iJournals: International Journal of Software & Hardware Research in Engineering
ISSN:2347-4890, ijournals.in/ijshre
Volume 13 Issue 10, October 2025

Deep Learning for Stock Market Prediction: A Comparative Study on Nepal's Commercial Banking Sector

Author: Shantanu Bahadur Karki

Affiliation: Nankai University

E-mail: shantanukarki@gmail.com

DOI: 10.26821/IJSHRE.13.10.2025.131009

ABSTRACT

The prediction of stock market trends remains complicated in emerging markets like Nepal because of frequent market volatility and numerous economic influences. This research explores how machine learning and deep learning algorithms function in predicting stock market values for commercial banks that trade on the Nepal Stock Exchange (NEPSE). Daily historical stock data were collected from the time frame 2019-2024, along with external financial indicators such as gold prices, exchange rates, fuel prices, inflation rates, and interest rates, which were also collected from the same period and were preprocessed through normalization, missing value imputation, and interpolation for non-daily indicators. The research tests Long Short-term Memory (LSTM) along with Transformer and TimesNet using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE) evaluation methods.

The TimesNet model achieves superior performance than both LSTM and Transformer by delivering a 30–50% reduction in RMSE, MSE, and MAE across various commercial banks. These findings highlight the advantages of foundation models in financial forecasting, offering more accurate and stable predictions compared to traditional deep learning methods. The research findings provide essential knowledge to investors, policymakers, and financial analysts who need it for informed decision-making in the Nepalese financial sector. Future studies could explore hybrid models and additional. macroeconomic variables to enhance predictive performance further.

Keywords: Deep Learning, NEPSE, LSTM, Transformer, TimesNet.

1. INTRODUCTION

1.1 Background

The stock market represents a dynamic ecosystem where shares of publicly listed companies are traded, reflecting the overall economic health and spending capacity of a nation. In Nepal, the Nepal Stock Exchange (NEPSE) serves as the primary platform for stock trading. NEPSE was established on 14 January 1994 and has since grown to be an integral part of the Nepalese economy. NEPSE provides a platform for hydropower, IT companies, commercial banks, microfinance, etc., to sell their shares.

The commercial banking sector holds particular significance within this landscape. According to the International Monetary Fund and the World Bank, commercial banks in Nepal control 13% of the total market capitalization according to IMF[1] and contribute 10% to the nation's GDP according to World Bank[2]. This central role makes their stock prices a critical indicator for investors, policymakers, and financial analysts seeking to understand broader financial trends and make informed decisions. However, accurately predicting these prices is exceptionally difficult due to the inherent volatility of the market, which is susceptible to both domestic and international economic factors.

Traditional forecasting methods, such as fundamental and technical analysis, have long been the standard approach for assessing market movements. Fundamental analysis evaluates a stock's intrinsic value by examining economic data and financial statements, while technical analysis relies on historical price patterns and indicators to forecast future trends. While these methods offer valuable perspectives, they often prove insufficient for capturing the complex, non-linear, and dynamic patterns that characterize emerging markets like Nepal. The limitations of these traditional approaches necessitate the exploration of more sophisticated computational techniques that can leverage large datasets to uncover hidden patterns and improve predictive accuracy.

1.2 Problem Statement

The stock prices of commercial banks in Nepal are influenced by multiple factors, including economic indicators, financial performance, and market sentiment. Investors often rely on traditional forecasting methods, which may not be sufficient to capture the complex and dynamic nature of the stock market.

The primary problem addressed in this research is:

"How can machine learning and deep learning models be used to accurately predict the stock prices of commercial banks in Nepal?"

Formally, the stock prediction task can be defined as follows:

- A historical dataset of stock prices for commercial banks in Nepal, represented as:
  X_t = Open,Close,High,Low,Volume
- External financial indicators such as gold prices, exchange rates, fuel prices, inflation rates, and interest rates, denoted as:
  E_t = Gold Price, Exchange Rate, Fuel Price, Inflation Rate, Interest Rate

The objective is to learn a function f such that:

$$Y_{t+1} = f(X_t, E_t)$$

Where Y_{t+1} represents the predicted stock price for the next time step $t+1$.

