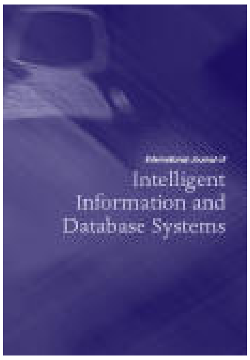


International Journal of Intelligent Information and Database Systems

ISSN online: 1751-5866 - ISSN print: 1751-5858
https://www.inderscience.com/ijiids

Updated deep long short-term memory with Namib beetle Henry optimisation for sentiment-based stock market prediction

Nital Adikane, V. Nirmalrani

DOI: 10.1504/IJIIDS.2024.10063202

Article History:

Received:	06 September 2023
Last revised:	04 December 2023
Accepted:	04 December 2023
Published online:	02 April 2024

Updated deep long short-term memory with Namib beetle Henry optimisation for sentiment-based stock market prediction

Nital Adikane* and V. Nirmalrani

Department of School of Computing,
Sathyabama Institute of Science and Technology,
Semmancheri, Chennai, Tamil Nadu 600119, India
Email: neetal.adikane9@gmail.com
Email: nirmalrani.it@sathyabama.ac.in
*Corresponding author

Abstract: Stock price prediction is a challenging and promising area of research due to the volatile nature of stock markets influenced by factors like investor sentiment and market rumours. Developing accurate prediction models is difficult, given the complexity of stock data. Long short-term memory (LSTM) models have proven effective in uncovering hidden patterns, enabling precise predictions. Therefore, in this research work, an innovative approach called updated deep LSTM (UDLSTM) combined with Namib beetle Henry optimisation (BH-UDLSTM) is proposed and applied to historical stock market and sentiment analysis data. The UDLSTM model enhances prediction performance, offering stability during training and increased data accuracy. By incorporating Namib beetle and Henry gas algorithms, BH-UDLSTM further improves prediction accuracy by striking a balance between exploration and exploitation. The evaluation against existing methods demonstrates that the proposed approach achieves a higher accuracy rate (92.45%) in stock price prediction compared to state-of-the-art techniques.

Keywords: stock price prediction; SPP; deep learning; DL; sentiment analysis; UDLSTM; Namib beetle algorithm; NBA; Henry gas solubility optimisation.

Reference to this paper should be made as follows: Adikane, N. and Nirmalrani, V. (2024) 'Updated deep long short-term memory with Namib beetle Henry optimisation for sentiment-based stock market prediction', *Int. J. Intelligent Information and Database Systems*, Vol. 16, No. 3, pp.316–344.

Biographical notes: Nital Adikane is working as an Assistant Professor in MITWPU Pune. She has more than 12 years of teaching experience and two years of research experience. She has published papers in national and international journals.

V. Nirmalrani is working as an Associate Professor in Sathyabama University Chennai. Her areas of interest are data analytics, data science and data mining. She has more than 20 years of teaching experience. She has published many papers in national and international journals.

1 Introduction

The stock market is a venue for the exchange, trading, and circulation of stocks. The stock market is viewed as a gauge of the monetary and economic health of a nation or territory (Lu et al., 2020). For a long time, scholars have been interested in making predictions movement of stock values. Therefore, strategies to anticipate stock values by looking at the layout over the former several decades could out to be very useful for determining stock price movements (Agrawal et al., 2019). These strategies would maximise profits and reduce losses. Due to their long-term unpredictability, stock value predictions are a difficult problem to solve (Yu and Yan, 2020; Mehtab and Sen, 2019). Common stock market issues that stockholders pay the greatest consideration to the erratic pattern of stock prices. Stock values are affected by several reasons, such as modifications to national legislation, changes in the domestic and international economies, current global events, etc. It's common for stock price changes to be nonlinear. It has always been important for economists to be able to predict changes in stock values. Investors' investment risk can be significantly decreased by developing an accurate stock price prediction (SPP) model. Such a projection enables depositors to comprise the anticipated stock value into their savings approach and aids investors in maximising their profits (Rezaei et al., 2021; Stoean et al., 2019; Chen et al., 2021).

