Journal of Global Research in Electronics and Communication (JGREC)
Volume 2 | Issue 7 | July 2026 | Pages 08-14
DOI: https://doi.org/xxxxxxxxxxxxx
e-ISSN: 2321-3175
RESEARCH ARTICLE OPEN ACCESS
Available online: https://jgrec.info/index.php/jgrec/index

Hybrid Machine Learning and Deep Learning Framework for Accurate Stock Price Forecasting

Dr. Sunita Dixit
Professor
Department of Computer Science & Engineering in St. Andrews Institute of Technology & Management
Gurgaon
bhardwajsunita23@gmail.com
Corresponding Author* Sunita Dixit (bhardwajsunita23@gmail.com)

Received: 5 July 2026 | **Revised:** 10 July 2026 | **Accepted:** 25 July 2026

***Abstract*—Accurate stock price prediction and analysis are now essential for businesses and investors in today's market. It has been shown that the traditional methods for predicting and visualizing stock information are not accurate enough for capturing complex stock market trends and future data. This work proposes a novel hybrid approach that utilizes machine learning (ML) and deep learning (DL) techniques for intelligent stock price prediction and analysis. The main purpose of this study is to create a prediction model that can effectively investigate the historical stock data and accurately predict stock. The proposed model uses the Historical Stock Market Dataset having 9,909 records and 7 attributes, and data preprocessing, normalization, balancing using SMOTE class, along with suitable data splitting techniques are adopted to boost the reliability of prediction results. The Hybrid model presented in the paper is a combination of XGBoost+LSTM which extracts the nonlinear characteristics of both feature learning and temporal sequence in the stock market. The experimental results show that the proposed model has high predictive performance with an accuracy (ACC) of 93.9%, precision (PRE) of 93.7%, recall (REC) of 93.8%, and F1 score (F1) of 93.6%. The results demonstrate the feasibility of the proposed hybrid model in capturing complex relationships and sequential patterns in historical stock data, which has the potential to generate accurate predictions of stock prices and facilitate informed investment decisions.**

Keywords*—*Deep Learning, Temporal Data Analysis, Stock Market, Price Prediction, Artificial Intelligence, Hybrid Models, Financial Data.

I. Introduction

Financial forecasting is broadly utilized in the financial services sector, such as banking, cryptocurrency, investment, insurance, trading, portfolio management, and risk assessment, where precise predictions of stock prices enable[1] well-informed financial choices and economic planning [2][3]. The stock market is an instrument that allocates capital, allows firms to borrow money, and enables wealth creation to be of benefit to investors, financial institutions, and policymakers, thus, forecasting ACC is essential. Stock prices are volatile, dynamic, and nonlinear, and influenced by several factors, such as price history, macroeconomic factors, company performance, geopolitical events, and investor sentiment. Such properties make the forecasting of future stock prices a difficult problem [4][5].

Historical price patterns and statistical techniques are the main traditional methods used in stock forecasting and determining future market behaviours [6]. Market trends are often detected using technical indicators like opening price, closing price, high price, low price, moving averages, trading volume, and momentum. Using only one type of financial data, however, can miss some of the important relationships that can affect stock prices. The complexity and volatility of the markets have increased further with the rise of computer-based trading, algorithmic trading and high-frequency trading. Under varying market conditions, it is important to take into account several financial characteristics and the nonlinear relationships in forecasting models in order to ensure that they provide reliable predictions [7][8][9].

The increasing availability of structured financial data has created new opportunities for developing intelligent stock forecasting systems. Financial data may include past prices, volume, technical indicators, and fundamental factors that help reveal trends that can be linked to future changes in prices[10] [11][12]. Furthermore, the incorporation of various financial data can enhance market dynamics representation and minimize the limitations of traditional forecasting methods. Accurate stock price forecasting, as well as the ability to make educated investing and risk management decisions, depends on a framework that can understand complicated correlations from market-related and historical information [13][14][15].

Stock price forecasting has recently seen the rise of ML and DL as potent methods due to their capacity to represent intricate nonlinear linkages and temporal interdependence [16]. While DL approaches are great at learning sequential patterns and temporal dependencies from past stock-price data, ML methods are great at identifying complicated patterns and correlations among technical and financial aspects. However, using either approach independently may limit the ability to simultaneously exploit nonlinear feature relationships and sequential market patterns [17][18][19][20]. Hence, in this study, an ML-DL hybrid framework for accurate stock price forecasting is proposed to ensure

DOI: https://doi.org/xxxxxxxxxxxxx

maximum ACC and strength of the stock price forecasting technique by incorporating the complementary advantages of both ML and DL [21].

A. *Motivation and Contribution*

This study is initiated by the high volatility, nonlinear behaviour and dynamism of the stock markets which make forecasting stock prices a difficult task. The relationships among financial data, and the temporal patterns, are often complex, and traditional forecasting methods are unable to capture them. As more market data becomes available, opportunities arise to use ML and DL techniques better to predict the market. Therefore, this study develops a hybrid ML-DL framework that combines nonlinear feature learning with sequential pattern analysis to improve the ACC and reliability of stock price forecasting. This research offers several key contributions as listed below:

- Utilized the Historical Stock Market Dataset containing 9,909 records and 7 attributes for stock price prediction.
- Used preprocessing methods such as label encoding, noise reduction, duplication removal, missing-value handling, and StandardAero normalization.
- Applied SMOTE to achieve racial and economic parity and strengthen underrepresented groups' visibility.
- Developed a Hybrid XGBoost+LSTM model that combines nonlinear feature learning with temporal sequence modelling.
- Evaluated the proposed model using ACC, PRE, REC, and F1.
- Compared the proposed model with LR, DNN, BiLSTM, and MLP, showing superior overall predictive performance.

