ADVANCED MODEL FOR STOCK PRICE FORECASTING WITH ENHANCED FEATURE ENGINEERING AND ADAPTIVE PARAMETER TUNING

Dr. Jackulin
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai,
Tamil Nadu, India.
chin.jackulin@gmail.com

Dr. Kavitha Subramani
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai,
Tamil Nadu, India.
kavitha.pec2022@gmail.com

R. Vaishnavi
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai,
Tamil Nadu, India.
raghunathvaishnavi7@gmail.com

D. Divyashree
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai,
Tamil Nadu, India.
dshreenaidu@gmail.com

K. Krithika
Department of Computer Science and Engineering
Panimalar Engineering College, Chennai,
Tamil Nadu, India.
kkrithika001@gmail.com

Abstract- **The stock market is essential to the health of the economy as a whole in capital market that can promote better and faster economic development, optimize asset allocation, and stimulate capital flow. The stock market is a great place for investors to make investments as well as a center for government regulation of economic trends. The government is always alert of economic catastrophes, and investors are more concerned with maximizing returns while minimizing risks. By adding a number of significant improvements, this study suggests a new framework for predicting stock prices that expands on current models. The suggested model will capture intricate temporal correlations in stock price data by leveraging Transformer architectures and other sophisticated deep learning models are at the forefront of contemporary AI development or more complex LSTM variations.**

Keywords- Accuracy, comparison, random forest, LSTM, ARIMA.

I. INTRODUCTION

AI-powered stock prediction is a new discipline that forecasts future stock and financial asset prices by fusing data science, financial analysis, and sophisticated machine learning algorithms. Developing algorithms that can examine past market data, identify trends, and offer insights to assist traders and investors in making wise selections is the purpose of stock performance forecasting. Because of the vast amount of data accessible, improvements in processing capacity, and the growing complexity of financial markets, this strategy has been increasingly popular in recent years. Traditionally, stock prediction relied on While technical analysis concentrates on examining patterns in trade volume and stock price trends, A company's entire performance and financial soundness are evaluated through fundamental analysis. While these methods has proven useful over time, they can be limited process of managing extensive datasets or complex relationships within the market. AI, especially making Employing techniques from both deep learning and machine learning enables a more sophisticated way of addressing these challenges by enabling systems to learn from large datasets without explicit programming for every possible scenario. AI- powered stock prediction represents a transformative approach to understanding and navigating the complexities of financial markets. While there are challenges to overcome, the continued development of AI techniques holds great potential for reshaping the future of investing and trading.

A. Objectives and Literature Review

This work focuses on developing and testing Predictive stock market models that utilize machine learning techniques using Python. ML and DL techniques are applied to estimate stock values by analyzing historical and current data. The background work survey and the main conclusions are displayed in the table. I.

B. Contribution

1. Novel Algorithmic Approaches
Improve stock forecasting is performed employing the computed values of using deep learning methods like the Long Short-Term Memory (LSTM) network with conventional statistical models like ARIMA, create an integrated framework.

2. Engineering Features:
Generating additional elements including technical analysis indicators (like moving averages and RSI), using data that already exists. Finding pertinent factors that can affect stock prices.