Several algorithms and methods for predicting stock prices have been developed in recent years (Vijh et al., 2020). Machine learning (ML) aims to recognise the original purpose from which data are created and understand the linear and nonlinear approaches that exist in the data by finding novel outlines in historical data. One of the ML techniques that have recently undergone innovation is deep learning (DL) (Wu et al., 2021). Deep neural network (DNN) models are one of DL techniques, and they are well-liked since they can infer responses from unknowable facts and identify patterns in data. These DL methods offer a more effective means of resolving nonlinear issues as compared to traditional ML techniques. When it comes to forecasting nonlinear data in various time series prediction issues, the DL Method outperforms all other models (Li et al., 2021). Therefore, in this research manuscript, a novel DL method called BH-UDLSTM is proposed. The key contributions are enumerated below.

1.1 Contributions

- Sentiment analysis with UDLSTM tuned with Namib beetle's Henry optimisation is proposed for accurate prediction of stock price, i.e., rising (positive) or falling (negative).
- The UDLSTM model's main aim is to accomplish accurate time-series prediction for data with various distributions, which makes the proposed scheme more flexible to various kinds of SPPs.
- The hybridisation of the Namib beetle algorithm (NBA) with the Henry gas solubility (HGS) optimisation algorithm is performed to maximise the weight matrix and minimise the error threshold from the input to each of the gates of UDLSTM for improving the accuracy of SPP.

- The performance evaluation with the state-of-the-art is done by using metrics like mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), correlation, and accuracy. Also, a statistical validation is done for proving the effectiveness of the proposed model.

The rest of this research manuscript are designed as: Existing works are summarised in Section 2. The proposed SPP is described in Section 3. The outcomes of the proposed prediction technique are elaborated on and explained in Section 4. At last, Section 5 concludes the proposed technique.

2 Summary of previous works

Among the various research on stock market prediction and forecasting, few of them are enumerated below.

Several ML models are used in SPP. In 2022, Polamuri SR et al. presented a hybrid of reinforcement learning and Bayesian optimisation for predicting stock value. Stock-generative adversarial network (GAN), a DL approach built on GAN, is executed with a discriminator and a generator. A GAN-based hybrid prediction algorithm (GAN-HPA) predicted the stock market values by setting a proper selection of hyperparameters but this method does not contract with expanded forms of stock data. In order to achieve sentiment analysis based SPP, a novel case-based reasoning approach (CRA) was suggested by Chun and Jang in 2022 which predicts stock values such as Airbnb, Zoom, and Twitter, and also the prediction of the Dow-Jones-Industrial-Average (DJIA) that makes use of traders' expertise to choose comparable historical trends among the closest neighbours' generated from a conventional case of the intellectual machine. This interactive approach makes use of nearby examples' comparable time patterns. The process of selecting best patterns is considered a difficult task. In 2022, Albahli et al. presented a forecast of closing stock prices using 1D Dense Net, an auto-encoder, and STIs data from Yahoo Finance. The estimated STIs were loaded into the 1D Dense Net together with the Yahoo finance data after being input for the autoencoder for dimension minimisation. To anticipate closing stock levels for small-, average-, and long-term horizons, the 1D Dense Net's resultant features were then fed into the SoftMax layer as input. Prediction with a higher error rate is the major issue of this method. Another model of sentiment analysis based multi-source aggregated classification (MAC) for predicting the movement of stock values is implemented by Ma et al. in 2023. This method incorporates the target stock's market driven news sentiments and arithmetic features for prediction. The news is fit to real stock price rising and falling by pretrained embedding feature generator for achieving real market sentiment form news. In order to obtain the related companies' news effect on the target stock, a graph convolutional network is used. Although this method poses several drawbacks like: failure to detect whether the news is real, lack of understanding in computational language. These ML models still poses various challenges. Therefore, DL methods are preferred recently because of the automatic feature learning and complex data handling capability. In 2023, Gülmez developed a specialised deep long short-term memory (LSTM) network with artificial rabbits optimisation (ARO) known as LSTM-ARO to forecast stock prices, utilising DJIA index stocks as the dataset. In the comparative analysis, LSTM-ARO was assessed against an artificial neural network (ANN) model, three distinct LSTM models, and an

LSTM model optimised through genetic algorithm (GA). Nevertheless, the error rate of this model is slightly higher.