By leveraging historical stock data, financial reports, and relevant economic indicators, this study seeks to develop a predictive model that provides better accuracy than traditional methods.

2. LITERATURE REVIEW

Research on predicting the Nepalese stock market has been growing, often paralleling the methodologies used globally. Early computational efforts in Nepal, such as a study by KC [3], utilized Support Vector Regression (SVR) to predict stock prices. Other analyses have focused on comparative risk assessments using statistical measures like standard deviation.

More recently, deep learning has been applied to the NEPSE index. Pokhrel et al. [4] and Gurung et al. [5] conducted studies using LSTM, Gated Recurrent Unit (GRU), and Convolutional Neural Networks (CNNs) for stock prediction, with their findings consistently suggesting that LSTM models provided the highest predictive accuracy.

While these studies have successfully introduced deep learning to the Nepalese context, a clear gap in the existing literature is the exploration of more advanced, time-series specific foundation models. Prior research has primarily relied on older architectures like LSTM and GRU. No published work has yet applied a cutting-edge model like TimesNet to the Nepalese banking sector. This provides a clear path for this study to contribute a novel and more sophisticated approach to the problem, aiming to establish a new benchmark for predictive accuracy.

3. METHODOLOGY

3.1 Data

For this study, we obtained daily stock price data for 19 commercial banks listed on the Nepal Stock Exchange (NEPSE) over five years (2019–2024). The dataset comprises two primary categories of variables:

1) Historical Stock Data: This includes daily open, close, high, low, and trading volume for each bank, collected from the official NEPSE website.
2) External Financial Indicators: These macroeconomic variables were included to account for broader market influences. Data sources and their respective frequencies are detailed in Table I. These include daily gold prices (Gahana Online), daily USD exchange

rates (Exchange Rates website), fortnightly fuel prices (Nepal Oil Corporation), and monthly inflation and interest rates (Trading Economics and Nepal Bank)

TABLE I: DATA SOURCES AND FREQUENCY OF COLLECTION

Data	Frequency	Source	Details
Open Price	Daily	Nepal Stock	Opening price of stock when the market opens.
Close Price	Daily	Nepal Stock	Last recorded price at market close.
High	Daily	Nepal Stock	Highest price during the trading session.
Low	Daily	Nepal Stock	Lowest price during the trading session.
Volume	Daily	Nepal Stock	Number of shares traded per day.
Gold Rate	Daily	Gahana Online	Price per tola (11.6638 g).
Dollar Rate	Daily	Exchang e Rates	1 USD to NPR.
Petrol Price	Fortnight	NOC	Retail price of 1 litre petrol.
Diesel Price	Fortnight	NOC	Retail price of 1 litre diesel.
Inflation Rate	Monthly	Trading Economi cs	National inflation rate.
Interest Rate	Monthly	Nepal Bank	Loan base interest rate.

3.2 Data Preprocessing

The raw data collected required several preprocessing steps to prepare it for use in the deep learning models.

- First, missing data and inconsistencies were addressed. The stock market is closed on weekends and public holidays, leading to gaps in the daily data.1 Corresponding gold data for these non-trading days was discarded, and for the few instances where the stock market was open but gold trading was closed, the gold price from the previous day was used for imputation.
- Second, non-daily data for fuel prices, inflation rates, and interest rates were converted into a daily format through interpolation.1 For petrol and diesel prices, which are updated fortnightly, the values remained constant until the next update. For inflation and interest rates, which are monthly averages, Cubic Spline Interpolation was used to create a smoother, continuous daily series.
- Third, feature transformation was applied to the gold price data. Raw gold prices showed an exponential increase over the study period, which would not establish a meaningful linear relationship with the stock prices. To make the data more useful for analysis, it was transformed to represent the daily ”rate of change”.1 This was calculated by subtracting yesterday’s price from today’s price and dividing the result by yesterday’s price.
- Fourth, all data were scaled to a fixed range using a MinMax scaler. This normalization is critical for deep learning models, as it helps to ensure that no single feature with a large numerical value dominates the learning process, leading to improved performance and faster convergence.
- Finally, the preprocessed dataset was split into training and testing sets, with 70% of the data used for training and the remaining 30% reserved for testing the model’s performance. A subset of the testing data (10%) was also used for validation during model training.