Table I ***Previous Literature Survey***

S.NO	SOURCES	YEAR	METHODOLOGY	KEY FINDINGS
1.	N. Naik et al. [1]	2021	HYBRID (HFS, XGBOOST, DNN)	The significant level of market fluctuations complicates the identification of a stock market crisis.
2.	S. S. Alotaibi et al. [2]	2021	RDA, GWO	The RD-GWO method chosen features more effectively than RDA and GWO employed alone, greatly lowering dimensionality while keeping important information.
3.	K. Huang et al. [3]	2022	LSTM, CNN	In terms of prediction accuracy and trading results, ML-GAT performs better than single-layer GNNs (like GCN and GAT) and traditional time-series models (like LSTM and CNN).
4.	G. Mu et al. [4]	2023	HYBRID (CNN-LSTM)	Models that incorporate investor emotion offer noticeably higher predicted accuracy than those that only use historical data.
5.	R. Zhang et al. [5]	2024	LSTM	Emotion-based models perform better in regression and classification tests than conventional models that use historical pricing.
6.	Y. Li et al. [6]	2025	HYBRID (LSTM, Bi LSTM, Transformer)	All benchmark models were outperformed by the LSTM with Attention mechanism in terms of prediction accuracy and volatility clustering detection.
7.	J. You et al. [7]	2024	NLP	Stocks with a surge in positive sentiment and high post volume were more likely to experience price increases the following day.
8.	K. Alam et al. [8]	2024	LSTM, DNN	The proposed model consistently outperformed standalone LSTM, DNN, and traditional models across most datasets. Average classification accuracy exceeded 78%, while regression models showed an average $R^2 > 0.85$.
9.	S. Wang et al. [9]	2023	BiLSTM	The BiLSTM + Improved Transformer with a better R2 score and a lower RMSE, the model outperformed all baseline techniques in regression tasks. Classification accuracy improved by 6–12% over standard LSTM models.
10.	Y. Fund et al. [10]	2023	ARIMA, ES	ARIMA showed superior performance on stock series that were well-differenced and stationary, achieving lower RMSE and MAPE in most cases.
11.	S. Srisuay et al. [11]	2024	SVM, WLP	The Wilson Loop Perceptron achieved higher directional prediction accuracy across multiple datasets, particularly in volatile market conditions, with an average accuracy improvement of 3–7% over SVM.
12.	X. Lu et al. [12]	2024	MACD	The hybrid model combining machine learning and human-like reasoning outperformed traditional models, showing a 15-20% improvement in classification accuracy (directional predictions) and price change predictions.

13.	J. Liu et al. [13]	2020	MEHACN	The MEHACN model outperformed traditional stock prediction models, such as LSTM and ARIMA, in terms of both classification (directional prediction) and regression (price prediction) tasks. It showed higher accuracy and lower prediction errors.
14.	M. Patel et al. [14]	2025	LSTM, ARIMA	LSTM provided the lowest error in predicting future stock prices because of how well it can identify temporal patterns in time series data.
15.	T. Liu et al. [15]	2025	FD-GRNet	The FD-GRNet model was rigorously evaluated against several benchmark algorithms across major global stock market indices.

II. Methods

Using a variety of Stock market forecasting is undergoing a transformation thanks to machine learning, deep learning, and artificial intelligence approaches. Trends in historical stock data are found using conventional methods like linear regression, support vector machines (SVM), and random forest classifiers. More sophisticated models, such as Long Short-Term Memory (LSTM) networks and Recurrent Neural Networks (RNNs), provide better performance and are excellent at simulating sequential data do very well in time-series forecasting. Candlestick chart technical indicators and patterns are also interpreted by Convolutional Neural Networks (CNNs). Simultaneously, sentiment analysis relies on obtaining valuable insights from news articles and social media sites with the use of natural language processing, or NLP. Through ongoing learning and decision-making, reinforcement learning is showing promise as a method for creating automated trading strategies. In order to increase forecast reliability and efficiently react to turbulent markets, many contemporary systems include various methodologies into hybrid models. AI-powered models provide quicker and more accurate insights than conventional methods, enabling more intelligent investment choices in a trading environment that is becoming more complicated. The model uses a hybrid deep learning strategy that combines CNN layers to extract localized features from technical indicators and LSTM networks for assessing time- dependent sequences in order to efficiently identify complicated patterns in market dynamics.

Mean Squared Error (MSE)- Loss Function

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

Where:

