

Multi-Head Attention Based LSTM Model for Forecasting Bangladeshi Stock Market Prices

Md. Khurshid Jahan
Dept. of ECE

North South University
Dhaka, Bangladesh
khurshid.jahan@northsouth.edu

Md. Estehad Chowdhury
Dept. of Business Administration
Shahjalal University of Science & Technology
Sylhet, Bangladesh
estehadchow@gmail.com

Saroar Jahan Shuba
Dept. of MSCQM
Lamar University
Texas, USA
email.shouvo@gmail.com

Fardin Sabahat Khan
Dept. of Accounting & MIS
University of Delaware
Delaware, USA
fardinsk@udel.edu

Md Mahafuzur Rahman
Dept. of MPE
Ahsanullah University of Science & Technology
Dhaka, Bangladesh
m.mahafuzur.ra@gmail.com

Md Omar Faruk
Dept. of CSE
Ahsanullah University of Science & Technology
Dhaka, Bangladesh
f.omarfaruk01@gmail.com

Md Ashfaquul Azam Chowdhury
Dept. of CSE
Ahsanullah University of Science & Technology
Dhaka, Bangladesh
azamashfaquul12@gmail.com

S. M. Jahidul Islam
Dept. of Computer Science & Mathematics
Bangladesh Agricultural University
Mymensingh, Bangladesh
jahidul.ict@bau.edu.bd

Mithila Arman
Dept. of CSE
BRAC University
Dhaka, Bangladesh
mithila.arman@g.bracu.ac.bd

Abstract—Forecasting stock prices is crucial for investment and risk management, offering insights into market trends and aiding investors in making informed decisions. Traditional time-series models, such as LSTM and GRU, have demonstrated promise in capturing temporal dependencies within stock market data. However, these models often fall short in identifying complex patterns due to limited attention mechanisms, particularly in emerging markets with unique economic factors like the Bangla stock market. Addressing this gap, we propose a novel Multi-Head Attention and LSTM-based architecture that combines the sequential processing capabilities of LSTM with multi-head attention layers to better capture contextual dependencies and subtle fluctuations in stock price movements. Our approach is evaluated on datasets from six major Bangladeshi banks, including Islami Bank, Dhaka Bank, Arab Bangladesh (AB) Bank, Bank Asia, Eastern Bank Limited (EBL), and City Bank. Results demonstrate that our model outperforms traditional models across all metrics. The integration of multi-head attention allows for nuanced attention allocation across time steps, significantly enhancing forecasting accuracy. This enhanced model can serve as a valuable tool for investors and financial analysts, supporting better forecasting for risk assessment and

investment strategy formulation. The framework demonstrates potential scalability across other emerging markets, offering an adaptable solution for financial forecasting in regions with complex market dynamics.

Index Terms—Stock Price Prediction, LSTM, Multi-Head Attention

I. INTRODUCTION

The stock market serves as a vital component of the global economy, reflecting the financial health and stability of a nation. It provides a platform for businesses to raise capital and for investors to earn returns, making accurate stock price forecasting critical for informed decision-making in investments and risk management. However, predicting stock market prices remains a challenging task due to the dynamic and volatile nature of financial markets. In Bangladesh, the stock market has gained significant attention as a growing economic hub in South Asia. With institutions like the Dhaka Stock Exchange playing a crucial role, the market exhibits unique characteristics influenced by regional

economic factors and global trends. According to the Dhaka Stock Exchange official site, it hosts over 397 listed companies, making it a key area for financial analysis and forecasting [1]. Despite its importance, the Bangladeshi stock market has been less explored in terms of adopting advanced computational models for forecasting. Technological advancements have introduced various methods for stock price prediction, ranging from traditional statistical models like ARIMA to advanced machine learning (ML) approaches, including LSTM, GRU, and transformer models. These computer-aided systems analyze vast amounts of historical and real-time data to uncover hidden patterns, offering significant improvements over manual analysis. Machine learning-based systems, in particular, have shown remarkable potential in handling complex temporal and contextual dependencies, providing more accurate forecasts. However, challenges such as data quality issues, over-fitting, and computational costs remain barriers to widespread adoption. Furthermore, reliance on historical data alone can limit the predictive accuracy of these systems in volatile markets. Despite these limitations, computer-aided systems, particularly ML-driven approaches, hold great promise in transforming stock price prediction. With continued advancements, these systems are poised to become indispensable tools for investors and analysts, bridging the gap between market complexity and decision-making.

