ISSN: 2182-2069 / E-ISSN: 2182-2077

Time Series Data Analysis and Prediction of Stock Prices Using Deep Learning Models

K.V.K. Sasikanth[1*], B. Sujatha[2], and D. Haritha[3]

[1]Research Scholar, JNTUK, Kakinada, Andhra Pradesh, India. sasikanthkolanu@gmail.com, https://orcid.org/0000-0002-5250-9701

[2]Professor, Department of CSE, Godavari Institute of Engineering & Technology (A), Rajahmundry, Andhra Pradesh, India. birudusujatha@gmail.com, https://orcid.org/0000-0001-9433-3647

[3]Professor, Department of CSE, University College of Engineering, Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh, India. harithadasari9@yahoo.com, https://orcid.org/0000-0003-4627-6005

Received: February 11, 2025; Revised: March 24, 2025; Accepted: May 05, 2025; Published: May 30, 2025

Abstract

The objective of this paper is to develop a reliable method for predicting NSE (National Stock Exchange) stock prices in India using deep learning models. The purpose of this model is to enhance the precision of time series data management by developing an advanced hybrid multistep-ahead CNN-LSTM model, which is capable of real-time time series forecasting with high accuracy. Learnable sets are trained and tested using commonly recognized error measures, such as accuracy, and the model is fine-tuned using 5-minute interval NSE stock data. A thorough examination of deep learning models against traditional time series models is necessary to establish the advantages and disadvantages of deep learning models. Accordingly, even though deep learning models appear to exhibit a promising trend in time series prediction and analysis, a comprehensive evaluation is required. This combined approach of CNNs and LSTMs in the hybrid model has a higher chance of detecting true changes in stock price movements of the NSE market in real-time.

Keywords: Time Series Analysis and Prediction, Deep Learning, CNN, LSTM, Error Metrics.

1 Introduction

Time series, or TS, is the entire set of sequential observations that are collected, taking the timing of the respective recorded observations into considerable attention. A mathematical notation of a time series is represented by $\{X(t)t=0, \pm1, \pm2\ ,..., T\}$, where variable 't' is equivalent to time index or timestamp. Dem X(t) vectors are the random numbers that are representing the real time value of the time series variable in each moment "t". Thus, "T" denotes the total description number or time set in the series, and hence, a TS can be expressed as: $\{X(1), X(2), ..., X(t), \cdots..., X(T)\}$. Thus, in both cases, it is possible to refer to both discrete and continuous values of X(t); prices of stock over time and monthly automobile sales are examples of the former type.

Journal of Internet Services and Information Security (JISIS), volume: 15, number: 2 (May), pp. 499-511.
DOI: *10.58346/JISIS.2025.I2.035*

*Corresponding author: Research Scholar, JNTUK, Kakinada, Andhra Pradesh, India.

While univariate and multivariate time series data share similarities, they also exhibit profound differences that make them truly distinct. Shifts over time, cyclicality, determinism, autocorrelation, and robustness are all examples of these concepts (Chatterjee et al., 2021). The future estimated values of the series are examined, and Time Series Analysis (TSA) is simply a method for understanding them (Jasim, 2022). One could conclude that TSA is a technique useful for finding data patterns and predicting future data through the analysis of these patterns (Sachdeva et al., 2019). These are finance, agriculture, healthcare, marketing, economics, environment, and technology, which are not only a significant sector but also a high demand area for TSA.As in the case of machine learning and deep learning, classical methods and the techniques in TSA involve both practical and theoretical parts to master and make use of the cutting-edge tools for a holistic understanding Mahmoud & Mohammed (2021). Classical approaches to time series analysis, such as traditional statistical methods for analyzing and modeling time series data, are complemented by various modern techniques. Different forecasting methods, such as ARIMA Malik, et al., (2022); Sonkavde et al. (2023) seasonal ARIMA (SARIMA) Winata et al., (2021), and Exponential smoothing models, have commonly been used to provide estimates Zhang et al. (2022). Traditionally, ARIMA, SARIMA, and Exponential smoothing models have been considered effective models that utilize historical data to forecast and identify time series patterns. Despite this, these methods are not free from limitations. They fail to model the seasonality properly and are also unable to manage the high-dimensional and nonlinear relationships. Therefore, more advanced time series modeling models like machine learning based models, have been developed which are more flexible and can easily handle the complex data of a high dimension directly and non- traditional relationships as well Sonkavde et al. (2023). Classical time series methods are still used and efficient enough for moderate and well-known data, but modern approaches appear especially useful for complex and unknown series Mehtab & Sen (2020).

Many machine learning algorithms, like MLP, RNN, SVM, Kernel Methods, and Ensemble Methods, being presented in the scientific literature, are intended to deal with short-term forecasting tasks (Choy, et al., 2021; Zhang et al., 2022). In another sense, models employ a different set of methods to forecast from time series data. However, it is a challenging task, as obtaining constant training is expected to consume a substantial amount of resources and large volumes of data. Moreover, the problem of overfitting may be widespread in ML models, as a result of the exact tuning and validation are essential for making accurate forecasts.

Forecasting of time series data and deep learning models have become more popular due to the recent emergence (Alzaidi, 2024). This concept demonstrates how RL agents are capable of carrying out complex interactions that are often non-linear and characteristic of real-time data. While techniques such as regression are applied to big data that contain many input variables, it is more efficient to use machine learning models in this case. One does note that the capability of Long Short-Term Memory and its varieties to capture the long-run dependencies in the time series data has increased the usage of LSTM in many areas, e.g. Voice Recognition, Language Modeling, and Stock Price Prediction among many other fields (Palash & Dhurvey, 2024).