y_i = Actual stock price

$\hat{y}_i$ = Predicted stock price

III. Flowchart Diagram

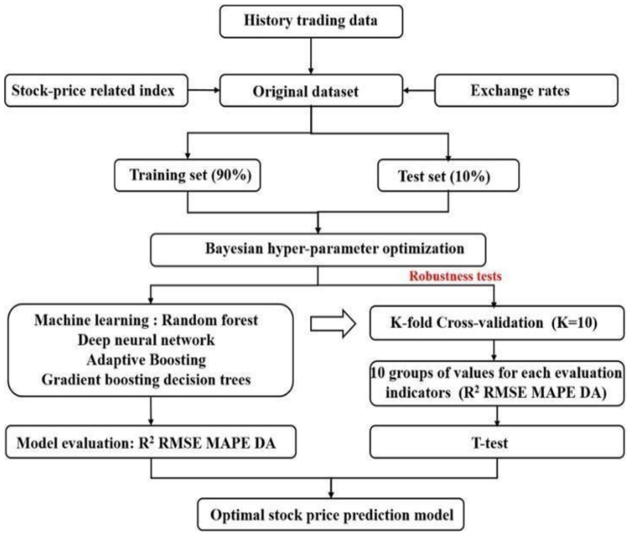

Fig.1. Flowchart Diagram

The above Figure.1. diagram provides a thorough summary of framework for creating an ideal stock price prediction model by combining machine learning and sophisticated optimization approaches is shown in the flowchart. To create the original dataset, historical trade data is first gathered and enhanced using currency rates and stock-related indexes. To aid in the creation and validation of the model, this dataset is 90% of the data is used for training, while 10% is used for testing. The model parameters are efficiently adjusted by the use of Bayesian hyperparameter optimization. Gradient boosting decision trees, adaptive boosting, random forest, and deep neural networks are among the machine learning methods that are trained using the optimum parameters. To make sure the model is resilient, a 10-fold cross-validation procedure is employed guarantee robustness and dependability, producing several performance indicators for every model, including R2, RMSE, MAPE, and directional accuracy (DA). The findings are validated and model performances are compared using a statistical t-test. The model that offers the highest accuracy and resilience is ultimately chosen as the ultimate stock price forecasting solution.

A. LSTM

The capacity of the Long Short-Term Memory (LSTM) neural network is a powerful tool for stock price forecasting because it can effectively handle intricate temporal connections across time. In contrast to conventional models, LSTM is highly effective in forecasting stock prices that show both long-term trends and short-term variations. The main characteristic of LSTM is the gating mechanism, which is made up of the input gate, forget gate, and output gate. Together, these gates control the information flow throughout the network. The input gate regulates when fresh data is added, guaranteeing that the forecast makes efficient use of pertinent, current knowledge. The forget gate allows the model to concentrate on the most crucial history data by deciding which previous data should be ignored. Lastly, the output gate determines how the processed data is converted into predictions, controlling the ultimate output.

B. ARIMA

When predicting time series data, it can identify patterns and trends in past price data, the Auto Regressive Integrated Moving Average (ARIMA) model is frequently employed for stock price forecasting. Three essential elements are included in ARIMA: autoregression (AR), differencing (I), and moving averages (MA). By simulating the correlation between the present stock price and its historical values, the autoregressive component enables the model to generate predictions based on historical data. By differentiating the time series, the integrated component assists in eliminating trends, rendering it stationary and simpler to model. The moving average component smoothes out random fluctuations by concentrating on the link between the observed values and historical forecast mistakes. When paired with expanded feature engineering, which adds more variables to the model such as trade volume, market mood, and economic indicators, ARIMA can be especially useful in stock price forecasting. These outside characteristics can offer insightful background information that raises prediction accuracy.

IV. *Model Building*

Data Collection: Compile historical stock data from Quandl, Alpha Vantage, and Yahoo Finance. Utilize web scraping technologies or APIs to get sentiment data from social media and news websites.

Data Preprocessing: Clean up the data and convert it into a machine learning-ready format (normalization, addressing missing numbers, etc.).

Feature Selection: Choose relevant features such as moving averages, stock volumes, or external data like economic indicators.

Model Selection: Choose an appropriate AI model (e.g., LSTM, SVM, or Random Forest).

Training and Testing: Use historical data to train the model, and use validation methods (like cross-validation) to assess it.

Model Evaluation: Use measurements such as accuracy, To measure the effectiveness, Root Mean Squared Error and Mean Absolute Error were used to gauge performance.

