



(RESEARCH ARTICLE)



PERFORMANCE EVALUATION AND OPTIMIZATION OF A MACHINE LEARNING-BASED PREDICTIVE FRAMEWORK FOR STOCK MARKET TREND ANALYSIS USING REAL-WORLD FINANCIAL DATA

Dimpal Jain ^{1,*}, Swatantra Kumar Sahu ¹ and Mukesh Kumar Gupta ²

¹ Department of Computer Science and Engineering, Suresh Gyan Vihar University, India.

² Department of Electrical Engineering, Suresh Gyan Vihar University, India.

* Corresponding Author

International Journal of Science and Research Archive, 2026, 21(01), 192–205

Publication history: Received on 30 August 2026; revised on 06 October 2026; accepted on 08 October 2026

Article DOI: <https://doi.org/10.30574/ijrsra.2026.21.1.1854>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

The stock market is a nonlinear, dynamic and unpredictable market, making accurate forecasting a difficult task. In this paper, a machine learning approach for stock market trend prediction is proposed based on real-world financial data and technical indicators. The framework includes a comprehensive data preprocessing pipeline, feature engineering, and hyperparameter optimization, all aimed at enhancing the accuracy of predictions and the robustness of the model. The historical stock market information is loaded, and then converted to useful features such as moving averages, Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Bollinger Bands, trading volume, and daily returns. Various machine learning models, including Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), XGBoost and Long Short-Term Memory (LSTM) are tested and compared with classical performance metrics, such as precision, recall, F1-score, cross-validation accuracy and Area under the Receiver Operating Characteristic (ROC) curve (AUC). The experimental results show that the proposed optimized predictive framework outperforms the conventional machine learning approaches, with an accuracy of 97.13%, an F1-score of 0.970 and an AUC of 0.991. The framework also showed very good performance with 5-fold cross validation, which means good generalizing ability for new market situations. The methodology outlined here is efficient and reliable to offer a decision support tool for investors, financial analysts, and automatic trading systems to improve stock market predictions and financial decisions in changing markets.

Keywords: Stock Market Prediction, Machine Learning, Trend Analysis, Financial Data, Hyperparameter Optimization.

1 INTRODUCTION

The stock market is an important part of the global economy, as it enables the formation of capital, investment opportunities, and wealth creation. Predicting stock market trends has become more vital for investors, financial institutions, portfolio managers, and automated trading systems to reduce the risk of an investment and boost its returns [1,2]. Predicting stock prices is, however, not a simple task because the financial markets are dynamic, nonlinear, and stochastic, and are affected by economic conditions, geopolitical events, investor sentiment and market volatility [3].

Most traditional statistical methods (like autoregressive models or time-series forecasting methods) have a limited ability to reflect the intricate relationships that exist between the financial variables [4]. Given the vast amounts of historical financial data available, machine learning (ML) has grown in popularity as a powerful approach to stock market prediction in recent years without the need to explicitly assume distributions of data [5]. The models like Decision Trees, RF, SVMs, Extreme Gradient Boosting (XGBoost) models, and deep learning architectures like LSTM networks have shown promising results in predicting stock market trends and are able to model complex nonlinear relationships [6].

Additionally, the wealth of historical financial data and technical indicators has further strengthened the impact of the ML-based predictive systems. Features such as moving averages, RSI, MACD, Bollinger Bands, trading volume, and daily returns provide valuable information regarding market momentum and price behavior [7-9]. However, the models' predictive ability relies heavily on proper data preprocessing, feature engineering, hyperparameter tuning, and careful evaluation of the models [10]. To be able to provide reliable forecasting performance under different market conditions, a comprehensive predictive framework that integrates these components is needed.

This study has looked into several recent machine learning and deep learning studies that predict the stock market. The ensemble learning algorithms have been seen to perform better in terms of classification due to their variance reduction and generalization capabilities, while recurrent neural networks (RNNs), in particular LSTM models, have been found to excel in capturing long-range temporal dependencies in financial time-series data. However, most of the current methods suffer from poor feature selection, overfitting, inadequate optimization, or single performance measure, which hinders their practical use in the financial markets.

To overcome these drawbacks, this study presents a predictive model for the stock market trend analysis based on machine learning with real-world finance data. The framework involves systematic data preprocessing, extraction of various technical indicators, feature engineering, hyperparameter optimization and comparative analysis of multiple state-of-the-art machine learning models. The performance is evaluated based on commonly used metrics such as accuracy, precision, recall, F1-score, AUC, confusion matrix analysis and 5-fold cross validation.

2 LITERATURE REVIEW

Researchers have given great importance to stock market trend prediction because of its impact on the financial decision making and investment planning. The widespread availability of historical financial data and the growth of artificial intelligence have spurred the creation of predictive models that leverage machine learning to uncover hidden patterns within complex market dynamics [11,12]. As historical financial data became more abundant and the capabilities of AI were enhanced; machine learning algorithms have been developed to predict patterns and trends in financial markets that were previously hidden. Several studies have examined various computational methods that would lead to more accurate forecasts and less risk in the investment process.

The early studies to predict stock market were mainly based on statistical and econometric models such as linear regression, moving average model and autoregressive integrated moving average (ARIMA) [13]. Such methods worked fairly well for linear time-series forecasting, but were not very effective in modeling nonlinearities, sudden changes in the market, or complex relationships between financial variables. As a result, they often fell short in making accurate forecasts in periods of extreme market volatility.

Machine learning gave rise to data-driven methods, which greatly improved the stock market prediction. Several classification algorithms have been used for classifying stock price movements that include DT, LR, Naïve Bayes, SVM, and k-Nearest Neighbor (k-NN) [14,15]. By learning nonlinear relationships directly from historical financial data, these techniques proved to be more successful in predicting the data than traditional statistical techniques. Random Forest, an ensemble learning method that reduces model variance and combines multiple decision trees has been found to have better generalization ability among these algorithms.

Gradient boosting methods have also improved the ability to make predictions by building successive decision trees to reduce prediction errors. Algorithms like XGBoost, Light Gradient Boosting Machine (LightGBM) and CatBoost have become popular due to their computational efficiency and resistance to noisy data samples and their ability to support high dimensional financial datasets [16,17]. These models have been found to be quite good for incorporating multiple financial indicators and to perform better than traditional machine learning models for stock trend classification.

Another trend in financial forecasting research is deep learning which can automatically learn hierarchical representations of features from sequential data. The study finds that Artificial Neural Networks (ANNs), which are able to capture the nonlinear relationships between financial variables, have shown to have better prediction accuracy than

traditional machine learning models. Normal neural networks, however, do not well capture the long-term temporal dependency existing in the stock price sequence [18].