The subject of our project is to precisely model for stock exchange of India, NSE, the technique that would be reliable and could work for this aim. To achieve this aim, we present a system built on a deep learning technique that combines CNN, LSTM, and CNN-LSTM to enhance efficiency, stability, and accuracy Angel Merlin Suji & Anto Kumar (2022). The related framework we propose utilizes five-minute time intervals in the daily stock price information, allowing it to capture the minute-by-minute changes in these prices that models with less frequently sampled quantities of information would fail to notice.

By capturing the fundamental characteristics of the stock market, the suggested framework aims to provide traders accurate and dependable short-term predictions of stock price fluctuations. Rather of concentrating on anticipating movements in the long run, our approach is built to emphasize forecasts of short-term stock prices (Shen & Shafiq, 2020). To meet the demands of traders, this trade-oriented strategy provides them with accurate and timely stock price forecasts.

The research consists of six parts. Its goals are laid forth in the first part. Predicting and predicting stock prices is the subject of the second section's comprehensive literature study Sharma et al. (2023). Section 3 delves into the conversation of the profound learning models utilized in the exploration. The methodology, including preprocessing strategies, is presented in detail in Section 4. In Area 5, we thoroughly analyse the results of the DL models and examine how well they performed Deshmukh & Menon (2025). Finally, the sixth part summarizes the study's findings and contributions, discussing their implications for future research.

2 Existing Literature on Modeling and Predicting Stock Price

Stock price prediction accuracy has been the subject of scholarly debate, with several methods and factors being considered. This study employs a model professional approach across three different methodologies as an entity (Sharma & Jain, 2023). Due to the non-deterministic movement of the share price, the predictive power of the first group's OLS models appears to be limited (Sirimevan et al., 2019). The model is introduced in the second part using ML data, whereas the model and stock return prediction are made in the third part using ML, DL, and NLP (Bansal et al., 2022).

Deep learning methods have become increasingly popular in recent years for accurate stock market movement prediction, although they are complex functions characterized by their nonlinear nature. As an illustration, to calculate stock profits, Luo et al. (2022) implemented an LSTM model, assigning the optimal hyperparameters using the ASFL technique. They suggested a model that had a better relationship with others, such as ANN, SVM, and the simple LSTM, which was proven to be better during the performance comparison. A parallel approach to stock movement prediction, as proposed by Sen & Mehtab (2021), was conducted using convolutional neural networks on multivariate data sets. They discovered that the sensing of NIFTY movement direction was among the most accurate events in their device trials (Chen, 2023).

The research undertaken by Khattak et al. (2019) showed the wrongness of utilizing social media stock price data, which is mostly incomplete and could not be made appropriate based on the predictors chosen for incorporation in the prediction models and hence the prediction produced would be inaccurate. To address this issue, Rahman's model is an MLR that is oriented towards the selection of predictors in a systematic way, thereby garnering accurate predictions. Although it is easier to apply, the model faces a few limitations when dealing with TS data that shows a tremendous change in stock price. Through the combination of MLR and BPNN, proposed a hybrid model to forecast stock prices. The performance of the BPNN model was found to be superior to that of the MLR model when assessing the impact of changes in stock prices (Zhao & Cheng, 2022).

As you tried to interpret the time series data, full of sharp movements in the BSE index, it is doubtful that the BPNN model seems appropriate. The method demonstrated an excellent level of efficiency when tested using daily stock price data for estimating prices. Meanwhile, this tool may be too slow to be useful for those who wish to make real-time investment decisions because it takes time to process and analyze (Yeo & Jiang, 2024).

Perhaps the most glaring weakness of the existing studies on stock prices movements is their low accuracy of predicting even short-term price variations. The learning capability of profound brain networks in stock cost development, gauging, and modeling was the goal of this research. The model we propose in this research is intended to overcome the limitations of previous deep learning models in addressing this issue. We achieved this by the strategy above, and hence, we can boast of the ability to develop a more accurate, precise, and reliable method for stock valuation, especially over shorter periods.

3 Deep Learning Models

While there exist different deep learning models for the same purpose, one of the most effective classes in this context is the convolutional neural network for time series data forecasting, along with other reputable models, such as the LSTM. Data analysis is where CNNs shine, but LSTM networks shine when it comes to time series data. A stock prediction model that utilizes the best features of both models is created by merging them in the research. Financial data analysis, natural language handling, and voice recognition are a few examples of the many non-visual applications that have demonstrated CNN effectiveness Kumar et al. (2020).

3.1 Convolutional Neural Network (CNN)

The ability of CNN to immediately produce a multi-step vector and automatically learn features is a significant advantage when using it for time series prediction. Using the univariate data, the research use CNN to forecast several future time steps in a time series (Li et al., 2022). The two primary building blocks are the convolutional and pooling layers. Central neural network (CNN) architectures rely on convolutional layers. To extract features from the input data typically an image or a sequence of data it performs a convolution operation. It involves a set of layers that slide over the input data. When a kernel is convolved with the input data, it generates a feature map that represents a modified understanding of the input information. The resulting feature map is then mapped onto a filter map. To extract essential features, pooling layers employ techniques such as signal averaging or max pooling, which help condense the feature maps. These layers are repeated at different depths to create multiple levels of abstraction. Lastly, the last layer is transmitted to one or more fully connected layers, which decode the internal representation and map it to a class value. The process of multi-step ahead 1D-CNN is depicted in Figure 1.

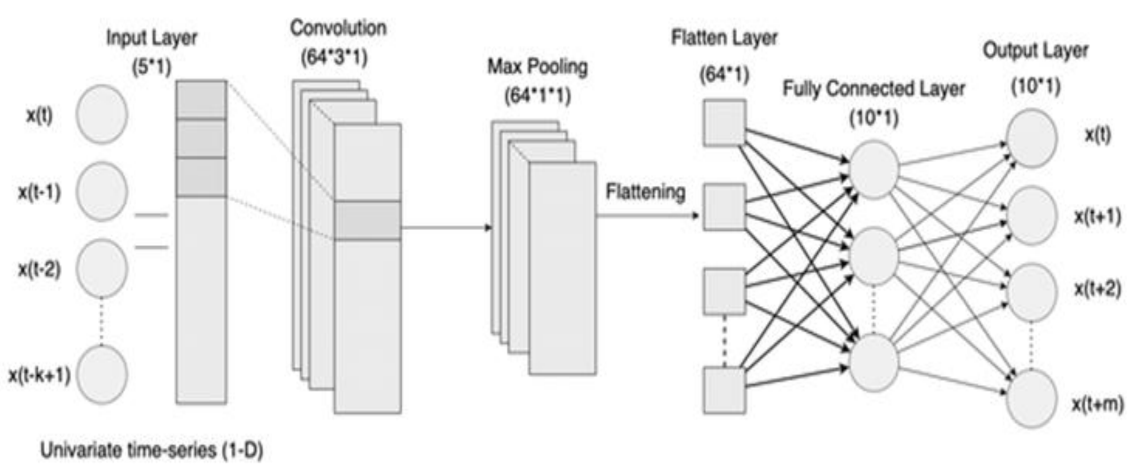

Figure 1: The Process of Multi-step ahead 1D-CNN

Through the use of a CNN, it is possible to extract relevant features and gain knowledge from a TS consisting of only the open values. With this data, we may foretell several time steps in the future. An auto-regression technique underpins the method, which utilizes historical data to project values for a specified future period.

3.2 LSTM (Long Short-Term Memory)

A recurrent neural network, LSTM is specifically designed to handle time-series data. Several forecasting domains have achieved success with it due to its ability to prevent gradient vanishing and identify input data with long-term relationships. A variety of gates are incorporated into the architectural design of the LSTM network to regulate information flow through the network. Such parking gates are seen on Figure 2. The duration of these gates, hence, allows for storing and retrieving data even over long periods, thereby enabling the network to remember long-term dependencies. They are also important factors in deciding what data to output or manipulate.