V. *Experimental Result Analysis*

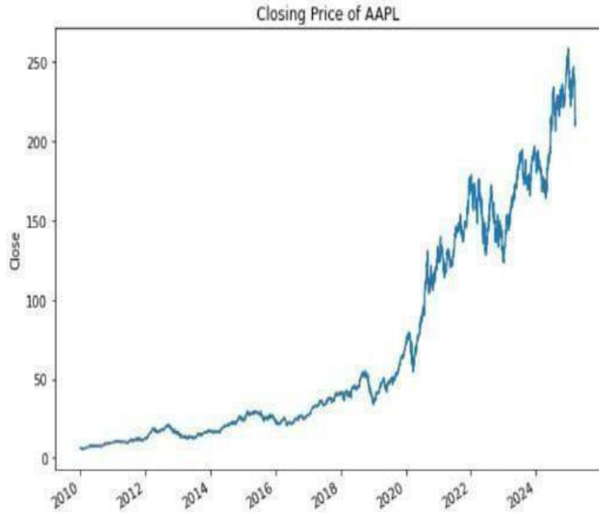

Fig.2. Apple Closing Price

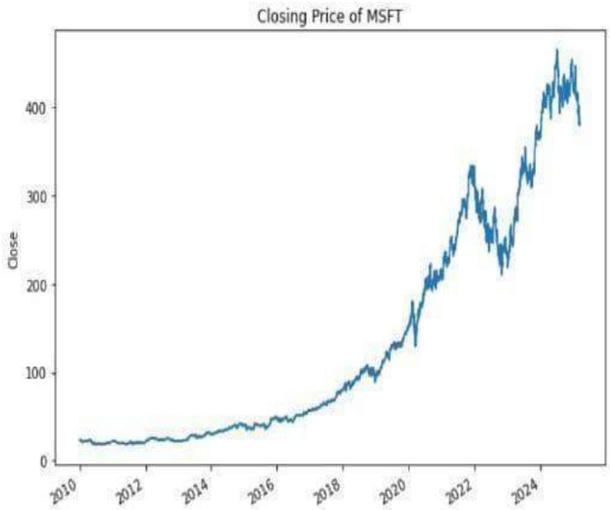

Fig.3. Microsoft Closing Price

The above Figure.2 and Figure.3 shows the historical closing prices of Apple Inc. (AAPL) and Microsoft Inc. (MSFT) from about 2010 to 2025 are depicted in the line graph. This graphic emphasizes the necessity for algorithms that can accurately efficiently record both long-term patterns and short-term variations within the parameters of an advanced stock price forecasting model. Improved feature engineering, such as adding external indicators (like interest rates, trading volume, or sentiment analysis), is advantage for models like LSTM, and ARIMA.

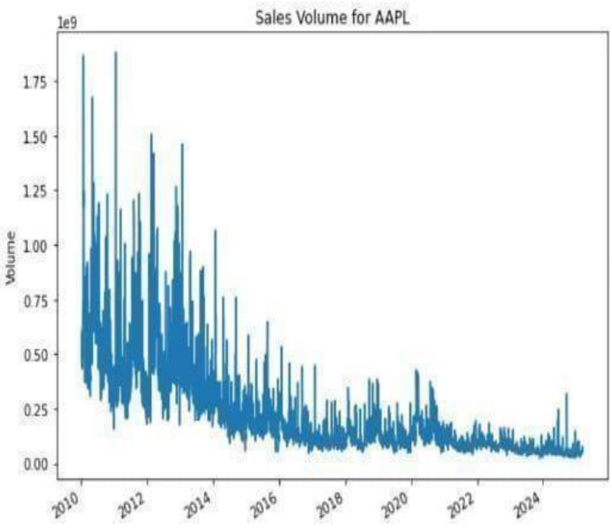

Fig.4. Apple Sales Volume

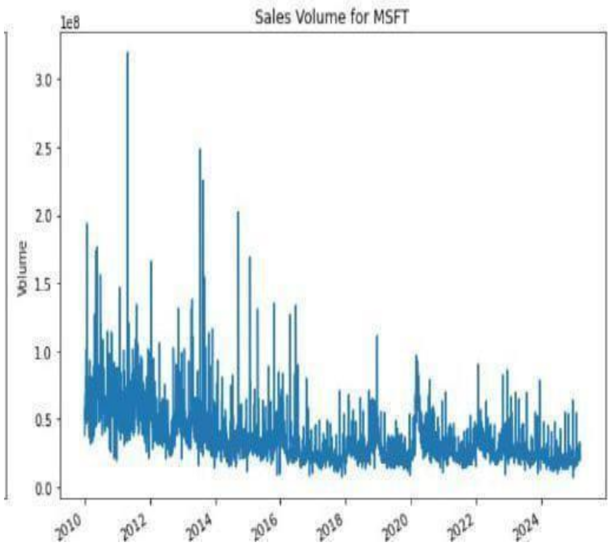

Fig.5. Microsoft Sales Volume

The above Figure.4 and Figure.5 shows the sales (trading) volume of Apple Inc. (AAPL) and Microsoft Inc. (MSFT) from 2010 to 2025 is shown in the chart. With high levels of trading activity in the early 2010s, reaching a peak of over 1.75 billion shares in certain cases, it exhibits a notable fall in volume over time. Trading volume is a key indicator of market interest, investor mood, and liquidity in the context of stock price forecasting. While smaller volumes can suggest less investor activity or market stability, high volumes frequently precede significant market movements, announcements, or events.