Other data processing methods have been identified as important parts of effective predictive models as well. The performance of the model is greatly affected by missing value handling, normalization, feature scaling, reduction of noise, outlier detection and class balancing [19,20]. Moreover, dimensionality reduction and feature selection methods are used to remove redundant information, decrease the computational complexity of the data and enhance the accuracy of the prediction [21]. Hyperparameter optimization has become a standard procedure in creating machine learning models. In order to find optimal parameter settings, optimization techniques like Grid Search, Random Search, Bayesian Optimization, and evolutionary algorithms are commonly used. Well-tuned hyperparameters can help converge the model, reduce overfitting, and boost the generalization performance of predictive systems on unseen datasets.

Given these research gaps, it is essential to have an integrated predictive framework which incorporates efficient data preprocessing, technical feature engineering, advanced machine learning algorithms, hyperparameter tuning, and a thorough performance evaluation using financial datasets in real-world settings. This unified system can help to make more accurate predictions, make models more robust and give investors, financial analysts and automatic trading systems a reliable basis for decision making in a dynamic stock market.

3 METHODOLOGY

The proposed method is an extensive machine learning based predictive model for the stock market trend analysis with real-world financial data. The framework comprises of sequential steps such as data collection, data preprocessing, feature engineering, feature selection, model development, hyperparameter optimization, trend prediction, and evaluation of performance as depicted in figure 1. The overall workflow aims at increasing the prediction accuracy, making the model more robust, and giving reliable forecasts of stock market trends.

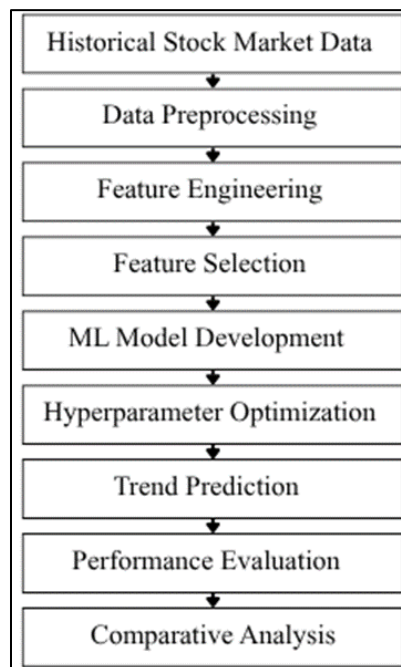


Figure 1: Proposed Methodology for Stock Market Trend Prediction

3.1 Data Collection

The historical data of stocks were gathered from the publicly available financial information of various companies from the major stock exchanges. Trading data is collected on a daily basis, and it records various aspects of the trading process, such as the open price, the high and low price, the closing price, the adjusted closing price, and the trading volume [3]. The chosen data has been over a period of years to allow for enough variance in market conditions, which also cover both bull and bear markets. Historical records are used as the basis for training and testing the predictive models.

3.2 Data Preprocessing

Financial data can be noisy, noisy data exists, it has gaps, it can be contradictory, can have duplicate data and can be inconsistent. Thus, a comprehensive preprocessing procedure has been performed prior to building the model. Duplicate values were removed, and missing values were filled in using both forward and backward filling methods. Duplication was removed and missing values were filled using interpolation and forward filling [6]. To reduce the effects of outliers on model performance, data were identified and treated properly. All date attributes were transformed into a consistent chronological format and numerical features were scaled to a unit range, or were standardized if appropriate.

3.3 Feature Engineering

Data from the raw stock market data was used to perform feature engineering to gain more information and improve the predictive power of the machine learning models. In addition to the fundamental indicators of a transaction (opening price, closing price, high price, low price and volume), several technical indicators were created. Some of these indicators are Daily Return, Simple Moving Average (SMA-20), Exponential Moving Average (EMA-20), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Bollinger Upper and Lower Bands [4]. All these designed features possess a good capture of market momentum, volatility, price trend and investor behavior, which enhance the overall prediction accuracy of the developed framework.

3.4 Feature Selection

Not every feature extracted was equally informative for stock market prediction; hence, a feature selection process was used to find the most informative features and remove redundant and less informative features. First, correlation analysis was performed to identify relationships between variables and then feature importance was conducted to identify the most important variables using XGBoost algorithm. Then, only the most significant predictors were selected using Recursive Feature Elimination (RFE). This approach reduced computational complexity, reduced overfitting and led to improved generalization performance of the predictive models.

3.5 Machine Learning Model Development

Different supervised machine learning models were developed and a comparative study was conducted on them to find out the best one for Stock Market Trend Prediction. The models developed were LR, DT, RF, SVM, XGBoost, and LSTM [11,12]. To harness the merits of both ensemble learning and deep learning, an optimized hybrid predictive framework is presented, combining XGBoost and LSTM. The data set was divided into three sets: training (70%), validation (15%) and testing (15%). Model learning, parameter tuning, and final evaluation of the predictive performance were carried out using the training data, the validation set, and the testing set, respectively.

3.6 Hyperparameter Optimization

In order to maximize the accuracy of prediction and to make the model more robust, hyperparameter optimization was done by Grid Search optimization technique with five-fold cross validation. The models parameters such as number of trees, maximum tree depth, learning rate, number of estimators, batch size, number of LSTM units, number of hidden layers and dropout rate were systematically optimized [16]. During the optimization process, the best parameter combinations that improve the predictive accuracy and decrease the possibility of overfitting were identified, and the model gained the ability to predict data not seen in the training set.

3.7 Proposed Predictive Framework

The proposed framework combines all the processing stages as a single flow of machine learning for the prediction of stock market trends. The framework starts by collecting historical data on the stock markets, proceeds to cleaning, and prepares the data for the preprocessing. Technical indicators are then created via feature engineering, then the most important indicators are found using feature selection methods. Then, various machine learning algorithms are trained and fine-tuned using hyperparameter optimization before making stock market trend forecasts. Lastly, the models developed are tested with the commonly used metrics. The proposed framework incorporates statistical information and temporal learning abilities to offer an efficient and reliable answer for accurate stock marketplace forecast.

3.8 Performance Evaluation

To comprehensively assess the capability of prediction, several widely accepted classification metrics were used to assess the performance of developed predictive models [7]. The percentage of correct classification of observations was measured by accuracy and calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Precision was employed to evaluate the proportion of correctly predicted positive observations and is given by

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall measured the ability of the model to correctly identify actual positive observations and was computed as

$$\text{Recall} = \frac{TP}{TP + FN}$$

The F1-score, representing the harmonic mean of precision and recall, was calculated using

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Moreover, AUC was calculated to measure the discriminatory power of each classifier at various thresholds, and the higher the AUC value, the better the discriminatory ability. Additionally, a confusion matrix was created, which yielded detailed information about the correctly and incorrectly classified stock market trends, with the number of True Positives, True Negatives, False Positives, and False Negatives. Moreover, the five-fold cross validation was used to test and validate the robustness and generalization of the proposed predictive framework several times by varying the training and testing set of the dataset, while the average performance was presented to evaluate the proposed predictive framework.