3.3 Model Architecture

This study implemented and trained three distinct deep learning architectures.

1) LSTM: The LSTM model was designed with sequential architecture featuring multiple layers to capture temporal dependencies. The architecture consists of an input layer that receives the normalized sequential data, followed by two stacked LSTM layers, each with 50 units, as shown in Fig. 1. The first

LSTM layer is configured to return sequences, passing a detailed output to the second layer, which in turn returns a single output per sequence. Dropout regularization (20%) was applied after each LSTM layer to prevent overfitting. A fully connected dense layer with 25 units processes the extracted features before a final dense output layer with a linear activation function provides the predicted stock price.

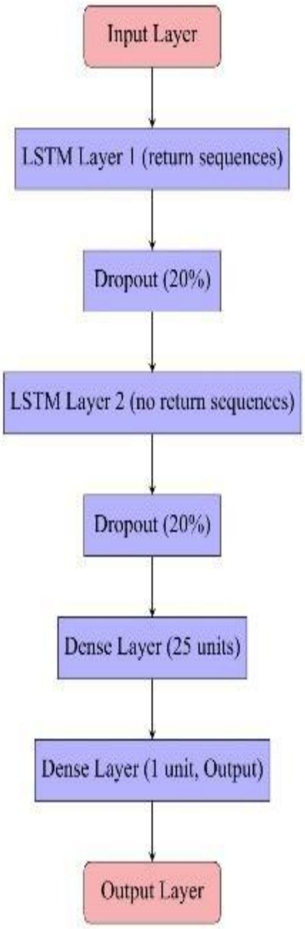

Fig 1. LSTM Architecture

2) Transformer Model: The Transformer model utilizes an attention-based mechanism to process sequential data. The architecture consists of an input layer, which feeds into a multi-head self-attention layer as shown in Fig. 2. This mechanism allows the model to weigh the importance of different time steps simultaneously. Residual connections and layer normalization are incorporated to stabilize training and preserve information. The output of the attention layer is passed to a position-wise feedforward network, which refines the extracted features. Global average pooling is applied before the final dense layers to reduce the sequence dimension. The model's final output is a linear-activated dense layer that predicts the future stock price.