II. RELATED WORK

Forecasting stock market prices in Bangladesh has gained traction through various machine learning approaches. The use of Long Short-Term Memory (LSTM) models for predicting stock market prices in Bangladesh has become popular because they can capture complex patterns in financial data over time. Studies have shown that LSTM models work well for stock price prediction, especially when combined with optimization techniques. Huang et al. developed an LSTM model with an Evolutionary Operating-weights strategy. It has greatly improved investment returns by accurately forecasting future prices [2]. Similarly, Bukhari et al. created a hybrid ARFIMA-LSTM model that mixes autoregressive fractional integrated moving average filtering with LSTM. This approach reduced volatility and overfitting issues, resulting in better accuracy [3]. Optimization algorithms like the Adaptive Genetic Algorithm and Artificial Rabbits Optimization have also been used to adjust LSTM hyperparameters, giving better results than traditional methods [4], [5]. External factors such as news sentiment and economic indicators has been found to improve prediction accuracy by about 24% when using multivariate LSTM models [?], [6]. Other

advanced methods, like transformer models and multi-modal deep learning approaches, have also shown good results. These methods handle the volatility of stock prices better and perform better than traditional models like ARIMA [7]. Together, these advancements highlight the potential of machine learning to provide reliable stock price predictions, helping investors make better decisions [8]. Research also shows that LSTM models are highly effective for predicting stock prices because they can identify complex patterns in historical data. This ability is crucial given the volatile nature of stock markets [9], [10]. Studies also highlight that stacking LSTM layers improves prediction accuracy by capturing different levels of information [11]. Hybrid models combining LSTM with Convolutional Neural Networks (CNN) have performed well in predicting prices on the Dhaka Stock Exchange [12]. Additionally, preprocessing methods like data scaling and feature engineering help optimize LSTM performance. These techniques provide better insights into market trends, supporting investors in making decisions [13].

III. RESEARCH BACKGROUND

In this section, we provide a detailed description of our approach for stock price prediction using LSTM and Multi-Head Attention. The entire process, including data preprocessing, model construction, and evaluation, is outlined in Algorithm 1. This algorithm highlights the key steps involved in training and assessing the model's performance.

A. Dataset Description

The dataset for this study was collected from the official website of the Dhaka Stock Exchange (DSE) and includes financial records of various banks listed on the exchange. Data collection spans from 2012 to 2024, covering more than a decade of financial activity across different banks, providing comprehensive insights into trends and fluctuations within the sector over time. Table I summarizes the dataset, detailing the distribution of total, training, and testing samples across six banks, including Islami Bank, Dhaka Bank, AB Bank, Bank Asia, EBL Bank, and City Bank. The dataset includes sufficient samples for each bank, ensuring robust model training and testing. The dataset is publicly accessible on Figshare [5] for further research and analysis in related financial forecasting and machine learning studies.

B. Dataset Preprocessing

The dataset used in this study contains five columns which are Date, Open, High, Low, Close, and Volume. For this paper, only the Close price is used to predict the stock price, as the closing price is a widely used metric in financial forecasting. To prepare the data for model

Algorithm 1 LSTM and Multi-Head Attention Model for Stock Price Prediction

Require: Dataset \mathcal{D} =
 {Date, Open, High, Low, Close, Volume}
1: Select only Close price: $X = \mathcal{D}[\text{Close}]$
2: Normalize Close price: $X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$
3: Split dataset: 80% train, 20% test
4: Create sliding window: $X_i = [X_i, \dots, X_{i+49}], y_i = X_{i+50}$
5: Reshape: $X_{\text{train}}, X_{\text{test}}$ →
 [samples, time_steps, features]
6: Build model:
7: Input layer: Input(shape = (50, 1))
8: Add LSTM(128, return_sequences=True)
9: Add LSTM(64, return_sequences=True)
10: Add MultiHeadAttention(n_head=3,
 d_model=300, d_k=64, d_v=64)
11: Add GlobalAveragePooling1D
12: Add GlobalMaxPooling1D
13: Concatenate(GAP, GMP)
14: Add Dense(64, activation='relu')
15: Add Dense(1, activation='sigmoid')
16: Compile: optimizer = RMSprop, loss = MAE
17: Train: epochs = 50, batch size = 32
18: Predict on test data
19: Inverse transform: X_{pred} =
 scaler.inverse_transform(y_{pred})
20: Calculate metrics: MSE, MAE, RMSE, R^2
21: Plot: Actual vs Predicted
22: **return** Trained model