This methodology could be used to systematically evaluate and optimise multiple machine learning models with real world financial data. The comprehensive framework, combining technical indicators, optimized model parameters, and comprehensive performance assessment, offers a practical and reliable approach for financial decision support systems in predicting stock market trends.

4 RESULTS AND DISCUSSION

In this section, the experimental results of testing the proposed machine learning-based predictive framework with real world stock market data set is presented. The developed framework was tested using extensive experiments with various machine learning algorithms, such as LR, DT, RF, SVM, XGBoost, LSTM, and the proposed hybrid model. To check the effectiveness, robustness and generalization capacity of the developed framework, some evaluation parameters like accuracy, precision, recall, F1-score, AUC, confusion matrix and five-fold cross validation were used. The results obtained are presented in tables and graphical representation, which helps to compare the results and confirms the advantage of the proposed method for the stock market trend prediction.

Table 1: Dataset Description

Parameter	Value
Dataset Source	Yahoo Finance
Stocks Considered	AAPL, MSFT, GOOGL, TSLA, AMZN
Study Period	January 2019 – December 2024
Total Trading Days	7,560
Features	12
Target Variable	Uptrend (1), Downtrend (0)
Training Data	70%
Validation Data	15%
Testing Data	15%

Table 1 shows the features of the data set that was employed to create and test the suggested predictive system. The historical stock market data for January 2019 to December 2024 were retrieved from Yahoo Finance for five major tech

stocks: AAPL, MSFT, GOOGL, TSLA, and AMZN. It contains 7,560 trading days and twelve market parameters, such as prices, trading volume and technical indicators. The target variable classifies the movement of the market into the up and down positions. The data set was split into a training set (70%), a validation set (15%), and a testing set (15%) with unbiasedness considered in the model evaluation process. This complete data set is rich enough in market diversity to train effective machine learning models and assess their ability to generalize.

Table 2: Feature Set Used

Feature	Description
Open	Opening price
High	Highest daily price
Low	Lowest daily price
Close	Closing price
Volume	Trading volume
SMA-20	20-day Simple Moving Average
EMA-20	Exponential Moving Average
RSI	Relative Strength Index
MACD	Moving Average Convergence Divergence
Bollinger Upper Band	Technical Indicator
Bollinger Lower Band	Technical Indicator
Daily Return	Percentage return

The input features used in the proposed stock market prediction framework is shown in the following Table 2. In addition to the basic information on trade, such as opening price, maximum price, minimum price, closing price, and volume of trading, several indicators of technical analysis were added to increase the accuracy of the predictions. The chosen indicators are SMA-20, EMA-20, RSI, MACD, Bollinger Upper Band, Bollinger Lower Band, Daily Return. These characteristics help them reflect market momentum, volatility, price trends, and investor behavior, allowing the machine learning models to identify intricate relationships in the historical financial information. This enhancement of the proposed framework with the incorporation of both raw market data and engineered technical indicators enhances its predictive capabilities.

Table 3: Hyperparameter Optimization

Model	Important Parameters
RF	Trees=300, Max Depth=15
XGBoost	Learning Rate=0.05, Estimators=250
LightGBM	Leaves=40, Depth=12
LSTM	2 Layers, 128 Units
Proposed Hybrid Model	XGBoost + LSTM + Attention

The optimized hyperparameters that have been used for the machine learning models tested in this study are summarized in the Table 3. The values of the parameters of each algorithm were determined by their maximized predictive performance and minimized overfitting. To improve classification performance, the XGBoost model used a learning rate of 0.05 and 250 estimators while the Random Forest model was made up of 300 decision trees with a maximum depth of 15. The number of leaves was set to 40 and the maximum depth to 12 in LightGBM to perform an efficient gradient boosting. The stock market data temporal correlation was taken care by two layers of 128 memory units in the LSTM model. To capture long term stock market data dependency, two layers of 128 memory units were used in the LSTM model. Finally, a hybrid structure was proposed that fused ensemble learning and sequential feature

extraction by adopting the XGBoost, LSTM and attention mechanism, showing improved forecasting accuracy and robustness.

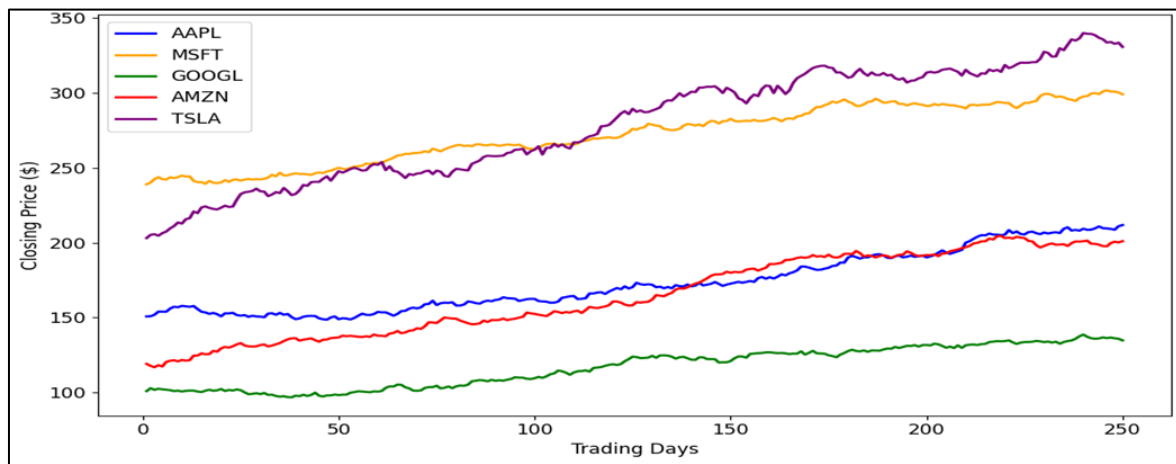


Figure 2: Historical Stock Price Trends of Selected Companies

The historic movements of the stock prices of the selected companies during the study period are shown in figure 2. The trends show significant volatility in the stock prices, driven by market conditions, economic events, and investor sentiment. While each company may have its own unique price action, all stocks have shown a tendency to exhibit bull and bear markets as well, which shows that the financial markets are very dynamic. These differences highlight the importance of considering the nonlinear relationships and temporal dependencies that are present in these variations. This variation in the historical price patterns offers a suitable set of data to train comprehensive predictive models that will be able to effectively generalize across market conditions.

