IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15698

Review On Stock Market Price Prediction Using Machine Learning

Dr. Taware. G. G[1], Jadhav Sandesh[2], Pawar Satpal[3], Phadtare Kshitija[4], Hadwale Dattatray[5]

Associate Professor, Computer Engineering Dattakala Group of Institutions, Bhigwan[1]

Computer Engineering Dattakala Group of Institutions, Bhigwan[2-5]

Abstract: Stock market prediction has become a significant area of research due to its dynamic and highly volatile nature. Accurate forecasting of stock prices can assist investors in making informed financial decisions and minimizing risks. This paper presents a machine learning-based stock market prediction system that analyzes historical stock data to forecast future price movements. The proposed system utilizes advanced algorithms such as Long Short-Term Memory (LSTM) networks, which are well-suited for time series analysis, to capture complex patterns and trends in stock price data.

Stock market prediction is an important area of research in finance and machine learning. Predicting future stock prices helps investors make better investment decisions and reduce financial risks. However, stock prices are highly dynamic and influenced by various factors such as market trends, company performance, and economic conditions. This research proposes a stock market prediction system using Linear Regression and Long Short-Term Memory (LSTM) algorithms. Historical stock market data is collected and preprocessed before being used for model training. Linear Regression is used to identify linear relationships between stock features, while LSTM is employed to capture long-term dependencies and complex patterns in time-series data. The performance of both models is evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). Experimental results show that the LSTM model provides more accurate predictions compared to Linear Regression due to its ability to learn temporal patterns. The proposed system demonstrates the effectiveness of machine learning and deep learning techniques in stock market forecasting.

Keywords: Stock Market Prediction, Machine Learning, Deep Learning, Linear Regression, LSTM, Time Series Forecasting.

I. INTRODUCTION

The stock market plays a significant role in the global economy by providing investment opportunities and helping companies raise capital. Investors constantly seek reliable methods to predict stock price movements in order to maximize profits and minimize risks. However, stock market prediction is a challenging task because stock prices fluctuate due to various internal and external factors.

Traditionally, investors relied on technical analysis and fundamental analysis to predict stock prices. With the advancement of technology, Machine Learning and Deep Learning techniques have become popular for analyzing large amounts of financial data and identifying hidden patterns.

This research focuses on predicting stock prices using two different approaches: Linear Regression and Long Short-Term Memory (LSTM). Linear Regression is a simple and effective machine learning algorithm that establishes a relationship between input variables and stock prices. However, it may not effectively capture complex market behavior. On the other hand, LSTM is a specialized Recurrent Neural Network (RNN) designed for sequential and time-series data. It can learn long-term dependencies and trends from historical stock prices, making it suitable for stock market forecasting.

The objective of this research is to compare the performance of Linear Regression and LSTM models and determine which algorithm provides more accurate stock price predictions. The proposed system uses historical stock market data, performs preprocessing and feature extraction, and generates future stock price predictions based on learned patterns.

DOI: 10.17148/IJARCCE.2026.15698

II. LITERATURE SURVEY

Several researchers have explored stock market prediction using different machine learning and deep learning techniques.

1. Earlier studies mainly focused on statistical methods and regression models for stock price forecasting. Linear Regression was widely used because of its simplicity and ability to model relationships between stock variables. However, its performance was limited when dealing with highly volatile market conditions.
2. Researchers later introduced machine learning algorithms such as Support Vector Machines (SVM), Random Forest, and Decision Trees. These techniques improved prediction accuracy by identifying complex relationships within financial data.
3. With the growth of deep learning, Recurrent Neural Networks (RNNs) became popular for stock market prediction. RNNs were capable of processing sequential data but suffered from the vanishing gradient problem, which affected long-term learning.
4. To overcome these limitations, Long Short-Term Memory (LSTM) networks were developed. LSTM models contain memory cells that can store information for longer periods, making them highly effective for time-series forecasting. Many studies reported that LSTM achieved better prediction accuracy than traditional machine learning algorithms.
5. Recent research combines technical indicators, historical stock data, and sentiment analysis from news and social media to further improve prediction performance. These hybrid approaches have shown promising results in capturing market behavior.