TABLE I: Distribution of Total, Training, and Testing Samples Across Different Banks

Dataset	Total Sample	Train Data	Test Data
Islami Bank	2661	2128	533
Dhaka Bank	2666	2132	534
AB Bank	2670	2136	534
Bank Asia	2633	2106	527
EBL Bank	2670	2136	534
City Bank	2674	2139	535

training, Min-Max normalization is applied to the closing prices. This technique scales the data to a range between 0 and 1, improving the performance and convergence of machine learning models. The Min-Max normalization is applied by calculating the minimum and maximum values of the dataset and transforming each value using the following formula:

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

where X represents an individual closing price, X_{\min} is the minimum closing price in the dataset, and X_{\max} is

the maximum closing price. Once the data is normalized, the dataset is split into training and testing sets, with 80% of the data used for training the model and the remaining 20% reserved for testing. This division ensures that the model is trained on a substantial portion of the data while also being evaluated on unseen data for performance assessment.

C. Time Series Dataset Preparation

For the time series forecasting task, we use a sliding window approach to structure the dataset, where each instance consists of a sequence of past values used as input to predict the next value in the series. This method is illustrated in Figure 1, where overlapping sequences are created to capture temporal dependencies in stock prices. In this approach, we define a parameter called *time_step* set to 50, which represents the number of consecutive days used as input for predicting the following day's stock price. Starting from day 1, each instance includes $r = 50$ days of input data and one target day (day $r + 1$) as output. The next instance begins from day 2 and includes days 2 to $r + 1$ as input, predicting day $r + 2$, and so forth. This process continues until the end of the dataset, creating sequences that cover days $N - r$ to N , where N represents the total number of days in the dataset. The training and test datasets generated are reshaped to fit the input requirements of the LSTM model. Specifically, each input instance is formatted as a 3D array with dimensions [samples, time steps, features]. For example, in our AB bank dataset, this reshaping results in training data with dimensions (2085, 50, 1) and test data with dimensions (483, 50, 1). Here, "samples" represents the number of training or test instances, "time steps" corresponds to the *time_step* of 50 days, and "features" is set to 1, as we are working with univariate stock price data. As illustrated in Figure 1, each sequence captures a window of 50 consecutive days, with the last day in each window serving as the target output for forecasting. This sliding window method allows the model to learn from the temporal structure of stock price movements, enhancing its ability to recognize patterns and trends, ultimately leading to more accurate predictions.

D. Methodology

This paper proposes a multi-head attention mechanism combined with LSTM layers to enhance temporal feature extraction and capture complex dependencies in sequential data. The methodology begins by taking an input tensor of shape (50, 1), where 50 represents the sequence length corresponding to the time steps, and 1 represents the feature dimension. After preprocessing, the input tensor is passed into the first LSTM layer with 128 units and *return_sequences = True* to capture

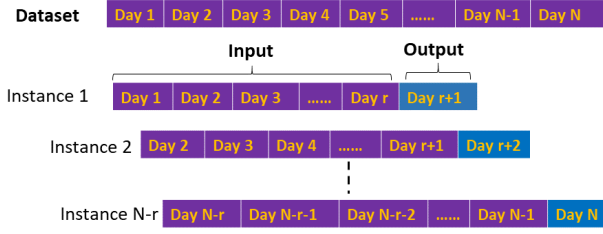


Fig. 1: Time series dataset preparation with sliding window approach where each instance includes 50 days of historical data as input to forecast the following day's price.

initial temporal dependencies and generate a transformed feature set. The output of this layer is then passed into a second LSTM layer with 64 units, also with *return_sequences = True*, which further refines the temporal feature representation, enhancing the model's ability to capture more complex sequential patterns.