The justification for this study is based on the limitations of conventional forecasting methods in capturing the complex, nonlinear, and temporal characteristics of stock-market data. The approach proposed combines ML and DL methods, leveraging their complementary predictive power. The novelty of its design is the integration of nonlinear feature extraction and sequential pattern learning in a single Hybrid XGBoost+LSTM framework. To enhance data quality and learning effectiveness, the model is also subjected to systematic preprocessing and normalization techniques, as well as SMOTE class balancing. This integrated framework provides a robust and accurate approach for stock-price forecasting compared with individual ML or DL models.

B. *Structure of the Proposed Study*

The rest of this paper is organized as follows: Section I presents the introduction and literature review, Section II describes the research methodology, Section III discusses experimental results and comparative analysis, Section IV presents the conclusion and future study, followed by the references.

II. Literature Review

This study involved a detailed review and critical analysis of recently published studies that focused on accurate stock price forecasting to determine key methodologies, findings, and research gaps. Table I summarizes these studies by highlighting the proposed approaches, datasets employed, significant findings, and key limitations.

Kouakan Adanin and Balungu (2025) aim to compare ARIMA and SARIMAX with advanced ML models, specifically multivariate LSTM and LightGBM, for predicting U.S. wheat stock prices from 2021 to February 21, 2025. The dataset includes historical prices and statistical measures from lagged values over 1, 7, and 30 days. Results show that multivariate LSTM achieved the best performance, with MSE of 0.0015 and RMSE of 0.0389, outperforming LightGBM and tuned SARIMAX. Its superior performance is attributed to its ability to capture complex temporal dependencies and nonlinear patterns in financial time series data [22].

Anand and Yadav (2025) propose a real-time stock market prediction framework using LSTM networks and prominent technical indicators. Yahoo Finance API data are enhanced with RSI, SMA, EMA, MACD, and Bollinger Bands. The LSTM model is compared with LR, SVM, and RF using MSE, MAE, and R^2. Results show that LSTM achieved superior ACC, with MSE of 0.004 and R^2 of 0.95, demonstrating its effectiveness in capturing temporal dependencies and nonlinear patterns in financial data [23].

Jakhar, Sharma and Ahmed (2024) proposed using 15 years of historical time-series data from 2009 to 2023, with 12 years for training and 3 years for testing. Day-wise high prices were predicted using open, low, and close prices. The study evaluated MLR, SVM with Linear, Polynomial, and RBF kernels, MLP (2,2), and XGBoost using MAE and RMSE. MLR achieved the best performance, with the lowest MAE of 11.57 and RMSE of 14.69 [24].

Kihtir and Oztoprak (2024) a comparison of hybrid KNN-SVM and KNN-LSTM models for stock price prediction using historical data from AAPL, GOOGL, and MSFT from 2020 to 2023. Results show varied performance across stocks, with KNN-SVM achieving the best results for GOOGL (RMSE: 2.11, R^2: 0.968), while recording RMSE values of 13.56 for AAPL and 25.39 for MSFT. The findings indicate that hybrid model effectiveness varies with stock characteristics and market conditions [25].

Lavanya and Gnanasskaran (2023) aim to identify patterns from historical data to predict future stock values. Using two years of Yahoo Stock Market data, the study employs feature engineering and ML to forecast stock price trends based on open, high, low, and close prices. The closing prices for the following five days are predicted using an LR model. With a 92% ACC rate, the suggested LR model outperformed all previous models. [26].

Baihaqi et al. (2023) proposed to compare DL models, including LSTM, Bi-LSTM, and GRU, for stock-price forecasting using historical stock market data. The study evaluates model ACC and the effect of training epochs using RMSE, MSE, and MAE. Results show that Bi-LSTM outperforms the other models, achieving RMSE, MSE, and MAE values of 0.00029, 0.00029, and 0.01, respectively [27].

Reddy and Magesh Kumar (2022) used GBM A stock values instead of the NB Algorithm, and wanted to assess the ACC of stock price predictions. To estimate the ACC of stock prices, use the NBs Algorithm iteratively using a sample size of 20 for the GBM Algorithm. The overall forecast error is mitigated by the Novel Loss Function of the GBM Algorithm, which is grounded on historical stock prices. The ACC of the GBM approach is 92.3%, which is much higher than the 87.7% ACC of the NBs algorithm [28].