Table 1 Issues of existing approaches

Techniques	*Technique used*	*Problems*
Polamuri et al. (2022)	• GAN-HPA	• Does not contract with expanded forms of stock data
Chun and Jang (2022)	• CRA	• Selecting best features is difficult
Albahli et al. (2022)	• Dense Net	•
	• Auto encoder	• Higher error rate
Ma et al. (2023)	• MAC	• News is not detected properly
	•	• Poor computational language understanding
Wu et al. (2022)	• S-I-LSTM	• Less attention on training labelled data
Mndawe et al. (2022)	• LSTM	
	• AE	• Inaccurate prediction
Asgarian et al. (2023)	• GAN	• Slow and unstable training
	• Price sentiment GAN	
	• Price sentiment Wasserstein GAN	
Gülmez (2023)	• LSTM-ARO	• Increased error rate

One of the DL models known as SPP incorporating LSTM (S-I-LSTM) that takes into account several data sources including market sentiment, is used for presenting a prediction of stock values by Wu et al. in 2022. Using convolutional neural networks, the investors' sentiment index is generated. Combining the sentimental index with technical analysis, historical transaction records, and the stock price of the China Shanghai A-share market, LSTM is used to forecast the stock price. In this method, training labelled data receives less attention. Similarly, in DL, sentiment analysis-based prediction of stock price is done by Mndawe et al. in 2022. This technique uses a sentiment classifier with a South African foundation to derive sentiment from recent articles and messages. Four DL approaches with LSTM and autoencoder (AE) are employed for basic analyses. The tests mine and gather data from online newspapers, Twitter comments, and Yahoo Finance. This approach does not offer enough accuracy. Also, a sentiment-based stock prediction with the opinions of public is developed by Asgarian et al. in 2023. This approach used GAN with only optimised features of prizes, price sentiment GAN and price sentiment Wasserstein GAN with both social media and optimised features of prizes. Also, this method has used English and Persian social media datasets. Even so, the training process of the neural network is slow and unstable.

From the above review of traditional topologies, some of the major problems were listed in Table 1.

2.1 Problem statement

For overcoming these problems of traditional techniques in the prediction of stock market values like difficulty in selecting features, insufficient accuracy, high rate of error and slow and unstable training, the proposed framework is implemented by performing stock market analyses and sentiment analysis data from the databases using BH-UDLSTM. Six companies are selected for the proposed prediction of rising and falling stock prices over the recent five years. The prediction is done by a UDLSTM neural network and a hybrid optimisation algorithm for improving the accuracy and error rate.

3 Framework of proposed SPP

This section introduces a novel method for SPP for improving accuracy and minimising the error rate of prediction. Figure 1 demonstrates the structure of the proposed prediction of stock value movements using BH-UDLSTM.

In the proposed method of stock value prediction, the input data i.e. stock market data is taken from Yahoo Finance and sentiment analysis data is taken from the real-time Twitter Data. Pre-processing of input data is done by iterative bicluster-based least squares (Bi-ILS) for data cleaning and handling the missing value. From the preprocessed output, the most relevant features are selected using Kumar-Hassebrook distance and Motyka similarity. By using the selected features, the stock prices are predicted by the proposed UDLSTM optimised with the Namib beetle Henry optimisation (NBHO) algorithm. NBA and HGS together form the NBHO.

3.1 Input acquisition

Initially, two source of input data i.e. stock market data and sentiment analysis data are taken. The stock market data during the period of January 1, 2018, to January 16, 2023, is gathered from Yahoo Finance (https://finance.yahoo.com/quote/%5ENSEI/history/) using finance-python from six companies including 'Bharti_airtel', 'Infosys', 'Adani_enterprises', 'Hcl_technologies', 'Tata_steel', 'SBI'. The data includes low, high, close, and open values of a given day. The sentiment analysis data is taken from real-time Twitter data (https://twitter.com) using scrape-python. For each tweet from a specific period, this includes the timestamp and tweet text. The tweets are divided by day using their timestamps because predictions are made every day.