The attention mechanism is a crucial component of this model, implemented through a multi-head attention layer (refer to Algorithm 1). This layer allows the model to attend to different temporal segments simultaneously, enhancing its capacity to learn long-term dependencies. To achieve this, the model creates query (Q), key (K), and value (V) vectors for each head, transforming them with learned weights. In this implementation, the model uses 3 attention heads, with each head using a key, query, and value dimension of 64. This multi-head attention mechanism computes scaled dot-product attention for each head, defined as:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{Q \cdot K^T}{\sqrt{d_k}} \right) \cdot V \quad (2)$$

where d_k is the dimension of the key vectors, and the scaling factor $\sqrt{d_k}$ prevents the dot products from growing excessively large. The 3 heads allow the model to capture diverse aspects of the temporal relationships in the data.

Each head independently computes its attention scores and then concatenates the heads' outputs to form a single tensor. This concatenated output undergoes a final linear transformation, yielding a tensor that represents attention-weighted temporal features. Layer normalization (LN) follows, which stabilizes training by normalizing inputs across feature dimensions. The normalization layer computes:

$$\text{LN}(x) = \frac{\gamma(x - \mu)}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (3)$$

where μ and σ are the mean and standard deviation of x , and γ and β are learnable parameters.

The aggregated output from the multi-head attention layer undergoes global pooling, where Global Average Pooling (GAP) and Global Max Pooling (GMP) extract robust representations by summarizing each temporal feature across time. The pooling operations yield a single, fixed-dimensional feature vector, which is subsequently concatenated to retain information from both pooling approaches. A dense layer with a sigmoid activation transforms this pooled representation to enable the final prediction output. The model is compiled with RMSprop as the optimizer, and mean squared error (MSE) and mean absolute error (MAE) as performance metrics. The loss function is optimized for MAE, which minimizes absolute differences between predictions and actual values.

The training and testing evaluation results are calculated on metrics such as MSE, Root Mean Squared Error (RMSE), MAE, and R^2 . These metrics evaluate both the accuracy and consistency of the model's predictions. After predictions, the inverse transformation of the scaled data returns results to their original scale, facilitating real-world interpretability.

IV. RESULT EVALUATION

The results section presents an in-depth analysis of the performance of the proposed multi-head attention-based LSTM model across six banks, including Islami Bank, Dhaka Bank, AB Bank, Bank Asia, EBL Bank, and City Bank. Each bank dataset was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 scores. These metrics assess the accuracy and reliability of the model's predictions, with lower MSE, RMSE, and MAE values reflecting better prediction accuracy, while higher R^2 scores indicate a stronger correlation between predicted and actual values.

In Table II, the multi-head attention-based LSTM consistently outperformed traditional LSTM, GRU, and Bi-LSTM models across all datasets. For Islami Bank, the proposed model achieved the lowest error rates, with an MSE of 0.0005, RMSE of 0.023, and MAE of 0.014, along with an impressive R^2 score of 0.98. Similar improvements were observed for Dhaka Bank, where the proposed model reached an MSE of 0.0001, RMSE of 0.010, and MAE of 0.006, achieving an R^2 value of 0.97, signifying high accuracy. For AB Bank, the multi-head attention-based LSTM model recorded an MSE of 0.0002, RMSE of 0.013, and MAE of 0.009 with an R^2 score of 0.92, showing a notable enhancement over the other models. In Bank Asia, this model achieved an MSE of 0.0009, RMSE of 0.031, and MAE of 0.023, with an R^2 value of 0.98. For EBL Bank, the multi-head attention-based LSTM performed with an MSE of 0.0002, RMSE of 0.013, MAE of 0.006, and R^2 of

0.97. Lastly, for City Bank, the model reached an MSE of 0.0001, RMSE of 0.010, and MAE of 0.008, along with an R^2 score of 0.83. These results consistently demonstrate the superiority of the multi-head attention-based LSTM model across all datasets.