TABLE I. SUMMARYY SUMMERY OF THE STUDY ON ACCURATE STOCK PRICE FORECASTING WITH HYBRID ML AND DL

Author	Proposed Work	Dataset	Key Findings	Limitations & Future Work
Kouakan Adanin and Balungu (2025)	ARIMA, SARIMAX, multivariate LSTM, and LightGBM	U.S. wheat stock data, 2021–2025	LSTM achieved the best performance with MSE = 0.0015 and RMSE = 0.0389; no accuracy percentage was reported.	Hybrid ML-DL integration was not explored. Future work can combine boosting and LSTM models.
Anand and Yadav (2025)	LSTM with RSI, SMA, EMA, MACD, and Bollinger Bands	Yahoo Finance historical stock data	R^2 = 0.95 = 95%, with MSE = 0.004.	Focuses mainly on LSTM; hybrid ML-DL architecture can be explored.
Jakhar, Sharma and Ahmed (2024)	MLR, SVM, MLP, and XGBoost	Stock data from 2009–2023	No percentage accuracy was reported; MLR obtained MAE = 11.57 and RMSE = 14.69.	Does not combine boosting with recurrent DL models.
Kihtir and Oztoprak (2024)	KNN-SVM and KNN-LSTM hybrid models	AAPL, GOOGL, MSFT, 2020–2023	GOOGL: R^2 = 0.968 = 96.8% with KNN-SVM; RMSE = 2.11.	Performance varies across stocks; more robust hybrid models are needed.
Lavanya and Gnanasskaran (2023)	Linear Regression with feature engineering	Yahoo Stock Market data	92% accuracy using Linear Regression.	Uses a relatively simple model and limited historical period.
Baihaqi et al. (2023)	LSTM, Bi-LSTM, and GRU	Historical stock-market data	Bi-LSTM achieved the best performance: RMSE = 0.00029, MSE = 0.00029, MAE = 0.01; no accuracy percentage reported.	Does not combine DL with boosting-based ML.
Reddy and Magesh Kumar (2022)	GBM vs Naive Bayes	Stock-price data	GBM = 92.3% accuracy, while Naive Bayes = 87.7%.	Limited reported sample size and no DL component.

Research Gaps: The methodology proposed is to use Kaggle's Historical Stock Market Dataset, which has 9,909 records and 7 attributes, followed by data pre-processing, normalization, balancing using SMOTE, and splitting the data. The Hybrid XGBoost+LSTM model combines nonlinear feature learning with temporal sequence modelling to achieve accurate stock-market price prediction. Finally, the performance of the model is evaluated using REC, ACC, PRE, F1, confusion matrix, and ROC analysis.

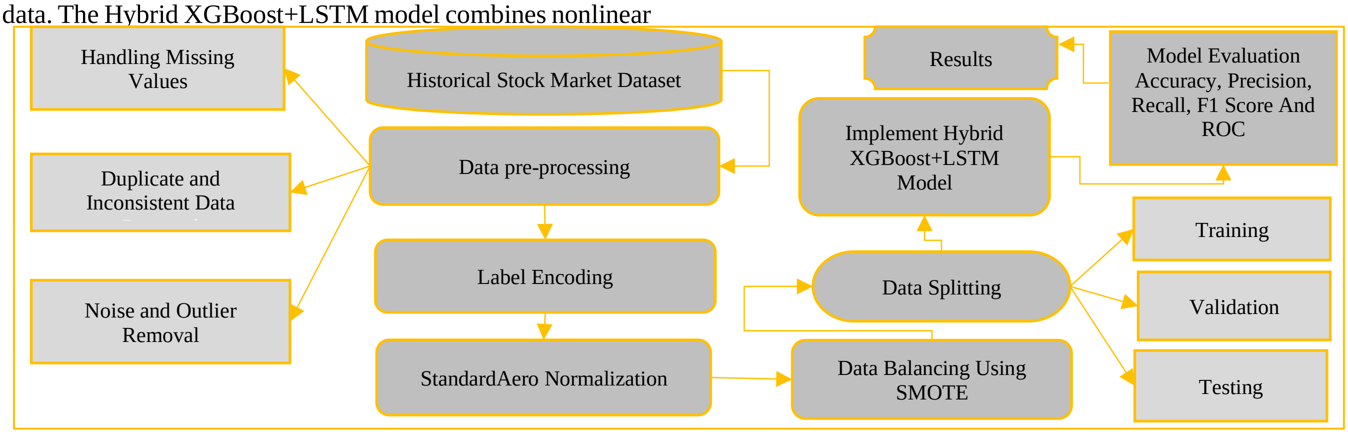

Fig. 1. Proposed Flowchart for Accurate Stock Market Price Prediction

III. RESEARCH METHODOLOGY

The methodology proposed is to use Kaggle's Historical Stock Market Dataset which has 9,909 records and 7 attributes, followed by data pre-processing, normalization, balancing using SMOTE, and splitting the data. The Hybrid XGBoost+LSTM model combines nonlinear feature learning with temporal sequence modelling to achieve accurate stock-market price prediction. Finally, the performance of the model is evaluated using REC, ACC, PRE, F1, confusion matrix, and ROC analysis. Fig. 1 illustrates the proposed flowchart for Accurate Stock Price Forecasting.

The following section presents a detailed explanation of each step involved in the proposed methodology:

A. *Data Gathering and Analysis*

The Historical Stock Market Dataset was obtained from Kaggle and comprises 9,909 records with 7 attributes, containing key historical market information such as date, opening, closing, highest, lowest, adjusted price, and trading volume. The distributions and correlations of the set of data variables were analyzed using data visualization techniques such as bar plots and heat maps:

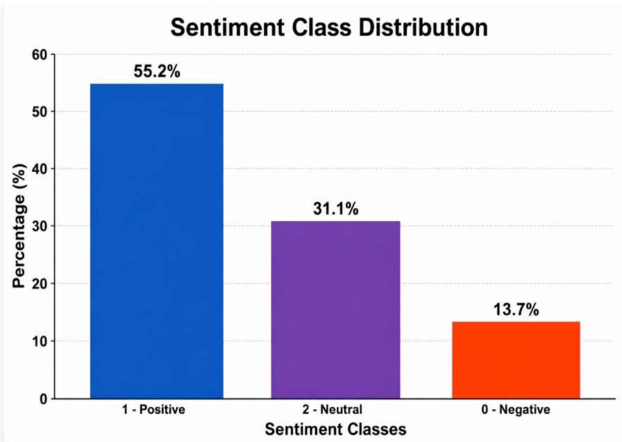

Fig. 2. Bar Graph of Class Distribution

The results of the three classes of sentiment are presented in Fig. 2, and the Positive sentiment class is the largest at 55.2%. The neutral sentiment represents 31.1%, with the lowest sentiment being negative at 13.7%. Overall, the distribution indicates a clear positive sentiment class with a less frequent negative sentiment class and neutral sentiment class.

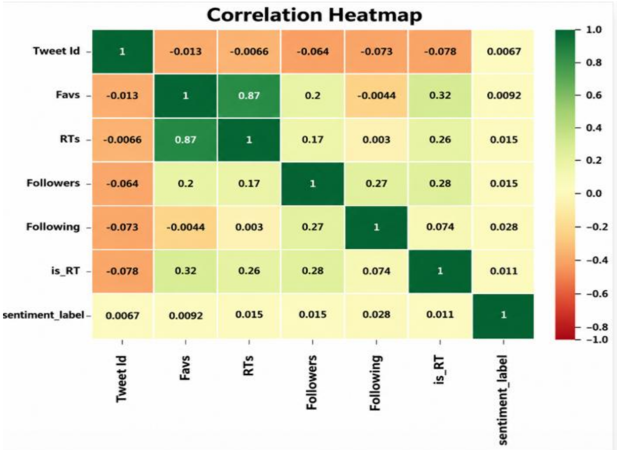

Fig. 3. Correlation Matrix Heatmap

The correlation heatmap (Fig. 3) illustrates the relationships between the features of the dataset, with Favs and RTs having the highest positive correlation (0.87). The other variables have relatively low to moderate correlations, and sentiment label has low correlations with the remaining variables. The heatmap reveals that there is minimal multicollinearity between the majority of the features, suggesting they are appropriate for subsequent analysis and predictive modelling.

B. Data Pre-Processing

Data preparation included concatenation of data from the CICIDS 2017 dataset, cleansing and feature engineering. The pre-processing steps include missing value handling, duplication elimination, noise and redundant value elimination, and data labelling and normalization. The main pre-processing actions are briefly summarized below:

- **Handling Missing Values:** Records of the stock exchange are identified as missing or incomplete data, and then interpolated or imputed with the mean values. This allows for a full picture and prevents loss of observation from influencing model training.
- **Duplicate and Inconsistent Data Removal:** Duplicate data and inconsistent values are detected and eliminated to ensure data quality and reliability. This ensures that each observation is unique and suitable for accurate stock-price forecasting.
- **Noise and Outlier Removal:** Appropriate filtering techniques are used to detect and remove noise and abnormal fluctuations in historical stock prices. The method used here is called Variational Mode Decomposition (VMD) which extracts meaningful price patterns from unwanted noise.
- **Label Encoding:** Label Encoding transforms categorical values into discrete labels, enabling ML models to handle categorical data efficiently. This transformation does not add any further dimensions to the data and maintains the identity of the categories.

C. Standardscaler Normalization

The dataset was transformed using the StandardScaler function to normalize the data, resulting in a distribution with a mean of zero and a standard deviation of one. This was done to account for the varying scales of each descriptor. Achieving this transformation is as simple as taking the standard deviation from each observation and dividing it by the mean, as seen in Equation (1):

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

where z represents the feature's modified value, x stands for each descriptor's original value, μ denotes the feature's mean, and σ denotes its standard deviation within the dataset.

D. Data Balancing using Synthetic Minority Over-Sampling Technique (SMOTE)

Data balancing reduces class imbalance by boosting the representation of underrepresented classes in the dataset. Synthetic minority over-sampling technique (SMOTE) was used to create synthetic samples in this investigation. Using Euclidean distance, SMOTE finds the instances in the minority class that are closest to each other, and then it generates fresh samples by interpolating between those examples and their neighbours. With the synthetic observations added to the training data, the class distribution is more evenly distributed, and the ML model is able to learn minority-class patterns more successfully.

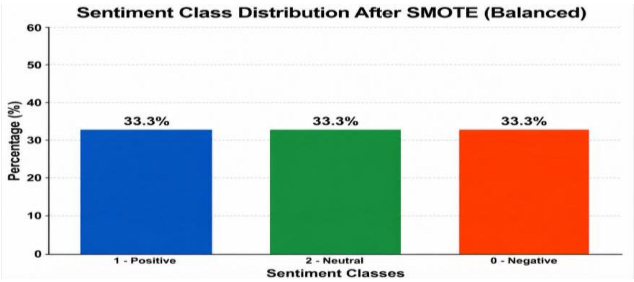

Fig. 4. Bar Graph of Class Distribution After SMOTE

Fig. 4. The bar graph illustrates the class distribution after applying SMOTE, where the Positive, Neutral, and Negative classes are equally represented at 33.3% each. This optimal distribution helps to minimize class imbalance and avoid majority class bias. Therefore, the predictive model is able to learn patterns of all sentiment classes better.

