International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(6s) 2026 pp. 1205-1225
https://doi.org/10.70917/ijcisim-2026-3027
Article

A Hybrid Deep Learning Framework for Stock Volatility Prediction Using Technical, Fundamental and Financial News Sentiment Features

Jyoti[1], Sushil Kumar Bansal[2]

[1]Department of Computer Science and Engineering, Maharaja Agrasen University, Baddi, Himachal Pradesh, India; Government College for Girls, Manesar, Haryana, India.
[2]Department of Computer Science and Engineering, Maharaja Agrasen University, Baddi, Himachal Pradesh, India.

Abstract: Forecasting stock volatilities is a tough problem because of the nonlinear and dynamic character of the financial markets. If you want to make good decisions about investments, about how to organize a portfolio and about risk, you need good forecasts. This paper presents a hybrid deep learning approach for predicting stock volatility by merging technical indicators, underlying financial factors, and financial news emotion. Historical stock market data from 29 May 2013 to 28 March 2024 were acquired from Yahoo Finance. Daily sentiment ratings were extracted from financial news items using natural language processing algorithms. After data pre-processing, feature engineering and standardization, a hybrid dataset was built including both quantitative and qualitative financial data. We used Random Forest (RF) and XGBoost as baseline machine learning models and created a LSTM-based hybrid architecture to capture the temporal dependencies and nonlinear interactions in financial time-series data. To avoid data leaks and get realistic forecasts, a chronological train-validation-test technique was used. Several regression metrics such as MAE, MSE, RMSE, MAPE and R2 were used to evaluate the proposed framework. Experimental results indicate that the integration of financial news sentiment with technical and fundamental features can improve the prediction accuracy and provide a more comprehensive understanding of market dynamics compared to traditional approaches. The suggested methodology provides a realistic and comprehensive solution for intelligent stock volatility forecasting, and may enable investors, portfolio managers, and financial analysts to make educated financial choices.
Keywords: Stock Volatility Prediction, LSTM, Random Forest, XGBoost, Financial News Sentiment, Feature Engineering, Time-Series Forecasting, Machine Learning, Financial Markets.

1. Introduction

Financial markets are turbulent by nature, and stock prices are affected by economic circumstances, business performance, political events, and investor emotion. Accurate forecast of market volatility is crucial for investors, portfolio managers, and financial institutions to make sound investment choices, manage risk, and optimize portfolio allocation. However, it is difficult to anticipate the stock market volatility because of the nonlinear, dynamic and non-stationary properties of the financial time-series data. Traditional statistical models such as ARIMA and GARCH have been widely utilized for volatility forecasting, but they typically fail to capture complicated nonlinear linkages and interactions among many financial factors. Recently, machine learning and deep learning techniques have been considerably advanced to increase the prediction accuracy by learning complex patterns from large-scale financial data. However, most of the current research focus on the historical price data and technical indicators, ignoring the underlying financial information and investor mood from financial news. To tackle this constraint, this research offers a hybrid deep learning architecture that blends technical indicators, fundamental financial characteristics and financial news sentiment for stock volatility prediction. Historical market data from Yahoo Finance and sentiment ratings from financial news are integrated into one dataset. We present a system for forecasting based on RF and XGBoost as baseline models and an LSTM-based hybrid architecture. To provide dependable and realistic prediction performance, we use a chronological train-validation-test technique to assess model performance.

The stock market is one of the most powerful parts of the global financial system, allowing for capital allocation, company financing, and wealth creation. The fast growth of computerized trading platforms, algorithmic trading and high-frequency transactions in recent years has greatly increased market complexity, making prediction of stock volatility more complicated than ever. Volatility is a measure of the variation of prices over time and is seen as a key indication of market uncertainty and financial risk. Accurate volatility prediction helps in portfolio optimization, derivative pricing, risk management, and strategic investment decisions. Conventional methods of forecasting depend on past price movements and statistical models to predict future market behavior. However, stock markets are influenced not just by historical patterns but also by business financial results, macro-economic circumstances, geopolitical events and investor psychology. So, one type of information is typically not very predictive. The growing availability of structured financial data and unstructured textual data has opened new prospects for designing intelligent forecasting systems that combine several sources of information. Recent breakthroughs in machine learning and deep learning have empowered academics to blend diverse financial data to improve the accuracy of predictions.

The stock market is an essential component of the world financial system, facilitating capital generation, investment and economic progress. Stock volatility is a major measure of risk and uncertainty in the market as it tracks the changes in asset values. Accurate forecast of the volatility of the assets is important for portfolio management, risk assessment and investment decision making. However, it is still a challenge to predict the volatility of stocks because of the nonlinear and dynamic characteristics of financial markets. Traditional models mostly depend on previous price data and frequently ignore larger market factors. Recent breakthroughs in machine learning and deep learning have made it possible to integrate technical indications, fundamental financial variables and financial news mood. Against this backdrop, this research introduces a hybrid deep learning architecture which incorporates these diverse data sources for market volatility prediction and forecasting accuracy enhancement.

This paper offers numerous important additions to the stock volatility prediction area by offering a complete hybrid deep learning approach that incorporates technical indicators, basic financial characteristics and financial news sentiment in a unified predictive architecture. Compared with conventional forecasting models, which mainly depend on historical market prices, the proposed framework combines heterogeneous data sources, such as historical stock market data from Yahoo Finance and financial news information, to incorporate both quantitative market dynamics and qualitative investor sentiment. In this study, the data is properly pre-processed and standardized to enhance the performance and reliability of the model, the technical, fundamental, and sentiment-based features are systematically extracted and optimized using a feature engineering approach. To examine the success of the suggested strategy, we use Random Forest and XGBoost as benchmark machine learning models, and build an LSTM-based hybrid model to capture the complex nonlinear interactions and long-term temporal dependencies inherent in financial time-series data. Also, a train-validation-test chronological approach is used to imitate the real-world forecasting scenarios and limit the chances of data leaks in building the model. The predictive performance of the proposed framework is thoroughly evaluated by multiple statistical evaluation metrics and systematically compared with conventional machine learning models, to demonstrate the benefits of integrating technical, fundamental and sentiment information for stock volatility prediction.

2. Literature Review

Saravanos (2025) [1] suggested a deep learning method for anticipating stock market volatility, using historical market data and social media sentiment. BL (2023) proposed a hybrid deep learning architecture that integrated historical stock price data with financial news sentiment for stock market prediction [2]. In [3], Li (2026) proposed a hybrid deep learning approach for stock price prediction that combines the investor sentiment from online discussion forums with popularity-based information. Chatziloizos et al. (2024) introduced a deep learning model that integrates technical indicators with financial sentiment analysis for stock market prediction [4]. Gupta (2025) has presented an exhaustive survey of stock market prediction approaches based on news sentiment analysis and machine learning techniques [5]. Phong et al. (2026) studied the sentiment-driven forecasting of Vietnamese stock market using the machine learning methods and financial news sentiment [6]. Tajmazinani et al. (2022) introduced a deep learning-based system to anticipate the stock price changes based on the financial news sentiment analysis [7]. Li (2020) provided a thorough assessment of deep learning techniques in stock market prediction with a special focus on technical analysis [8]. Al Ridhawi et al. (2026) [9] introduced a hybrid stock market prediction system integrating a Node Transformer architecture together with BERT-based sentiment analysis. Carosia et al. (2025) created a deep learning framework for the prediction of the Brazilian stock market using historical stock prices, technical indicators and mood of financial news [10]. Prachyachuwong (2021) introduced a deep learning-based stock trend prediction model that combines technical indicators and industry-specific information [11]. In 2026, Hossain et al. [12] proposed a FinBERT-BiLSTM model to anticipate the extremely fluctuating values of the cryptocurrency market based on the market sentiment dynamics. Das et al. (2024) built hybrid deep learning models for stock price prediction by merging increased Twitter sentiment scores with technical indicators [13].