TABLE II: Result of all dataset

Dataset	Model	MSE	RMSE	MAE	R^2
Islami Bank	LSTM	0.0008	0.028	0.021	0.97
	GRU	0.0018	0.042	0.029	0.94
	Bi-LSTM	0.0007	0.026	0.019	0.97
	Multi-Head Attention Based LSTM	0.0005	0.023	0.014	0.98
Dhaka Bank	LSTM	0.0001	0.014	0.011	0.94
	GRU	0.0003	0.019	0.017	0.88
	Bi-LSTM	0.0009	0.030	0.022	0.95
	Multi-Head Attention Based LSTM	0.0001	0.010	0.006	0.97
AB Bank	LSTM	0.0003	0.017	0.016	0.86
	GRU	0.0003	0.017	0.015	0.86
	Bi-LSTM	0.0007	0.027	0.019	0.79
	Multi-Head Attention Based LSTM	0.0002	0.013	0.009	0.92
Bank Asia	LSTM	0.0019	0.043	0.035	0.97
	GRU	0.0023	0.048	0.039	0.95
	Bi-LSTM	0.0016	0.040	0.033	0.97
	Multi-Head Attention Based LSTM	0.0009	0.031	0.023	0.98
EBL Bank	LSTM	0.0005	0.022	0.009	0.91
	GRU	0.0006	0.024	0.022	0.89
	Bi-LSTM	0.0035	0.059	0.058	0.74
	Multi-Head Attention Based LSTM	0.0002	0.013	0.006	0.97
City Bank	LSTM	0.0001	0.011	0.009	0.80
	GRU	0.0001	0.011	0.010	0.78
	Bi-LSTM	0.0002	0.014	0.012	0.76
	Multi-Head Attention Based LSTM	0.0001	0.010	0.008	0.83

Comparing the performance of other models, the traditional LSTM, GRU, and Bi-LSTM models generally showed higher error rates and lower R^2 scores than the proposed model. For example, in the case of Islami Bank, the traditional LSTM had an MSE of 0.0008, RMSE of 0.028, and MAE of 0.021, with an R^2 score of 0.97, which is slightly lower than the proposed model's performance. Similarly, for Dhaka Bank, the GRU model produced an MSE of 0.0003, RMSE of 0.019, and MAE of 0.017, with an R^2 score of 0.88, which is notably lower than the multi-head attention-based LSTM's results. Bi-LSTM, while performing slightly better than LSTM and GRU in some cases, still could not match the accuracy achieved by the proposed model. Across all banks, the proposed multi-head attention-based LSTM consistently provided lower MSE, RMSE, and MAE values along with higher R^2 scores, illustrating its effectiveness in capturing the underlying patterns in financial time series data more accurately than other approaches.

Figure 2 shows the training and testing results of the proposed model for each bank, highlighting the model's ability to capture stock price trends with high accuracy. For Islami Bank, Figure 2a shows a close alignment between the actual and predicted values in both the training and testing phases, indicating the model's robustness. A

similar pattern is observed in Dhaka Bank (Figure 2b), where the predictions closely follow the actual stock prices, demonstrating the model's ability to generalize effectively on new data. In AB Bank (Figure 2c, the multi-head attention-based LSTM captures the stock price variations with minimal deviation. For Bank Asia (Figure 2d, EBL Bank (Figure 2e), and City Bank (Figure 2f), the model consistently follows the actual stock price trends, showcasing its strong predictive performance across diverse datasets. The close alignment between predicted and actual values across all banks demonstrates the model's ability to handle the complexity of financial time series data.

Figure 3 illustrates the forecasting results for the proposed model across the six banks, validating its effectiveness in predicting future stock price trends. In Islami Bank (Figure 3a) and Dhaka Bank (Figure 3b), the model successfully captures both the general trends and short-term fluctuations in the stock prices, demonstrating its reliability for future predictions. For AB Bank, as shown in Figure 3c, the model accurately forecasts upcoming stock price trends, confirming its effectiveness in capturing relevant patterns. Similarly, in Bank Asia (Figure 3d), EBL Bank (Figure 3e), and City Bank (Figure 3f), the multi-head attention-based LSTM model provides forecasts that closely align with actual stock price movements, indicating a strong potential for real-world application. The forecasting results in Figure 3 highlight the model's capability to manage complex dependencies and patterns in financial data, making it a robust tool for stock price forecasting.