E. Data Splitting

The original dataset was divided into three sets: 72% for training, 8% for validation, and 20% for testing, providing a proper distribution for developing, tuning, and testing the model.

F. Proposed Hybrid XGBoost+LSTM Model

The capabilities of XGBoost and LSTM networks are combined in the Proposed Hybrid XGBoost+LSTM Model to provide precise stock market price prediction. When it comes to financial characteristics, XGBoost is great at capturing nonlinear correlations, whereas LSTM is great at learning sequential patterns and temporal dependencies. The XGBoost component first generates informative predictions/features, which are then provided to the LSTM component to model time-dependent behaviour and produce the final prediction. The overall hybrid prediction process is represented by Equation (2).

$$F_t = XGBoost(X_t) \tag{2}$$

DOI: https://doi.org/xxxxxxxxxxxxx

where X_t represents the input stock-market features at time t, and F_t denotes the feature representation or prediction generated by the XGBoost component.

The extracted XGBoost representation is subsequently arranged as a temporal sequence and supplied to the LSTM network. The LSTM uses its memory mechanism to retain relevant information from previous time steps while processing the current input. The LSTM operations are defined using the input gate, forget gate, output gate, and cell-state update, as given in Equations (3) to (6).

$$i_t = \sigma(W_i F_t + U_i h_{t-1} + b_i \quad (3)$$

$$f_t = \sigma(W_f F_t + U_f h_{t-1} + b_f \quad (4)$$

$$o_t = \sigma(W_o F_t + U_o h_{t-1} + b_o \quad (5)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh\,(W_c F_t + U_c h_{t-1} + b_c \quad (6)$$

here, i_t, f_t, and o_t stand for the input, forget, and output gates, respectively; C_t is the current cell state; h_{t-1} is the prior hidden state; W and U are weight matrices; b is the bias term; σ is the sigmoid activation function; and $\odot$ is element-wise multiplication.

A fully connected output layer is used to acquire the final stock-price forecast from the hidden representation generated by the LSTM. This last forecast combines LSTM's temporal learning capabilities with XGBoost's nonlinear feature-learning capabilities. The suggested hybrid model's ultimate output is given by Equation (7).

$$\hat{Y}_t = W_y h_t + b_y \quad (7)$$

where $\hat{Y}_t$ represents the predicted stock price, h_t is the current LSTM hidden state, and W_y and b_y denote the output-layer weights and bias, respectively. The proposed Hybrid XGBoost+LSTM Model is a unified model that can capture complex feature relationships and sequential stock-market patterns, thereby enhancing the reliability of the stock price prediction. The proposed Hybrid XGBoost+LSTM model was set with XGBoost: 200 estimators, learning rate=0.05, max depth=6, and subsample=0.8, and LSTM with 2 layers, 128 hidden units, dropout=0.2, batch size=32, epochs=50, learning rate=0.001. To ensure stability and efficient convergence, LSTM training was carried out using the Adam optimizer. For stable and efficient convergence, the Adam optimizer was used for LSTM training.

G. Evaluation Metrics

The proposed model was tested using different standard classification metrics. The numbers of correct and incorrect classifications per class were presented in a confusion matrix. The matrix was then used to identify the True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), which were then used to calculate the ACC, PRE, REC, and F1 as described below in Equations (8) to (11):

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \quad (8)$$

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \quad (9)$$

$$Recall = \frac{\text{TP}}{TP+FN} \quad (10)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

The ACC is the number of predictions that are correct divided by the total number of instances in the data set. PRE is the ratio of actual positive predictions to total positive predictions and measures the ACC of the positive predictions. REC is the percentage of positive instances that are correctly identified by the model. The F1, which falls in the range of 0 to 1, is the harmonic mean of PRE and REC. The ROC curve is a plot that compares the TPR with FPR across various decision thresholds of the model.

IV. Results and Discussion

This Section provides a brief description of the experimental setup and analyzes the performance of the suggested model at each stage, demonstrating its effectiveness, dependability, and computing efficiency. Using the data processing, graphing, and analysis packages Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, and Python 3.10.6, the model was developed. Operating System: Windows 11 Pro, Hardware: Intel Core i9-13900K CPU (3.0 GHz), Graphics: NVIDIA RTX 4090 with 24GB VRAM, and Random Access Memory: 64 GB were all used to conduct the studies. Training and testing of the Hybrid model were conducted using the Historical Stock Market Dataset available on the Kaggle website. The model was evaluated using their performance measures, which include ACC, PRE, REC, and F1. There is consistent classification performance for accurate stock market price prediction using the suggested strategy, which yielded 93.9% ACC, 93.7% PRE, 93.8% REC, and 93.6% F1 (Table II).

TABLE II. Classification Results of the Proposed Approach for Accurate Stock Market Price Prediction

Matrix	Proposed Hybrid XGBoost+LSTM
Accuracy	93.9
Precision	93.7
Recall	93.8
F1-score	93.6

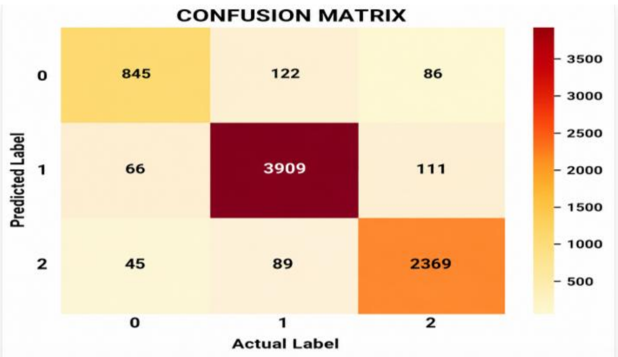

Fig. 5. Confusion Matrix for the Proposed Hybrid XGBoost+LSTM Model

Fig 5 shows the confusion matrix for the three classes (0, 1, and 2) of the proposed hybrid XGBoost+LSTM model. The diagonal values of 845, 3909, and 2369 represent correctly classified instances for the respective classes. High values on the diagonal represent high classification accuracy, and low off-diagonal values show that there are few misclassifications.