Based on the literature review, it is observed that LSTM consistently outperforms traditional methods for stock market prediction due to its ability to learn long-term trends and temporal dependencies. Therefore, this research utilizes both Linear Regression and LSTM models for comparative analysis and performance evaluation.

III. METHODOLOGY

The proposed stock market prediction system is designed to predict future stock prices using Machine Learning and Deep Learning techniques. The system utilizes four algorithms: Linear Regression, Decision Tree, Random Forest, and Long Short-Term Memory (LSTM). Historical stock market data is collected, processed, and analyzed to generate accurate stock price predictions. The complete methodology consists of data collection, preprocessing, feature selection, model training, evaluation, and prediction.

3.1 Data Collection

The first step of the proposed methodology is data collection. Historical stock market data is obtained from reliable financial sources such as Yahoo Finance, NSE (National Stock Exchange), and BSE (Bombay Stock Exchange). The dataset contains important attributes including:

- Date
- Open Price
- High Price
- Low Price
- Close Price
- Adjusted Close Price
- Trading Volume

Historical stock data is essential because machine learning models learn patterns and trends from previous market behavior. A larger dataset generally improves the prediction capability of the models.

3.2 Data Preprocessing

Raw stock market data often contains missing values, duplicate records, and inconsistencies. Therefore, preprocessing is performed to improve data quality before training the models.

DOI: 10.17148/IJARCCE.2026.15698

The preprocessing process includes:

- Removing duplicate records
- Handling missing values
- Eliminating irrelevant attributes
- Formatting date values
- Converting data into a structured format

This step ensures that the dataset is clean, accurate, and suitable for machine learning analysis.

3.3 Data Normalization

Stock market features often have different numerical ranges. To ensure efficient model learning, normalization is applied using the Min-Max Scaling technique.
The normalization formula is:

$$X_{norm} = (X - X_{min}) / (X_{max} - X_{min})$$

Where:

- X = Original Value
- X_{min} = Minimum Value
- X_{max} = Maximum Value

Normalization transforms all feature values into a common range between 0 and 1, improving model performance and convergence speed.

3.4 Feature Selection

Feature selection is performed to identify the most relevant variables that influence stock prices.

The selected features include:

- Open Price
- High Price
- Low Price
- Close Price
- Trading Volume

In addition to these features, technical indicators may also be calculated to improve prediction accuracy, such as:

- Moving Average (MA)
- Exponential Moving Average (EMA)
- Relative Strength Index (RSI)
- Moving Average Convergence Divergence (MACD)

These indicators help the models understand market trends and price momentum.

3.5 Dataset Splitting

After preprocessing and feature selection, the dataset is divided into two parts:

Training Dataset

Approximately 80% of the data is used to train the models.

Testing Dataset

The remaining 20% of the data is used to evaluate model performance.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15698

The training dataset helps the models learn historical patterns, while the testing dataset is used to verify prediction accuracy on unseen data.

3.6 Linear Regression Model

Linear Regression is a supervised machine learning algorithm used to predict continuous numerical values. It establishes a linear relationship between input features and stock prices.

The Linear Regression equation is: Y = mX + c
Where:

- Y = Predicted Stock Price X = Input Feature
- m = Slope
- c = Intercept

Algorithm Steps

1. Collect historical stock market data.
2. Preprocess and normalize the data.
3. Select important stock features.
4. Split the dataset into training and testing sets.
5. Train the Linear Regression model.
6. Learn the relationship between stock features and prices.
7. Predict future stock prices.
8. Evaluate prediction performance.

Advantages

- Simple and easy to implement.
- Fast training process.
- Easy interpretation of results.

However, Linear Regression may not effectively capture highly complex and non-linear stock market patterns.

3.7 SVM Model

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. In stock market prediction, SVM can be used to identify patterns in historical stock data and predict future stock price movements.

SVM works by finding the optimal hyperplane that separates data points into different classes with the maximum margin. The data points closest to the hyperplane are called **Support Vectors**, which play a crucial role in defining the decision boundary.