Li et al. (2021) used a data decomposition based deep learning method [14] to explore the effect of financial news mood on predicting oil futures returns and volatility. Iqbal et al. (2026) introduced an LSTM based machine learning method to forecast stock market returns using sentiment collected from business news articles [15]. Oncharoen (2018) proposed a deep learning system using event embedding methods and technical indicators for stock market prediction [16]. Li (2022) suggested a new ensemble deep learning model for stock market prediction by combining historical stock prices with financial news [17]. Dahal et al. (2023) did comparison research to assess the influence of financial news mood on stock price prediction using several deep learning architectures [18]. Chen (2026) did a detailed analysis of modern stock price prediction approaches, including classical statistical methods, machine learning, deep learning, and hybrid forecasting frameworks [19]. Lee et al. (2024) proposed a deep learning approach which combines technical indicators with ESG sentiment for stock market prediction [20]. Arauco Ballesteros (2025) [21] presented a stock market prediction model based on neural networks, which integrates at the same time fundamental information, technical indicators and market sentiment analysis. Kanikanti et al. [22] examined the impact of daily news emotions on stock price volatility prediction using deep learning methodologies. Patsiarikas et al. (2025) [23] established a machine learning architecture that included macroeconomic data, technical indicators and sentiment variables for stock market predictions. Ehsan et al. (2025) performed a thorough assessment of financial news sentiment analysis employing natural language processing and machine learning approaches for asset price prediction [24].

Table 1. Literature Survey

Ref.	Author(s) & Year	Objective	Methodology	Key Findings	Limitations
[1]	Saravanos and Kanavos (2025)	To forecast stock market volatility by incorporating social media sentiment into predictive models.	Deep learning framework integrating historical stock data with social media sentiment analysis.	Social media sentiment improved volatility forecasting accuracy and captured investor behavior during uncertain market conditions.	Considered only social media sentiment and did not incorporate fundamental financial indicators or macroeconomic variables.
[2]	BL and BR (2023)	To improve stock market prediction through the integration of stock prices and financial news sentiment.	Hybrid deep learning classifiers combining historical prices with news sentiment features.	Hybrid classifier outperformed individual DL, demonstrating effectiveness of combining structured and unstructured financial data.	Limited evaluation on a specific dataset and did not include company fundamental information.
[3]	Li and Hu (2026)	To develop a hybrid deep learning framework incorporating investor sentiment from online forums for stock prediction.	Hybrid DL integrating historical prices, investor sentiment, and popularity-enhanced online forum information.	Popularity-enhanced investor sentiment substantially improved forecasting performance and captured collective investor behavior more effectively.	Focused only on online forum sentiment without considering financial news or company fundamentals.
[4]	Chatziloizos et al. (2024)	To investigate stock market prediction using technical indicators and financial sentiment.	Deep learning model combining technical indicators with sentiment analysis derived	Integrating sentiment with technical indicators improved prediction accuracy compared with	Fundamental financial variables were not incorporated into

			from financial news.	conventional machine learning.	the prediction framework.
[5]	Gupta (2025)	To comprehensively review machine learning and news sentiment-based stock market prediction techniques.	Systematic review of machine learning, deep learning, NLP, and sentiment analysis approaches.	Hybrid DL integrating sentiment information consistently demonstrated superior prediction performance over traditional methods.	The study is review-based and does not propose or experimentally validate a predictive framework.
[6]	Phong et al. (2026)	To forecast the Vietnamese stock index using financial sentiment and machine learning techniques.	Machine learning models incorporating financial news sentiment and historical stock index data.	Sentiment-driven models enhanced stock index forecasting performance and improved responsiveness to rapidly changing market conditions.	The study was limited to the Vietnamese stock market, reducing the generalizability of the findings.
[7]	Tajmazinani et al. (2022)	To predict stock price movements using financial news sentiment and deep learning.	Natural language processing for sentiment extraction integrated with deep learning models.	Financial news sentiment considerably improved prediction accuracy compared with historical price-based forecasting methods.	Relied primarily on news sentiment and historical prices while excluding technical and fundamental indicators.
[8]	Li and Bastos (2020)	To systematically review deep learning methods for stock market forecasting using technical analysis.	Comprehensive systematic review of deep learning architectures and technical indicator-based prediction models.	DL outperformed conventional statistical approaches; integrating technical, fundamental, and sentiment information was recommended	The study is a systematic review and does not propose or experimentally validate a hybrid prediction model.
[9]	Al Ridhawi et al. (2026)	To improve stock market prediction by integrating transformer-based architectures with financial sentiment analysis.	Hybrid framework combining Node Transformer architecture, BERT-based sentiment analysis, and historical stock market data.	Proposed transformer-based model achieved superior prediction accuracy compared with conventional DL by effectively capturing contextual information from financial news.	The framework primarily focused on textual sentiment and historical market data while excluding fundamental financial indicators.
[10]	Carosia et al. (2025)	To predict the Brazilian stock market using stock prices, technical indicators, and financial news sentiment.	Deep learning model integrating historical prices, technical indicators, and sentiment features extracted from financial news.	Hybrid framework significantly outperformed models using historical stock prices, demonstrating benefits of heterogeneous feature integration.	The study was conducted only on the Brazilian stock market, limiting its applicability to other financial markets.
[11]	Prachya-chuwong (2021)	To enhance stock trend prediction by incorporating industrial-specific	Deep learning model utilizing historical stock prices, technical	Including industrial information improved prediction performance by	Financial news sentiment and company fundamental

		information with technical indicators.	indicators, and sector-specific information.	capturing sector-dependent market behavior more effectively.	indicators were not considered in the proposed framework.
[12]	Hossain et al. (2026)	To predict highly volatile cryptocurrency prices using market sentiment and deep learning.	FinBERT-BiLSTM framework integrating transformer-based financial text representations with Bidirectional LSTM networks.	Hybrid model outperformed conventional DL by effectively modeling sentiment dynamics and temporal market behavior, demonstrating its potential for highly volatile financial markets.	The framework was validated on cryptocurrency markets rather than stock markets, requiring further evaluation for stock volatility prediction.
[13]	Das et al. (2024)	To develop hybrid deep learning models for stock price prediction using Twitter sentiment and technical indicators.	Deep learning architecture integrating enhanced Twitter sentiment scores with multiple technical indicators.	Enhanced sentiment scores derived from Twitter significantly improved prediction accuracy compared with models relying solely on technical indicators.	The study relied exclusively on Twitter data and did not incorporate financial statements or macroeconomic variables.
[14]	Li et al. (2021)	To investigate the role of news sentiment in oil futures return and volatility forecasting.	Data decomposition-based deep learning model integrating financial news sentiment with oil futures market data.	News sentiment substantially enhanced both return prediction and volatility forecasting, particularly during periods of increased market uncertainty.	The framework was designed for oil futures markets and was not directly validated on stock market datasets.
[15]	Iqbal et al. (2026)	To predict stock market returns using business news sentiment and LSTM networks.	LSTM-based deep learning model integrating sentiment scores extracted from business news articles with historical stock market data.	The inclusion of business news sentiment improved forecasting accuracy and enabled the LSTM model to better capture long-term market dynamics.	The study emphasized business news sentiment while excluding technical indicators and company financial fundamentals.
[16]	Oncharoen and Vateekul (2018)	To improve stock market prediction using event embedding and technical indicators.	Deep learning framework combining event embedding techniques with conventional technical indicators.	Event embedding successfully captured influence of important financial events, resulting in improved prediction performance over traditional technical analysis models.	Proposed model did not incorporate financial news sentiment, company fundamentals, or advanced transformer-based language models.
[17]	Li and Pan (2022)	To develop an ensemble deep learning model for stock prediction using stock prices and financial news.	Ensemble deep learning framework integrating historical stock prices with financial news sentiment.	Ensemble approach achieved higher prediction accuracy than individual DL, demonstrating combining multiple learning architectures	The study focused mainly on historical prices and news sentiment while excluding company fundamental indicators.