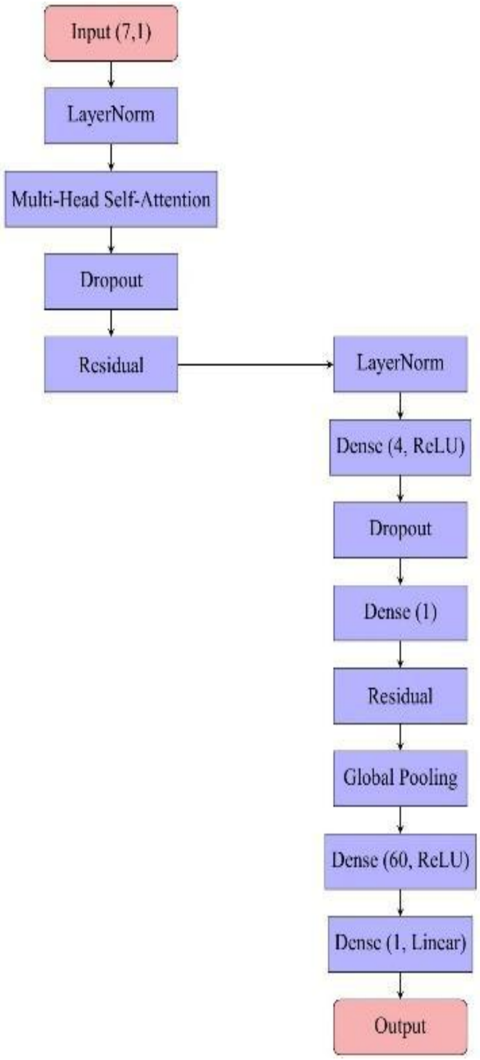

Fig 2. Transformer Architecture

3) TimesNet Model: TimesNet [6] is a time-series foundation model that is specifically optimized for forecasting. Its architecture is built around Time-Series Tokenization, which converts 1D time-series data into a 2D representation to capture temporal dependencies better. This is followed by a stack of TimesBlock layers, which are designed to capture patterns at multiple temporal scales, from short-term daily fluctuations to long-term market trends. The final architecture comprises a fully connected prediction head that refines the extracted multi-

scale features, followed by a final linear layer that provides the single-step-ahead forecast.

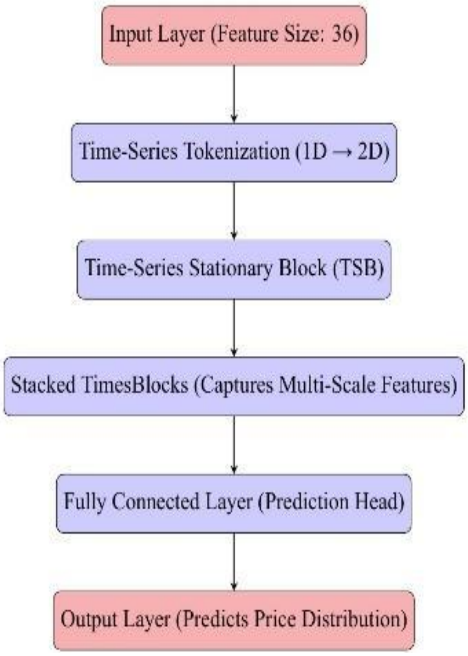

Fig 3. TimesNet Architecture

3.4 Evaluation Metrics

To provide a robust evaluation, the models' predictive performance was assessed using three widely adopted regression metrics for time-series forecasting. RMSE, MAE, MSE, and MAPE are the generally used metrics for stock market prediction [7] [8], with each metric seeing 20%, 16%, 21%, and 11% use in the most dominant journals for stock market forecasting respectively [9].

1) RMSE: RMSE calculates the standard deviation of the residuals, providing a measure of the average deviation of predicted values from actual values in the same unit as the stock prices. Its formula is:

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

2) MAE: MAE measures the average of the absolute differences between predicted and actual values. Its formula is:

$$MAE = \left(\frac{1}{n}\right) \sum_{\{i=1\}}^{\{n\}} |y\_i - \hat{y}\_i|$$

3) MSE: MSE is the average of the squared differences between predicted and actual values. Its formula is:

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

TABLE II: RMSE VS MAE VS MSE VS MAPE

Metric	Best For	Limitations
RMSE	Balancing small and large errors while keeping original units.	Still sensitive to outliers.
MAE	Simple interpretation, robust to outliers.	doesn't penalize large errors more than small ones.
MSE	Minimizing large errors in short-term prediction	Sensitive to outliers, not interpretable in stock prices.
MAPE	Comparing models across stocks and markets	Cannot handle zero or very small actual values.