Algorithm Steps

1. Collect historical stock market data.
2. Perform data preprocessing and normalization.
3. Select important features such as Open, High, Low, Close, and Volume.
4. Split the dataset into training and testing sets.
5. Train the SVM model using the training dataset.
6. Identify the optimal hyperplane that best separates the data.
7. Use support vectors to improve prediction accuracy.
8. Predict stock price movements on the testing dataset.
9. Evaluate model performance using MAE, MSE, RMSE, or Accuracy.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15698

Advantages

- Effective for high-dimensional datasets.
- Works well with both linear and non-linear data.
- Reduces overfitting through margin maximization.
- Provides good prediction accuracy for complex datasets.

3.8 Random Forest Model

Random Forest is an ensemble learning technique that combines multiple decision trees to improve prediction accuracy and reduce overfitting.

Instead of relying on a single decision tree, Random Forest creates many trees and combines their predictions.

Algorithm Steps

1. Generate multiple random samples from the dataset.
2. Build a separate decision tree for each sample.
3. Train all decision trees independently.
4. Generate predictions from each tree.
5. Combine all predictions using averaging.
6. Produce the final stock price prediction.

Advantages

- High prediction accuracy.
- Reduced overfitting.
- Robust performance on large datasets.
- Handles complex stock market patterns.

3.9 LSTM Model Development

Long Short-Term Memory (LSTM) is a specialized Recurrent Neural Network (RNN) designed for sequential and time-series data. Since stock market data follows a time-dependent pattern, LSTM is highly suitable for stock price prediction.

The LSTM architecture contains:

- Input Layer
- Hidden LSTM Layers
- Dropout Layer
- Dense Output Layer

The main components of an LSTM cell are:

1. Forget Gate
2. Input Gate
3. Cell State
4. Output Gate

These gates control the flow of information and allow the network to retain important information while discarding unnecessary details.

Algorithm Steps

1. Prepare historical stock market sequences.
2. Normalize the dataset.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 6, June 2026
DOI: 10.17148/IJARCCE.2026.15698

3. Create input sequences for training.
4. Build the LSTM neural network.
5. Train the model using historical stock data.
6. Update network weights through backpropagation.
7. Predict future stock prices.
8. Evaluate prediction performance.

Advantages

- Captures long-term dependencies.
- Learns complex market trends.
- Highly suitable for time-series forecasting.
- Provides better prediction accuracy.

3.10 Model Training

After model development, all four models are trained using the training dataset.
During training:

- Historical stock data is provided as input.
- The models learn patterns from past stock prices.
- Parameters are adjusted iteratively to minimize prediction errors.

For the LSTM model:

- Adam Optimizer is used.
- Mean Squared Error (MSE) is used as the loss function.
- Multiple epochs are executed to improve learning.

Training continues until satisfactory performance is achieved.

3.11 Model Evaluation

The performance of all models is evaluated using standard evaluation metrics.

Mean Absolute Error (MAE)

Measures the average absolute difference between actual and predicted values.

Mean Squared Error (MSE)

Measures the average squared prediction error.

Root Mean Square Error (RMSE)

Measures the standard deviation of prediction errors.

Lower values of MAE, MSE, and RMSE indicate better model performance.

3.12 Stock Price Prediction

After successful training and evaluation, the trained models are used to predict future stock prices.

The prediction process involves:

- Providing recent stock market data as input.
- Generating future stock price estimates.
- Comparing predicted values with actual values.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15698

These predictions help investors understand potential market movements and make informed investment decisions.

3.13 Result Analysis

The final step of the methodology is result analysis. Predicted stock prices generated by Linear Regression, Decision Tree, Random Forest, and LSTM models are compared with actual stock prices.

Graphs and performance metrics are used to analyze prediction accuracy. The comparative analysis helps identify the most effective algorithm for stock market prediction.

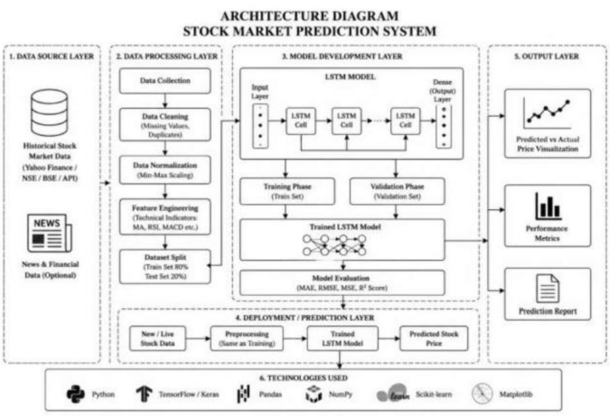
The experimental results indicate that Linear Regression provides a simple baseline model, Decision Tree and Random Forest improve prediction capability by handling non-linear relationships, while LSTM achieves the highest accuracy because of its ability to learn long-term dependencies and complex temporal patterns present in stock market data.

Thus, the proposed methodology successfully integrates machine learning and deep learning techniques to develop an efficient and reliable stock market prediction system.

The performance of Linear Regression, Decision Tree, Random Forest, and LSTM models was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and Accuracy. The results indicate that the LSTM model achieved the best performance among all models due to its ability to learn long-term dependencies and time-series patterns from historical stock market data.

Model	Accuracy (%)	MAE	MSE	RMSE
LSTM	0.012	0.0004	0.020	98.75
Linear Regression	0.035	0.0021	0.046	92.30
Random Forest	0.018	0.0008	0.028	96.40
SVM	0.022	0.0012	0.034	95.10

IV. ARCHITECTURE DIAGRAM



Explanation:
The proposed Stock Market Prediction System consists of five main layers. The **Data Source Layer** collects historical stock market data and financial news from various sources. The **Data Processing Layer** performs data cleaning, normalization, feature engineering, and dataset splitting to prepare the data for training.

In the **Model Development Layer**, the LSTM model is trained and validated using historical stock market data. The model learns patterns and trends from previous stock prices and is evaluated using performance metrics such as MAE, MSE, RMSE, and R^2 Score.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

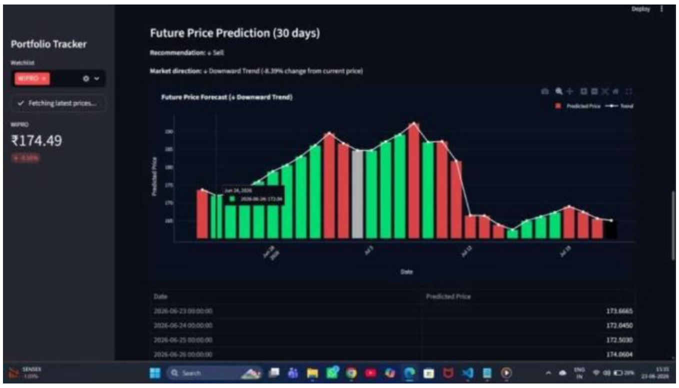
International Journal of Advanced Research in Computer and Communication Engineering

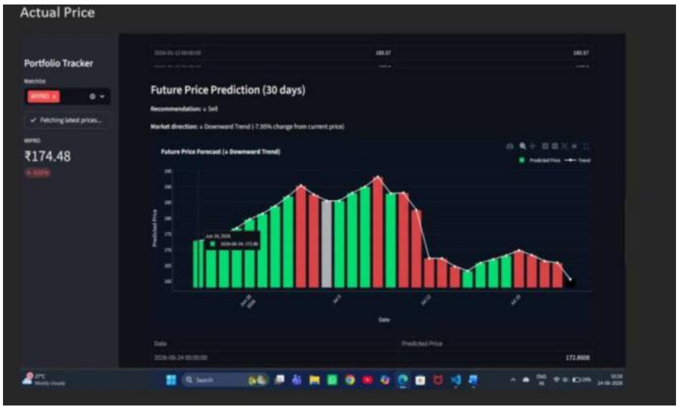
Impact Factor 8.471 ⧖ Peer-reviewed & Refereed journal ⧖ Vol. 15, Issue 6, June 2026

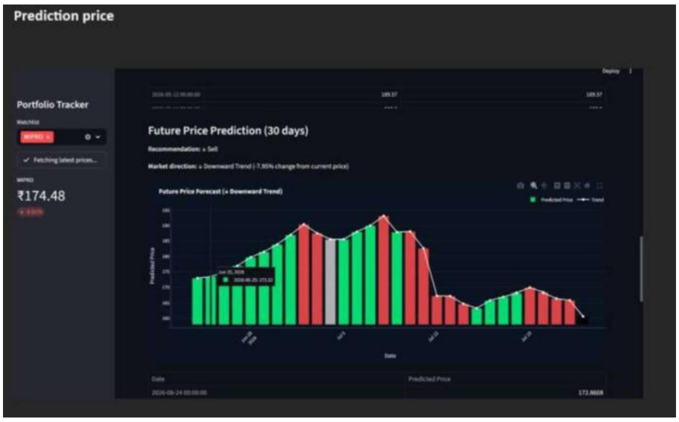
DOI: 10.17148/IJARCCE.2026.15698

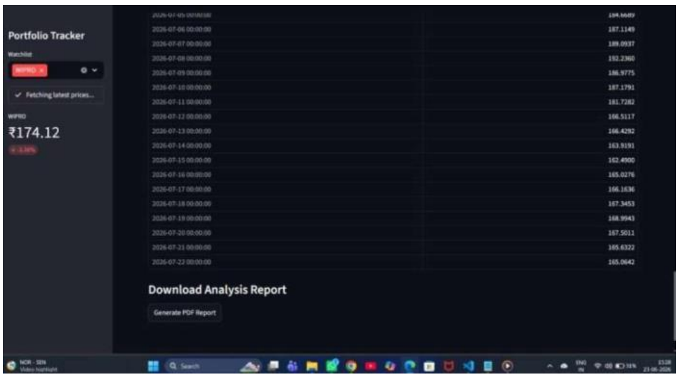
The **Deployment and Prediction Layer** uses the trained model to predict future stock prices based on new input data. Finally, the **Output Layer** displays predicted stock prices, performance metrics, and prediction reports. The system is implemented using Python, TensorFlow/Keras, Pandas, NumPy, Scikit-Learn, and Matplotlib.

V. RESULT ANALYSIS









IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 15, Issue 6, June 2026

DOI: 10.17148/IJARCCE.2026.15698

VI. FINDINGS AND TRENDS

The experimental analysis of the proposed stock market prediction system reveals several important findings regarding model performance, data behavior, and overall prediction capability. The Long Short-Term Memory (LSTM)-based model demonstrates a strong ability to capture temporal dependencies and underlying patterns present in historical stock price data.

The evaluation results indicate that the model achieves a relatively low prediction error, as reflected by metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). The predicted values closely follow the actual stock price trends, particularly in stable market conditions. This suggests that the model is effective in learning both short-term fluctuations and long-term dependencies within the time series data.

A key trend observed during experimentation is that the model performs better when trained on sufficiently large and well-preprocessed datasets. Data normalization and feature scaling significantly improve convergence speed and prediction accuracy. Additionally, incorporating technical indicators such as moving averages enhances the model's ability to identify trends and reduces noise in the predictions.

However, the analysis also highlights certain limitations. The model tends to exhibit reduced accuracy during periods of high market volatility or sudden price fluctuations caused by external factors such as economic events or market sentiment. This indicates that while the model effectively captures historical patterns, it may struggle to account for unpredictable real-world influences.

Another notable trend is the impact of hyperparameter tuning on model performance. Adjustments in parameters such as the number of LSTM layers, number of neurons, batch size, and learning rate significantly influence prediction accuracy. Optimal parameter selection leads to improved generalization and reduced overfitting.

Overall, the findings demonstrate that deep learning-based approaches, particularly LSTM networks, provide a promising solution for stock market prediction. The system shows strong potential for assisting investors and analysts in understanding market behavior, although further enhancements such as integrating sentiment analysis and real-time data can improve robustness and predictive reliability.

VII. FUTURE DIRECTIONS

Although the proposed stock market prediction system demonstrates promising results using deep learning techniques, there are several avenues for future research and improvement to enhance its accuracy, robustness, and real-world applicability.

One important direction is the integration of alternative data sources, such as financial news, social media sentiment, and macroeconomic indicators. Incorporating Natural Language Processing (NLP) techniques to analyze textual data can provide valuable insights into market sentiment and external factors that influence stock price movements, thereby improving prediction performance.

Another potential enhancement involves the adoption of hybrid and ensemble learning models. Combining multiple approaches, such as statistical models, machine learning algorithms, and deep learning architectures, can help capture both linear and non-linear patterns more effectively. Ensemble techniques can also improve model stability and reduce prediction variance.

The application of advanced deep learning architectures, such as Transformer-based models and attention mechanisms, represents a promising area for future exploration. These models have shown superior performance in sequence modeling tasks and may further improve the ability to capture long-range dependencies in financial time series data.

Real-time prediction and deployment is another critical area for development. Future systems can be designed to process streaming data and provide live stock price forecasts, enabling practical use in trading environments. This requires addressing challenges related to computational efficiency, latency, and scalability.

Additionally, expanding the system to perform multi-stock and portfolio-level prediction can provide a more comprehensive understanding of market dynamics. Modeling correlations and interdependencies between different stocks and sectors can enhance decision-making for investors.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 6, June 2026
DOI: 10.17148/IJARCCE.2026.15698

Finally, improving model interpretability and explainability is essential for gaining user trust and ensuring transparency in financial decision-making. Techniques such as explainable AI (XAI) can be incorporated to provide insights into how predictions are generated.Overall, future research should focus on developing more adaptive, scalable, and intelligent systems that can effectively handle the complexity and uncertainty of financial markets while delivering reliable and actionable predictions.

VIII. CONCLUSION

The research presented a stock market prediction system using Linear Regression and LSTM algorithms. Historical stock market data was collected, preprocessed, and used for training both models. The Linear Regression model provided a simple approach for identifying relationships between stock variables, while the LSTM model effectively captured long-term dependencies and time-series patterns.

The experimental results indicated that LSTM achieved higher prediction accuracy and lower error rates compared to Linear Regression. This demonstrates that deep learning techniques are more suitable for stock market forecasting due to their ability to handle complex and dynamic market behavior.

The proposed system can assist investors and analysts in making informed decisions based on predicted stock trends. Future enhancements may include sentiment analysis, hybrid models, and real-time stock market data integration to further improve prediction performance.

REFERENCES

[1] Prof. Samatha R. Swamy, Sarah Rukhmuddin Rajgoli and Tejas Hegde, "Stock Market Prediction with Machine Learning: A Comprehensive Review," 2024.
[2] S. Agarwal, S. Sharma, K. N. Faisal and R. R. Sharma, "Time-Series Forecasting Using SVMD-LSTM: A Hybrid Approach for Stock Market Prediction," Journal of Probability and Statistics, Wiley, vol. 2025, Article ID 9464938, 2025.
[3] Chin Yang Lin and João Alexandre Lobo Marques, "Stock Market Prediction Using Artificial Intelligence: A Systematic Review of Systematic Reviews," Social Sciences & Humanities Open, vol. 9, no. 1, Elsevier, 2024.
[4] F. M. P. Fozap et al., "Hybrid Machine Learning Models for Long-Term Stock Market Prediction," Journal of Risk and Financial Management, vol. 18, no. 4, 2025.
[5] S. Joshi, "Integrating LSTM and CNN for Stock Market Prediction," Journal of Artificial Intelligence and Technology, 2025.
[6] Rajneesh Chaudhary, "Advanced Stock Market Prediction Using Long Short Term Memory Networks: A Comprehensive Deep Learning Framework," arXiv, 2025.
[7] A. Rohan, "Artificial Intelligence in Financial Market Prediction: A Review of Modern Approaches," Frontiers in Artificial Intelligence, 2025.
[8] A. Chauhan, P. Mayur, Y. S. Gokarakonda, P. Jamie and N. Mehrotra, "Indian Stock Market Prediction Using Augmented Financial Intelligence ML," arXiv, 2024.
[9] K. Attaluri, M. Tripathi, S. Reddy and S. Shivendra, "News-Driven Stock Price Forecasting in Indian Markets: A Comparative Study of Advanced Deep Learning Models," arXiv, 2024.
[10] C. Angie, S. Achmad and A. Kurniawan, "Machine Learning Approaches in Stock Market Prediction: A Systematic Literature Review," Procedia Computer Science, vol. 216, pp. 96–102, 2023.
[11] C. Y. Lin and J. A. Marques, "Stock Market Prediction Using Artificial Intelligence," SSRN Electronic Journal, 2023.