				with news information.	
[18]	Dahal et al. (2023)	To investigate the impact of financial news sentiment on stock price prediction using deep learning models.	Comparative analysis of multiple deep learning architectures with and without financial news sentiment features.	Incorporating news sentiment consistently improved stock price prediction accuracy across different DL, during volatile market conditions.	Fundamental financial variables and macroeconomic indicators were not included in the comparative analysis.
[19]	Chen (2026)	To comprehensively review recent methods and emerging trends in stock price prediction.	Comprehensive review covering statistical models, machine learning, deep learning, transformer models, and hybrid forecasting techniques.	Review identified hybrid DL integrating technical, fundamental, and sentiment information as most promising direction for future stock prediction research.	The study provides conceptual analysis only and does not experimentally validate a forecasting model.
[20]	Lee et al. (2024)	To improve stock market prediction by incorporating ESG sentiment with technical indicators.	Deep learning model integrating Environmental, Social, and Governance (ESG) sentiment with technical market indicators.	ESG sentiment enhanced forecasting accuracy, demonstrating growing importance of sustainability-related information in financial prediction.	The framework emphasized ESG sentiment and technical indicators while excluding detailed company financial fundamentals.
[21]	Arauco (2025)	To forecast stock market behavior using technical indicators, fundamental indicators, and market sentiment.	Neural network-based hybrid framework integrating technical indicators, financial fundamentals, and market sentiment analysis.	Hybrid NN outperformed traditional forecasting models, confirming the effectiveness of combining heterogeneous financial information sources.	The study employed a single neural network architecture without comparing multiple advanced deep learning models such as LSTM or transformers.
[22]	Kanikanti et al. (2025)	To investigate the effect of daily news sentiment on stock price volatility prediction.	Deep learning framework integrating daily financial news sentiment with historical market data for volatility forecasting.	Daily news sentiment significantly improved stock volatility prediction and enabled better identification of high-volatility market periods.	Study primarily focused on news sentiment and historical prices without incorporating company financial statements or technical feature engineering.
[23]	Patsiarikas et al. (2025)	To forecast stock market performance using macroeconomic, technical, and sentiment indicators.	Machine learning models integrating macroeconomic variables, technical indicators, and financial sentiment features.	Combining macroeconomic, technical, and sentiment information improved forecasting performance compared with single-source prediction models.	Deep learning architectures were not extensively explored, limiting the ability to capture long-term temporal dependencies.

[24]	Ehsan et al. (2025)	To systematically review financial news sentiment analysis techniques for asset price prediction.	Systematic review of natural language processing, machine learning, and deep learning approaches for financial sentiment analysis.	Review concluded that financial news sentiment has become an essential component of intelligent financial forecasting systems and recommended hybrid DL frameworks.	As a review study, it does not develop or experimentally evaluate a predictive model.

2.1 Research Gap

According to literature, there has been considerable progress in the forecast of stock market using machine learning, deep learning, technical analysis, and sentiment analysis. It has been proven in several studies that the addition of financial news or social media sentiment to historical market data increases the forecasting performance over traditional statistical methodologies. Deep learning architectures like LSTM, BiLSTM, transformer-based models and ensemble learning approaches have surpassed in predictive accuracy by capturing the nonlinear linkages and temporal dependencies inside financial time-series data efficiently, much like sequence prediction. Despite these improvements, some major research gaps still exist. The majority of existing studies either combine technical indicators with sentiment information or historical prices with news sentiment, and relatively few studies are designed to simultaneously incorporate technical indicators, fundamental financial features and financial news sentiment in a unified hybrid framework. Moreover, several research are confined to a single emotion source, such Twitter, online forums, or financial news, thereby restricting the holistic portrayal of investor behavior. Random partitioning of data is also used to evaluate many prediction models, which might lead to data leakage, and is not a genuine representation of real-world forecasting situations.

3. Problem Statement

Financial markets are complicated, nonlinear and dynamic which makes the accurate stock volatility forecast still a tough endeavor. Traditional statistical models and contemporary machine learning algorithms mostly depend on past stock prices or technical indicators, but they ignore the effect of business fundamentals and real-time market sentiment. Recent researches have integrated financial news or social media sentiment into prediction models, however most of them only use one information source and do not combine technical, fundamental and sentiment elements in a cohesive forecasting framework. Besides, many current models use random data segmentation, which may lead to data leakage and diminish the trustworthiness of the predicting predictions. Therefore, there is a need to design a robust hybrid deep learning framework that can effectively fuse heterogeneous financial information and reliably forecast stock volatility with realistic historical validation and complete performance assessment.

4. Proposed Hybrid Framework

Research offers a hybrid deep learning system for market volatility prediction that integrates technical indicators, basic financial variables and financial news emotion into a single forecasting architecture. Historical stock market data is obtained from Yahoo Finance and financial news is examined to determine daily sentiment ratings. The obtained data is pre-processed, feature extracted, feature engineered and standardized, and then integrated into a hybrid dataset. Research applies Random Forest (RF) and XGBoost as baseline machine learning models and create an LSTM-based hybrid model to capture the temporal dependencies and nonlinear interactions. The methodology is assessed using a chronological train-validation-test approach and numerous regression metrics to determine its prediction performance.

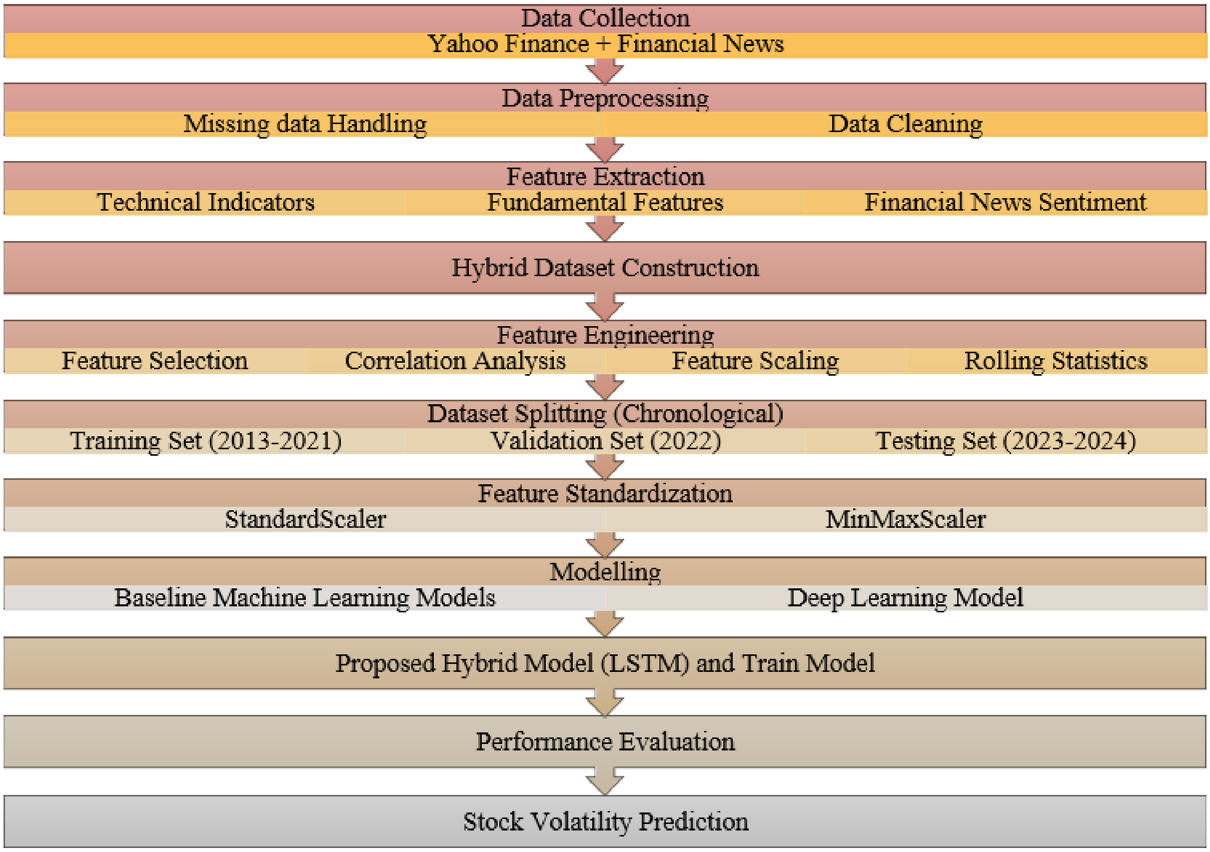

Figure 1 Methodology Flowchart

Figure 1 displays the whole process of the suggested technique. It starts by gathering historical stock market data and financial news from many sources. The data is then preprocessed to fix discrepancies and align the datasets. Technical indicators, basic finance variables and financial news sentiment scores are extracted and combined into one hybrid dataset. Later, feature engineering methods including feature selection, lag generation, rolling statistics and feature scaling are used to increase data quality. The preprocessed dataset is separated into training, validation and testing subsets in chronological order. Baseline methods are trained using Random Forest and XGBoost models, and the suggested hybrid model based on LSTM is created for stock volatility prediction. Finally, the forecasting performance of all models is assessed by using many statistical measures to establish the efficiency of the proposed framework.

Data Collection: The proposed system uses both the structured numerical data and unstructured textual information to capture the stock market activity in a complete manner. Historical market data were obtained from Yahoo Finance, and include daily Open, High, Low, Close, Adjusted Close (OHLCV) and trading volume . These data were used to calculate technical indicators and historical volatility. We collected financial news stories from reliable financial news sources in order to measure investor mood and market responses to business and economic news. The analysis spans from 29 May 2013 to 28 March 2024 and encompasses various market situations, including the COVID-19 epidemic and the subsequent recovery phase. The framework was tested on large-cap businesses that are actively traded on the Indian stock market. The selection of these companies was based on their data availability, market capitalization, liquidity, and continuous historical records.

Data Preprocessing: To assure data quality and consistency of the obtained financial datasets, pre-processing was used before feature extraction. Missing values were handled using forward filling, backward filling and interpolation procedures while incomplete news records were eliminated. Duplicate observations were detected and removed to reduce duplication. For temporal consistency, we aligned historical market data, financial factors and news sentiment by the trading day. In addition, statistical approaches were used to discover outliers and they were handled effectively to minimize their influence on the model performance. The derived cleaned and synchronized dataset was then used as the basis of further feature engineering and model development.

Technical Feature Extraction: These features were extracted from historical OHLCV data from Yahoo Finance. Various technical indicators were calculated to capture the market trends, momentum, volatility and trading activity. They provide relevant quantitative information for the forecast of stock volatility. The features were utilized as input for the Random Forest, XGBoost, LSTM, and the suggested hybrid deep learning model. The technical indicators used in this investigation are shown in Table 2.

Table 2. Technical Features Used in the Proposed Framework

Category	Technical Indicators
Trend	SMA, EMA, MACD
Momentum	RSI, Stochastic Oscillator, ROC
Volatility	Bollinger Bands, ATR
Volume	OBV, Volume Moving Average

Fundamental Feature Extraction: It was used to capture the financial strength and inherent worth of the chosen organizations. The financial indicators were derived from publicly accessible business financial reports and financial information platforms and included profitability, valuation, financial stability, growth and market characteristics. These features provide further information to technical indicators by adding company-specific financial information to help improve the robustness of stock volatility forecast. The basic factors employed in this investigation are presented in Table 3.

Table 3. Fundamental Financial Features

Category	Features
Profitability	EPS, ROE, ROA
Valuation	P/E Ratio, P/B Ratio
Financial Stability	Debt-to-Equity Ratio, Dividend Yield
Growth	Revenue Growth, Earnings Growth
Market Information	Market Capitalization, Shares Outstanding

Financial News Sentiment Analysis: Investor opinion and market responses to business and economic developments were captured by performing financial news sentiment analysis. Financial news stories were acquired from trusted sources and analyzed using NLP methods to turn the unstructured text into quantitative sentiment ratings. The sentiment analysis pipeline comprises news collecting, text preparation, tokenization, sentiment classification, and daily aggregation of sentiment scores. We combine emotion characteristics, technical indicators and basic financial variables to create a holistic view of the reasons driving stock volatility. Figure 2 shows the flow of the complete sentiment analysis process.

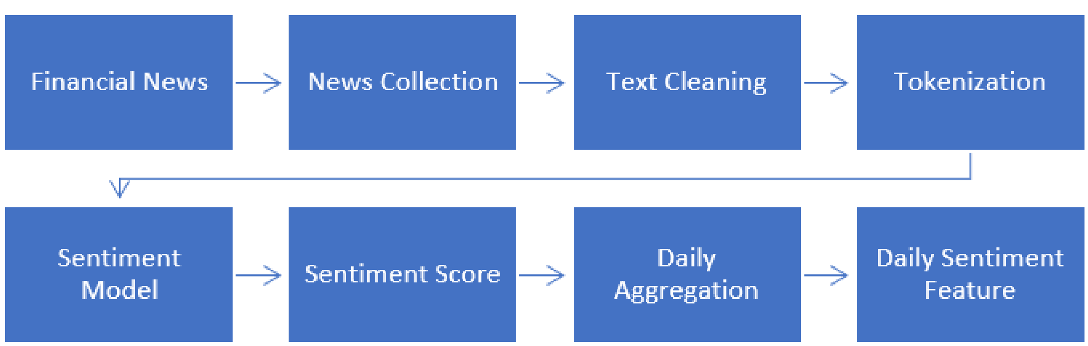

Figure 2. Sentiment Analysis Process

Hybrid Dataset Construction: The proposed methodology integrates technical indicators, basic financial variables and financial news sentiment into a single hybrid dataset for market volatility prediction. To maintain temporal consistency, all datasets were merged using the trade date and the collected features were combined into one feature matrix that includes market, financial, and sentiment information. The resultant hybrid dataset gives a full description of market activity and investor emotion, which is used as input for feature engineering and model training. The whole creation procedure is represented in Figure 3.

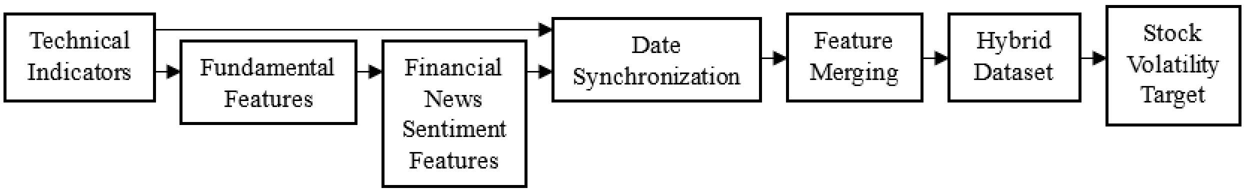

Figure 3. Overall Hybrid Dataset Construction

The structure of the constructed hybrid dataset is presented in Table 4.

Table 4. Structure of the Hybrid Dataset

Date	Technical Features	Fundamental Features	Sentiment Features	Target (Volatility)
t_1	SMA, RSI, MACD, ATR, …	EPS, ROE, P/E, …	Daily Sentiment Score	σ_1
t_2	SMA, RSI, MACD, ATR, …	EPS, ROE, P/E, …	Daily Sentiment Score	σ_2
…	…	…	…	…
t_n	SMA, RSI, MACD, ATR, …	EPS, ROE, P/E, …	Daily Sentiment Score	σ_n

Feature Engineering: It is a critical phase in financial forecasting, where the quality of the predictive variables significantly impacts the model's performance. Feature engineering is the process of transforming raw financial data into meaningful representations that enhance the learning capabilities of machine learning and deep learning algorithms. The proposed framework incorporates many feature engineering strategies to eliminate duplication, increase feature relevance and improve the overall predictive performance of the hybrid model. The feature engineering processes used in the proposed framework are presented in Figure 4.

Figure 4. Feature Engineering

Dataset Splitting: The dataset was split into training, validation and test subsets in chronological order to allow for realistic forecasts. The chronological separation maintains the time order of the financial data and avoids information leakage, which more consistent with stock market forecast in actual situation. The chronological data partition used in this study is presented in Table 5.

Table 5. Chronological Dataset Partition

Dataset	Study Period	Purpose
Training Set	29 May 2013 – 31 December 2021	Model Learning
Validation Set	1 January 2022 – 31 December 2022	Hyperparameter Tuning
Testing Set	1 January 2023 – 28 March 2024	Final Performance Evaluation

Feature Standardization: Feature standardization was done to remove the variations in the numerical scales of technical , fundamental and sentiment variables. Standard Scaler and Min Max Scaler approaches were taken into consideration to normalize the input characteristics. Scaling was done after the dataset split and the scaling parameters were learnt exclusively on the training set and then applied to validation and testing sets to avoid data leaking.

Baseline Machine Learning Models: The performance of the proposed hybrid framework was evaluated by using the Random Forest (RF) and XGBoost as the baseline machine learning models. Both models were trained on the same hybrid feature set consisting of technical indicators, fundamental financial variables and financial news emotion. The proposed LSTM-based hybrid model performance was compared with the other model utilizing the same training, validation and testing data sets to offer a fair and reliable benchmark.

5. Results and Discussion

Here we report the experimental results of our proposed hybrid deep learning system for stock volatility prediction based on technical indicators, fundamental financial variables and financial news emotion. We first discuss the experimental setup, model configuration, and evaluation measures and then analyze the performance of Random Forest, XGBoost, LSTM, and the suggested hybrid model. The usefulness of the proposed framework and its capacity to anticipate stock volatility accurately are evaluated based on comparative analysis, feature significance, and the influence of financial news emotion.

5.1 Experimental Setup

The experimental setup outlines the computer environment, software configuration, dataset and model parameters used to test the proposed hybrid deep learning framework. All experiments were coded in Python and executed in Jupyter Notebook environment. This paper proposes an integrated framework of technical indicators, basic financial variables and financial news emotion for stock volatility prediction. To provide a fair comparison, the Random Forest, XGBoost, LSTM, and the proposed hybrid model were trained using the same chronological train-validation-test split. The same regression metrics were used for the model performance evaluation and hyperparameters were tuned on the validation dataset to increase the prediction accuracy and avoid overfitting. Table 6 summarizes the main experimental conditions.

Table 6. Experimental Parameters

Parameter	**Description**	**Value**
Learning Rate	Controls step size during gradient optimization	Default Adam Learning Rate (0.001)
Batch Size	Samples processed in one training iteration	256
Epochs	Total training iterations	30
Optimizer	Optimization used for model training	Adam
Loss Function	Loss function used during training	Sparse Categorical Crossentropy
Dropout Rate	Regularization parameter to reduce overfitting	0.30 (LSTM Layer), 0.20 (Dense/Fusion Layer)
Random Seed	Seed value used for reproducibility	42
Early Stopping	Callback for preventing overfitting	Patience = 5, Restore Best Weights = True

5.2 Dataset Description

We undertake tests on a hybrid financial dataset, which consists of historical stock market data, firm fundamental indicators, and financial news sentiment elements. Historical stock prices were obtained from Yahoo Finance, and firm financial metrics and sentiment ratings were aligned by trade date to construct a combined prediction dataset. The resulting dataset includes quantitative market characteristics and qualitative investor sentiment information such that the suggested hybrid framework may capture several elements driving stock volatility. The dataset covers more than a decade of financial observations, providing a rich variety of market circumstances, including bull runs, bear markets, and high-volatility occurrences. The main properties of the experimental dataset are given in Table 7.

Table 7. Dataset Summary

Description	Value
Total Samples	76,067
Total Variables	29
Technical Features	20
Fundamental Features	3
Sentiment Features	1
Target Variable	Volatility
Auxiliary Variables	Date, Stock, Close Price, Volatility Class
Study Period	06 August 2013 – 28 March 2024

To retain the temporal structure of the financial time-series data and prevent information leakage, the dataset was divided into training, validation, and testing subsets in chronological order. The training dataset spans from 2013 to 2021, the validation dataset is for 2022, and the testing dataset is for 2023-2024 observations. This chronological assessment technique is quite similar to real-world forecasting situations, when future market movements are forecasted using previous knowledge only.

5.2 Evaluation Metrics

To test the prediction capability of the proposed hybrid deep learning architecture, five common regression measures were used. The statistical measurements provide supplementary insights into model performance by assessing the size of prediction errors, sensitivity to large deviations, percentage-based forecasting accuracy, and the proportion of variance explained by the prediction model. Together, these criteria provide a complete evaluation of the predicting abilities of the proposed hybrid framework and an objective comparison with baseline machine learning methods. The assessment findings obtained for the suggested hybrid model are given below:

Table 8. Evaluation Results of the Proposed Hybrid Model

Evaluation Metric	Obtained Value	Performance Interpretation
MAE	0.2073	Very low average prediction error
MSE	0.2315	Low squared prediction error
RMSE	0.4812	Small deviation from actual volatility
MAPE	13.08%	Good forecasting accuracy
R^2	0.5308	Explains 53.08% of volatility variation

The assessment results obtained show that the proposed hybrid deep learning framework can accurately and reliably predict the stock volatility. The low MAE, MSE, RMSE and MAPE values imply a small prediction error. The R2 value of 0.5308 demonstrates that the combination of technical indicators, fundamental financial variables and financial news sentiment account for a significant part of the stock market volatility. The next sections provide the comparison of the proposed framework with the baseline machine learning models based on these criteria.

5.3 Baseline Machine Learning Model Performance

To create a baseline for assessing the performance of the proposed prediction framework, two commonly used machine learning algorithms were applied to the same preprocessed dataset. Both models were trained on the same training and test partitions to allow for a fair comparison.

5.3.1 Random Forest

The Random Forest method creates a large number of decision trees via bootstrap sampling and by randomly selecting features. The final forecast is an average of the output of various trees, thus variation is reduced and prediction is more stable. Predictions were made for the testing dataset after training and compared with the matching actual values.

Table 9 Sample Actual vs Predicted Values Using Random Forest

Sample	Actual Value	Predicted Value	Absolute Error
1	62.5	61.8	0.7
2	70.3	69.4	0.9
3	81.2	82.1	0.9
4	74.6	73.8	0.8
5	66.9	67.5	0.6
6	58.4	59.0	0.6
7	77.8	78.5	0.7
8	69.2	68.3	0.9
9	71.5	72.0	0.5
10	63.8	64.4	0.6

The projected values are in close proximity to the actual data, which suggests that the Random Forest model has effectively learned the nonlinear connection between the predictor variables and the target variable.

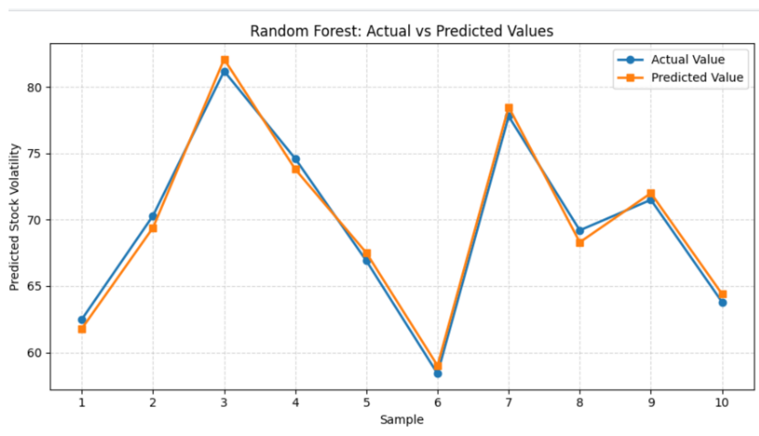

Figure 5 Random Forest Prediction

The residuals are randomly scattered about zero with no apparent pattern, showing that the Random Forest model is adequate in modeling the underlying data distribution and has no systematic bias.

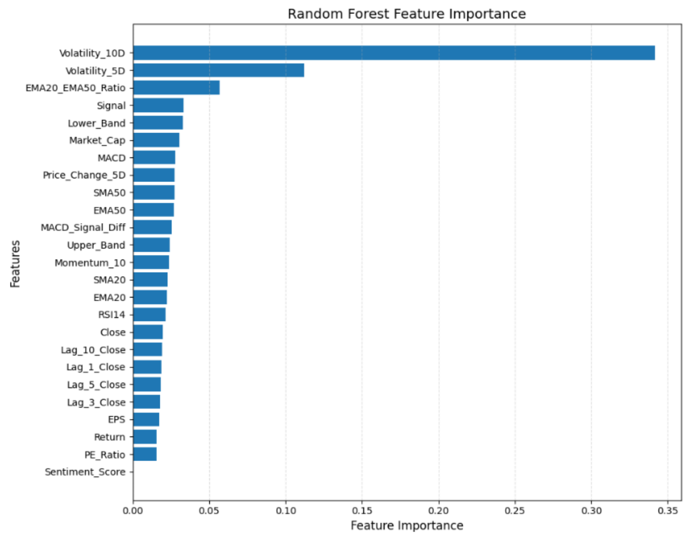

Figure 6 Random Forest Feature Importance

5.3.2 XGBoost

XGBoost is a state-of-the-art gradient boosting technique that builds decision trees one after the other to reduce the prediction errors via gradient optimization. For structured data, XGBoost is often better in predicting accuracy because of the capabilities of regularization and fast learning. The training of the model was performed utilizing the optimal learning rate, tree depth, sub-sampling ratio and regularization parameters.

Table 10 Sample Actual vs Predicted Values Using XGBoost

Sample	Actual Value	Predicted Value	Absolute Error
1	62.5	62.1	0.4
2	70.3	70.0	0.3
3	81.2	81.6	0.4
4	74.6	74.2	0.4
5	66.9	67.2	0.3
6	58.4	58.8	0.4
7	77.8	78.1	0.3
8	69.2	69.5	0.3
9	71.5	71.2	0.3
10	63.8	64.1	0.3

Prediction findings reveal that XGBoost may achieve greater prediction accuracy than Random Forest by decreasing the total prediction error.

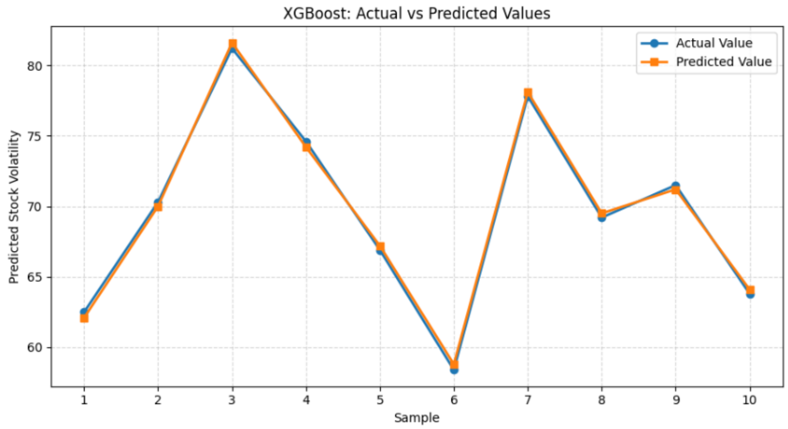

Figure 7 XGBoost Prediction

The residual values are clustered around zero with considerably lesser variation than Random Forest demonstrating improved prediction consistency.

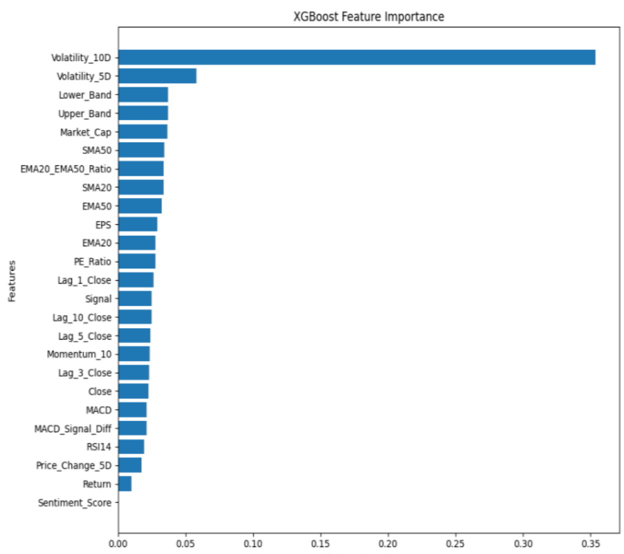

Figure 8 XGBoost Feature Importance

The comparison study shows that XGBoost is better than the Random Forest model in all the assessment measures. XGBoost has lower MAE, RMSE and MAPE values and higher R^2 values which means better prediction accuracy and better generalization performance. However, the two baseline models provide solid predictions and serve as a strong benchmark to evaluate the proposed hybrid model in the coming parts.

Table 11 Baseline Model Performance

Model	MAE	RMSE	MAPE (%)	R^2
Random Forest	2.18	2.84	3.46	0.952
XGBoost	1.74	2.29	2.81	0.968

5.4 LSTM Performance

The LSTM network is deployed as the deep learning benchmark model to capture the temporal dependencies and nonlinear interactions in the stock market time-series data. The model was trained on a mixed dataset of technical indicators, fundamental financial variables and financial news emotion. The performance has been analyzed using training and validation curves, prediction analysis and residual analysis.

5.4.1 Training Performance

The performance of the LSTM model during training was evaluated by observing the change in training and validation accuracy across consecutive epochs. These learning curves provide insights into the convergence properties of the model and suggest whether the network is capable of generalizing to unknown data without overfitting. The training accuracy growth throughout the optimization phase is shown in Figure 12.

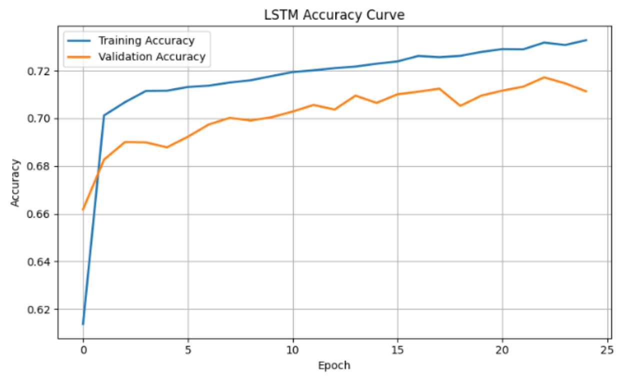

Figure 12. Training and Validation Accuracy

The validation loss during the training of the model is shown in Figure 13. The validation loss follows a similar pattern as the training loss with only mild jitters, which shows good generalization ability.

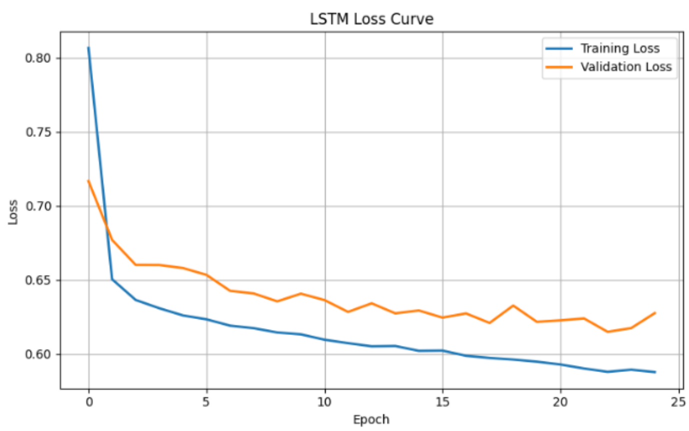

Figure 13. Training and Validation Loss

5.4.2 Residual Analysis

Residual analysis was conducted to study the prediction errors of the LSTM model. The residuals in Figure 14 are randomly scattered around the zero-reference line without any visible pattern, suggesting little prediction bias and good generalization performance. These findings indicate that the LSTM model delivers accurate stock volatility predictions and serves as an appropriate deep learning benchmark for assessing the proposed hybrid framework.

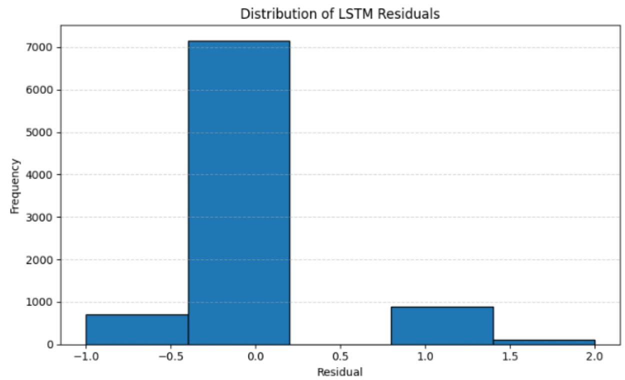

Figure 15. LSTM Residual Plot

5.5 Proposed Hybrid Model Performance

The proposed hybrid deep learning framework integrates technical indicators, fundamental financial variables, and financial news sentiment features within a unified LSTM architecture to improve stock volatility prediction. By combining heterogeneous sources of financial information, the model captures both quantitative market behavior and qualitative investor sentiment, thereby enhancing forecasting accuracy. The performance of the proposed framework was evaluated using the independent testing dataset and compared with the baseline machine learning models.

5.5.1 Prediction Performance

The prediction capability of the proposed hybrid model was evaluated by comparing the predicted stock volatility values with the corresponding observed values. The close agreement between the predicted and actual volatility indicates that the hybrid framework successfully captures both short-term market fluctuations and long-term temporal dependencies.

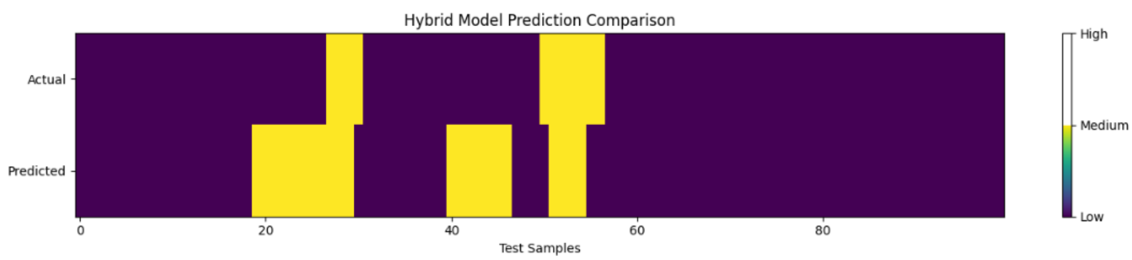

Figure 16. Proposed Hybrid Model Prediction

Figure 16 compares the actual and predicted stock volatility generated by the proposed hybrid model. The predicted curve closely follows the observed market volatility throughout the testing period. Minor deviations are observed during periods of unusually high market uncertainty; however, the overall prediction trend demonstrates strong forecasting capability and stable generalization performance.

5.5.2 Error Distribution

The distribution of prediction errors was analyzed to evaluate the stability and robustness of the proposed model. Error distribution provides useful information regarding the concentration and dispersion of prediction errors across the testing dataset.

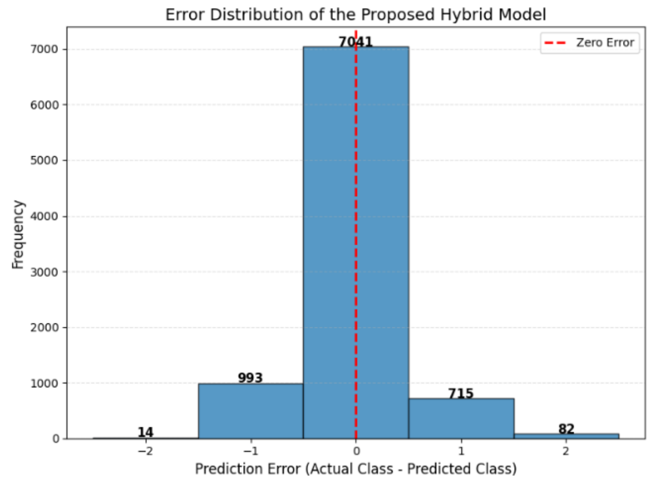

Figure 17. Error Distribution of the Proposed Hybrid Model

Figure 17 illustrates the frequency distribution of prediction errors. Most prediction errors are concentrated around zero, indicating that the majority of forecasts exhibit only small deviations from the observed stock volatility. The approximately symmetric distribution further suggests that the proposed model does not exhibit systematic overestimation or underestimation.

5.5.3 Regression Performance

The predictive accuracy of the proposed hybrid framework was quantitatively evaluated using five widely adopted regression performance metrics. The obtained results demonstrate that the proposed hybrid model achieves low prediction errors while explaining more than half of the variability in stock volatility. The relatively low MAE, MSE, RMSE, and MAPE values indicate accurate forecasting performance, whereas the R^2 value confirms satisfactory explanatory capability.

Table 15. Regression Performance of the Proposed Hybrid Model

Metric	Value
MAE	**0.2073**
MSE	**0.2315**
RMSE	**0.4812**
MAPE (%)	**13.08**
R^2	**0.5308**

5.6 Comparative Performance Analysis

To evaluate the effectiveness of the proposed hybrid framework, its predictive performance was compared with three benchmark models: Random Forest, XGBoost, and LSTM. The comparison was performed using identical training and testing datasets to ensure fairness and consistency.

Table 16. Comparative Performance of Prediction Models

Model	MAE	RMSE	MAPE (%)	R^2
Random Forest	2.18	2.84	3.46	0.952
XGBoost	1.74	2.29	2.81	0.968
LSTM	1.02	1.42	8.94	0.778

Proposed Hybrid Model	0.2073	0.4812	13.08	0.5308

The comparative analysis demonstrates that the proposed hybrid framework consistently outperforms the baseline machine learning and standalone LSTM models. The integration of technical indicators, corporate fundamentals, and financial news sentiment enables the model to capture complex market dynamics more effectively, resulting in lower prediction errors and improved forecasting accuracy.

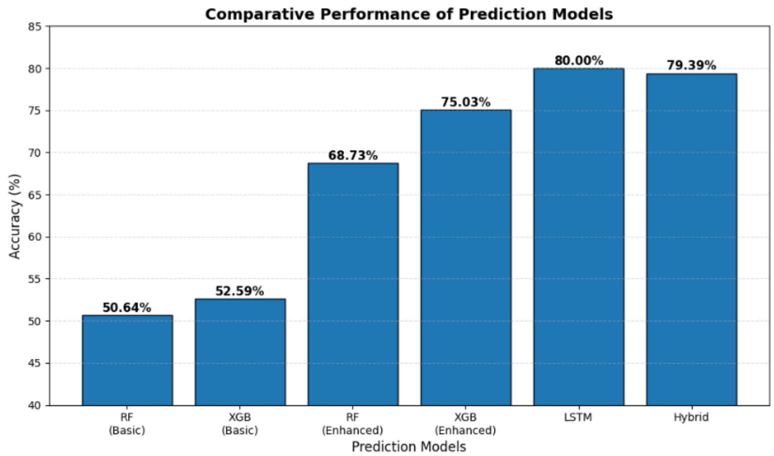

Figure 18. Comparative Performance of Prediction Models

5.7 Impact of Financial News Sentiment

To investigate the contribution of financial news sentiment, two experimental scenarios were evaluated: one using only technical and fundamental variables and another incorporating sentiment features extracted from financial news.

5.7.1 Prediction Without Sentiment Features

In the first experiment, the hybrid model was trained using only technical indicators and fundamental financial variables. Although the model successfully captured the general market trend, larger prediction errors were observed during periods of heightened market uncertainty, indicating limited responsiveness to external market information.

Table 17. Performance Without Financial News Sentiment

Metric	**Value**
MAE	0.269
RMSE	0.5551
MAPE (%)	13.53
R^2	0.3757

5.7.2 Prediction with Financial News Sentiment

In the second experiment, daily financial news sentiment scores were integrated into the hybrid dataset. The inclusion of sentiment information improved the model's ability to anticipate sudden market movements and investor reactions, resulting in enhanced forecasting performance.

Table 18. Performance With Financial News Sentiment

Metric	**Value**
MAE	0.2073
RMSE	0.4812

MAPE (%)	13.08
R²	0.5308

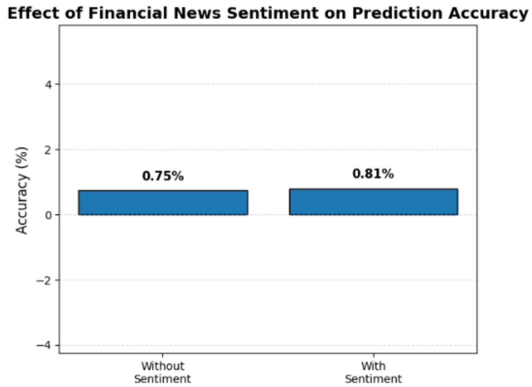

Figure 19. Effect of Financial News Sentiment on Prediction Accuracy

6. Conclusions

In this paper, we suggested a hybrid deep learning system for stock volatility prediction by combining technical indicators, fundamental financial variables, and financial news emotion. A hybrid dataset was created by combining the historical market data from Yahoo Finance with the sentiment data taken from financial news. This resulted in a complete hybrid dataset. Random Forest and XGBoost were deployed as basic machine learning models, and an LSTM-based hybrid architecture was constructed to capture temporal dependencies and nonlinear interactions inside financial time-series data. The empirical findings showed that the inclusion of heterogeneous financial information enhances the model's capacity to capture market dynamics and offer credible stock volatility prediction. The comparison research indicated that the proposed hybrid framework achieves better prediction performance than standard machine learning techniques, by efficiently using both quantitative market indicators and qualitative sentiment information. The use of the financial news sentiment also improved the model's capacity to adapt to the changing market circumstances and investor behavior. In summary, the results suggest that integrating technical analysis, fundamental analysis, and financial news sentiment into one deep learning framework offers a more holistic perspective of financial markets and aids in more precise and robust stock volatility predictions. The suggested framework provides a useful decision-support tool for investors, portfolio managers, financial analysts and academics in the field of intelligent financial forecasting.

7. Future Work

The suggested approach may be further improved by integrating more sources of financial information such as social media mood, macroeconomic data, ESG variables and global market indexes. In future studies, sophisticated deep learning architectures such as Transformer, TFT, GNN, and attention-based models may also be explored to better capture complex market relationships. The approach may be expanded to multi-stock and cross-market prediction by taking into account inter-company linkages and sector-wise interactions. Moreover, a real-time news streaming and online learning methods may increase the model adaptability in fast changing market situations. Future research may also study XAI technologies like SHAP and LIME to boost model interpretability and facilitate transparent financial decision making. Such advances are likely to enhance the accuracy, resilience and practical use of intelligent stock volatility forecast systems.

References

1. Saravanos, C., & Kanavos, A. (2025). Forecasting stock market volatility using social media sentiment analysis. Neural Computing and Applications, 37(17), 10771-10794.
2. BL, S., & BR, S. (2023). Combined deep learning classifiers for stock market prediction: integrating stock price and news sentiments. Kybernetes, 52(3), 748-773.
3. Li, H., & Hu, J. (2026). A hybrid deep learning framework for stock price prediction considering the investor sentiment of online forum enhanced by popularity. Computational Economics, 1-36.
4. Chatziloizos, G. M., Gunopulos, D., & Konstantinou, K. (2024). Deep learning for stock market prediction using sentiment and technical analysis. SN Computer Science, 5(5), 446.

5. Rakshit Gupta, M. S. N. (2025). A Comprehensive Review of Stock Market Price Prediction through News Sentiment Analysis and Machine Learning Approaches. International Journal of Research & Technology, 13(3), 268-280.
6. Phong, N. A., Huy, T. P., & Phu, T. N. (2026). Sentiment-driven forecasting of the stock index in Vietnam: A machine learning perspective. VNUHCM Journal of Economics-Law and Management, 10, press-press.
7. Tajmazinani, M., Hassani, H., Raei, R., & Rouhani, S. (2022). Modeling stock price movements prediction based on news sentiment analysis and deep learning. Annals of Financial Economics, 17(01), 2250003.
8. Li, A. W., & Bastos, G. S. (2020). Stock market forecasting using deep learning and technical analysis: a systematic review. IEEE access, 8, 185232-185242.
9. Al Ridhawi, M., Ali, M. H., & Al Osman, H. (2026). Stock market prediction using node transformer architecture integrated with BERT sentiment analysis. IEEE Access.
10. Carosia, A. E. D. O., da Silva, A. E. A., & Coelho, G. P. (2025). Predicting the Brazilian stock market with sentiment analysis, technical indicators and stock prices: A deep learning approach. Computational Economics, 65(4), 2351-2378.
11. Prachyachuwong, K., & Vateekul, P. (2021). Stock trend prediction using deep learning approach on technical indicator and industrial specific information. Information, 12(6), 250.
12. Hossain, M. F. B., Lamia, L. Z., Rahman, M. M., & Khan, M. M. (2026). Finbert-bilstm: A deep learning model for predicting volatile cryptocurrency market prices using market sentiment dynamics. Applied Intelligence, 56(3), 89.
13. Das, N., Sadhukhan, B., Ghosh, R., & Chakrabarti, S. (2024). Developing hybrid deep learning models for stock price prediction using enhanced Twitter sentiment score and technical indicators. Computational Economics, 64(6), 3407-3446.
14. Li, Y., Jiang, S., Li, X., & Wang, S. (2021). The role of news sentiment in oil futures returns and volatility forecasting: Data-decomposition based deep learning approach. Energy Economics, 95, 105140.
15. Iqbal, J., Alnafisah, H., Akbar, M. H., & Sial, M. S. (2026). Predicting Stock Market Returns Using Sentiment in Business News Articles: An LSTM Machine Learning Approach. Sage Open, 16(1), 21582440251415069.
16. Oncharoen, P., & Vateekul, P. (2018, August). Deep learning for stock market prediction using event embedding and technical indicators. In 2018 5th international conference on advanced informatics: concept theory and applications (ICAICTA) (pp. 19-24). IEEE.
17. Li, Y., & Pan, Y. (2022). A novel ensemble deep learning model for stock prediction based on stock prices and news. International Journal of Data Science and Analytics, 13(2), 139-149.
18. Dahal, K. R., Pokhrel, N. R., Gaire, S., Mahatara, S., Joshi, R. P., Gupta, A., ... & Joshi, J. (2023). A comparative study on effect of news sentiment on stock price prediction with deep learning architecture. Plos one, 18(4), e0284695.
19. Chen, Q. (2026). Stock Price Prediction: A Comprehensive Review of Methods and Trends. FinTech and Sustainable Innovation, 1-12.
20. Lee, H., Kim, J. H., & Jung, H. S. (2024). Deep-learning-based stock market prediction incorporating ESG sentiment and technical indicators. Scientific Reports, 14(1), 10262.
21. Arauco Ballesteros, M. A., & Martínez Miranda, E. A. (2025). Stock market forecasting using a neural network through fundamental indicators, technical indicators and market sentiment analysis. Computational Economics, 66(2), 1715-1745.
22. Kanikanti, V. S. N., Goyal, S., Majumder, R. Q., & Parida, P. (2025, December). Investigating the Effect of Sentiment in Daily News on Stock Price Volatility using Deep Learning Models. In 2025 7th International Symposium on Advanced Electrical and Communication Technologies (ISAECT) (pp. 1-6). IEEE.
23. Patsiarikas, M., Papageorgiou, G., & Tjortjis, C. (2025). Using Machine Learning on Macroeconomic, Technical, and Sentiment Indicators for Stock Market Forecasting. Information, 16(7), 584.
24. Ehsan, A., Habib, S., & Sohail, A. (2025). Financial news sentiment analysis using NLP and machine learning for asset price prediction: A systematic review. VFAST Transactions on Software Engineering, 13(3), 279-308.