As shown in Table II, each evaluation has its strengths and limitations. But in our paper, we shall be omitting MAPE since it's best used for cross-market comparison. Since we are only looking at the NEPSE, we shall not be using it coupled with the fact that it is the weakest of the four-evaluation metrics, with only 11% usage [9]

To statistically validate the performance differences, a two-step framework was employed. First, the Shapiro-Wilk test was conducted to check if the performance scores of the models were normally distributed. This is a crucial step for determining the appropriate statistical tests to use. The test's p-value was then compared to a significance level of 0.05. If the data were not normally distributed, non-parametric tests would be required. The Paired t-test and ANOVA would be used for comparison if the data were normally distributed, while the Wilcoxon Signed-Rank Test and the Kruskal-Wallis Test would be used for non-parametric analysis. This rigorous approach ensures that any conclusions about model superiority are statistically sound and not based on violated assumptions.

iJournals: International Journal of Software & Hardware Research in Engineering
ISSN:2347-4890, ijournals.in/ijshre
Volume 13 Issue 10, October 2025

4. RESULTS & DISCUSSION

4.1 Comparative Performance of Predictive Models

The experimental results demonstrate a clear hierarchy of performance among the three deep learning models. A comparison of the aggregate performance metrics across all 13 commercial banks can be seen in Table III.

TABLE III: MODEL EVALUATION METRICS

Model	LSTM	Transformer	TimesNet
Train RMSE	20.7781	28.7257	13.7381
Test RMSE	17.5474	16.6454	14.0098
Train MSE	431.7311	825.1713	188.7376
Test MSE	307.9145	277.0708	196.2751
Train MAE	20.5216	18.7949	6.6707
Test MAE	15.7036	12.4313	4.5833

TimesNet consistently delivered the lowest error scores across all three metrics (RMSE, MSE, and MAE) on both the training and testing datasets. This is quantitatively represented by its significantly lower RMSE and MAE on the test data compared to LSTM and Transformer.

The superiority of the TimesNet model is qualitatively evident in the individual bank simulations. For example, the NICA bank simulation (Figure 4 in the source material) shows a sharp, non-linear drop in the stock price in late 2023. The TimesNet model's predictions closely track this significant price movement, demonstrating its ability to capture abrupt market shifts. In contrast, while the LSTM model also attempts to follow this trend, its predictions are less accurate, and the Transformer model struggles to adapt to the high-variance period. This observation aligns with TimesNet's theoretical design, which leverages time-series specific tokenization and a multi-scale approach to effectively model complex temporal variations. The Transformer model, on the other hand, performs best on data with lower variance but high frequency, but it becomes unstable when faced with significant, rapid price fluctuations. The LSTM model, while robust, appears to be less capable of maintaining accurate tracking of extended trends compared to TimesNet.

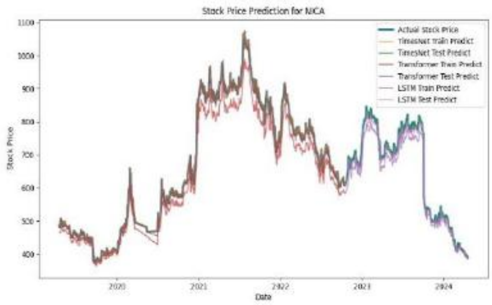

Fig 3. NICA Simulations

4.2 Statistical Validation of Model Performance

To move beyond a simple comparison of error scores, a rigorous statistical analysis was performed to determine if the observed performance differences were statistically significant.

First, a Shapiro-Wilk test was conducted on the performance scores of each model to check for normality. The results, presented in Table IV, showed that the p-values for all models were well below the 0.05 significance level, indicating that the data were not normally distributed.

TABLE IV: SHAPIRO-WILK TEST

Model	W-statistic	p-value	Normal? ($p > 0.05$)
LSTM	0.7077	0.0075	No
Transformer	0.6744	0.0033	No
TimesNet	0.6814	0.0039	No

This finding is crucial as it invalidates the use of parametric tests like the Paired t-test and ANOVA, which assume a normal distribution. Consequently, the results of the Paired t-test, which found no statistically significant difference between the models (p-value > 0.05), cannot be considered reliable.