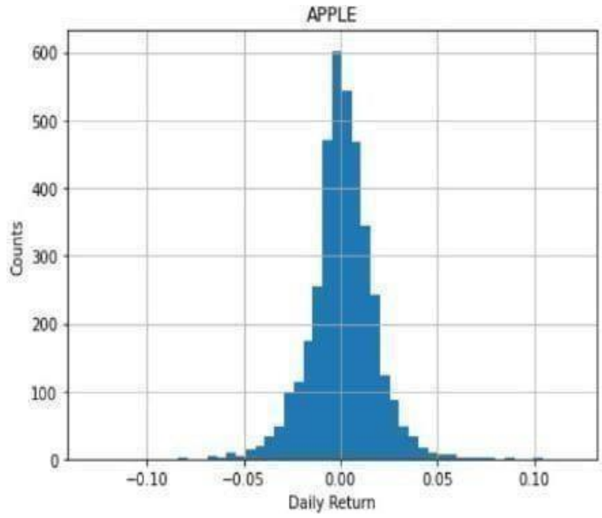

Fig.6. Apple Daily Returns

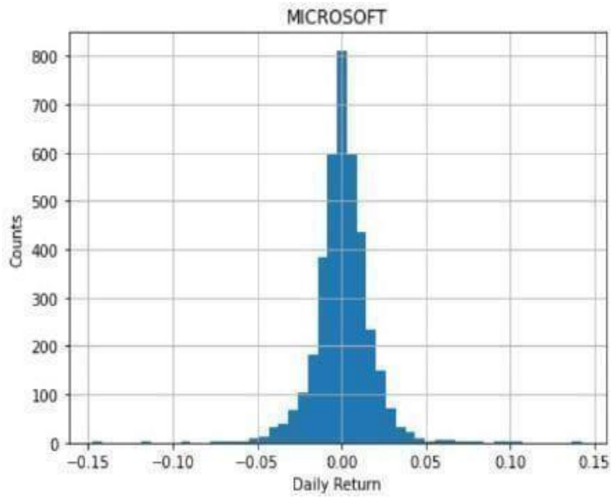

Fig.7. Microsoft Daily Returns

The above Figure.6 and Figure.7 shows how most of the swings lying between -2% and +2%, the daily returns are closely focused around zero. Although most price fluctuations are tiny, there are sporadic greater swings, as shown by the distribution's moderate tails and relatively narrow peak. With a sharper peak and narrower tails, Microsoft's distribution is even more concentrated than Apple's, indicating that its stock fluctuates less every day. The bulk of Microsoft's returns are concentrated at zero, indicating a generally lower level of daily volatility.

TABLE II. Accuracy Analysis of the Algorithms

Classifier	Accuracy (%)
ARIMA	85.7%
LSTM	94.3%

The above Table II shows the accuracy analysis of the algorithms model's success is indicated by these accuracy numbers, which are based on how well the model predicts real stock prices across a test set. Because it can simulate temporal relationships, the LSTM model obtains the best accuracy of (94.3%). Due to its limits in capturing intricate nonlinear patterns, ARIMA performs rather poorly even if it is good for linear trends (85.7%).

VI. CONCLUSION

This study investigated a comprehensive method to stock price forecasting that combined the capabilities of improved feature engineering and adaptive parameter tuning across several forecasting models such as LSTM and ARIMA. The objective was to increase model resilience and forecast accuracy in the naturally noisy and unpredictable world of financial markets. Raw stock market data was converted into a collection of extremely useful inputs through sophisticated feature engineering. Together with normalization and time-series decomposition, techniques like the creation of technical indicators, moving averages, and lag characteristics allowed the models to better comprehend volatility trends and temporal patterns. For precise predictions across several models, this preprocessed data provided a strong basis. Three different forecasting models were used in the project; each has advantages of its own. Because of its solid statistical foundation, ARIMA was able to identify seasonality and linear patterns in historical data. Long-term temporal linkages and non-linear dependencies were modeled using LSTM to guarantee optimal performance for every unique dataset and circumstance, these models were adjusted utilizing adaptive parameter optimization approaches including grid search and learning rate tuning.

REFERENCES

[1] N. Naik and B. R. Mohan, "Novel Stock Crisis Prediction Technique—A Study on Indian Stock Market," in IEEE Access, vol. 9, pp. 86230-86242, 2021, doi: 10.1109/ACCESS.2021.3088999.

[2] S. S. Alotaibi, "Ensemble Technique With Optimal Feature Selection for Saudi Stock Market Prediction: A Novel Hybrid Red Deer-Grey Algorithm," in IEEE Access, vol. 9, pp. 64929-64944, 2021,doi: 10.1109/ACCESS.2021.3073507.

[3] K. Huang, X. Li, F. Liu, X. Yang and W. Yu, "ML-GAT: A Multilevel Graph Attention Model for Stock Prediction," in IEEE Access, vol. 10, pp. 86408-86422, 2022, doi:10.1109/ACCESS.2022.3199008.

[4] G. Mu, N. Gao, Y. Wang and L. Dai, "A Stock Price Prediction Model Based on Investor Sentiment and Optimized Deep Learning," in IEEE Access, vol. 11, pp. 51353-51367, 2023, doi: 10.1109/ACCESS.2023.3278790.

[5] R. Zhang and V. Y. Mariano, "Integration of Emotional Factors With GAN Algorithm in Stock Price Prediction Method Research," in IEEE Access, vol. 12, pp. 77368-77378, 2024, doi: 10.1109/ACCESS.2024.3406223.

[6] Y. Li and C. Long, "Innovative Study on Volatility Prediction Model for New Energy Stock Indices," in IEEE Access, vol. 13, pp. 29754-29777, 2025, doi:10.1109/ACCESS.2025.3535584.

[7] J. You, H. Jang, M. Kang, S. -B. Yang and S. -H. Yoon, "Leveraging Stock Discussion Forum Posts for Stock Price Predictions: Focusing on the Secondary Battery Sector," in IEEE Access, vol. 12, pp. 153537-153549, 2024, doi: 10.1109/ACCESS.2024.3481442.

[8] K. Alam, M. H. Bhuiyan, I. U. Haque, M. F. Monir and T. Ahmed, "Enhancing Stock Market Prediction: A Robust LSTM-DNN Model Analysis on 26 Real-Life Datasets," in IEEE Access, vol. 12, pp. 122757-122768, 2024, doi: 10.1109/ACCESS.2024.3434524.

[9] S. Wang, "A Stock Price Prediction Method Based on BiLSTM and Improved Transformer," in IEEE Access, vol. 11, pp. 104211-104223, 2023, doi: 10.1109/ACCESS.2023.3296308.

[10] Y. Fund and A. Damani, ‘‘Comparison of ARIMA and exponential smoothing models in prediction of stock prices,’’ J. Predict. Markets, vol. 17, no. 1, pp. 21–38, Jul. 2023.

[11] S. Srisuay, K. Kanjamapornkul, and W. Fongngen, ‘‘Directional prediction of financial time series using SVM and Wilson loop perceptron,’’ Social Network Compute. Sci., vol. 5, no. 4, p. 395, Apr. 2024.

[12] X. Lu, J. Poon and M. Khushi, "Bridging the Gap Between Machine and Human in Stock Prediction: Addressing Heterogeneity in Stock Market," in IEEE Access, vol. 12, pp. 186171-186185, 2024, doi: 10.1109/ACCESS.2024.3511613.

[13] J. Liu, H. Lin, L. Yang, B. Xu and D. Wen, "Multi-Element Hierarchical Attention Capsule Network for Stock Prediction," in IEEE Access, vol. 8, pp. 143114-143123, 2020, doi: 10.1109/ACCESS.2020.3014506.

[14] M. Patel, K. Jariwala and C. Chattopadhyay, "An Approach Toward Stock Market Prediction and Portfolio Optimization in Indian Financial Sectors," in IEEE Transactions on Computational Social Systems, vol. 12, no. 1, pp. 128-139, Feb. 2025, doi: 10.1109/TCSS.2024.3450291.

[15] T. Liu, J. Li, Z. Zhang, H. Yu and S. Gao, "FD-GRNet: A Dendritic-Driven GRU Framework for Advanced Stock Market Prediction," in IEEE Access, vol. 13, pp. 28265-28279, 2025, doi: 10.1109/ACCESS.2025.3541074.