The proposed Hybrid XGBoost+LSTM model for Class 0, Class 1 and Class 2 is still in the top left corner of the ROC curve, as illustrated in Fig. 6. providing high TPRs and low FPRs. The AUC values are 0.94, 0.97 and 0.98, which reflect very good class discrimination. Overall, the curves validate the good and robust classification performance of the model.

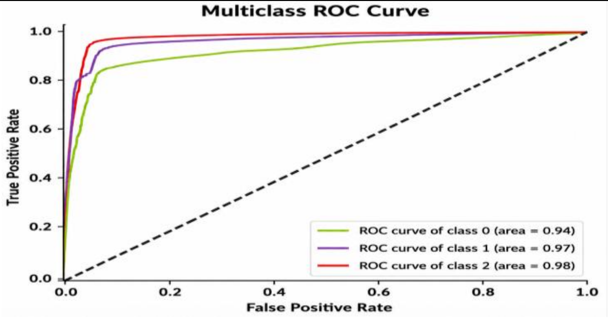

Fig. 6. ROC Curve for Proposed Hybrid XGBoost+LSTM Model

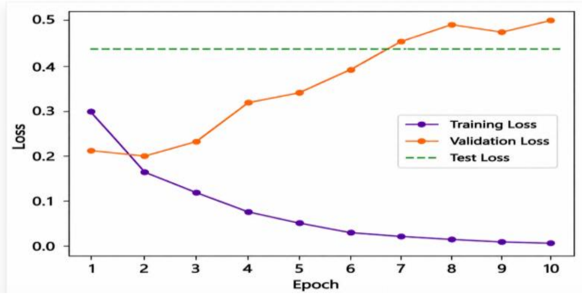

Fig. 7. Training, Validation, and Test Loss Curves for the Proposed Hybrid XGBoost+LSTM Model

The training loss gradually decreases with epochs, demonstrating successful model learning as visualized in Fig 7. Initially the validation loss drops, but then slowly rises, whereas the test loss is almost constant. The learning behaviour of the model is showcased in these trends and they indicate the model's generalization performance.

A. *Comparative Analysis*

The effectiveness of the proposed approach was compared to a number of existing ML and DL models as listed in Table III. The proposed hybrid method has the highest overall ACC (93.9%), PRE (93.7%), REC (93.8%) and F1 (93.6%). In comparison, the existing models achieved lower ACC values, including 72.7% for LR, 80.7% for DNN, 88.1% for BiLSTM, and 90.0% for MLP. The proposed approach also illustrates balanced PRE, REC, and F1, which makes it capable of identifying the movement of the stock market prices. The results prove that the suggested hybrid method yields better and stable classification results for the accurate prediction of the stock market prices.

TABLE III. COMPARISON OF DIFFERENT MACHINE LEARNING AND DEEP LEARNING MODELS FOR ACCURATE STOCK MARKET PRICE PREDICTION

Model	Accuracy	Precision	Recall	F1-score
LR[29]	72.7	72.9	72.7	72.6
DNN[30]	80.7	-	-	-
BiLSTM[31]	88.1	-	-	87.5
MLP[32]	90	-	92	90
Proposed Hybrid XGBoost+LSTM	93.9	93.7	93.8	93.6

The combination of XGBoost and LSTM in the proposed hybrid model is expected to yield more accurate predictions, harnessing the strengths of both ML and DL techniques. It is quite useful for handling the complex relationships of features and sequences in the stock market. The model can predict with a greater ACC of 93.9%, which is better than the existing approaches. Thus, it delivers a dependable and effective model for forecasting stock market prices accurately.

V. CONCLUSION AND FUTURE STUDY

The accurate prediction of stock price movements is crucial to informed investment decisions and financial risk management in highly dynamic and non-linear markets. This work proposed Hybrid XGBoost+LSTM framework with the Kaggle Historical Stock Market Dataset that has 9,909 rows and 7 attributes. The methodology involved data preprocessing, normalization using StandardScaler, class balancing with SMOTE and performance evaluation. By combining XGBoost with nonlinear feature learning and LSTM for temporal sequence modelling, the framework was able to capture complementary patterns within the historical financial data. The use of XGBoost for nonlinear feature learning and LSTM for temporal sequence modelling allowed the framework to learn complementary patterns within the historical financial data. The proposed model delivers an ACC of 93.9%, PRE of 93.7%, REC of 93.8% and F1 of 93.6%, which is better than LR, DNN, BiLSTM and MLP. Moreover, the ROC-AUC values of 0.94, 0.97, and 0.98 indicated excellent multi-class discrimination. Overall, the results validate the proposed hybrid model as an efficient and effective technique for predicting stock prices and assist in making informed trading decisions.