Instead, the non-parametric Wilcoxon Signed-Rank Test was used to compare the models pairwise, as this test does not rely on the assumption of normality. The results of this test provide a statistically valid conclusion.

TABLE V: WILCOXON SIGNED-RANK TEST

Comparison	W-statistic	p-value	Significant? ($p > 0.05$)
LSTM vs. Transformer	10.0	1.0000	No
LSTM vs. TimesNet	0.0	0.0312	Yes
Transformer vs. TimesNet	0.0	0.0312	Yes

The Wilcoxon test revealed that while there was no significant difference between the performance of LSTM and the Transformer model ($p = 1.0000$), TimesNet's performance was statistically significantly better than both LSTM ($p = 0.0312$) and the Transformer model ($p = 0.0312$). This result confirms that TimesNet's superior performance is not a random occurrence but a statistically robust finding.

4.3 Interpretation of Results

To better understand the factors driving the predictions, Pearson's correlation coefficient was used to analyze the linear relationship between the input features and the predicted stock prices for each model.

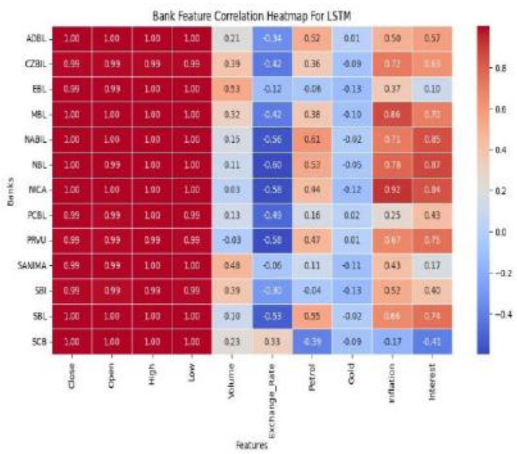

Fig. 5. LSTM correlation

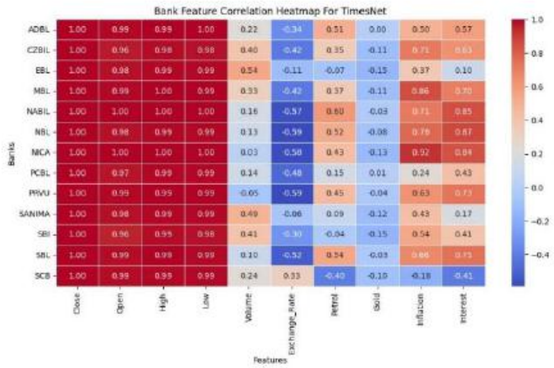

Fig. 6. TimesNet correlation

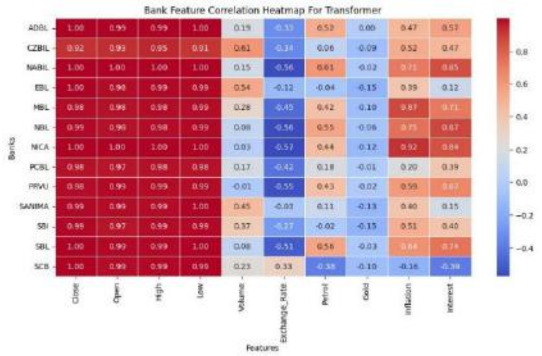

Fig. 7. Transformer correlation

The findings, visualized in the correlation heatmaps (Figures 5, 6, and 7), provided several key financial insights.

Unsurprisingly, the core stock market features—Close, Open, High, and Low—showed a very high correlation (approaching 1.00) with the predicted output, as these values are inherently interdependent. Trading volume, on the other hand, showed a mixed correlation, ranging from moderate to weak depending on the bank.

The analysis of external macroeconomic factors yielded particularly interesting results. The USD exchange rate, inflation, and interest rates generally exhibited a strong positive correlation with the predicted stock prices for most domestic commercial banks. This suggests that an increase in the stock values of these institutions often accompanies a rise in these macroeconomic indicators.

