DSE Stock Price Simulation: A Comparative Study

S M Rakib Hasan
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
sm.rakib.hasan@g.bracu.ac.bd

Farah Binta Haque
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
farah.binta.haque@g.bracu.ac.bd

Md Humaion Kabir Mehedi
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
humaion.kabir.mehedi@g.bracu.ac.bd

Annajiat Alim Rasel
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
annajiat@gmail.com

***Abstract*—This research paper presents an innovative simulation model for forecasting the stock prices of the Dhaka Stock Exchange (DSE). The model employs a multi-faceted approach, integrating advanced machine learning algorithms, including deep neural networks, along with generic machine learning techniques. The primary objective is to explore the effectiveness of these varied computational models in accurately predicting stock prices in the DSE context. The paper meticulously compares the performance of the simulation models, assessing their predictive accuracy, computational efficiency, and robustness in handling the volatile nature of stock market data. The findings of this study are expected to provide significant insights into the applicability of cutting-edge AI methodologies in financial market analysis and contribute to the development of more reliable and efficient tools for financial forecasting. For reproducibility and future research endeavours, the code is made open source and can be found at https://bit.ly/3QVLVTN**

***Index Terms*—RNN, Machine Learning, LSTM, GRU, Stock Price Prediction, Financial Forecasting**

I. Introduction

The realm of financial forecasting has always been a fertile ground for the application of advanced computational techniques. With the advent of more sophisticated machine learning (ML) algorithms, the potential to revolutionize stock market predictions has grown exponentially. The Dhaka Stock Exchange (DSE), representing an emerging financial market, presents a unique and challenging landscape for such technological interventions. This study is rooted in the exploration and application of advanced ML techniques - specifically, deep neural networks, in contrast to traditional ML algorithms - to model and predict stock prices within the DSE.

The unpredictability and volatility inherent in stock markets make them complex systems to model. Traditional financial models often fall short of capturing the nonlinear and dynamic nature of market movements. The advent of machine learning and AI has opened new avenues for financial analysts and researchers, offering tools that can learn from vast amounts of data and identify patterns that are often imperceptible to human analysis.

In this paper, we delve into the application of cutting-edge ML techniques, which have shown exceptional capabilities in other domains like natural language processing and image recognition. Their application in financial forecasting, particularly in a less-studied market like the DSE, is both innovative and challenging. Alongside these, we also explore the performance of generic ML algorithms, which have been the backbone of many predictive models in finance. The comparison aims to not only highlight the strengths and weaknesses of each approach but also to shed light on their suitability and adaptability to a market environment like the DSE.

The core contribution of this paper lies in its comprehensive approach to modeling stock prices in the DSE using a range of ML techniques, and in its analytical comparison of these methods. By doing so, it seeks to provide insights into the effectiveness of advanced AI methods in stock price prediction and to pave the way for future research and practical applications in financial market analysis.

II. Previous Works

The rapid advancement in machine learning (ML) and deep learning has led to significant breakthroughs in financial forecasting, particularly in stock market prediction. This literature review focuses on a selection of studies that exemplify the diverse methodologies and findings in this dynamic field.

The study [1] conducted a comprehensive study on the effectiveness of Long Short-Term Memory (LSTM) models for predicting stock prices. Their research provided a detailed analysis of LSTM's ability to model the temporal dependencies of stock market prices. A significant aspect of their work was the emphasis on hyperparameter tuning and its impact on the model's performance. They demonstrated that with proper configuration, LSTM models could significantly outperform traditional time-series prediction models, offering a more nuanced understanding of market dynamics.

Another groundbreaking study involved the use of Deep Convolutional Generative Adversarial Networks (DCGAN) for forecasting stock prices [2]. This research took an innovative approach by applying DCGANs, a class of AI models renowned in image processing, to the financial domain. The study highlighted the model's proficiency in dealing with the noisy and volatile nature of stock market data, achieving superior forecasting accuracy. The empirical performance of the DCGAN model was tested on the FTSE MIB index, showing that it outperformed standard linear and non-linear forecasting methods in both single-step and multi-step predictions. This research opened new avenues in financial forecasting, suggesting the untapped potential of complex neural network architectures in this field.

