  $f_t, i_t, o_t \in R^m$  are the forget, input, and output gates respectively,  $\tilde{c}_t \in R^m$  is the candidate cell state,  $c_t \in R^m$  is the updated cell state, and all other notations are as previously defined.

When it comes to capturing the non-linear relationships and temporal dependencies that influence market movements, LSTM networks have proven to perform better. LSTM has been used in numerous studies to forecast stock prices. In terms of predictive accuracy, Fischer and Krauss (2018) showed that LSTM performs better than conventional models like GARCH and ARIMA. LSTM is the preferred model for stock market forecasting because of its capacity to identify complex patterns in historical data. With the ability to capture characteristics that represent market circumstances, LSTM-based techniques have proven very effective in forecasting both short-term and long-term price patterns.

Even in volatile market settings, Chen et al. (2018) demonstrate that LSTM is successful at predicting stock movements and market trends.

### *The Weighted LSTM*

Weighted LSTM is an enhancement of the standard LSTM network where different weights are assigned to the input data to emphasize more important features in the model. These weights are often derived from the relative importance of different data points or features, allowing the model to focus on more influential inputs for better predictions as depicted by Figure 3. Weighted LSTM models have been used to enhance the forecasting ability of stock price prediction models by integrating domain knowledge or expert opinion into the learning process. In stock market prediction, feature weights can be assigned based on the significance of certain technical indicators or past market conditions.

In W-LSTM, each hidden state output  $h_t$  from the LSTM is multiplied by a corresponding weight  $\alpha_t$ , which may be learned during training or set based on domain knowledge. This allows the model to focus more on relevant time steps and reduce the influence of less important ones. The final prediction is typically a weighted sum of all hidden states over the sequence. W-LSTM improves interpretability and performance, especially in time series forecasting and sequence modeling tasks where not all time points contribute equally.

$$h_t^{(w)} = \alpha_t \cdot h_t \quad (11)$$

where:  $h_t^{(w)} \in \mathbb{R}^m$  is the weighted hidden state, and  $\alpha_t \in [0,1]$  is a scalar or vector representing the learned importance weight for time step  $t$ .

In order to increase prediction accuracy, Liu et al. (2020) employed a weighted LSTM technique to combine several characteristics, including sentiment data, volume, and stock prices. In financial applications, weighted LSTM models have demonstrated promise, particularly when paired with other engineering techniques. Weighted LSTM can increase the model's ability to identify significant stock market trends and produce more accurate forecasts by giving greater weights to more pertinent elements. Selecting the right weight values is the primary difficulty with Weighted LSTM models. The predictive power of the model may be impacted by overfitting or underfitting caused by improper weight allocations. Furthermore, Weighted LSTM models can become computationally demanding, particularly when working with complicated feature sets or huge datasets.

### *LSTM with Attention*

In deep learning, the attention mechanism is a technique that enables the model to concentrate on the most pertinent portions of the input sequence during the prediction process. By giving distinct attention weights to various time steps in the input sequence, the LSTM with Attention model is able to concentrate on particular time periods or characteristics that have a greater bearing on stock price prediction. The three primary components of the attention mechanism layer are the feature reconstruction layer, important weight allocation layer with weight softmax, and feature similarity calculation layer, as seen in Figure 4. The correlation is investigated using the feature similarity computation layer. The feature similarity calculation layer explores the correlation between known different layer depth sampling data and the target predicted layer by taking the outputs  $h_{t-1}$ ,  $h_t$ , and  $h_{t+1}$  of the residual module as inputs, and outputs the correlation coefficient  $r$ .

Stock price forecasting has significantly improved when attention mechanisms were added to LSTM networks. By enabling the network to concentrate on the most significant characteristics in the time-series data, attention-based LSTM models enhance accuracy and interpretability. According to Bahdanau et al. (2014), attention processes have the potential to greatly improve LSTM networks' performance in sequence-to-sequence tasks. In order to enhance the model's capacity to capture the temporal dependencies and volatility in stock prices, LSTM with Attention has been used for stock price prediction. The network can assign greater weight to time steps or market situations that are more predictive of future price movements when using attention-based LSTM models. Zhang et al. (2021) demonstrated how LSTM with Attention enhanced the model's capacity to forecast stock prices in times of market volatility when historical trends might not be enough to provide adequate predictive power. The primary difficulty with LSTM with Attention is the model's increased complexity. By adding more factors, the attention mechanism raises the computational expense and overfitting risk. Furthermore, it can be difficult to adjust attention weights, and in order to prevent poor generalization to novel, untested market situations, the model could need a large amount of training data.

$$e_{t,i} = v^T \tanh(W_1 h_i + W_2 h_t) \tag{12}$$

$$\alpha_{t,i} = \frac{\exp(e_{t,i})}{\sum_j \exp(e_{t,j})} \tag{13}$$

$$c_t = \sum_i \alpha_{t,i} h_i \tag{14}$$

where:  $e_{t,i}$  is the attention energy score between the current hidden state  $h_t$  and encoder hidden state  $h_i$ ,  $\alpha_{t,i} \in [0,1]$  is the attention weight computed via softmax,  $c_t$  is the context vector (a weighted sum of hidden states),  $\hat{h}_t$  is the final attention-enhanced hidden state,  $v, W_1, W_2, W_c$  are learnable parameters, and  $[c_t; h_t]$  denotes the concatenation of  $c_t$  and  $h_t$ .

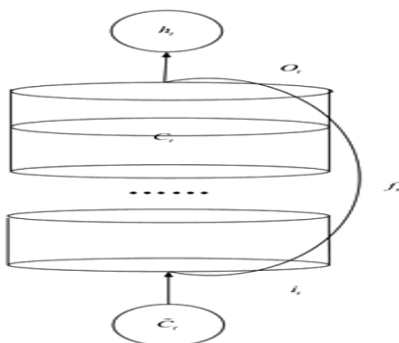


Figure 1: LSTM neural network structure.

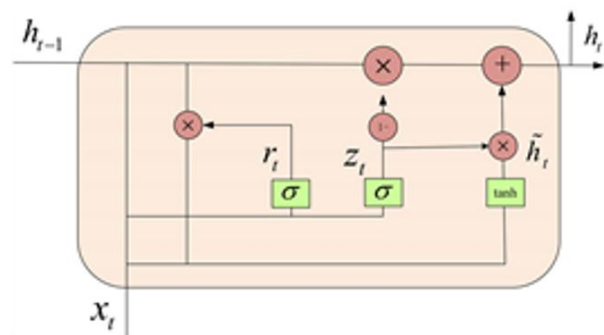
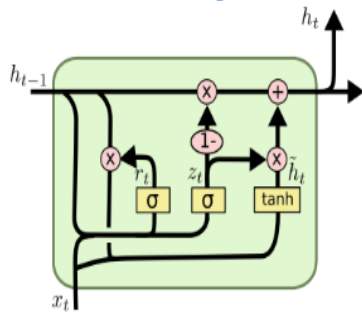


Figure 2: GRU model diagram



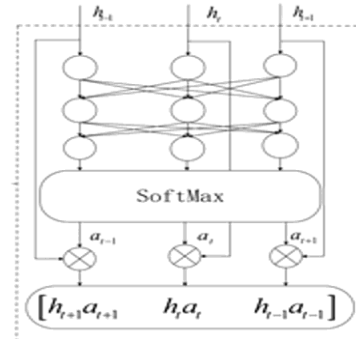
**Figure 3:** Weighted LSTM module.

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

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

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t])$$

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



**Figure 4:** Attention mechanism

### Data and Evaluation

To do this, we used the stock prices of four African countries, including Nigeria, Tanzania, Kenya, Morocco, and South Africa, and preprocessed the data to make sure there were no missing or void data points. We used 80% of the data for training and 20% for testing, totaling 762 data points. In addition, the input variables included data, and the output variable was close.

A statistical or machine learning model's performance and efficacy are evaluated using evaluation metrics, which are quantifiable measurements. These metrics aid in comparing various models or algorithms and offer insights into how well the model is operating. To measure and compare the algorithms' performance, we used four assessment indicators. The average of the squared discrepancies between the actual values  $y_i$  and the projected values  $\hat{y}_i$  is determined by the Mean Squared Error (MSE). Larger mistakes are penalized more harshly.

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

$y_i$  is the Actual value for the  $i^{\text{th}}$  observation,  $\hat{y}_i$  is Predicted value for the  $i^{\text{th}}$  observation and  $n$  is Total number of observations.

The Mean Absolute Error (MAE) measures the average magnitude of the errors in a set of predictions, without considering their direction. It is less sensitive to outliers than MSE.

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

Where  $|y_i - \hat{y}_i|$  is the absolute error for the  $i^{\text{th}}$  observation

Root Mean Squared Error (RMSE) is the square root of MSE. It provides errors in the same units as the target variable and is more sensitive to large errors due to the squaring.

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

R-Squared Score (RSCORE or  $R^2$ ) measures the proportion of variance in the actual values that is predictable from the predicted values. It indicates how well the model explains the variability of the response data.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ : Mean of actual values,  $R^2 = 1$ : Perfect prediction  $R^2 = 0$  is the model predicts no better than the mean and  $R^2 < 0$ : Model performs worse than using the mean

## RESULTS AND DISCUSSION

Deep learning models for four African stock markets - South Africa, Nigeria, Kenya, Morocco, and Tanzania - are reported in this work. Four models were evaluated: Weighted LSTM (W-LSTM), Gated Recurrent Unit (GRU), Simple Long Short-Term Memory (LSTM), and LSTM with Attention. Common forecasting metrics, such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R<sup>2</sup> Score (RSCORE), showed that the GRU and Simple LSTM models were more predictive. These results imply that more straightforward, optimized architecture would be more appropriate. than complex alternatives for stock price prediction in frontier markets. Figure 1 shows the correlation between the stock prices of the countries. The outcome shows some degree of relations amongst the stock prices in Africa.

In South Africa, the GRU model emerged as the most effective, with an R<sup>2</sup> of 0.94 - the highest among all models and countries - alongside MSE and RMSE values of 0.87. This suggests a strong alignment between predicted and actual stock price patterns. Although the LSTM with Attention and Weighted LSTM models produced slightly lower RMSE values (0.82 and 0.83, respectively), they underperformed in explanatory power, with R<sup>2</sup> values of 0.83 and 0.84. This implies that, while capable of reducing absolute error, their complexity did not translate into a better overall fit. The standard LSTM model performed moderately well, achieving an R<sup>2</sup> of 0.89, confirming GRU's relative advantage in this market.

In the Kenyan stock market, the Simple LSTM model performed best, with an R<sup>2</sup> score of 0.92, the lowest MAE (0.83), and RMSE (0.84), confirming its strength in capturing longer dependencies effectively. The GRU model followed, with slightly higher error metrics and a lower R<sup>2</sup> of 0.87. Interestingly, both LSTM with Attention and Weighted LSTM performed competitively (R<sup>2</sup> = 0.88 and 0.89 respectively), but not as effectively as Simple LSTM. This suggests that in markets like Kenya, where stock price sequences may require retention of longer patterns, traditional LSTM has a structural advantage over more complex or compressed architectures.

**Table 1:** Summary Statistics

	Kenya	Nigeria	Morocco	South Africa	Tanzania
Mode	762.00	762.00	762.00	762.00	762.00
Mean	1963.06	1488.86	11779.64	58163.05	1896.91
Std	267.54	309.13	1176.37	6656.33	90.13
Min	1563.68	871.26	8987.89	36490.00	1697.65
0.25	1811.47	1136.86	11071.46	52850.00	1823.09
0.5	1889.58	1594.00	11955.63	60113.00	1884.85
0.75	1976.20	1746.22	12611.36	62934.50	1953.40
Max	2766.77	2065.62	13991.47	71034.00	2136.52

Source: Author (2025)

In Morocco, a similar trend was observed. Simple LSTM achieved the highest R<sup>2</sup> of 0.91, along with the lowest MSE (0.85), MAE (0.83), and RMSE (0.84). GRU again performed well (R<sup>2</sup> = 0.87), indicating its consistent reliability across different market structures. However, the performance of the Weighted LSTM and Attention-based LSTM models declined, particularly in terms of error metrics. The W-LSTM model reported the highest MSE (0.91) and RMSE (0.89), while LSTM with Attention recorded a slightly better but still suboptimal R<sup>2</sup> of 0.86. These findings reinforce the conclusion that added architectural complexity does not always improve forecasting accuracy in relatively stable or moderately volatile markets.

Moreso, Tanzania made an incredible result with the GRU model exhibited the best performance, by making MSE of 0.94, MAE of 0.90, RMSE of 0.92, and an R<sup>2</sup> score of 0.92. These results indicate that GRU achieved the highest accuracy and lowest prediction error, demonstrating its superior capability in modeling the temporal dependencies of the dataset. The standard LSTM model followed with an MSE of 0.86, MAE of 0.92, RMSE of 0.89, and R<sup>2</sup> score of 0.87. While it maintained competitive error rates, its R<sup>2</sup> score was lower than that of GRU, suggesting a slightly reduced ability to explain the variance in the target variable. Weighted LSTM and LSTM with Attention recorded lower MSE and MAE values in some instances - 0.84 and 0.82 for MSE, and 0.87 and 0.84 for MAE, respectively - but their R<sup>2</sup> scores (0.84 and 0.85) were also lower than that of GRU. This suggests that although these enhanced LSTM variants captured certain patterns effectively, their overall predictive power was less robust in the Tanzanian context. Overall, GRU proved to be the most reliable and accurate model across all evaluation metrics.

Lastly, Nigeria presented another clear case for the GRU model’s dominance. With an R<sup>2</sup> of 0.93, it demonstrated remarkable forecasting strength, despite posting a slightly higher RMSE (0.92). Both MAE and MSE were also competitively low (0.89 and 0.90 respectively), placing GRU ahead of other models. Weighted LSTM (R<sup>2</sup> = 0.87) showed moderate performance, while both LSTM with Attention and Simple LSTM achieved R<sup>2</sup> values of 0.86. Interestingly, although the Attention-based LSTM had slightly better error metrics than the standard LSTM, neither model surpassed GRU, highlighting that efficient sequence modeling can outperform models that attempt to emphasize temporal focus through attention. The visualization of the actual and the prediction of each country were shown the Figure 5.

The findings of this study offer important implications for policy-makers, financial regulators, investors, and researchers. For governments and regulators, the deployment of accurate predictive models can improve market surveillance, risk assessment, and the timely detection of market anomalies or bubbles. Investors and capital market can utilize these tools to enhance investment decision-making, asset allocation, and risk management. The availability of accurate and scalable deep learning models can democratize access to forecasting tools for retail investors, promoting financial inclusion and better portfolio outcomes in emerging markets. For researchers, these results underscore the need to develop models that are not only accurate but also interpretable and adaptable to diverse market environments. Ballings et al. (2015) found that the random forest is the most effective model, significantly outperforming all models except SVM for the firm-level financial data and economic indicators used to forecast stock price movements of 5767 European firms. Labiad et al. (2016) also shows that the random forest and GBM outperformed SVM, with testing accuracies of 95%, 94%, and 90%, respectively. Imandoust and Bolandraftar (2014) show that the decision tree slightly approach outperformed the others with a hit rate of 80.08%, compared to 78.39% and 72.15% for random forest and naive bayes, respectively.

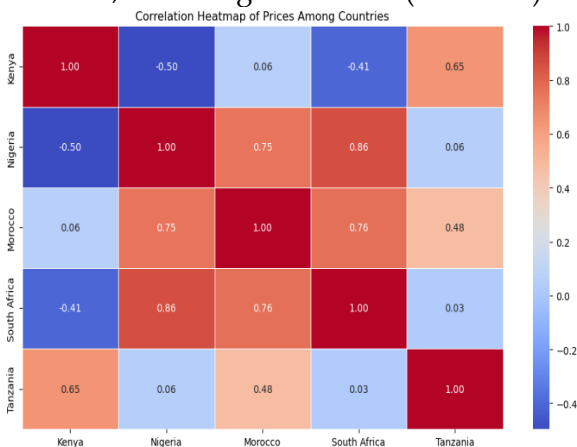
**Table 2:** Evaluation Metrics

Countries	Models	MSE	MAE	RMSE	RSCORE
Kenya	GRU	0.85	0.86	0.9	0.87
	LSTM	0.85	0.83	0.84	0.92
	Weighted LSTM	0.89	0.87	0.88	0.89
	LSTM with Attention	0.88	0.86	0.87	0.88
Nigeria	GRU	0.91	0.89	0.92	0.93
	LSTM	0.88	0.89	0.87	0.86

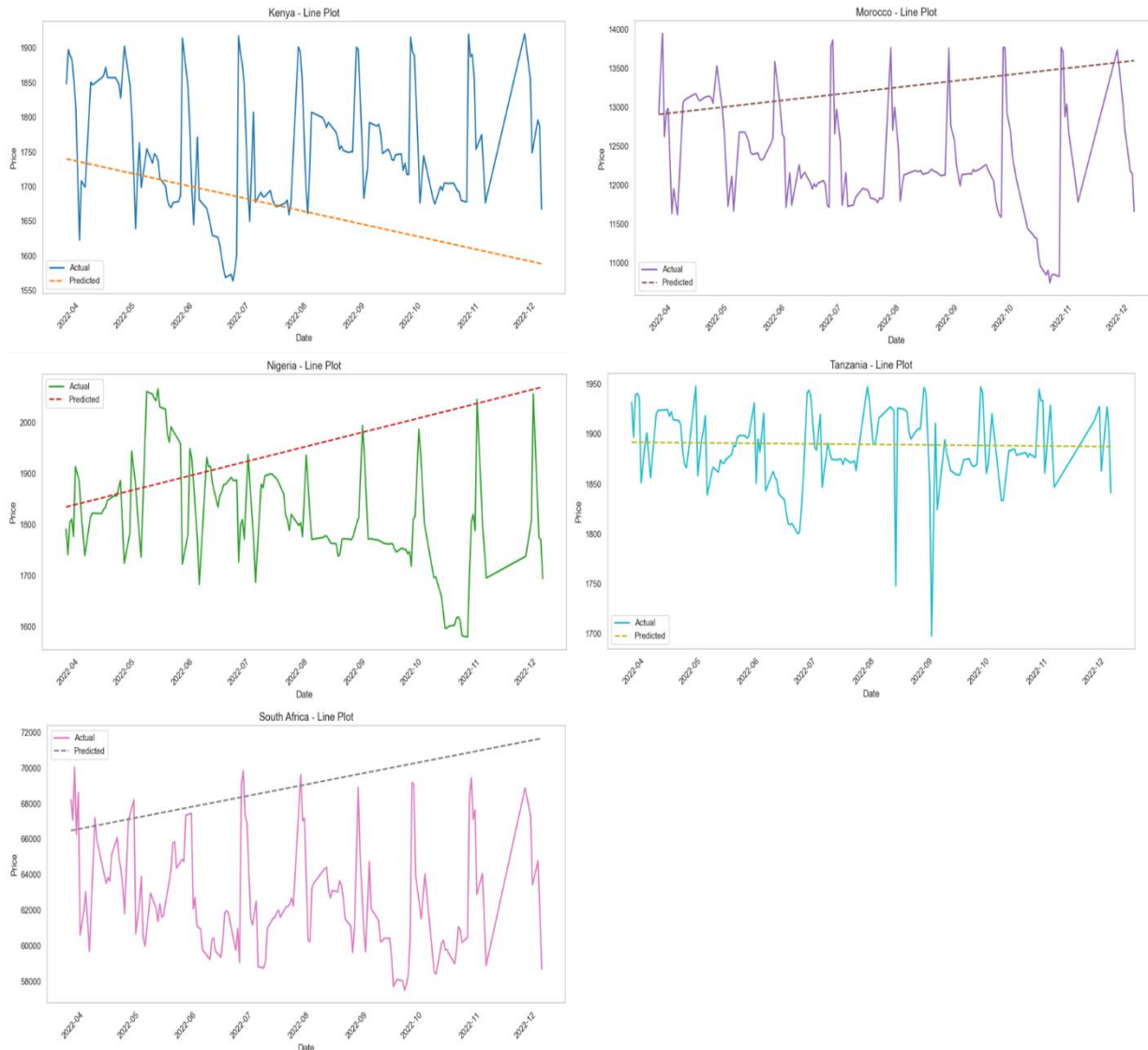
	Weighted LSTM	0.87	0.85	0.86	0.87
	LSTM with Attention	0.86	0.84	0.85	0.86
Morocco	GRU	0.87	0.85	0.86	0.87
	LSTM	0.85	0.83	0.84	0.91
	Weighted LSTM	0.91	0.87	0.89	0.89
	LSTM with Attention	0.89	0.86	0.88	0.86
South Africa	GRU	0.87	0.89	0.87	0.94
	LSTM	0.85	0.85	0.89	0.89
	WLSTM	0.85	0.85	0.83	0.84
	LSTM with Attention	0.84	0.84	0.82	0.83
Tanzania	GRU	0.94	0.90	0.92	0.92
	LSTM	0.86	0.92	0.89	0.87
	Weighted LSTM	0.84	0.87	0.86	0.84
	LSTM with Attention	0.82	0.84	0.88	0.85

**Source:** Author (2025)

**Note:** Gated Recurrent Unit (GRU), Simple Long Short-Term Memory (LSTM), LSTM with Attention, and Weighted LSTM (W-LSTM)



**Figure 5:** Stock Price correlation heatmap 5.



**Figure 6:** Stock prediction results of the models.

### CONCLUSIONS AND RECOMMENDATION

This study set out to evaluate the predictive performance of various deep learning models - namely Gated Recurrent Unit (GRU), Simple Long Short-Term Memory (LSTM), LSTM with Attention, and Weighted LSTM (W-LSTM) - in forecasting stock prices across four African markets: Uganda, Nigeria, Kenya, and Morocco. The findings demonstrate that GRU achieved the highest predictive performance in Uganda and Nigeria, attaining  $R^2$  scores of 0.94 and 0.93, respectively. These results suggest that GRU's simpler architecture and effective handling of short-term dependencies make it particularly well-suited for capturing stock patterns in relatively less complex markets. Meanwhile, the Simple LSTM model delivered better results in Kenya ( $R^2 = 0.92$ ) and Morocco ( $R^2 = 0.91$ ), indicating that slightly longer memory retention may be beneficial in those contexts.

Although LSTM with Attention was designed to enhance the model's ability to focus on significant time steps, its performance did not exceed that of the GRU or Simple LSTM models in

any of the countries studied. Similarly, the Weighted LSTM model, despite its theoretical strength in learning the relative importance of different time steps, failed to show significant improvement and recorded lower  $R^2$  scores in Uganda (0.85) and Nigeria (0.87). This points to a possible trade-off between architectural complexity and predictive performance in emerging and frontier markets, where data quality, market efficiency, and noise may limit the benefits of more complex modeling techniques. The overall results favor lightweight yet robust architectures, which deliver high performance with lower computational cost and model complexity.

Given the consistent performance of GRU and Simple LSTM models, it is recommended that financial institutions, capital market authorities, and fintech innovators in Africa integrate these models into stock price forecasting and portfolio management systems. These models can also support regulatory authorities and central banks in anticipating market trends, enhancing economic policy formulation, and strengthening financial stability. Training programs and academic curricula should emphasize the application of deep learning models in financial forecasting to develop technical expertise among analysts, data scientists, and financial engineers across the continent. Fintech startups can leverage these findings to develop low-cost, high-impact financial tools tailored to local markets. The study is not without limitations. First, the research focused on only four African countries, which may limit the generalizability of the results across other regions, particularly more developed or highly volatile markets. Second, the models used in this study primarily relied on historical stock price data, excluding macroeconomic indicators, market sentiment, geopolitical factors, and high-frequency trading signals that could significantly affect model performance. Third, while  $R^2$  was the primary evaluation metric for model comparison, incorporating additional measures such as directional accuracy, confusion matrices for classification-based forecasting, or even profit-and-loss simulations could provide a more holistic performance assessment.

Future studies should expand the geographical scope to include more African stock markets - such as South Africa, Egypt, and Ghana - to enhance the external validity of the findings. Researchers are also encouraged to develop hybrid forecasting frameworks by integrating GRU or LSTM with exogenous macroeconomic variables, technical indicators, and sentiment analysis from news and social media. Additionally, exploring ensemble learning methods, such as combining GRU and LSTM models with boosting or bagging techniques, may yield more robust predictive models. Another avenue of exploration is the use of interpretable and explainable AI frameworks (e.g., SHAP or LIME) to make these models more transparent and acceptable to stakeholders in the financial domain.

### References

- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1412.3555*.
- Ballings, M., Van den Poel, D., Hespels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20), 7046–7056.
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669.
- Gu, S., Kelly, B. T., & Xiu, D. (2018). Empirical asset pricing via machine learning. *Chicago Booth Research Paper*, No. 18-04.

- Oyasor, E. I. (2025). Forecasting Global Stock Prices Using Gated Recurrent Unit, Long Short-Term Memory Weighted-LSTM and LSTM with Attention. *Malete Journal of Accounting and Finance*, 6 (1), 147 – 161
- Gündüz, H., Çataltepe, Z., & Yaslan, Y. (2017). Stock daily return prediction using expanded features and feature selection. *Turkish Journal of Electrical Engineering & Computer Sciences*, 25(6), 4829–4840.
- Gururaj, V., Shriya, V. R., & Ashwini, K. (2019). Stock market prediction using linear regression and support vector machines. *International Journal of Applied Engineering Research*, 14, 1931–1934.
- Hamzac,ebi, C., Akay, D., & Kutay, F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert Systems with Applications*, 36(2), 3839–3844.
- Imandoust, S. B., & Bolandraftar, M. (2014). Forecasting the direction of stock market index movement using three data mining techniques: The case of Tehran Stock Exchange. *International Journal of Engineering Research and Applications*, 4(6), 106–117.
- Kolte, G., Kini, V., Nair, H., & Babu, K. S. S. (2022). Stock market prediction using deep learning. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 10(4). <https://doi.org/10.22214/ijraset.2022.41159>.
- Kumar, M., & Thenmozhi, M. (2014). Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *International Journal of Banking Accounting and Finance*, 5, 284–308.
- Kyriakidis, M., de Winter, J. C. F., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., Martens, M. H., Bengler, K., Andersson, J., Merat, N., Reed, N., Flament, M., & Hagenzieker, M. (2019). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, 20(3), 223–249. <https://doi.org/10.1080/1463922X.2017.1293187>
- Li, L., Wu, Y., Ou, Y., Li, Q., Zhou, Y., & Chen, D. (2017). Research on machine learning algorithms and feature extraction for time series. *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 1–5.
- Mann, J., & Kutz, J. N. (2016). Dynamic mode decomposition for financial trading strategies. *Quantitative Finance*, 16, 1643–1655.
- Polamuri, S. R., Srinivas, K., & Mohan, A. K. (2019). Stock market prices prediction using random forest and extra tree regression. *International Journal of Recent Technology and Engineering*, 8, 1224–1228.
- Pilli, S. K., Nallathambi, B., George, S. J., & Diwanji, V. (2014). eAGROBOT: A robot for early crop disease detection using image processing. *IEEE International Conference on Electronics and Communication Systems (ICECS)*.
- Parray, I. R., Khurana, S. S., Kumar, M., & Altalbe, A. A. (2020). Time series data analysis of stock price movement using machine learning techniques. *Soft Computing*, 24, 16509–16517.
- Seber, G. A. F., & Lee, A. J. (2012). *Linear regression analysis*. John Wiley & Sons.
- Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017). Stock price prediction using LSTM, RNN, and CNN-sliding window model. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1643–1647.
- Soni, P., Tewari, Y., & Krishnan, D. (2022). Machine learning approaches in stock price prediction: A systematic review. *Journal of Physics Conference Series*, 2161, 012065.
- Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation*, 5, 1–20.