3.2 Preprocessing

The input data are preprocessed by using bi-ILS for data cleaning and handling the missing data. The input stock market data which has missing values and conditions is known as the goal data. The goal data d_t^T is expressed in equation (1) as,

$$d_t^T = (a \quad b) \tag{1}$$

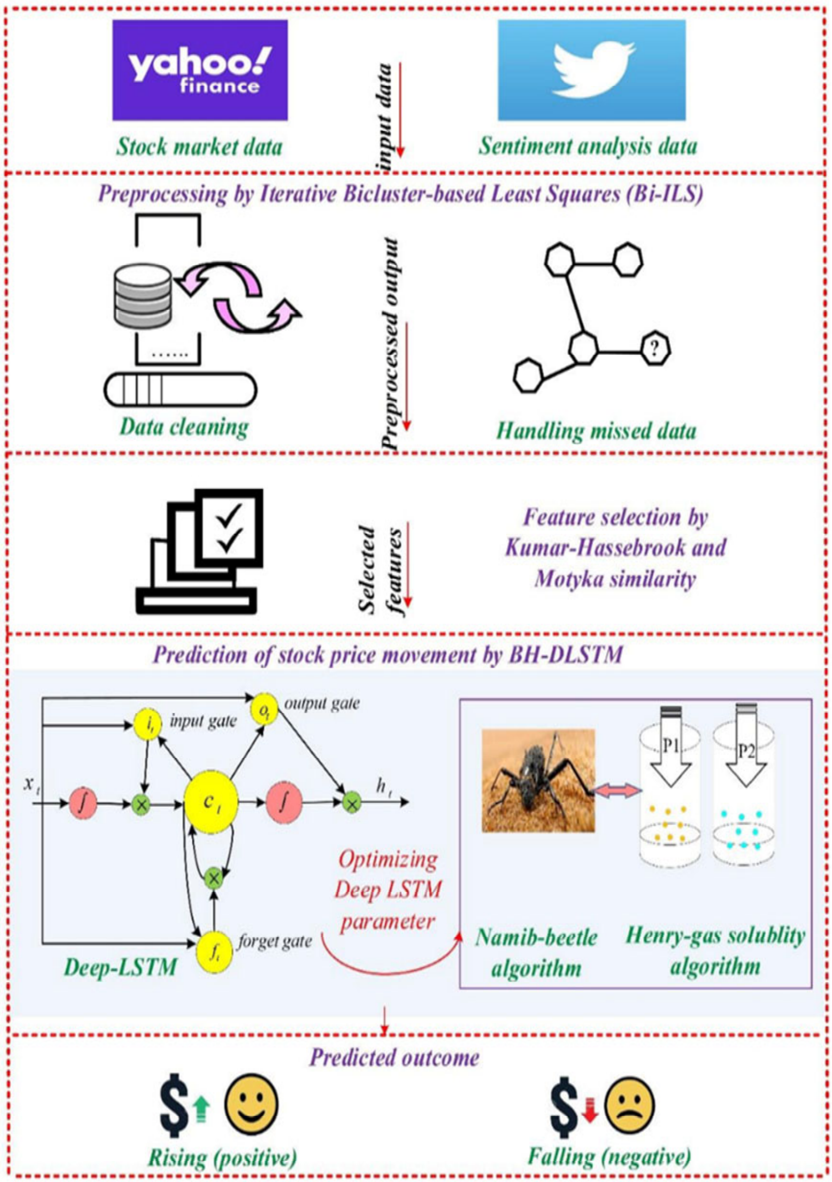
where the vector $n = 1 \times (c - m)$ consisting of the goal-data's b non-missing values and vector $a = 1 \times m$, i.e., $a = \textit{missing values}$ comprising gene-goal's m missing values.

Equation (2) is used to estimate the gene-goal's j^{th} missing value a_j.

$$a_j = X_j^T (Y^T)_j^+ b_j^T \tag{2}$$

where the j^{th} data's columns are represented as X_j^T, b_j^T is the j^{th} non-missing values, a matrix made up of the desired columns of a similar set of data is represented as Y_j^T and its pseudoinverse is $(Y^T)_j^+$. Thus, the preprocessed outcome from equation (2) is given to the feature selection phase (Soemartojo et al., 2022).

Figure 1 Structure of proposed BH-UDLSTM for SPP (see online version for colours)



3.3 Feature selection

This feature selection strategy uses two similarity metrics, such as Kumar-Hassebrook and Motyka (Karimi et al., 2019), which are restricted to a specific domain and are bounded in the range of [0, 1], to remove the relevant features and choose the most suitable sentiment and stock market features S_{KH} and S_M for prediction of the stock price with greater accuracy.

$$S_{KH} = \frac{\sum_{i=1}^{d} Ds_i Dc_i}{\sum_{i=1}^{d} Ds_i^2 + \sum_{i=1}^{d} Dc_i^2 - \sum_{i=1}^{d} Ds_i Dc_i} \tag{3}$$

$$S_M = \frac{\sum_{i=1}^{d} MIN\left(Ds_i \cdot Dc_i\right)}{\sum_{i=1}^{d} MIN\left(Ds_i + Dc_i\right)} \tag{4}$$

where Dc_i and Ds_i is the probability densities of the calculated and simulated values, d represents the total data. The highest-ranking features with the uppermost values are chosen as ideal features after computing Kumar-Hassebrook and Motyka. By using equations (3) and (4), the optimal features are selected which are listed below (Karimi et al., 2019).

features = [‘Open’, ‘High’,‘Low’, ‘Close’, ‘Volume’, ‘Likes’, ‘Quotes’,
‘Retweets’,’Activity’,‘Difference’, ‘Negative’, ‘Neutral’, ‘Positive’]

These features are given to the BH-UDLSTM for predicting the stock price.

3.4 Prediction of the stock price using UDLSTM

The selected features are given to the UDLSTM for predicting the stock value over the last five years of the stock market. The UDLSTM variant of the RNN family is intended to solve gradient vanishing difficulties. The UDLSTM layer has several series connections between the brain units, and irregular movement is employed to collect the data during the prediction process. UDLSTM uses memory cells and gate modules to control the memory at each input. An UDLSTM node is composed of dynamic gate structures like input, forget, and output gate frameworks. Figure 2 provides a detailed depiction of the architecture of the proposed network within the context of predicting stock prices. It elucidates the structural components and interconnections of the network as it engages in the decision-making process specifically designed for SPP.

A cell C_t, input gate I_t, forget gate F_t, output gate O_t, and the hidden condition H_t are all components of a UDLSTM cell and are all specified in the equations (5), (6), (7), (8), and (9) that follow.

$$I_t = \sigma\left(W_{xi}F_t + W_{hi}C_{t-1} + B_i\right) \tag{5}$$

$$F_t = \sigma\left(W_{xf} F_t + W_{hf} H_{t-1} + W_{cf} C_{t-1} + B_f\right) \tag{6}$$

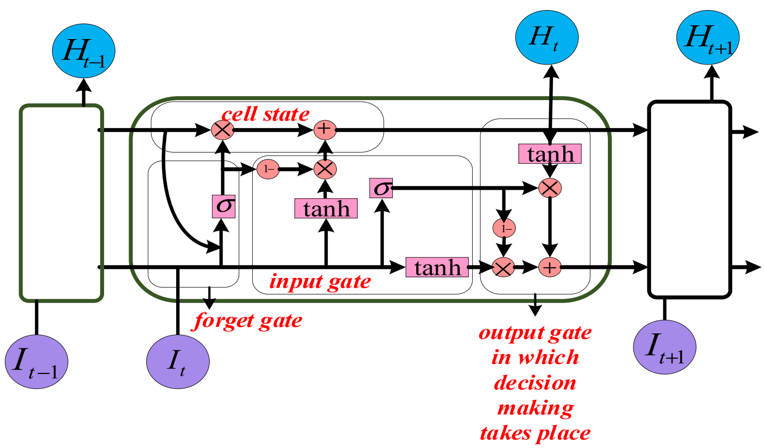
$$C_t = F_t \otimes C_{t-1} + I_t \otimes \tanh\left(W_{xc} F_t + W_{hc} H_{t-1} + B_c\right) \tag{7}$$

$$O_t = \sigma\left(W_{xo} F_t + W_{ho} H_{t-1} + W_{co} C_t + B_o\right) \tag{8}$$

$$H_t = O_t \otimes \tanh\left(C_t\right) \tag{9}$$

where B_β signifies the bias of i, f, c, o which is denoted as $\beta \in \{i, f, c, o\}$. The weight matrix is represented as W_β, $\otimes$, symbolises the factor wise product, $\sigma(x)$ is the sigmoid role which has the value of $\dfrac{1}{(1+e^{-x})}$.

Figure 2 Architecture of proposed network (see online version for colours)



The following changes are made to the LSTM in order to attain accurate time-series prediction for stock data with various distributions (Wang et al., 2022).

In contrast to the traditional LSTM, the proposed model connects the input and the forget gates. It is estimated that only the portions of the cell conditions that new elements is added is capable of being altered. