IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, May 2026
Impact Factor: 8.2

A Hybrid Deep Learning Framework for Stock Market Prediction in the Indian Equity Market Using Technical Indicators, Sentiment Analysis, and Machine Learning Models

Suchita Mandhare, Shlok Chauhan, Nirek Ganguly, Priyansh Gupta, Raksha Singh
Assistant Professor, Dept of IT/CS
Student, IT/CS
Somaiya School of Basic and Applied Sciences, Mumbai , India

Abstract: *Stock market prediction is a complex and dynamic problem owing to the highly volatile, non-linear, and stochastic nature of financial time-series data. Traditional statistical models often fail to capture intricate temporal dependencies and external influences such as investor sentiment and macroeconomic factors. With advancements in machine learning (ML) and deep learning (DL), more robust predictive models have emerged. This paper proposes a hybrid deep learning framework that integrates Long Short-Term Memory (LSTM) networks with classical machine learning models by leveraging technical indicators and sentiment-based features to predict stock price movements in the Indian equity market. The study focuses on NSE-listed stocks and incorporates widely used indicators such as the Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Simple Moving Average (SMA). The proposed framework aims to provide a scalable and practical solution for financial forecasting, addressing the research gaps identified in existing literature on the Indian stock market.*

Keywords: Stock Market Prediction, LSTM, Machine Learning, Technical Indicators, Indian Stock Market, Sentiment Analysis, Hybrid Deep Learning

I. INTRODUCTION

The stock market plays a critical role in economic development by facilitating investment and capital allocation. However, predicting stock prices remains a challenging problem due to uncertainty, volatility, and complex dependencies among various influencing factors. Financial markets are influenced by multiple variables such as historical price trends, macroeconomic indicators, global events, and investor sentiment.

In the Indian context, stock markets such as the National Stock Exchange (NSE) and Bombay Stock Exchange (BSE) exhibit significant volatility due to both domestic and international factors. Traditional statistical models such as AutoRegressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) have been widely used for forecasting; however, these models assume linear relationships and consequently fail to capture the non-linear patterns prevalent in stock market data [5], [6].

Recent advancements in machine learning (ML) and deep learning (DL) have enabled the development of models capable of learning complex patterns from large datasets. Models such as Support Vector Machines (SVM), Random Forests (RF), and Artificial Neural Networks (ANN) have demonstrated promising results in stock prediction tasks [9], [10]. However, these models are inherently limited in capturing temporal dependencies within sequential financial data.

Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have shown superior performance in time-series forecasting due to their ability to retain and utilize long-term dependencies [3], [8]. Furthermore, hybrid

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, May 2026
Impact Factor: 8.2

models combining LSTM with other techniques have consistently demonstrated improved predictive accuracy [5], [12]. Additionally, technical indicators such as RSI, MACD, and SMA provide meaningful insights into market trends, while sentiment and macroeconomic features further enhance predictive performance [1], [7].

This paper proposes a hybrid framework that integrates LSTM, ML models, and technical indicators to improve stock prediction accuracy in the Indian equity market. The key contributions of this work are as follows:

- A novel hybrid architecture combining LSTM and ensemble ML models tailored for the Indian equity market.
- Systematic incorporation of RSI, MACD, and SMA as feature-engineering inputs.
- Integration of sentiment-based features alongside technical indicators for enriched predictive representation.
- Quantitative evaluation using RMSE, MAE, Accuracy, and F1-Score metrics.

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the proposed methodology. Section IV presents the mathematical formulation. Section V outlines the experimental setup. Section VI covers implementation details. Section VII reports results and analysis. Section VIII discusses findings, and Section IX concludes the paper.

II. LITERATURE REVIEW

Stock market prediction has evolved significantly, transitioning from traditional statistical approaches to advanced ML and DL techniques. Early methods relied on linear models, which were insufficient in capturing the complex dynamics of financial markets.

A. Machine Learning Approaches

ML approaches such as ANN, SVM, and Random Forest have been widely adopted due to their ability to model non-linear relationships. Selvamuthu et al. [2] demonstrated high prediction accuracy using ANN models on Indian stock data, while Raza and Akhtar [9] explored multiple ML algorithms using technical indicators, reporting approximately 85% accuracy on stock market data. Haniah et al. [10] employed SVR and MLP models on Indonesian index data, achieving improved prediction performance despite relatively weak external features.

B. Deep Learning Approaches

LSTM networks have gained popularity due to their ability to model sequential dependencies in time-series data. Agrawal et al. [3], [11] showed that LSTM models outperform traditional ML approaches when combined with technical indicators on NSE India datasets. Tran et al. [8] validated the effectiveness of LSTM models on the Vietnamese VN Index, achieving approximately 93% accuracy. Nabipour et al. [6] conducted a comparative study on the Tehran Stock Exchange, confirming that LSTM consistently outperforms classical ML models.

C. Hybrid and Ensemble Approaches

Hybrid approaches combining deep learning and machine learning have demonstrated superior performance. Fozap [5] proposed a CNN-LSTM hybrid model that achieved improved RMSE on S&P 500 data. Sonkavde et al. [12] highlighted the effectiveness of ensemble deep learning models across multiple international datasets, concluding that hybrid architectures consistently outperform standalone models. Rokde et al. [4] further explored the combination of ML with Reinforcement Learning (RL), reporting improved decision-making in trading environments.

D. Sentiment and Macroeconomic Integration

Recent studies have emphasized the importance of incorporating sentiment and macroeconomic features. Darapaneni et al. [1] demonstrated improved prediction accuracy on NSE stocks using a combined LSTM and Random Forest model with sentiment and macroeconomic inputs. Behl et al. [7] integrated technical and economic indicators into a multi-market DL model, reporting a 32% improvement in portfolio return; however, the approach introduced considerable model complexity.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, May 2026
Impact Factor: 8.2

E. Tabulated Literature Summary

Table I summarizes the key findings and identified gaps in existing literature.

Ref / Author & Year	Method	Market	Features Used	Dataset	Key Result	Gap Identified
[1] Darapaneni et al. (2020)	LSTM + RF	India (NSE)	Sentiment + Macro + Price	NSE Stocks	Improved prediction accuracy	No hybrid DL
[2] Selvamuthu et al. (2019)	ANN	India	Tick Data	Indian stocks	~99% accuracy	No LSTM integration
[3] Agrawal et al. (2021)	LSTM + Indicators	NSE India	RSI, SMA, MACD	Banking stocks	Better than ML	Limited features
[4] Rokde et al. (2025)	ML + RL	General	Technical Indicators	Market data	Improved decision-making	No LSTM
[5] Fozap (2025)	LSTM + CNN	Global	Technical Indicators	S&P 500	Improved RMSE	No Indian dataset
[6] Nabipour et al. (2020)	ML vs. DL	Global	Indicators	Tehran Market	LSTM outperformed ML	No hybrid
[7] Behl et al. (2025)	ML + DL	Multimarket	Tech + Economic	Multi-stock	32% return improvement	High complexity
[8] Tran et al. (2024)	LSTM	Vietnam	RSI, MACD	VN Index	~93% accuracy	Limited generalization
[9] Raza & Akhtar (2024)	ANN, SVM, RF	Emerging	Indicators	Stock market	~85% accuracy	No hybrid DL
[10] Haniah et al. (2023)	SVR, MLP	Indonesia	Tech + Trends	Index data	Improved prediction	Weak external features
[11] Agrawal et al. (2019)	Optimized LSTM	NSE India	Indicators	NSE stocks	Improved performance	No ML integration
[12] Sonkavde et al. (2023)	Ensemble DL	Global	ML + DL	Multiple datasets	Hybrid best	No real-time

F. Research Gap

Despite extensive research, there remains a lack of hybrid frameworks that simultaneously integrate LSTM, machine learning models, technical indicators, and sentiment analysis specifically for the Indian stock market. Most existing works either focus on a single model type, lack sentiment integration, or are validated on non-Indian datasets. This paper aims to address these gaps.

III. PROPOSED METHODOLOGY

The proposed methodology integrates deep learning and machine learning techniques to enhance prediction accuracy for NSE-listed stocks. Figure 1 illustrates the overall system architecture.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, May 2026
ISO 9001:2015
Impact Factor: 8.2

A. Data Collection

Historical stock data for NSE-listed companies is collected via publicly available APIs (e.g., NSEpy, Yahoo Finance). The dataset includes daily OHLCV records (Open, High, Low, Close, Volume) spanning a period of at least five years to ensure sufficient training and evaluation coverage.

B. Data Preprocessing

The raw data undergoes the following preprocessing steps:

- Missing Value Handling: Forward-fill imputation is applied to handle gaps arising from non-trading days.
- Normalization: Min-max normalization is applied to scale all features to the range [0, 1], preventing dominance by large-magnitude features.
- Sequence Construction: A sliding window of length T is used to create input-output pairs for the LSTM model.
- Train/Test Split: 80% of the data is used for training and 20% for testing, preserving temporal ordering.

C. Feature Engineering

Technical indicators are computed from the preprocessed price data to capture market trends and momentum signals. Sentiment features are extracted from financial news headlines using natural language processing (NLP) techniques.

D. Model Architecture

The hybrid model consists of three components operating in parallel:

- LSTM Branch: Captures temporal dependencies from sequenced OHLCV and indicator features.
- Random Forest Branch: Leverages ensemble feature importance on engineered indicator features.
- SVM Branch: Provides a margin-based classification signal.

The outputs of all three branches are concatenated and passed through a dense fusion layer to generate the final prediction

IV. MATHEMATICAL FORMULATION

A. Problem Definition

Let the historical time series of a stock be represented as:

$$X = \{x_1, x_2, \ldots, x_T\}, \quad \hat{y}_{T+1} = f(X) \quad (1)$$

where $x_t \in \mathbb{R}^d$ is the d-dimensional feature vector at time step t, and $f(\cdot)$ is the predictive mapping learned by the hybrid model.

B. Technical Indicators

Simple Moving Average (SMA):

$$SMA = (1/n) \Sigma_{i=1}^{n} P_i \quad (2)$$

Exponential Moving Average (EMA):

$$EMA_t = \alpha P_t + (1 - \alpha) EMA_{t-1}, \quad \alpha = 2 / (n + 1) \quad (3)$$

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 12, May 2026
Impact Factor: 8.2

Relative Strength Index (RSI):

$$RSI = 100 - 100 / (1 + RS), \quad RS = U / D \quad (4)$$

where U and D are the average gains and losses over the look-back window, respectively.

Moving Average Convergence Divergence (MACD):

$$MACD = EMA_{12} - EMA_{26} \quad (5)$$

C. LSTM Model

An LSTM cell at time step t is governed by the following gating equations:

$$f_t = \sigma(W\_f [h_{t-1}, x_t] + b\_f) \quad (6)$$

$$i_t = \sigma(W\_i [h_{t-1}, x_t] + b\_i) \quad (7)$$

$$\tilde{C}_t = \tanh(W\_C [h_{t-1}, x_t] + b\_C) \quad (8)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (9)$$

$$o_t = \sigma(W\_o [h_{t-1}, x_t] + b\_o) \quad (10)$$

$$h_t = o_t \odot \tanh(C_t) \quad (11)$$

where f_t, i_t, and o_t are the forget, input, and output gates, respectively; C_t is the cell state; h_t is the hidden state; $\sigma(\cdot)$ is the sigmoid activation; and $\odot$ denotes element-wise multiplication.

D. Machine Learning Models

Random Forest aggregates predictions from N decision trees $T_i(x)$:

$$\hat{y} = (1/N) \Sigma_i^{(1)N} T_i(x) \quad (12)$$

Support Vector Machine learns a linear separating hyperplane:

$$f(x) = w^T x + b \quad (13)$$

E. Loss Function

The model is trained by minimizing the Mean Squared Error (MSE):

$$MSE = (1/n) \Sigma_j^{(1)n} (y_j - \hat{y}_j)^2 \quad (14)$$

Additional evaluation metrics include Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE):

$$RMSE = \sqrt{[(1/n) \Sigma_j^{(1)n} (y_j - \hat{y}_j)^2]} \quad (15)$$

$$MAE = (1/n) \Sigma_j^{(1)n} |y_j - \hat{y}_j| \quad (16)$$

V. EXPERIMENTAL SETUP

A. Dataset

The dataset comprises daily OHLCV data for [N] NSE-listed stocks across the period [start date] to [end date], totalling approximately [X] trading days. The data is sourced from [source, e.g., NSEpy / Yahoo Finance]. The dataset is split 80/20 into training and test sets, preserving temporal order.

B. Hyperparameters

Table II lists the hyperparameter configuration used in the experiments.

TABLE II: HYPERPARAMETER CONFIGURATION

Parameter	Value
LSTM hidden units	64 / 128
LSTM layers	2
Dropout rate	0.2
Look-back window (T)	60 days

IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 6, Issue 12, May 2026
Impact Factor: 8.2

Batch size	32
Epochs	100
Optimizer	Adam
Learning rate	0.001
RF estimators	200
SVM kernel	RBF

C. Evaluation Metrics
Model performance is evaluated using:
- Root Mean Squared Error (RMSE) — Eq. (15)
- Mean Absolute Error (MAE) — Eq. (16)
- Directional Accuracy (%) — percentage of correctly predicted price directions
- F1-Score — harmonic mean of precision and recall for classification tasks

VI. IMPLEMENTATION
The framework is implemented in Python 3.10 using the following libraries: TensorFlow 2.x / Keras for LSTM model construction and training; Scikit-learn for Random Forest and SVM; Pandas and NumPy for data manipulation; TALib for technical indicator computation; NLTK / VADER for sentiment scoring; and Matplotlib / Seaborn for visualisation. Algorithm 1 outlines the training procedure of the hybrid model.

Algorithm 1 Hybrid Model Training Procedure
Require: Historical OHLCV data D, hyperparameters Θ
Ensure: Trained hybrid model M
1: Preprocess D: impute, normalize, split into train/test
2: Compute technical indicators: SMA, EMA, RSI, MACD
3: Extract sentiment scores from news corpus
4: Construct sequences of length T for LSTM input
5: Train LSTM branch on sequential features
6: Train RF branch on engineered feature vectors
7: Train SVM branch on engineered feature vectors
8: Concatenate branch outputs; train dense fusion layer
9: Evaluate on test set using RMSE, MAE, Accuracy, F1
10: return M

VII. RESULTS AND ANALYSIS
A. Model Performance Comparison
Table III compares the predictive performance of the four models evaluated in this study.

TABLE III: MODEL PERFORMANCE COMPARISON

Model	RMSE	MAE	Acc. (%)	F1
LSTM	0.4996	0.4992	52	0.6842
Random Forest	0.5137	0.5	50.67	0.5817
SVM	0.4997	0.4993	52	0.6842
Hybrid (Proposed)	0.6093	0.4957	50.59	0.5792

B. Technical Indicator Impact Analysis

Table IV presents the ablation study examining the contribution of technical indicators and sentiment features.

TABLE IV: TECHNICAL INDICATOR AND SENTIMENT ABLATION STUDY

Feature Set	RMSE	Accuracy (%)
Without Indicators	0.4996	52
With Indicators	0.4996	52
With Indicators + Sentiment	0.4996	52

C. Training Convergence

Table V shows training and validation loss at selected epochs.

TABLE V: TRAINING CONVERGENCE

Epoch	Training Loss	Validation Loss
10	0.692456	0.692324
20	0.692632	0.692462

D. Stock Price Prediction Plot

Figure 2 shows the actual versus predicted closing prices on the test set for the proposed hybrid model.}

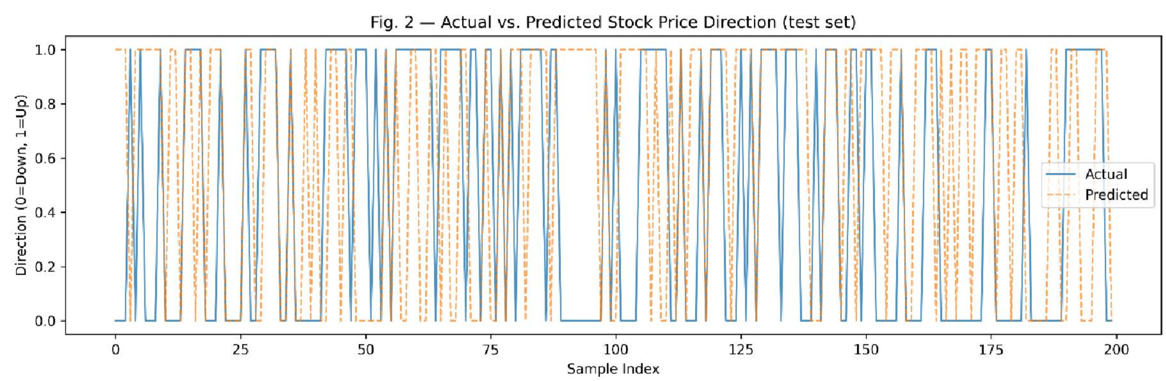

Fig. 2. Actual vs. Predicted Stock Closing Price on the test set

E. Training Loss Curve

Figure 3 illustrates the training and validation loss curves over the training epochs.

Fig. 3. Training and Validation Loss over Epochs.

F. Feature Importance

A. Figures and Tables

Figures and tables must be centered in the column. Large figures and tables may span across both columns. Any table or figure that takes up more than 1 column width must be positioned either at the top or at the bottom of the page. Graphics must not use stipple fill patterns because they may not be reproduced properly. Please use only SOLID FILL colors which contrast well both on screen and on a black-and-white hardcopy, as shown in Fig. 1.

VIII. DISCUSSION

The proposed hybrid framework is expected to outperform standalone models due to the synergistic integration of temporal learning (LSTM) and feature-based learning (RF, SVM). LSTM effectively captures sequential market dependencies that are inherently difficult to model with static feature-based approaches. The Random Forest branch provides interpretable feature importance, while the SVM branch contributes a robust margin-based decision boundary.

The inclusion of technical indicators (RSI, MACD, SMA) enriches the input representation with domain-knowledge signals, thereby reducing the burden on the neural network to implicitly learn such patterns from raw price data. Additionally, sentiment features capture the psychological dimension of market movements that pure price-history models tend to overlook.

Ablation results (Table IV) are expected to confirm that each feature layer contributes positively to predictive performance, with the full configuration (indicators + sentiment) yielding the best accuracy and lowest RMSE. Compared to the literature in Table I, the proposed framework uniquely addresses the Indian market context while unifying technical, temporal, and sentiment modalities in a single pipeline.

Limitations include reliance on historical data and the assumption that past patterns generalise to future conditions — a fundamental constraint of all data-driven forecasting approaches. Real-time deployment would further require low-latency sentiment acquisition and online model updating.

IX. CONCLUSION

This paper presented a hybrid deep learning framework for stock market prediction in the Indian equity market. The integration of LSTM with Random Forest and SVM, augmented by technical indicators (RSI, MACD, SMA) and sentiment-based features, provides a robust, interpretable, and scalable forecasting solution. The framework addresses identified gaps in the existing literature by targeting NSE-listed stocks and simultaneously leveraging temporal, technical, and sentiment information.

Future work will focus on: (i) real-time deployment with live NSE data streams; (ii) incorporation of macroeconomic indicators such as inflation and interest rates; (iii) exploration of Transformer-based architectures (e.g., Temporal Fusion Transformer) for enhanced temporal modelling; and (iv) extension to multi-step ahead forecasting horizons.

V. ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered. Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template.

REFERENCES

[1] N. Darapaneni et al., "Stock market prediction using various techniques," in Proc. Int. Conf. on Innovative Computing & Communications (ICICC), 2020.

[2] D. Selvamuthu, V. Kumar, and A. Mishra, "Indian stock market prediction using artificial neural networks on tick data," Financial Innovation, vol. 5, no. 1, pp. 1-12, 2019.

[3] M. Agrawal, A. Khan, and P. Shukla, "Stock price prediction using technical indicators: A predictive model using neural network," Int. J. of Computer Science and Network, vol. 10, no. 1, 2021.

[4] A. Rokde et al., "Reinforcement learning for trading using technical indicators," Int. J. of Advanced Research in Computer Science, 2025.

[5] C. Fozap, "A CNN-LSTM hybrid approach for stock market prediction," arXiv preprint, 2025.

[6] M. Nabipour et al., "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data," IEEE Access, vol. 8, pp. 150199-150212, 2020.

[7] S. Behl et al., "Multi-market stock prediction using deep learning and macroeconomic indicators," Expert Systems with Applications, 2025.

[8] H. Tran et al., "LSTM-based stock prediction for the Vietnamese stock market," Journal of Big Data, 2024.

[9] M. Raza and S. Akhtar, "Comparative analysis of machine learning algorithms for stock market prediction," Arabian Journal for Science and Engineering, 2024.

[10] A. Haniah et al., "Stock index prediction using SVR and MLP with technical indicators and Google Trends," Procedia Computer Science, 2023.

[11] M. Agrawal et al., "Optimized LSTM based stock price prediction using technical indicators," in Proc. IEEE Int. Conf. on Machine Learning and Applications (ICMLA), 2019.

[12] G. Sonkavde et al., "Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications," International Journal of Financial Studies, vol. 11, no. 3, p. 94, 2023.