In future, these analyses can be expanded by leveraging more comprehensive and varied datasets that include real-time market data, technical analysis indicators, macroeconomic data, and investor sentiment. To enhance the robustness of predictions, advanced architectures such as attention mechanisms, transformers, and ensemble methods can be explored. Cross-market generalisation, computational efficiency and real-time deployment under varying market conditions are additional areas that can be explored further.

- **Funding:** This research received no external funding.
- **Institutional Review Board Statement:** Not applicable.
- **Informed Consent Statement:** Not applicable.
- **Data Availability Statement:** The data supporting the findings of this study are available from the corresponding author upon reasonable request.
- **Conflicts of Interest:** The author declares no conflict of

How to cite this article:
S. Dixit, "Hybrid Machine Learning and Deep Learning Framework for Accurate Stock Price Forecasting," *Journal of Global Research in Electronics and Communication*, vol. 2, no. 7, pp. 08-14, Jul. 2026. Doi: XXXX

REFERENCES

[1] C. S. Pareek, "Testing Approaches in Life Insurance: Accelerated and Fluid less Underwriting Amidst Data-Driven Dynamics," *Int. J. Sci. Res. Publ.*, vol. 14, no. 11, pp. 8–10, Nov. 2024, doi: 10.29322/IJSRP.14.11.2024.p15503.

[2] V. P. Rambabu, J. Sreerama, and J. T. S. Singh, "AI-Driven Data Integration: Enhancing Risk Assessment in the Insurance Industry," *Aust. J. Mach. Learn. Res. Appl.*, vol. 2, no. 2, 2022.

[3] M. Beyene and K. R. Shekar, "Performance Analysis of Homomorphic Cryptosystem on Data Security in Cloud Computing," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India: IEEE, Jul. 2019, pp. 1–7. doi: 10.1109/ICCCNT45670.2019.8944837.

[4] D. Ajiga, R. Adeleye, O. F. Asuzu, and O. R. Owolabi, "Review Of Ai Techniques In Financial Forecasting: Applications In Stock Market Analysis," *Financ. Account. Res. J.*, vol. 6, no. 2, pp. 125–145, Feb. 2024, doi: 10.51594/farj.v6i2.784.

[5] A. E. El-Shora, M. M. Abbassy, W. M.Ead, A. I. Elfeky, and A.

DOI: https://doi.org/xxxxxxxxxxxxx

A. Aboalndr, “Hybrid Machine Learning and Data Analytics Framework to Improve Prediction of the Stock Market,” *J. Posthumanism*, vol. 4, no. 3, Dec. 2024, doi: 10.63332/joph.v4i3.989.

[6] N. Rouf *et al.*, “Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions,” *Electronics*, vol. 10, no. 21, p. 2717, Nov. 2021, doi: 10.3390/electronics10212717.

[7] Y. Ansari, “Multi-Cluster Graph (MCG): A Novel Clustering-Based Multi-Relation Graph Neural Networks for Stock Price Forecasting,” *IEEE Access*, vol. 12, pp. 154482–154502, 2024, doi: 10.1109/ACCESS.2024.3476159.

[8] L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan, “Machine learning approaches in stock market prediction: A systematic literature review,” *Procedia Comput. Sci.*, vol. 216, pp. 96–102, 2023, doi: 10.1016/j.procs.2022.12.115.

[9] W. M. Shaban, E. Ashraf, and A. E. Slama, “SMP-DL: a novel stock market prediction approach based on deep learning for effective trend forecasting,” *Neural Comput. Appl.*, vol. 36, no. 4, pp. 1849–1873, Feb. 2024, doi: 10.1007/s00521-023-09179-4.

[10] K. Ramkumar, P. Preethi, A. Nerella, S. Kilaru, G. G. Battu, and K. A, “A Temporal Graph Neural Network Approach for Deep Fraud Detection in Real-Time Financial Transactions,” in *The 16th International IEEE Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2025.

[11] P. Whig, P. Sharma, A. B. Bhatia, R. R. Nadikattu, and B. Bhatia, “Machine Learning and its Role in Stock Market Prediction,” in *Deep Learning Tools for Predicting Stock Market Movements*, Wiley, 2024, pp. 271–297. doi: 10.1002/9781394214334.ch12.

[12] D. K. Sohi and H. Singh, “Fostering through Digital Platforms Financial Literacy: Systematic Review of Literature,” *J. Appl. Manag.*, vol. 14, no. 1, pp. 98–112, 2022.

[13] M. Louisa, G. Darmawan, and B. Tantular, “Enhancing Stock Price Forecasting with CNN-BiGRU-Attention: A Case Study on INDY,” *Mathematics*, vol. 13, no. 13, p. 2148, Jun. 2025, doi: 10.3390/math13132148.

[14] F. K. Mirza, Ö. Pekcan, M. Hekimoğlu, and T. Baykaş, “Stock price forecasting through symbolic dynamics and state transition graphs with a convolutional recurrent neural network architecture,” *Neural Comput. Appl.*, vol. 37, no. 20, pp. 15855–15890, Jul. 2025, doi: 10.1007/s00521-025-11325-z.

[15] M. Ez-zaiym, Y. Senhaji, M. Rachid, K. El Moutaouakil, and V. Palade, “Fractional Optimizers for LSTM Networks in Financial Time Series Forecasting,” *Mathematics*, vol. 13, no. 13, p. 2068, Jun. 2025, doi: 10.3390/math13132068.