In the realm of convolutional neural networks (CNN), a notable study presented a 2-D CNN model specifically designed for stock trading [3]. The model leveraged CNN's prowess in detecting intricate patterns and trends, a crucial aspect of financial trading. This study showcased the versatility of CNNs, extending their application from image and video analysis to financial time series data. The model's success indicated that deep learning could offer novel insights into market trends, assisting in more informed trading decisions.

Another pivotal study focused on forecasting stock prices from high-frequency trading data using CNNs [4]. This research underscored the importance of incorporating high-frequency data, such as the limit order book information, in predictive models. By applying CNNs to this type of data, the study showed how deep learning could capture the intricacies and rapid movements within the stock market, providing a more accurate and timely prediction of stock prices. This approach highlighted the synergy between advanced ML techniques and the rich, complex datasets characteristic of financial markets.

The paper by [5] is a pioneering work that applies deep learning neural networks (DLNNs) to the challenging task of predicting short-term stock price movements in the Chinese market. The paper recognizes that the stock market is a complex and nonlinear system that requires advanced and flexible methods to capture its patterns and dynamics. The paper proposes to model stock price charts as images, which can preserve the spatial and temporal information of the price movements. The paper then uses DLNNs, which are powerful and versatile models that can learn from high-dimensional and complex data, to process the price chart images and generate predictions. It suggests that the DLNNs can be applied to other markets and time horizons and that future research can explore other types of inputs and outputs for the DLNNs.

The paper [6] proposes a novel approach to predict the future price of stocks using the Transformer model. It uses a new encoding technique called time2vec to represent the time series features of the stock data. The time2vec technique transforms the time series data into a high-dimensional vector space, where the temporal patterns and trends can be better captured by the Transformer model. It evaluates the performance of the proposed model on the Dhaka Stock Exchange (DSE), the leading stock exchange in Bangladesh, using four years of historical data from 2018 to 2021. The paper compares the proposed model with several baseline models, such as the LSTM, GRU, and ARIMA models, and shows that the proposed model achieves superior results in terms of accuracy, mean squared error, and mean absolute error. The paper also conducts an ablation study to analyze the impact of different components of the proposed model, such as the time2vec technique, the number of layers, and the number of attention heads.

The paper [7] examines the factors that affect the growth and performance of the stock market in Bangladesh, using both primary and secondary data sources. The paper uses various statistical methods, such as correlation and regression analysis, to test the relationship between the stock market development and the independent variables, such as the stock market size, liquidity, volatility, asset pricing, and regulatory and institutional indicators. The paper finds that all the independent variables have a positive and significant impact on the stock market development and that the stock market size and liquidity are the most influential factors. The paper also provides some policy implications and suggestions for enhancing the stock market development in Bangladesh, such as reducing the gap between the demand and supply of shares, developing the financial intermediaries, and creating awareness among the investors.

The paper [8] introduces a novel artificial neural network (ANN) model that combines technical and fundamental analysis to predict stock prices at stock exchange markets, such as the Nairobi Securities Exchange and the New York Stock Exchange. The paper develops a prototype and tests it on 2008-2012 data from these markets. The paper claims that the proposed model outperforms traditional methods, such as linear regression, moving average, and autoregressive integrated moving average, as well as existing ANN models, in terms of accuracy and efficiency. The paper uses a feedforward neural network with a backpropagation algorithm and a genetic algorithm to optimize the network parameters. The paper evaluates the performance of the model using mean absolute percentage error (MAPE) and root mean square error (RMSE) metrics.