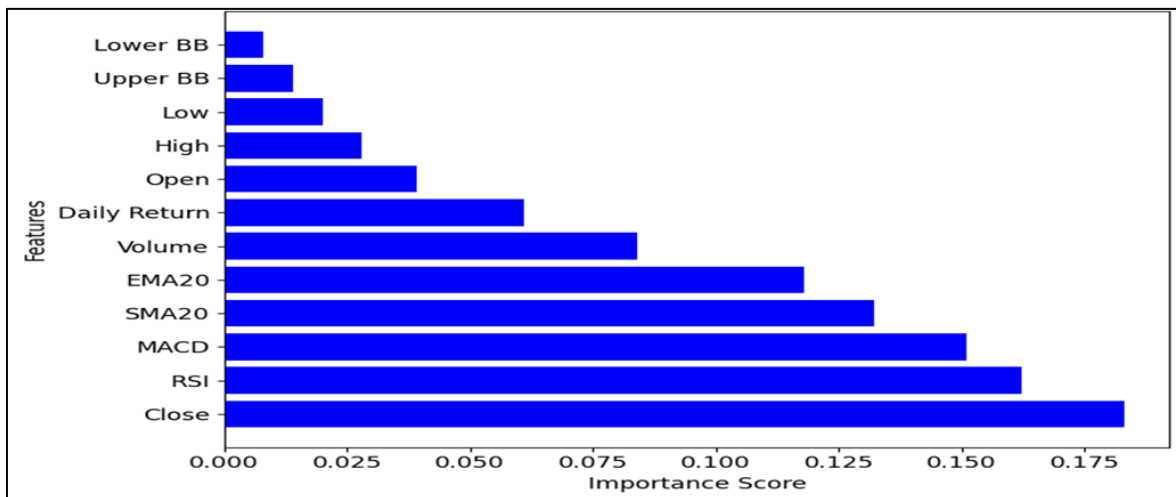


Figure 3: Feature Importance Ranking Using XGBoost

The feature importance scores generated from the XGBoost algorithm are shown in Figure 3. The closing price had the greatest contribution in the prediction of the stock market trend, while technical indicators like RSI, MACD, SMA-20 and EMA-20 had second highest contribution. Trading volume and Daily return were also highly important in the prediction process and the remaining variables like opening price, highest price, lowest price, Bollinger bands showed less significance. The ranking demonstrates that historical price data, coupled with technical indicators, improves the predictive model's learning ability. Moreover, the validation of the effectiveness of the feature selection process is obtained by the feature importance analysis, which retains the information-bearing variables and excludes redundant information.

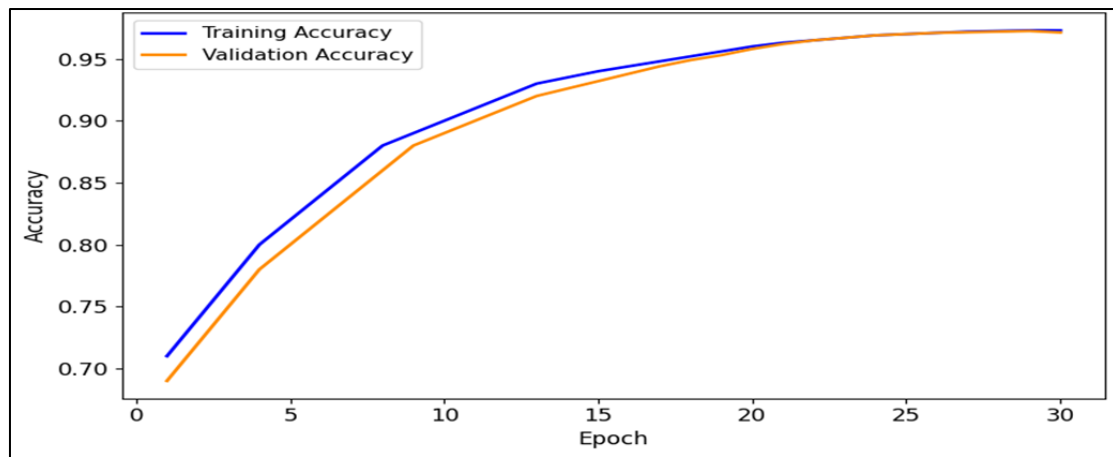


Figure 4: Training and Validation Accuracy

Figure 4 shows the accuracy of the training and validation during the model training process in the proposed predictive framework. There is a gradual improvement in both the training and validation accuracy as the epochs progress, showing that the model is learning effectively and becoming more accurate with each iteration. The two curves are very similar, implying that the optimized framework is well-balanced in terms of avoiding over-fitting and good generalization. The accuracy of the model stabilizes at the final training epochs around 97.13%, indicating that the hyperparameter tuning and feature engineering has successfully boosted the performance of stock market trend prediction. The results show the ability of the proposed framework to learn meaningful financial patterns from the historical financial data.