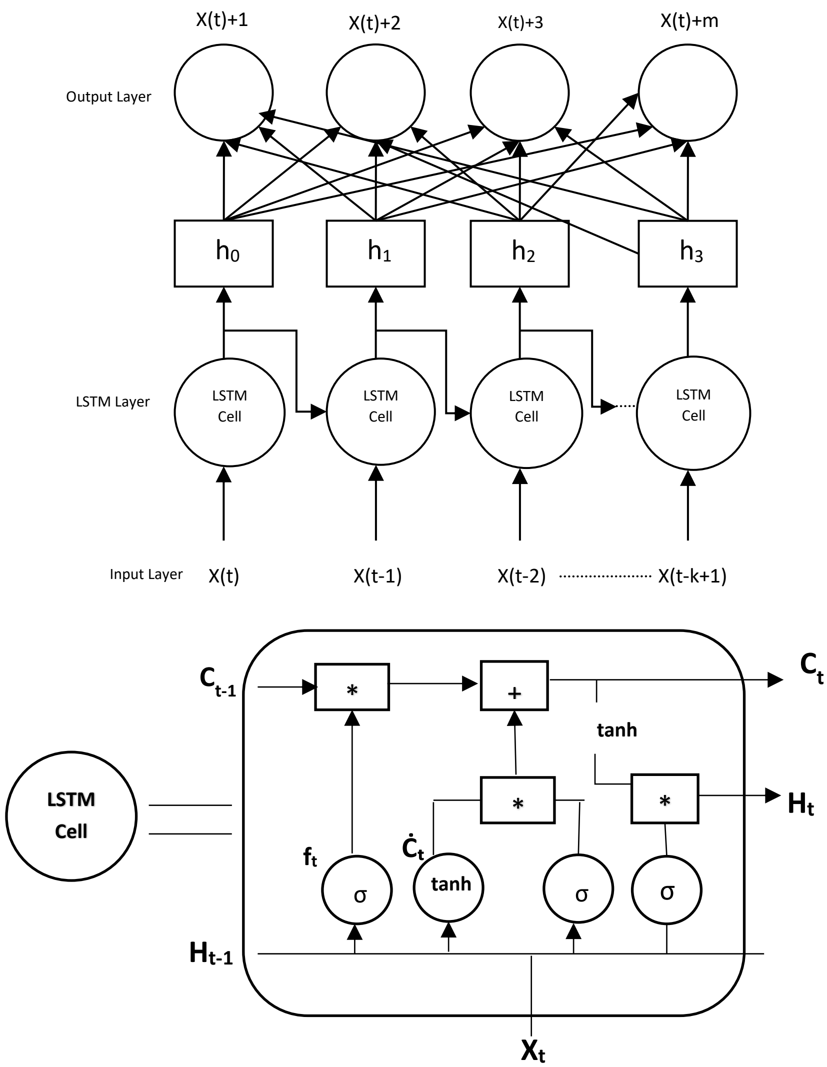

Figure 2: LSTM Neural Network

To regulate data transmission, the LSTM design utilizes multiple gates. Choosing which pieces of data to update is the input gate task, while choosing which pieces of data to remove is the forget gate job. The main function of the output gate and the second in command of data storage is the cell state, not what is stated in the latter role. In this architecture, the LSTM can be configured so that the cell state's activity influences the LSTM's decision on whether to retain or discard extra data from the previous time step. The importance of careful pre-processing algorithms when LTD is involved in obtaining input data cannot be overemphasized. At this specific moment, the activation gate selects the crucial information required for the subsequent hidden state and output vector.

3.3 Hybrid CNN-LSTM

The fact that only CNNs and LSTMs are applied, and even when they are carefully combined, may not be sufficient to forecast stock prices, is viewed as a roadblock for this proposed study; hence, the need to include a novel DL architecture. As shown in Figure 3, this architecture comprises two stages: an LSTM phase and a CNN pre-processing phase in the second stage. Before being trained by convolutional neural networks (CNNs), the univariate input data that will be used should first undergo feature extraction, 1-D convolution, and other processing steps to convert it into a multi-dimensional dataset. Automated decisions facilitate improvement in efficiency at this stage; however, problems arise in accurately predicting patterns and consequences. CNN layer's ability to make univariate data into multidimensional data is favoured, the LSTM neural network gets the inputs which are richer in information, and thus it is easier for the network to make more informative predictions The introduced hybrid deep learning model is intended to better forecast stock prices through combining both the most important aspects of CNNs and LSTM learning. The performance of the framework is aimed at improvement by incorporating the two models and attempting to identify more meaningful patterns within the datasets. As a result of employing advanced deep learning techniques, such as CNN and LSTM, there is a possibility that computerized trades on the stock market may have real-life consequences. This indicates that the technology has the potential to enhance the forecasting accuracy of stock prices. The research presented here demonstrates the potential for using ML methodology to achieve more precise stock price forecasting. That means extra value for investors and financial sharks. Our suggested CNN-LSTM model hybrid is depicted in Figure 3.

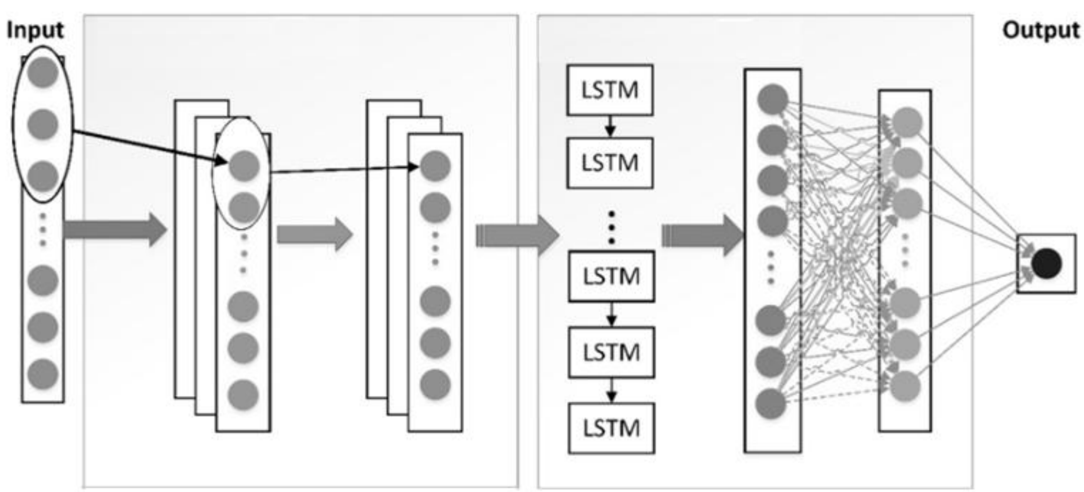

Figure 3: Our Suggested CNN-LSTM Model Hybrid

4 Proposed Method

Figure 4 illustrates the hybrid CNN-LSTM model, which is described in detail below. It consists of five separate phases.

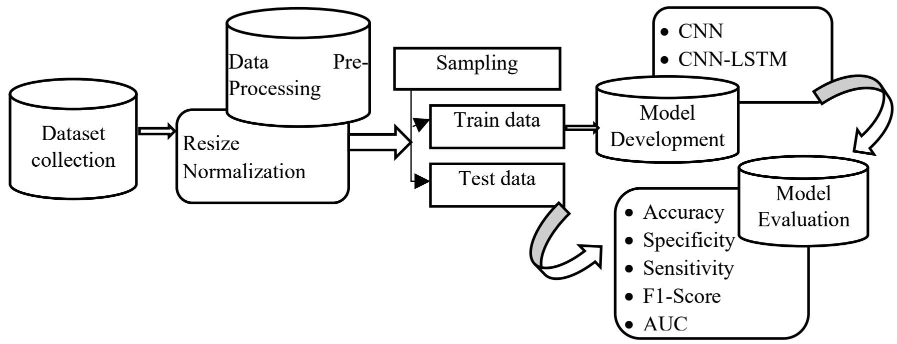

Figure 4: Time Series Prediction and Analysis Using the Suggested Deep Learning Model

Step 1: Data Collection: The Metastock tool is used to gather data, which includes variables such as the NIFTY index, date, time, open, high, low, close, and volume of the traded stock. Every day that NSE was operational in 2013 and 2014, data was collected at a rate of five minutes. The prediction was further refined by dividing the period into three distinct blocks: morning, afternoon, and night. Additionally, eleven derived variables were constructed to serve as input variables for stock price forecasting prediction models that employed both classification and regression techniques Investing.com (n.d.). The goal of including these variables is to improve the model's predictive accuracy.

Step 2: Data Preprocessing: Processing steps such as scaling, normalization, and feature selection may be required for this. To begin, we used the dataset's average to eliminate any missing values. As a second step, we eliminated or imputed outliers using median to cope with data that had them. Thirdly, we employed Min-Max normalization methods to shape and standardize the data, enabling it to be accurately entered into the models.

Step 3: Sampling: Time series analysis makes extensive use of regular sampling, in which data are gathered at regular intervals (e.g., hourly, daily, or weekly) for the most part. The data will be ready for analysis if it is uniform and consistent, which this process guarantees. To prepare and testing the CNN and CNN-LSTM models, the pre-processed data used in this work was also gathered at regular intervals and divided into two parts: a training set and a testing set.

Model Training: The selected deep learning model will next be trained using the backpropagation technique. To minimize the error in the training data, this strategy involves adjusting the model's parameters.

Model Testing: The next phase, after training, involves evaluating the model's predictive abilities using novel, unseen time series data. As part of this procedure, suitable measures like RMSE and MAE are used to assess how well the model predicts outcomes. This statistic measures the accuracy of the forecast by comparing it to the actual values in the test data; a lower number indicates a better prediction. The success and dependability of the deep learning model in practical settings can only be ascertained during the testing phase.

Step 4: Model Development: The study's proposed hybrid CNN-LSTM model is illustrated in Figure 3. This model consists of both the CNN pre-processing and LSTM stages. During the pre-processing phase of a convolutional neural network (CNN), 1D convolution is used to convert the original univariate time series data into multidimensional data, allowing for the extraction of important features. Afterwards, the model takes the pre-processed data into account during the LSTM phase, when the LSTM layer accurately predicts future outcomes by capturing the temporal relationships in the data. Stock price forecasts are the result of the model. In summary, the hybrid CNN-LSTM model enhances stock price prediction accuracy by leveraging the complementary capabilities of CNN and LSTM.

Step 5: Model Evaluation: The test set is a distinct dataset used to assess the effectiveness of the applied model in terms of its generalizability and accuracy in predicting new, unknown data. To assess the model's exhibition on the test information, RMSE and MAE are used.

As a time series model is not entirely accurate, it is necessary to convert it into a large number of aspects that make it more suitable for forecasting future values of the time series. The choice of model (for example, filter, layer), and varying model parameters (such as stride length, window size, or number of delays). The objective of fine-tuning is to achieve a perfect forecast of future time-series values by the model.

5 Model Evaluation and Performance Analysis

First, the research focuses on evaluating the system's performance based on classification and regression tasks, allowing us to measure the accuracy with which the system operates. Among these evaluations, the accuracy of the model and its potential for future improvement are the key factors that can be utilized. Specific sets of metrics are used to calculate the system's effectiveness in the task of classification. This metric consists of specificity, accuracy, sensitivity, and recall. When you put the total number of predictions over that which of particular guess was right, you obtain the accuracy measure. You can view the number of positive outcomes among all positive cases and negative cases that are generating false results by examining the SS ratio for sensitivity/recall and the SC ratio for specificity, respectively. During the effectiveness evaluation process and when seeking makeshift solutions, these phasers have great value. We will analyze and consider two measurements to quantify model's performance: AUC and PPV. The PPV model assesses a model's ability to correctly identify events within the specified group. In turn, the AUC is a metric of the model's overall performance because it represents the area under the ROC curve. Higher AUC values suggest maximum score, whereas PPV is expressed as a percentage and more serves as the ratio of true positives to the total of true positives and false positives. In this model, the False Discovery Rate, also known as FDR, is the negative counterpart of the Positive Predictive Value, or PPV; conversely, it defines how accurately the model discerns negative instances, which is referred to as the Negative Predictive Value, or NPV. The number of negative cases that are being skipped over by the model testing is then calculated using the False Omission Rate (FOR), also known as the negative predictive value (NPV). The F1 score is a statistic used for evaluating the accuracy of a model, taking into consideration not only precision and recall, but also varying between 0 and 1. A higher score will indicate good results. The formula to calculate the F1 Score is 2× (recall × precision) / (recall + precision).

Two regression measures that are used to evaluate the precision of forecasting models are MAPE and RMSE. While RMSE estimates the standard deviation of the residuals given in Eq. 2 and MAPE estimates the typical absolute rate of discrepancy between anticipated and actual values, respectively, they are shown in Eqs. 1 and 2. Both metrics provide valuable information about a model's ability to estimate future time series values and can be used to assess and compare the accuracy of various models.

$$MAPE = \frac{\sum \left(\frac{|actual\ values - predicted\ values|}{actual\ values}\right)}{n} \times 100\% \tag{1}$$

$$RMSE = \sqrt{\frac{\sum (actual\ values - predicted\ values)^2}{n}} \quad (2)$$

6 Experimental Analysis

This paper aims to develop a hybrid deep learning model for predicting daily fluctuations in the NIFTY 50. To achieve this, the data collected from Godrej Consumer Products Ltd. every 5 minutes during the operational days of the NSE from 2013 to 2015. The original data for each stock includes date, price, open, high, low, volume values which have been computed for each day. CNN, LSTM, and CNN-LSTM deep learning models are being developed for classification and regression to predict future market trends, based on data collected over four years. Then, the data is split into two types: one is training data and the other is testing data, as shown in Table 1.