III. Methodology

Our research methodology is grounded in a rigorous and systematic approach, designed to ensure the integrity and reproducibility of our results. The process encompasses a comprehensive suite of procedures, including meticulous data collection from the Dhaka Stock Exchange, thorough exploratory data analysis to uncover deep market insights, and careful data preparation to tailor the input for optimal machine learning model performance. Subsequently, we deployed a variety of advanced computational models, each subjected to a robust result collection protocol, followed by a detailed comparative analysis to elucidate the efficacy of each model in the context of financial forecasting. This section elucidates each step of our methodology, providing a transparent view of the mechanics of our research. An overview of our steps is depicted in Figure 1

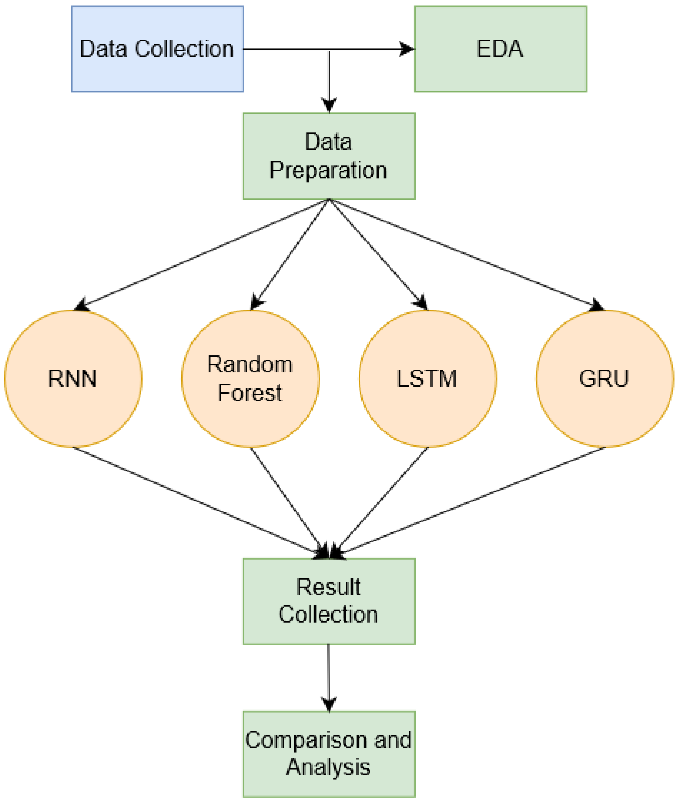

Fig. 1. Our work sequence.

A. Data Collection

The data collection process was a critical foundation of the study. For this research, we have accumulated the data from www.investing.com. We identified the DSE as the primary data source given its representation of an emerging financial market with unique characteristics. The dataset spans ten years, from January 2013 to November 2023, providing a comprehensive view of the market over different economic cycles. Historical stock prices, including opening, closing, high, low, and volume of trades, were extracted. This was done using direct download features from the DSE or third-party financial data providers. The collected data were verified for completeness and accuracy. Any missing dates or anomalies were checked against market events, holidays, or data source issues.

B. EDA

EDA was the next step, vital for understanding the data and informing the subsequent preparation steps. We employed various visualization techniques such as line charts for stock price trends, volume bars, and candlestick charts to assess the volatility and price movements over time. Descriptive statistics were used to summarize the data, including measures of central tendency and dispersion like mean, median, variance, and standard deviation. To identify potential relationships, we conducted a correlation analysis between different stocks and between stock prices and other market indicators. Important features that could influence stock prices, such as moving averages, were identified for model input. We searched for outliers or anomalies that could be due to market manipulation, data errors, or significant market events.

C. Data Preparation

The data preparation phase involved several steps to transform raw data into a format suitable for machine learning models. We addressed missing values, duplicates, or incorrect entries. Missing values were handled through imputation or deletion, depending on their significance and volume. Stock price data were normalized to bring them to a common scale without distorting differences in the ranges of values. This is crucial for models that are sensitive to the scale of input data. New features were derived from the existing data to improve model performance. These included technical indicators like moving averages, the Relative Strength Index (RSI), and others that could provide additional insights into the market dynamics. We transformed the time-series data into a supervised learning format, where past stock prices were used to predict future prices. This involved creating lag features and restructuring the dataset accordingly. The dataset was split into training, validation, and test sets, ensuring that the models would be trained on one subset of the data and validated and tested on unseen data to assess their generalization capabilities. For models like RNN, LSTM, and GRU, the data were further processed into sequences to capture temporal dependencies effectively.