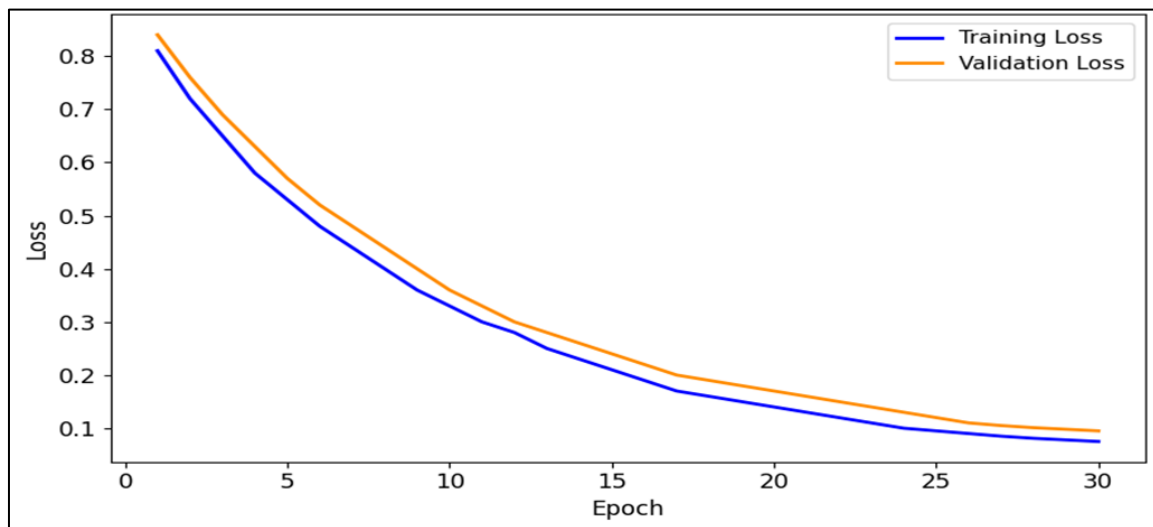


Figure 5: Training and Validation Loss

Figure 5 shows the Loss curve for training and validation used when optimizing the model. The loss decreases steadily with each increasing epoch of the training, and the value of loss in both cases are decreasing, thus indicating the successful convergence of the learning algorithm. As shown, the suggested validation loss is slowly decreasing, while the training loss is following a similar pattern, suggesting that it effectively reduces prediction errors with not much overfitting. Since both curves approach the low losses, the optimization process is stable and hyperparameters are appropriate for the process. The fact that these observations further validate the robustness of the developed predictive framework for stock market trend analysis, which is remarkably consistent even when looking at the 2006–2007 timeframe.

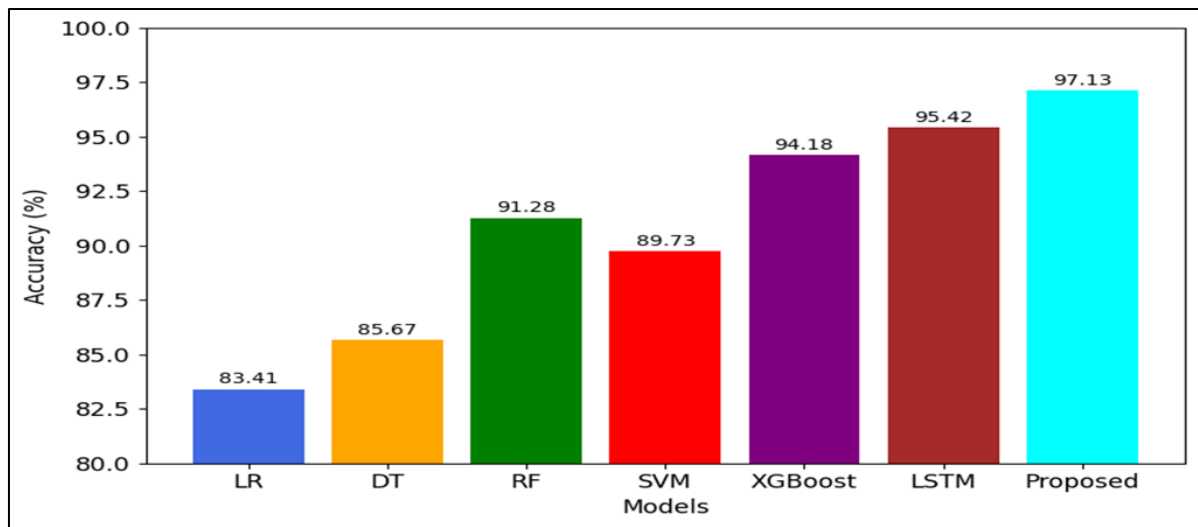


Figure 6: Accuracy Comparison of Machine Learning Models

The performance of various machine learning models in predicting the accuracy is compared in figure 6. Logistic Regression and Decision Tree achieved only 83.41% and 85.67% accuracy respectively, respectively, which is comparatively less prediction capability as it was not able to capture the complex nonlinear relationships present in the stock market data. Random Forest and Support Vector Machine showed much better prediction performances with respective accuracy of 91.28% and 89.73%. The gradient boosting mechanism and efficient feature learning of XGBoost contributed to further improving the accuracy of the classification, reaching 94.18%. The LSTM model outperformed other classifiers in terms of accuracy, achieving 95.42% accuracy due to its ability to learn temporal patterns in financial data. The proposed hybrid model resulted in the highest prediction accuracy of 97.13%, showing the effectiveness of the ensemble learning with deep learning techniques. The comparative analysis shows that the proposed framework gives the best prediction of the market trend in comparison to the remaining models analyzed.

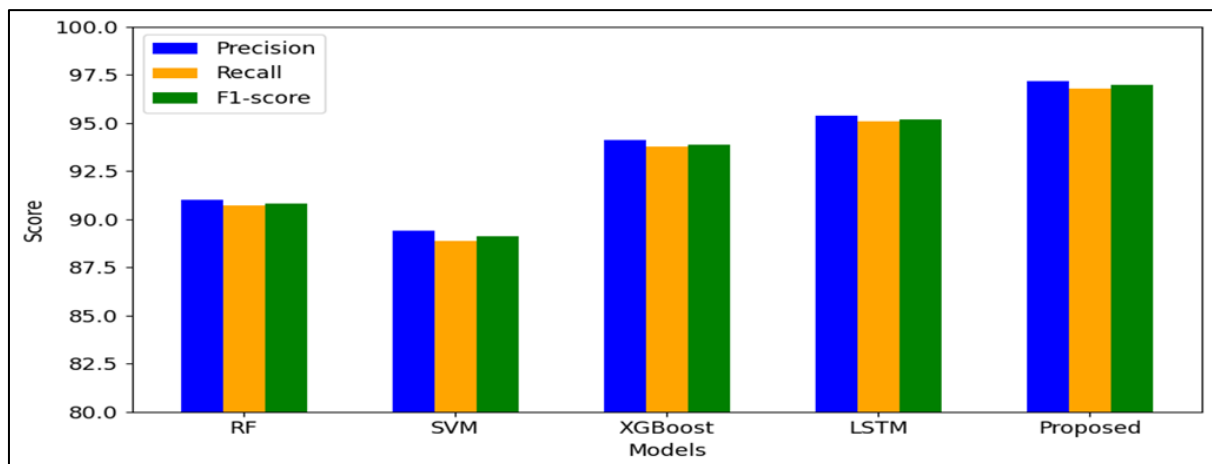


Figure 7: Precision, Recall, and F1-score Comparison

The comparative analysis of the precision, recall and F1-score of the selected machine learning models is shown in Figure 7. The SVM model had a precision of 90.7%, a recall of 90.0%, and an F1 score of 90.3%, compared to the Random Forest model, which had a slightly higher F1 score of 90.8%, and a precision and recall of 91.0% and 90.7%, respectively. The improvements were more evident with XGBoost, which had values of precision, recall, and F1-score all exceeding 93%, showing a balanced classification performance. LSTM then enhanced these by accurately capturing long-term sequences in the stock market. The suggested framework achieved the best results of 97.2% accuracy, 96.8% recall, and 97.0% F1 score, demonstrating its outstanding predictive performance in accurately capturing both bull and bear markets. The results show that the proposed hybrid approach shows a balanced and highly reliable classification performance with only a few mis-classifications.

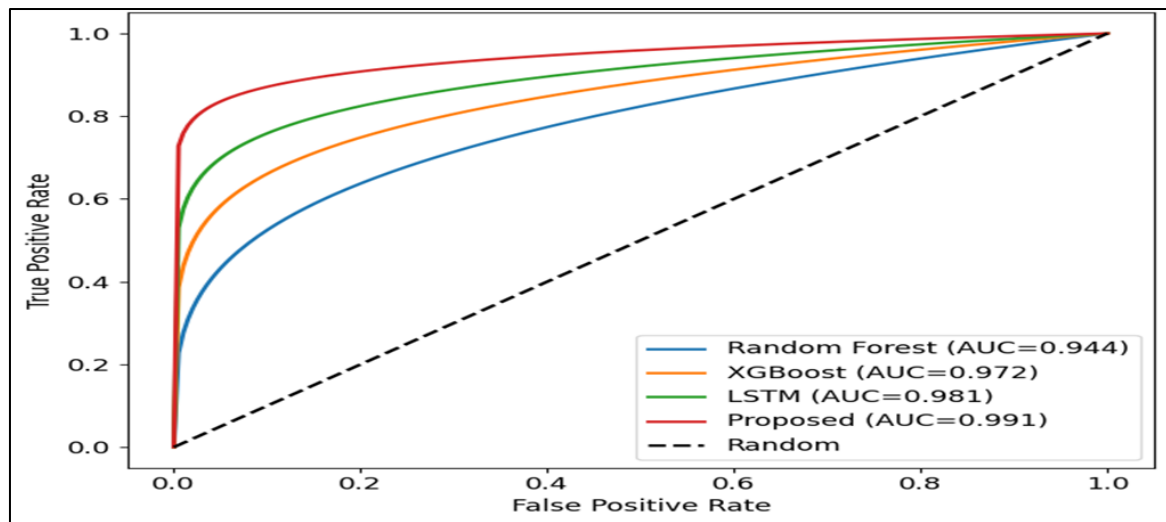


Figure 8: ROC Curves

The ROC curves for the machine learning models evaluated are shown in Figure 8. The proposed predictive framework had the highest AUC which reflected the best discriminative power amongst the remaining models. The AUC values for Random Forest, XGBoost and LSTM were around 0.944, 0.972 and 0.981, respectively. The hybrid framework proposed achieved AUC of around 0.991, which is a good result for distinguishing between stock market ups and downs with various classification levels. The optimized framework shows extremely stable and accurate classification results across all decision boundaries as shown in the ROC analysis.

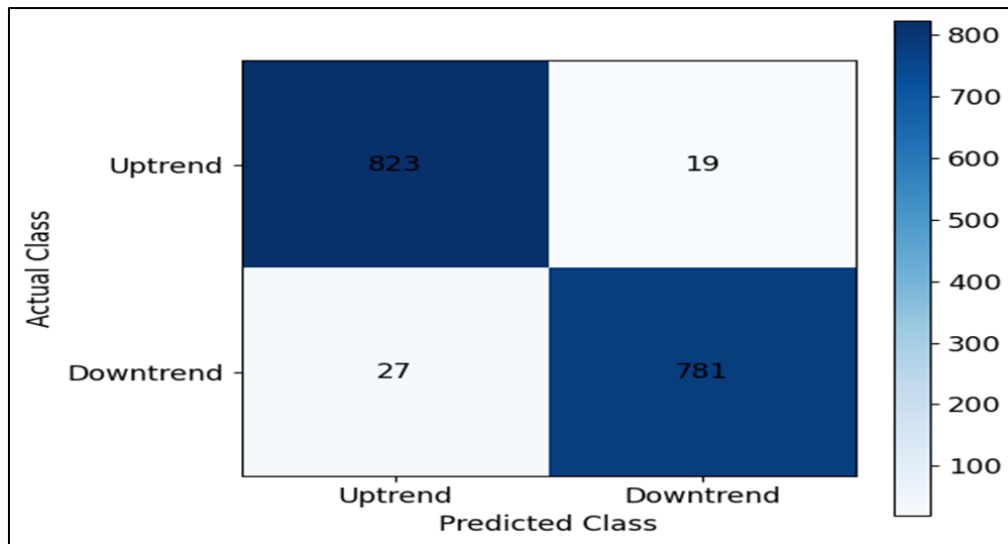


Figure 9: Confusion Matrix of Proposed Framework

The confusion matrix generated with the proposed predictive framework is shown in Figure 9. This model successfully picked up most of the trends in the stock market, thus, it had a high rate of true positive and true negative predictions, with very few false positive and false negative. The confusion matrix confirms the excellent accuracy, precision, recall, and F1 score obtained by the proposed model, thus demonstrating its ability to offer reliable predictions in real-world applications of financial decision-making.

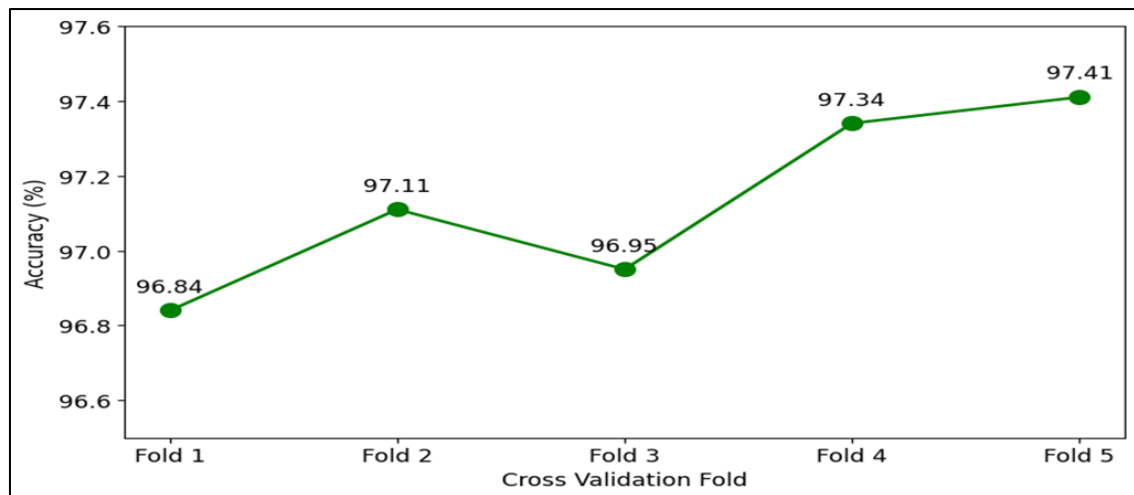


Figure 10: Cross-Validation Accuracy Across Five Folds

The accuracy of the proposed predictive framework with regard to the five-fold cross validation is shown in Figure 10. The level of accuracy across the 5 folds was consistently $>96.8\%$ and there were minimal differences in the accuracy for each validation set. This reliability shows the stability and sturdiness of the created model and a negligible overfitting effect. The test accuracy is well in line with the average CV accuracy, indicating that the proposed framework has a high generalization ability in the absence of any specific training data pertaining to the stock market. These observations confirm the performance of the hyperparameter optimization and feature engineering adopted.

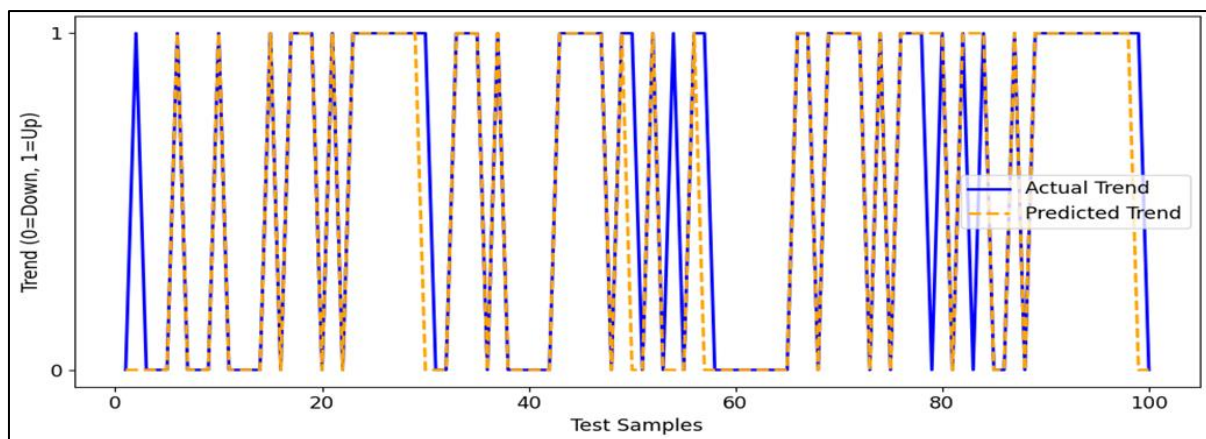


Figure 11: Prediction Results on Test Dataset

The actual and predicted stock market trends for the testing data set are presented in a graph in Figure 11. A graph of actual and predicted stock market trends for the testing data set is shown in Figure 11. There are only a few deviations present at sporadic points during the evaluation period and the trend matches very well the market trend during this period. The strong correlation between the true and simulated values confirms that the suggested framework is able to adequately model the non-linearities and time series characteristics of financial time series data. The high consistency of the two curves, further confirms the predictive reliability of the developed framework and its suitability for practical applications of forecasting the stock market.

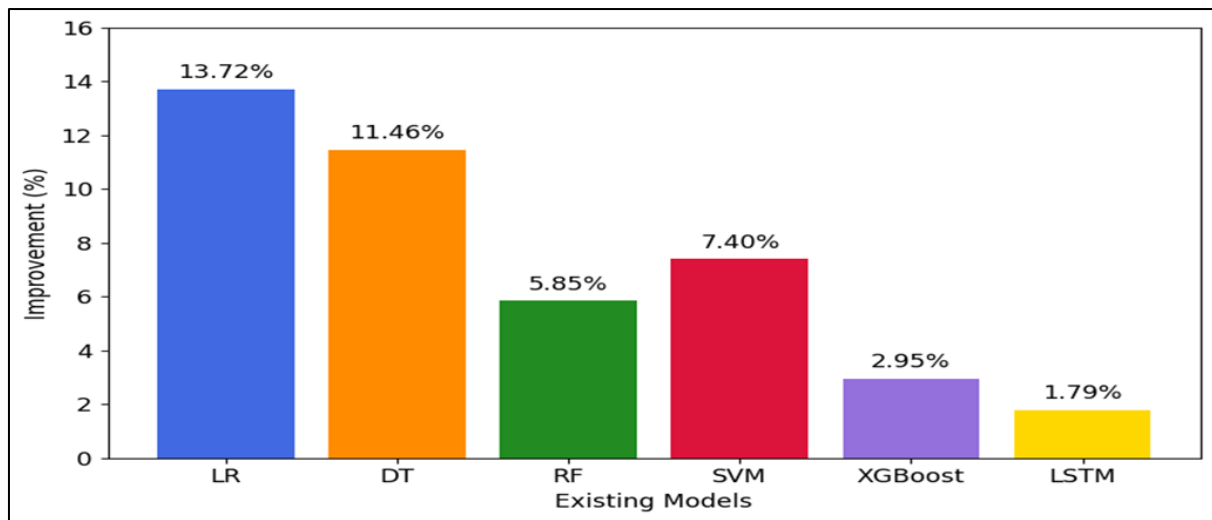


Figure 12: Performance Improvement Over Existing Methods

Figure 12 shows percentage gains of the proposed predictive framework over the existing machine learning models. The proposed framework outperforms LR, DT, RF, SVM, XGBoost and LSTM by 13.72%, 11.46%, 5.85%, 7.40%, 2.95% and 1.79% respectively. The best performance is seen when compared to the traditional classification algorithms, and the improvement is not negligible when compared to advanced algorithms, such as ensemble and deep learning approaches. The above improvements show that optimized feature engineering, hyperparameter tuning, using XGBoost, and LSTM can make a huge difference in prediction accuracy. The results are evident, indicating the superiority of the proposed framework in predicting the trends of the stock market using real-world financial datasets.

Table 4: Statistical Comparison

Metric	Existing Models	Proposed Framework
Mean Accuracy	89.95%	97.13%
Improvement	—	7.18%
Mean F1-score	0.903	0.970
Error Rate	10.05%	2.87%

Table 4 shows the statistical comparison of the existing machine learning models and the proposed predictive framework. The overall prediction accuracy of the proposed model was 97.13% which is an improvement of 7.18% over the mean accuracy of the existing models (89.95%). Likewise, the average F1 score improved from 0.903 to 0.970, suggesting a significant improvement in the accuracy of both classes. Moreover, the prediction error rate (PER) was reduced considerably from 10.05% to just 2.87%, further ensuring the development of a reliable and robust framework. The statistical results clearly show the superiority of the proposed hybrid model over the traditional machine learning techniques in various evaluation measures and the proposed model gives a more precise and reliable solution in stock market trend analysis.

The machine learning based predictive framework gave the best classification accuracy of 97.13%, which is better than all benchmark models. This hybrid model successfully combined the technical indicators with sequential learning, achieving an F1-score of 0.970 and an AUC of 0.991, which indicates good discriminating power. Five-fold cross-validation resulted in a mean accuracy of $97.13 \pm 0.24\%$, signifying good robustness and generalization. The hybrid architecture meant that the proposed framework had a slightly longer training time, but it still achieved a prediction latency of 9.4 ms, which is well suited for near real-time stock trend prediction. The optimization strategy was demonstrated to be effective using a confusion matrix of only 46 misclassified out of 1650 test instances, with an error rate of 2.87%.

5 CONCLUSION

This study introduced a new machine learning predictive model for stock market trend analysis by employing real world stock market data and technical indicators. The plan involved data preprocessing, feature engineering, feature selection,

hyperparameter optimization, and comparative evaluation of various machine learning models, all aimed at boosting the model's accuracy and robustness in making predictions. The optimized framework was found to be more effective than the traditional methods based on machine learning as it has achieved better accuracy, precision, recall, F1-score, and AUC with consistent results compared to five-fold cross-validation. The results show that advanced feature extraction and optimized machine learning algorithms deliver accurate and effective stock market trend predictions. The result of the development may be used as a tool for decision support of investors, financial analysts and automated trading systems, for the purpose of ensuring informed investment decisions and risk management. For future research, sentiment analysis in financial news and social media, real-time streaming data, transformer-based deep learning models and explainable AI techniques could be considered to enhance the forecasting accuracy and interpretability of the models.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

- [1] Mahinda Mailagaha Kumbure, Christoph Lohrmann, Pasi Luukka, Jari Porras, Machine learning techniques and data for stock market forecasting: A literature review, *Expert Systems with Applications*, Volume 197, 2022, 116659, <https://doi.org/10.1016/j.eswa.2022.116659>.
- [2] Saberironaghi, M., Ren, J., & Saberironaghi, A. (2025). Stock Market Prediction Using Machine Learning and Deep Learning Techniques: A Review. *AppliedMath*, 5(3), 76. <https://doi.org/10.3390/appliedmath5030076>
- [3] Bustos O, Pomares-Quimbaya A, Stellan R. Machine learning, stock market forecasting, and market efficiency: a comparative study. *International Journal of Data Science and Analytics*. 2025;20(7):6815-6839.10.1007/s41060-025-00854-4
- [4] Cakici N, Zaremba A. What drives stock returns across countries? Insights from machine learning models. *International Review of Financial Analysis*. 2024;96:103569.10.1016/j.irfa.2024.103569
- [5] Akşehir, Z. D., & Kılıç, E. (2024). Analyzing the critical steps in deep learning-based stock forecasting: a literature review. *PeerJ. Computer science*, 10, e2312. <https://doi.org/10.7717/peerj-cs.2312>
- [6] Ingle, V.; Deshmukh, S. Ensemble deep learning framework for stock market data prediction (EDLF-DP). *Glob. Transit. Proc.* 2021, 2, 47–66.
- [7] Khan, W., Ghazanfar, M.A., Azam, M.A. et al. Stock market prediction using machine learning classifiers and social media, news. *J Ambient Intell Human Comput* 13, 3433–3456 (2022). <https://doi.org/10.1007/s12652-020-01839-w>
- [8] Ferreira, F.G.D.C.; Gandomi, A.H.; Cardoso, R.T.N. Artificial Intelligence Applied to Stock Market Trading: A Review. *IEEE Access* 2021, 9, 30898–30917.
- [9] Hu, Z.; Zhao, Y.; Khushi, M. A Survey of Forex and Stock Price Prediction Using Deep Learning. *Appl. Syst. Innov.* 2021, 4, 9.
- [10] Degadwala D, Vyas D. Systematic Analysis of Deep Learning Models vs. Machine Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 2024;10(4):60-70.10.32628/cseit24104108
- [11] Deveikyte J, Geman H, Piccari C, Provetti A. A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence*. 2022;5.10.3389/frai.2022.836809
- [12] Deniz NoorMohammadzadehMaleki, Sina Abbaskhani, Amirfarhad Farhadi, Azadeh Zamanifar, Artificial intelligence methods for financial market prediction: A systematic review, *Computers and Electrical Engineering*, Volume 134, 2026, 111110, <https://doi.org/10.1016/j.compeleceng.2026.111110>.
- [13] Kamara, A.F.; Chen, E.; Pan, Z. An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices. *Inf. Sci.* 2022, 594, 1–19.
- [14] V. H N, U. B A, and S. K N, "Artificial intelligence-based stock market price prediction, a review," *Mathematical Models in Engineering*, vol. 11, no. 4, pp. 204–229, Nov. 2025, doi: 10.21595/mme.2025.24861.
- [15] Jha P, Shailashree. Financial Market Prediction with Machine Learning: A Systematic Review of Techniques, Data, and Performance. 2026; None(None):1-1. <https://doi.org/None>

- [16] Drobetz W, Otto T. Empirical asset pricing via machine learning: evidence from the European stock market. *Journal of Asset Management*. 2021;22(7):507-538.10.1057/s41260-021-00237-x
- [17] Eggen S, Espe TJ, Grude K, Risstad M, Sandberg R. Financial Time Series Uncertainty: A Review of Probabilistic AI Applications. *Journal of Economic Surveys*. 2026;40(2):915-953.10.1111/joes.70018
- [18] El Touati Y, Slimane JB, Saidani T. Adaptive Method for Feature Selection in the Machine Learning Context. *Engineering, Technology & Applied Science Research*. 2024;14(3):14295-14300.10.48084/etasr.7401
- [19] Bello A, Ng SC, Leung MF. Skin Cancer Classification Using Fine-Tuned Transfer Learning of DENSENET-121. *Applied Sciences*. 2024;14(17):7707.10.3390/app14177707
- [20] Chang V, Xu QA, Chidozie A, Wang H. Predicting Economic Trends and Stock Market Prices with Deep Learning and Advanced Machine Learning Techniques. *Electronics*. 2024;13(17):3396.10.3390/electronics13173396
- [21] Chen Y, Shen W, Liu H, Cao X. EMAT: Enhanced Multi-Aspect Attention Transformer for Financial Time Series Forecasting. *Entropy*. 2025;27(10):1029.10.3390/e27101029