In addition, for training and testing the proposed models, it uses the TS, where it is reshaped into the appropriate format for each DL model. The entire training set was utilized to forecast the value for the following hour at each time step until the end of the testing set was reached. A desktop computer with a 12th Gen Intel Core i7-12700 processor, 16 GB RAM, 64-bit operating system, and an x64-based processor was utilized to evaluate the model's performance. For this aim, Google Colab's Jupyter Python Language programming was entirely loaded onto the machine.

Table 1: Dataset preparation for model building, training and testing

Dataset		**Period/Year**	**No. of Records @ time**	**Purpose**
Godrej Consumer Products Ltd	Case-1	2013	19385 @ 5-min	model building, testing and analysis
	Case-2	2014	18, 972 @ 5-min	model building, testing and analysis
	Case-3	2013 & 2014	19385 @ 5-min and 18, 972 @ 5-min	2013-Training 2014-Testing

The performance of a deep learning model training depends on the selection of parameters is shown in Table 1. The suggested CNN-LSTM model's parameters are divided into two categories: super parameters, such as layer size, which are selected by varying parameters to achieve the best outcome, and elementary parameters, like the weight matrix and bias, which are determined by random initialization. The suggested model includes three LSTM layers: attention layer-1, dropout layer-3, and pooling and convolution layers. To avoid overfitting, dropout layers randomly remove half of the data after each processing step. The model's input window size is 1 × 4, and its learning rate is 0.0003, using the Adam optimizer. Furthermore, MSE was used as the loss function, and ReLU was used as the activation function.

To enhance the predicted performance of the suggested hybrid CNN-LSTM model, we have conducted experiments for comparison. The main parameters used for predictive models are shown in Table 2. This study suggests applying eight distinct regression and classification models as part of their forecasting framework. These models include RF (Random Forest), LR (Logistic Regression), DT (Decision Tree), BG (Bagging), BO (Boosting), ANN (Artificial Neural Network), KNN (K-Nearest Neighbor), and SVM (Support Vector Machines). When it comes to using past data to forecast future events, every model is different. The researchers aim to determine the predictive accuracy of each categorization system for stock prices, so they're using a variety of different ones. Researchers may use the results of this study to inform their future investment choices by determining which model (or combination of models) is most effective at forecasting stock prices. The models' results in the classification and regression tasks are shown in Tables 3 and 4, respectively. These results may help inform future financial investments.

Table 2: Parameters of the Predictive Model

Model	Batch Size	Learning Rate	Epochs	Layers	Hidden Layers
CNN	20	0.002	200	4	CNN (6, 3)
LSTM	20	0.001	250	4	LSTM (48,16)
CNN-LSTM	30	0.001	250	6	CNN (6,3), LSTM (24,4)

Table 3: Classification results (%) Performance Metrics of Godrej Consumer Products Ltd.

Test Cases	Classification Methods										
	Metrics	LR	KNN	DT	BG	BO	RF	ANN	SVM	CNN	LSTM	**Proposed CNN-LSTM**
Case-1	SS	94.9	89.7	95.2	95.2	89.97	94.6	95.5	94.6	97.31	95.29	98.93
	SC	97.7	96.5	98.2	98.2	95.09	97.7	97.7	95.7	96.1	98.29	98.74
	PPV	97	95.2	97.6	97.6	98.09	97	97	94.3	95.9	97.68	98.01
	NPV	96.1	92.4	96.4	96.4	97.48	98.2	96.6	98.2	96.51	96.45	97.15
	F1 Score	95.9	92.4	96.4	96.2	96.25	95.8	96.3	94.4	95.83	96.47	97.96
	CA	96.5	93.5	96.9	96.9	96.07	96.4	96.8	96.5	94.89	96.98	97.97
Case-2	SS	94.95	87.05	92.52	95.56	97.61	93.13	93.74	94.79	94.87	94.68	97.31
	SC	96.08	93.05	95.83	96.58	96.86	94.31	95.83	93.47	94.31	97.81	99.31
	PPV	95.24	91.2	94.82	95.85	98.08	93.13	94.89	91.91	90.9	97.06	98.77
	NPV	95.84	89.66	93.93	96.34	95.66	94.31	94.87	95.83	91.85	98.28	97.47
	F1 Score	95.09	89.08	93.66	95.7	96.24	93.13	94.31	93.33	95.83	95.86	97.49
	CA	95.57	90.33	94.33	96.12	92.34	93.78	94.88	94.05	94.89	96.44	98.90
Case-3	SS	92.22	84.62	90.09	90.09	92.22	91.31	99.82	93.93	94.87	97.72	98.71
	SC	89.51	49.11	92.54	92.54	93.55	93.05	34.72	90.31	94.31	96.99	99.83
	PPV	87.95	58.04	90.92	90.9	92.22	91.58	56	87.66	98.98	96.13	98.08
	NPV	93.28	79.3	91.85	91.85	93.55	92.82	99.4	95.32	97.6	95.94	99.30
	F1 Score	90.03	68.85	90.5	90.49	92.22	91.44	71.74	90.69	96.4	96.95	98.88
	CA	90.74	65.22	91.43	91.43	92.95	92.26	64.26	91.84	96.45	96.79	97.46

Table 4: Regression Results (%) of Godrej Consumer Products Ltd

Test Cases	Regression Methods									
	Metric	**MVR**	**DT**	**BG**	**BO**	**RF**	**ANN**	**SVM**	CNN	LSTM	**Proposed CNN-LSTM**
Case-1	RMSE	13.32	35.35	40.29	23.4	16.26	17.16	53.88	42.96	40.58	**10.07**
	MAPE	35.02	60.73	23.47	16.72	15.23	12.32	17.31	30.13	28.25	**9.48**
	% matching	81.33	86.58	95.3	91.19	99	94.79	92.73	96.5	97.5	**99.34**
Case-2	RMSE	18.84	37.04	25.7	17.35	10.82	31.39	27.92	19.43	41.00	**13.23**
	MAPE	89	78.37	29.32	21.34	19.35	25.72	13.59	19.47	28.72	**9.06**
	% matching	94.62	82.62	94.9	95.31	97.38	90.62	95.59	95.03	96.5	**98.22**
Case-3	RMSE	18.88	65.83	34.91	41.51	32.02	36.83	82.96	44.78	27.56	**10.53**
	MAPE	3.28	4.37	3.39	4.28	4.21	4.31	4.23	4.45	9.68	**9.33**
	% matching	94.76	52.28	90.76	93.1	93.52	89.1	86.81	94.3	95.8	**98.14**

Based on the results, the model's ability to forecast the time series improves with a lower Root Mean Squared Error (RMSE) value. For MAPE, a value below 10% is often considered good for stock price prediction. However, it is essential to note that MAPE is a relative measure and can be sensitive to the scale of the data. Lastly, we present an overview of all the existing algorithms, including different predictive models of CNN, LSTM, and the proposed CNN-LSTM. The efficiency of the classification and regression models is tabulated in Table 2 and Table 3. For every scenario and metric, the CNN-LSTM model outperformed the existing methods.

7 Conclusion

In this study, we developed an integrated deep learning model for stock price prediction by combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The data was collected at five-minute intervals using the Metastock tool. The models were trained and tested using historical stock prices. Several derived predictor variables are developed after preprocessing and transforming the data. The models were trained and tested using this pre-processed data. These two models are used for both classification and regression models. The proposed hybrid model is compared to eight machine learning models for classification and regression. These eight models are used for analyzing the prediction results. The results show that, on every metric, the DL model using a CNN-LSTM architecture performs better than the eight regression models. In terms of the ratio of RMSE to the mean of the actual values of a predicted variable, CNN models perform better than machine learning models and LSTM-based deep learning models.

References

[1] Alzaidi, E. R. (2024). Optimization of Deep Learning Models to Predict Lung Cancer Using Chest X-Ray Images. *International Academic Journal of Science and Engineering, 11*(1), 351–361. https://doi.org/10.9756/IAJSE/V11I1/IAJSE1140

[2] Angel Merlin Suji, M., & Anto Kumar, R. P. (2022). Leukaemia, Convolutional Neural Networks, White Blood Cell, Classification, Image Extraction, Machine Learning, Comparison. *International Journal of Advances in Engineering and Emerging Technology, 13*(1), 19–30.

[3] Bansal, M., Goyal, A., & Choudhary, A. (2022). Stock market prediction with high accuracy using machine learning techniques. *Procedia Computer Science*, *215*, 247-265. https://doi.org/10.1016/j.procs.2022.12.028.

[4] Chatterjee, A., Bhowmick, H., & Sen, J. (2021, October). Stock price prediction using time series, econometric, machine learning, and deep learning models. In *2021 IEEE mysore sub section international conference (MysuruCon)* (pp. 289-296). IEEE.

[5] Chen, X. (2023, October). Stock Prediction Using Evolutionary Attention-based LSTM. In *2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA)* (pp. 1062-1067). IEEE.

[6] Choy, Y. T., Hoo, M. H., & Khor, K. C. (2021, September). Price prediction using time-series algorithms for stocks listed on bursa malaysia. In *2021 2nd international conference on artificial intelligence and data sciences (AiDAS)* (pp. 1-5). IEEE.

[7] Deshmukh, S., & Menon, A. (2025). *Machine Learning in Malware Analysis and Prevention.* In Essentials in Cyber Defence (pp. 74-89). Periodic Series in Multidisciplinary Studies.

[8] https://in.investing.com/equities/godrej-consumer-products-historical-data.

[9] Jasim, S. S. (2022). Fisher Kernel Discriminant Analysis Under Indefinite Kernel Function*. International Academic Journal of Social Sciences, 9*(2), 52–60. https://doi.org/10.9756/IAJSS/V9I2/IAJSS0914

[10] Khattak, A. M., Ullah, H., Khalid, H. A., Habib, A., Asghar, M. Z., & Kundi, F. M. (2019, December). Stock market trend prediction using supervised learning. In *Proceedings of the 10th International Symposium on Information and Communication Technology* (pp. 85-91).

[11] Kumar, V. U., Krishna, A., Neelakanteswara, P., & Basha, C. Z. (2020, July). Advanced prediction of performance of a student in an university using machine learning techniques.

In *2020 international conference on electronics and sustainable communication systems (ICESC)* (pp. 121-126). IEEE.

[12] Li, Z., Luo, X., Liu, M., Cao, X., Du, S., & Sun, H. (2022). Short-term prediction of the power of a new wind turbine based on IAO-LSTM. *Energy Reports*, *8*, 9025-9037.

[13] Luo, J., Zhu, G., & Xiang, H. (2022). Artificial intelligent based day-ahead stock market profit forecasting. *Computers and Electrical Engineering*, *99*, 107837.

[14] Mahmoud, A., & Mohammed, A. (2020). A survey on deep learning for time-series forecasting. In *Machine learning and big data analytics paradigms: analysis, applications and challenges* (pp. 365-392). Cham: Springer International Publishing.

[15] Malik, A., Gupta, P., & Vijh, S. (2022, January). Towards a Stock Price Prediction on Time Series Data using Long-Short Term Memory Method. In *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 525-529). IEEE.

[16] Mehtab, S., & Sen, J. (2020). A time series analysis-based stock price prediction using machine learning and deep learning models. *International journal of business forecasting and marketing intelligence*, *6*(4), 272-335.

[17] Palash, P. S., & Dhurvey, P. (2024). Influence and Prediction of Sintered Aggregate Size Distribution on the Performance of Lightweight Alkali-Activated Concrete. *Archives for Technical Sciences, 2*(31), 49–56. https://doi.org/10.70102/afts.2024.1631.049

[18] Sachdeva, A., Jethwani, G., Manjunath, C., Balamurugan, M., & Krishna, A. V. (2019, March). An effective time series analysis for equity market prediction using deep learning model. In *2019 International Conference on Data Science and Communication (IconDSC)* (pp. 1-5). IEEE.

[19] Sen, J., & Mehtab, S. (2021). Design and analysis of robust deep learning models for stock price prediction.

[20] Sharma Y, Kumar A, Dubey V, Rai V. Stock Price Prediction Using LSTM. *In: 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT). Delhi, India; 2023.* p. 1-6.

[21] Sharma, P., & Jain, M. K. (2023, November). Stock market trends analysis using extreme gradient boosting (XGBoost). In *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)* (pp. 317-322). IEEE.

[22] Shen, J., & Shafiq, M. O. (2020). Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of big Data*, *7*(1), 66. https://doi.org/10.1186/s40537-020-00333-6.

[23] Sirimevan, N., Mamalgaha, I. G. U. H., Jayasekara, C., Mayuran, Y. S., & Jayawardena, C. (2019, December). Stock market prediction using machine learning techniques. In *2019 International Conference on Advancements in Computing (ICAC)* (pp. 192-197). IEEE.

[24] Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D., & Bhat, S. K. (2023). 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*, *11*(3), 94. https://doi.org/10.3390/ijfs11030094.

[25] Winata, A., Kumara, S., & Suhartono, D. (2021, October). Predicting stock market prices using time series sarima. In *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)* (Vol. 1, pp. 92-99). IEEE.

[26] Yeo, M., & Jiang, L. (2024). Thermal and Fluid Systems: Analysis, Design, and Optimization. *Association Journal of Interdisciplinary Technics in Engineering Mechanics*, *2*(1), 7-12.

[27] Zhang, Y., Ren, G., Liu, X., Gao, G., & Zhu, M. (2022). Ensemble learning-based modeling and short-term forecasting algorithm for time series with small sample. *Engineering Reports*, *4*(5), e12486.

[28] Zhao, A. B., & Cheng, T. (2022). Stock return prediction: stacking a variety of models. *Journal of Empirical Finance*, *67*, 288-317.

Authors Biography

K.V.K. Sasikanth, received M. Tech Degree in Information Technology from SRKR Engineering College, Bhimavaram in the year 2015 and pursuing Ph.D. in the field of Computer Science & Engineering from JNTUK, Kakinada. He is working as Assistant Professor in the Department of Information Technology, SRKR Engineering College, Bhimavaram, Andhra Pradesh, India. His research areas include Digital Image Processing, Computer Vision, and Machine Learning.

Dr.B. Sujatha is a Professor and Head of the Department of Computer Science and Engineering at Godavari Institute of Engineering & Technology (A), Rajahmundry. She holds a Ph.D. in Computer Science with a specialization in Digital Image Processing from the University of Mysore and an M.Tech. in Computer Science & Technology from Andhra University. With more than 23 years of teaching experience, she has guided numerous B.Tech., M.Tech., and Ph.D. scholars. Her research expertise spans Digital Image Processing, Machine Learning, and Medical Image Processing. She has published extensively in international journals and conferences, led funded research projects, and actively contributes to academic and research forums.

Dr.D. Haritha is a Professor in the Department of Computer Science & Engineering at JNTUK, Kakinada. She earned her Ph.D. in Computer Science and has been dedicated to teaching and research since 2003. Her courses include subjects like Unix Programming, Data Structures, Compiler Design, and Operating Systems. With over two decades of academic experience, Dr. Haritha focuses on Image Processing, tackling projects such as brain tumour segmentation and developing techniques for varying lighting conditions. She publishes her work in both national and international platforms and actively participates in workshops and seminars. Dr. Haritha has made significant contributions to her field through her research, teaching, and active participation in academic forums.