V. CONCLUSION

The study presented an advanced multi-head attention-based LSTM model for stock price prediction, applied to datasets from six banks, including Islami Bank, Dhaka Bank, AB Bank, Bank Asia, EBL Bank, and City Bank. Results demonstrated that the proposed model consistently outperformed traditional LSTM, GRU, and Bi-LSTM models across key evaluation metrics, such as MSE, RMSE, MAE, and R^2 scores. Specifically, the multi-head attention mechanism enhanced the model's ability to focus on relevant information, leading to more accurate predictions across diverse datasets. This improvement is particularly significant, given the volatility and complexity inherent in financial data, where even small enhancements in prediction accuracy can have substantial implications. The proposed model's superior performance highlights its potential applicability in financial forecasting, risk assessment, and investment decision-making. By outperforming conventional recurrent models, the multi-head attention-based LSTM demonstrates its suitability for handling complex time series data with intricate patterns. Future research could

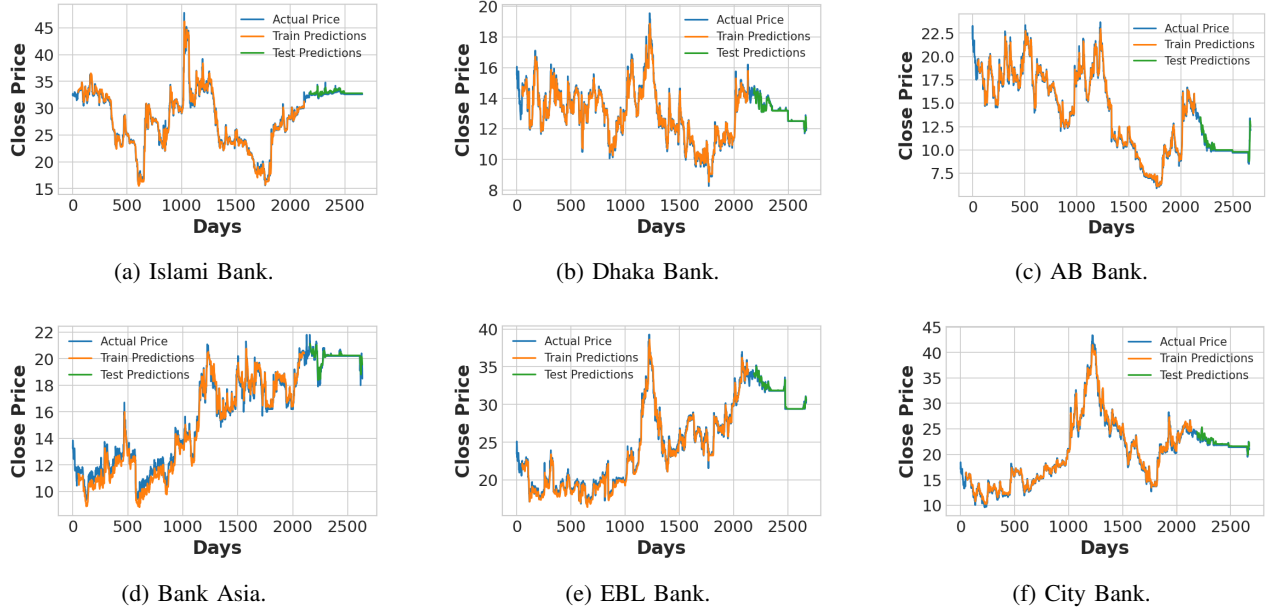


Fig. 2: Stock price training and testing dataset result of our proposed multi-head attention-based LSTM model.