However, a significant anomaly was observed in the case of Standard Chartered Bank Nepal (SCB). For this institution, the USD exchange rate, inflation, and interest rates consistently showed an inverse or negative correlation with its stock price across all three models. This unique behavior is likely attributable to SCB's international exposure, foreign currency holdings, and distinct risk management strategies. Unlike domestic banks that may benefit from the dynamics of a local economy, SCB's global financial linkages may see reduced profitability from the same domestic pressures. This finding demonstrates the necessity of a nuanced, institution specific understanding of market dynamics, as a single-sector model may not adequately capture the unique characteristics of every bank.

Furthermore, the analysis showed that gold prices had very little to no correlation with the stock price movements of any of the banks. This finding, which aligns with the conclusions of other studies on the NEPSE index, provides an important clarification to the initial hypothesis that gold prices could serve as a proxy for the health of Nepal's treasury. It indicates that the change in gold rate has a minimal effect on stock movements, which provides actionable information for future model development by allowing for the simplification of the feature set.

4.4 Computational Performance and Scalability

An evaluation of the computational performance of each model revealed a clear trade-off between predictive accuracy and resource consumption.

TABLE VI: MODEL TRAINING MEMORY USAGE

Model	Input Size	Training Time	Active GPU	Reserved GPU	Total Params
LSTM	16	3.5s	0.42 MB	4.19 MB	215 K
LSTM	32	3.6s	0.42 MB	4.19 MB	215 K
LSTM	64	3.7s	0.42 MB	4.19 MB	215 K
LSTM	128	4.0s	0.42 MB	4.19 MB	215 K
LSTM	256	4.5s	0.42 MB	4.19 MB	215 K
Transformer	16	9.1s	0.22 MB	6.29 MB	341 K
Transformer	32	9.0s	0.22 MB	6.29 MB	341 K
Transformer	64	12.4s	0.23 MB	10.49 MB	341 K
Transformer	128	19.4s	0.44 MB	10.49 MB	341 K
Transformer	256	36.9s	0.43 MB	8.39 MB	341 K
TimesNet	16	9.8s	3.10 MB	16.78 MB	587 K
TimesNet	32	10.5s	5.77 MB	18.87 MB	587 K
TimesNet	64	11.9s	11.12 MB	39.85 MB	587 K
TimesNet	128	28.7s	21.82 MB	41.94 MB	587 K
TimesNet	256	100.8s	44.79 MB	79.69 MB	587 K

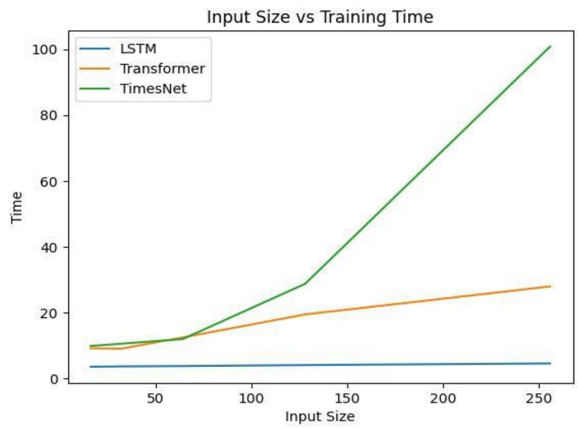

Fig 8. Input size vs GPU

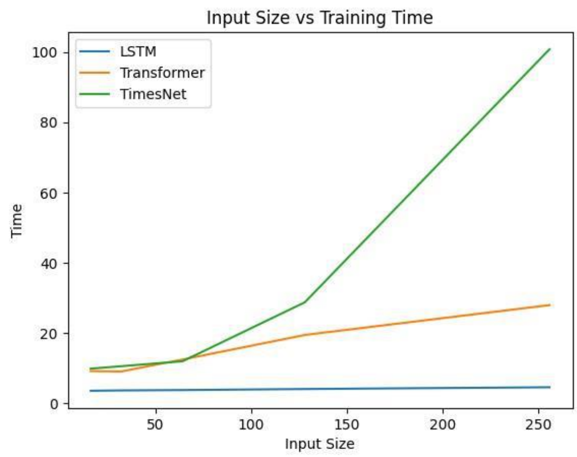

Fig 9. Input size Vs Training Time

The LSTM model was the most computationally efficient, requiring the least training time and memory, with a consistent GPU usage regardless of input size. The Transformer model required more time and GPU memory, with training time increasing significantly with larger input sizes. The TimesNet model, while the most accurate, was also the most computationally demanding. Its training time increased exponentially with larger input sizes, and it consumed the most GPU memory, a factor likely related to its higher number of total parameters. This computational intensity poses a critical consideration for practical deployment. While TimesNet is the optimal choice for a controlled research environment, its resource intensiveness would make it a less viable option for resource-constrained investors or real-time trading platforms without significant optimization.

5. CONCLUSION

This research systematically evaluated the predictive capabilities of three distinct deep learning models for forecasting stock prices in Nepal's commercial banking sector. The study leveraged a carefully curated dataset that included historical stock data and a range of external macroeconomic indicators.

The results unequivocally demonstrate that the TimesNet foundation model is the most accurate and robust model for this application, offering a substantial reduction in prediction errors compared to both LSTM and Transformer models. This conclusion was not only supported by a comparison of evaluation metrics but was also statistically validated through the Wilcoxon Signed-Rank Test, confirming the significance of TimesNet's superior performance.

Beyond the comparative analysis, the study provided novel insights into the specific influence of macroeconomic factors on the Nepalese banking sector. The unique inverse correlation observed between the stock price of Standard Chartered Bank Nepal and factors like exchange rates and inflation underscores the importance of considering institution-specific characteristics in financial models. Additionally, the empirical finding that gold prices have little to no correlation with stock movements provides a valuable clarification for future research and model development.

This work serves as a foundational study, providing a statistically sound and data-driven framework for financial forecasting in an emerging market. It validates the use of modern foundation models in a context where they have not been previously applied.

6. REFERENCES

[1]. International Monetary Fund, "Nepal: Staff Report for the 2023 Article IV Consultation," 2023.

[2]. World Bank, "Nepal Financial Sector Assessment," 2023.

[3]. H. KC, "Performance analysis and prediction of Nepal stock market (Nepse) for investment decision using machine learning techniques," International Journal of Computer Science Engineering (IJCSE), vol. 7, no. 1, pp. 15–27, 2018.

[4]. N. R. Pokhrel, K. R. Dahal, R. Rimal, H. N. Bhandari, R. K. C. Khatri, B. Rimal, and W. E. Hahn, "Predicting NEPSE index price using deep learning models," Machine Learning with Applications, vol. 9, p. 100385, 2022.

[5]. M. Gurung, P. Neupane, and S. Shrestha, "Stock market prediction in Nepali stock market using machine learning model," KEC Journal of Science and Engineering, vol. 8, no. 1, pp. 158–168, 2024.

[6]. H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TimesNet: Temporal 2D-variation modeling for general time series analysis," in Proc. Int. Conf. Learn. Representations (ICLR), 2023.

[7]. M. Mallikarjuna and R. P. Rao, "Evaluation of forecasting methods from selected stock market returns," Financial Innovation, vol. 5, no. 1, p. 40, 2019.

[8]. G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques–Part II: Soft computing methods," Expert Systems with Applications, vol. 36, no. 3, pp. 5932–5941, 2009.

[9]. D. Kumar, P. K. Sarangi, and R. Verma, "A systematic review of stock market prediction using machine learning and statistical techniques," Materials Today: Proceedings, vol. 49, pp. 3187–3191, 2022.