[16] B. Kantheti and C. A. P. De Araujo, “Confidential Trends: Assessing the Feasibility of Transformers in Private Time Series Forecasting,” in *Proceedings of the 2024 7th Artificial Intelligence and Cloud Computing Conference*, 2024, pp. 120–131. doi: 10.1145/3719384.3719400.

[17] X. Zeng, C. Liang, Q. Yang, F. Wang, and J. Cai, *Enhancing stock index prediction: A hybrid LSTM-PSO model for improved forecasting accuracy*, vol. 20, no. 1. 2025. doi: 10.1371/journal.pone.0310296.

[18] L. Sen Zhang, “Deep Learning-Based Optimization of Cloud Enterprise Resource Planning (ERP) Systems for Adaptive Decision Support and Management Effectiveness Analysis,” *IEEE Access*, vol. 12, no. December, pp. 193402–193415, 2024, doi: 10.1109/ACCESS.2024.3514879.

[19] A. Y. Fathi, I. A. El-Khodary, and M. Saafan, “A Hybrid Model Integrating Singular Spectrum Analysis and Backpropagation Neural Network for Stock Price Forecasting,” *Rev. d'Intelligence Artif.*, vol. 35, no. 6, pp. 483–488, Dec. 2021, doi: 10.18280/ria.350606.

[20] J. Xiao, X. Zhu, C. Huang, X. Yang, F. Wen, and M. Zhong, “A New Approach for Stock Price Analysis and Prediction Based on SSA and SVM,” *Int. J. Inf. Technol. Decis. Mak.*, vol. 18, no. 01, pp. 287–310, Jan. 2019, doi: 10.1142/S021962201841002X.

[21] H. J. Singh, K. Ghosh, and S. Sharma, “A Novel Multivariate Recurrent Neural Network based Analysis Model for Predicting Stock Prices,” in *2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT)*, IEEE, Dec. 2022, pp. 1–6. doi: 10.1109/CISCT55310.2022.10046540.

[22] N. A. Kouakan Adanin and D. M. Balungu, “Predicting Agricultural Commodity Stock Prices: A Comparison of Statistical Methods and Machine Learning Algorithms,” in *2025 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*, IEEE, May 2025, pp. 304–307. doi: 10.1109/USBEREIT65494.2025.11054062.

[23] P. Anand and S. Yadav, “A Time-Series Forecasting Framework for Stock Market Prediction Using LSTM and Technical Indicators,” in *2025 International Conference on Intelligent and Secure Engineering Solutions (CISES)*, IEEE, Aug. 2025, pp. 773–776. doi: 10.1109/CISES66934.2025.11265263.

[24] Y. K. Jakhar, P. Sharma, and B. Ahmed, “Stock Price Prediction by Using Machine Learning Techniques: A Study of TCS Ltd,” in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, IEEE, Jul. 2024, pp. 1256–1260. doi: 10.1109/ICSCSS60660.2024.10624829.

[25] F. Kihtir and K. Oztoprak, “A Comparative Analysis on Leveraging K-Nearest Neighbours for Accurate Stock Price Forecasting,” in *2024 IEEE International Conference on Big Data (BigData)*, IEEE, Dec. 2024, pp. 7061–7067. doi: 10.1109/BigData62323.2024.10825529.

[26] M. Lavanya and P. Gnanasskaran, “Stock Exchange Price Prediction Using Linear Regression Model,” in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, IEEE, Aug. 2023, pp. 1054–1059. doi: 10.1109/ICIRCA57980.2023.10220627.

[27] T. Baihaqi, M. A. Sugiyarto, R. P. Daksa, F. I. Kurniadi, M. Fakhruddin, and H. Erandi, “Unveling the Precision of Deep Learning Models for Stock Price Prediction: A Comparative Analysis of Bi-LSTM, LSTM, and GRU,” in *2023 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE)*, IEEE, Oct. 2023, pp. 61–64. doi: 10.1109/ICCTEIE60099.2023.10366646.

[28] P. V. Reddy and S. Magesh Kumar, “A Novel Approach to Improve Accuracy in Stock Price Prediction using Gradient Boosting Machines Algorithm compared with Naive Bayes Algorithm,” in *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, IEEE, Dec. 2022, pp. 695–699. doi: 10.1109/ICAC3N56670.2022.10074387.

[29] S. Mokhtari, K. K. Yen, and J. Liu, “Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning,” *Int. J. Comput. Appl.*, vol. 183, no. 7, pp. 1–8, 2021, doi: 10.5120/ijca2021921347.

[30] S. Agarwal and S. W. A. Rizvi, “Enhancing Stock Market Forecasting Using Transformer-Based Models,” *Int. J. Comput. Appl.*, vol. 187, no. 6, pp. 20–25, 2025, doi: 10.5120/ijca2025924892.

[31] S. Nowjiya and M. K. A, “A Hybrid Deep Learning Framework for Stock Market Prediction Using PPO-Based Sentiment Analysis and Transductive LSTM,” *Int. J. Sci. Res. Sci. Technol.*, vol. 13, no. 2, pp. 169–180, Mar. 2026, doi: 10.32628/IJSRST2613195.

[32] J. Shen and M. O. Shafiq, “Short-term stock market price trend prediction using a comprehensive deep learning system,” *J. Big Data*, vol. 7, no. 1, p. 66, Dec. 2020, doi: 10.1186/s40537-020-00333-6.