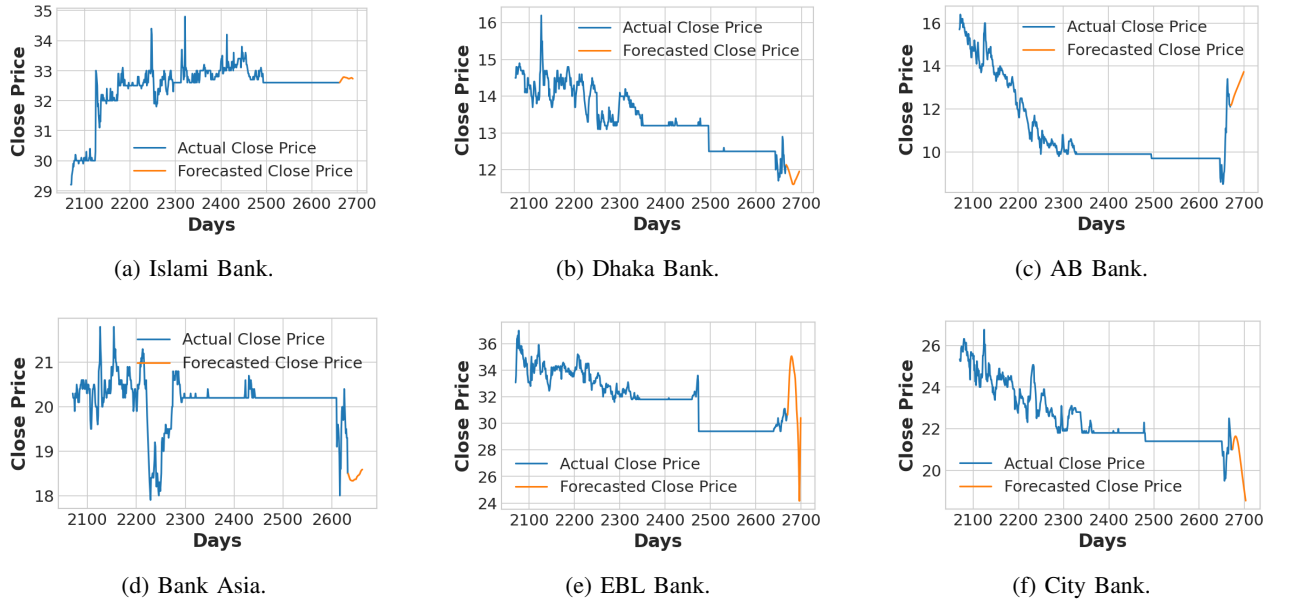


Fig. 3: Forecasting result of our proposed model on six different bank datasets.

explore the integration of additional external factors, such as macroeconomic indicators or sentiment analysis, to further enhance prediction accuracy. Overall, this study emphasizes the efficacy of attention mechanisms in deep learning models for financial forecasting and opens up new possibilities for more sophisticated applications in stock market analysis.

REFERENCES

- [1] "Dhaka Stock Exchange - Wikipedia — en.wikipedia.org," https://en.wikipedia.org/wiki/Dhaka_Stock_Exchange, [Accessed 25-11-2024].
- [2] X. Huang, C. Wu, X. Du, H. Wang, and M. Ye, "A novel stock trading utilizing long short term memory prediction and evolutionary operating-weights strategy," *Expert Systems With Applications*, 2024.
- [3] A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, and P. Kumam, "Fractional neuro-sequential arfima-lstm for fi-

- financial market forecasting,” *IEEE Access*, vol. 8, pp. 71 326–71 338, 2020.
- [4] Y. He, H. Li, and W. Wei, “Application of lstm model optimized based on adaptive genetic algorithm in stock forecasting,” Springer, Singapore, 2021, pp. 46–58.
 - [5] A. Rahman and M. J. Moon, “Historical Stock Price Data of Major Bangladeshi Banks on Dhaka Stock Exchange (2012-2014),” 11 2024. [Online]. Available: https://figshare.com/articles/dataset/Historical_Stock_Price_Data_of_Major_Bangladeshi_Banks_on_Dhaka_Stock_Exchange_2012-2014_/27693993
 - [6] M. A. Islam and A. Sattar, “Stock market prediction of bangladesh using multivariate long short-term memory with sentiment identification,” *International Journal of Power Electronics and Drive Systems*, vol. 13, no. 5, pp. 5696–5696, 2023.
 - [7] R. M. Blaber, “Transformer-based deep learning model for stock price prediction: A case study on bangladesh stock market,” *International Journal of Computational Intelligence and Applications*, 2023.
 - [8] A. Gupta, “Stock price prediction using machine learning,” *Indian Scientific Journal Of Research In Engineering And Management*, vol. 08, no. 05, pp. 1–5, 2024.
 - [9] A. Faheem, “Harnessing long short term memory networks for stock market forecasts,” 2024.
 - [10] R. Meghana, “Predicting stock prices using lstm networks: A comprehensive approach,” *International Journal For Science Technology And Engineering*, vol. 12, no. 7, pp. 1046–1057, 2024.
 - [11] I. Sahu, K. S. Paira, P. R. Bhoi, and S. Mohdiwale, “Dynamic long short-term memory model for stock market price forecasting,” Informa, 2024, pp. 189–198.
 - [12] M. Sakif, I. Yousuf, S.M.Rashidul, H. Nijhum, and M. I. Khan, “Hybstock: A hybrid deep learning-based model to predict stock market on dhaka stock exchange,” 2024, pp. 782–787.
 - [13] “Stock market prediction model using lstm,” *International Journal for Research in Applied Science and Engineering Technology*, 2023.