D. Result Collection and Analysis

Each model generated predictions for the test set, which were collected for comparison. The output included predicted stock prices for the given horizon, which, in this study, likely ranged from the next day to several days ahead. Models were benchmarked against each other as well as against a baseline model, which could be a simple statistical method like a moving average or a "naive" predictor that assumed the stock price would remain unchanged. Based on performance metrics, models were ranked to determine which algorithm was most effective in predicting stock prices for the DSE. Models were evaluated for robustness by checking their performance over different time periods and under varying market conditions. This helped in assessing their reliability and practical applicability.

By meticulously executing these steps, the study ensured that the subsequent machine learning models were trained and evaluated on high-quality, well-understood data, increasing the robustness and reliability of the forecasting results.

IV. Experimental Setup

The experimental framework of our study was meticulously designed to evaluate the predictive performance of a suite of advanced machine-learning models. We carefully selected and tailored each model to address the unique challenges posed by the stock price prediction task, leveraging their respective strengths in capturing temporal dependencies and complex patterns within the high-dimensional data from the Dhaka Stock Exchange (DSE).

A. Random Forest (RF)

The Random Forest algorithm leverages an ensemble of decision trees to make robust predictions and mitigate overfitting. It operates by constructing numerous trees during training and outputs the average prediction of the individual trees for regression tasks. The prediction of the RF model for a given input vector $\mathbf{x}$ is defined by the average of the predictions from all the individual trees $\{h(\mathbf{x}, \Theta_k), k = 1, ..., K\}$, where Θ_k are the random vectors independently sampled from a distribution Θ and K is the number of trees in the forest:

$$\hat{y} = \frac{1}{K} \sum_{k=1}^{K} h(\mathbf{x}, \Theta_k) \tag{1}$$

B. Long Short-Term Memory (LSTM)

Long Short-Term Memory networks are a type of recurrent neural network designed to learn order dependence in sequence prediction problems. LSTMs have a chain-like structure with repeating modules of neural networks, consisting of different interacting layers. The key equations governing an LSTM unit are as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4}$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5}$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$
$$h_t = o_t * \tanh(C_t) \tag{7}$$

We trained the model for 30 epochs using a learning rate of 0.001 and employed the Adam optimizer. The training process involved fitting the model to the training data with a batch size of 32, a mean squared error loss function, and a validation split of 20% to assess performance on a validation subset.

C. Gated Recurrent Unit (GRU)

Gated Recurrent Units simplify the LSTM design with a similar performance, merging the forget and input gates into a single update gate and combining the cell state and hidden state. The GRU's architecture is governed by the following equations:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{8}$$
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{9}$$
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{10}$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{11}$$

We trained the GRU model for 30 epochs using the same learning rate and optimizer as the LSTM. The model was built with a GRU layer consisting of 50 units, followed by a dropout layer with a dropout rate of 0.2 and a dense output layer with one unit

D. Recurrent Neural Networks (RNNs)

Recurrent Neural Networks are a basic form of neural network where connections between units form a directed cycle. This creates an internal state that allows the network to exhibit temporal dynamic behaviour. A simple RNN is described by the equation:

$$h_t = \tanh(W_{hh} \cdot h_{t-1} + W_{xh} \cdot x_t + b_h) \tag{12}$$
$$y_t = W_{hy} \cdot h_t + b_y \tag{13}$$

We trained the RNN model by constructing a Sequential model comprising a SimpleRNN layer with 50 units, followed by a dropout layer with a 0.2 dropout rate and a dense output layer with one unit. The model was compiled using the Adam optimizer with a learning rate of 0.001 and a mean squared error loss function.

E. Loss Functions

We have employed two loss functions or metrics for our assessment of performance in our study.

1) Mean Squared Error (MSE): The Mean Squared Error (MSE) is a commonly employed loss function in regression tasks, including neural network training. It measures the average squared difference between predicted ($\hat{Y}_i$) and actual (Y_i) values, computed across n samples. The MSE is defined by the equation:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

A lower MSE signifies better model performance, with zero indicating a perfect match between predictions and actual values.

2) Coefficient of Determination (R^2): The Coefficient of Determination, often denoted as R^2, is a metric used to assess the goodness of fit of a regression model. It quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables. R^2 is calculated using the following formula:

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

Here, Y_i represents the actual target value, $\hat{Y}_i$ is the predicted value, and $\bar{Y}$ is the mean of the actual values. The R^2 value ranges between 0 and 1, where 1 indicates a perfect fit, and values closer to 0 suggest poorer model performance. R^2 is a valuable metric for evaluating the predictive power of regression models.

V. Results and Discussion

This study employed various computational models, including Random Forest, LSTM, GRU, and RNN, for predicting stock prices. The data from the Dhaka Stock Exchange was analyzed, drawing from methodologies established in previous works [9], [10].

A. Overview of Analysis

The analysis conducted in this study utilized various computational models, including Random Forest, LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), and RNN (Recurrent Neural Network), to predict stock prices. Data from the Dhaka Stock Exchange was analyzed to gain insights into market trends, volatility, and predictability.

B. Volatility Analysis

Volatility in the stock market was analyzed using rolling standard deviation on the daily price changes. Figure 2 illustrates the 30-day rolling volatility, revealing significant fluctuations indicative of market instability, particularly during high-impact events such as the onset of the COVID-19 pandemic. Volatility was analyzed using a rolling standard deviation method on daily price changes, similar to the approach described by Schwert (1989) [11].

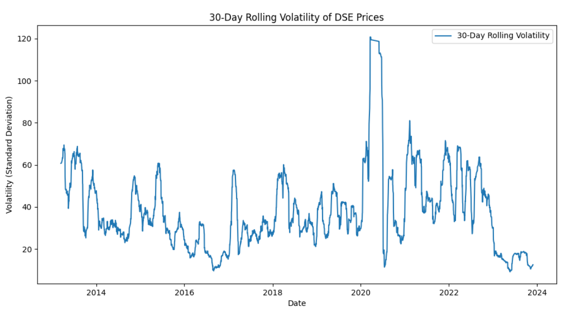

Fig. 2. 30-Day Rolling Volatility of Dhaka Stock Exchange Prices.

C. Model Performance

The performance metrics, MSE and R2, were used to evaluate model effectiveness, following the methodology outlined by Hyndman and Athanasopoulos (2018) [12]. Table I presents a summary of these metrics, highlighting the strengths and weaknesses of each model.

Model	MSE	R-squared
Random Forest	3204.81	0.982
LSTM	5698.70	0.948
GRU	3400.76	0.969
RNN	2907.93	0.974

TABLE I

PERFORMANCE METRICS OF DIFFERENT MODELS.

D. Stock Market Predictions

Predictive models offered insights into future market trends. Figures 3, 4, 5, and 6 depict the actual vs. predicted stock prices for each model, demonstrating their predictive capabilities.

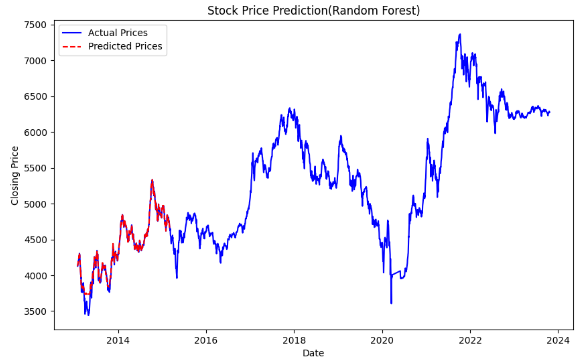

Fig. 3. Random Forest model prediction vs. actual stock prices.

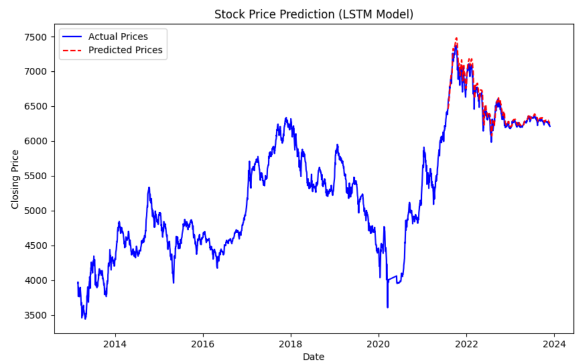

Fig. 4. LSTM prediction vs. actual stock prices.

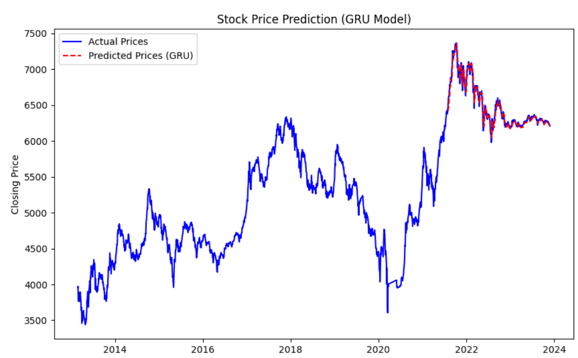

Fig. 5. GRU prediction vs. actual stock prices.

REFERENCES

[1] S. Selvin, V. Ravi, E. Gopalakrishnan, V. Menon, and S. Kp, "Stock price prediction using lstm, rnn and cnn-sliding window model," 09 2017, pp. 1643–1647.

[2] A. Staffini, "Stock price forecasting by a deep convolutional generative adversarial network," *Frontiers in Artificial Intelligence*, vol. 5, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frai.2022.837596

[3] U. Gudelek, A. Boluk, and M. Ozbayoglu, "A deep learning based stock trading model with 2-d cnn trend detection," 11 2017, pp. 1–8.

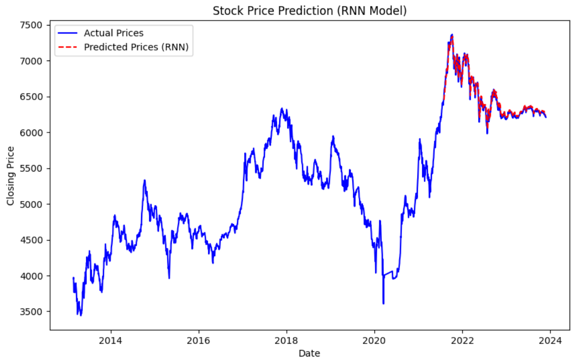

Fig. 6. RNN prediction vs. actual stock prices.

[4] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Forecasting stock prices from the limit order book using convolutional neural networks," 07 2017, pp. 7–12.

[5] Q. Liu, Z. Tao, Y. Tse, and C. Wang, "Stock market prediction with deep learning: The case of china," *Finance Research Letters*, vol. 46, p. 102209, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1544612321002762

[6] T. Muhammad, A. B. Aftab, M. Ibrahim, M. M. Ahsan, M. M. Muhu, S. I. Khan, and M. S. Alam, "Transformer-based deep learning model for stock price prediction: A case study on bangladesh stock market," *International Journal of Computational Intelligence and Applications*, vol. 22, no. 03, Apr. 2023. [Online]. Available: http://dx.doi.org/10.1142/S146902682350013X

[7] M. Z. Hosain and R. Nowreen, "The determinants of stock market development: A study on dhaka stock exchange limited (dse)," *European Journal of Business and Management*, vol. 7, pp. 85–93, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:55956942

[8] B. W. Wanjawa and L. Muchemi, "Ann model to predict stock prices at stock exchange markets," 2014.

[9] Y. Zhang and J. Wang, "Stock market prediction of s&p 500 via combination of improved bco approach and bp neural network," *Expert Systems with Applications*, vol. 77, pp. 60–69, 2017.

[10] J. Brownlee, *Deep Learning for Time Series Forecasting*. Machine Learning Mastery, 2018.

[11] G. W. Schwert, "Why does stock market volatility change over time?" *The Journal of Finance*, vol. 44, no. 5, pp. 1115–1153, 1989.

[12] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2018.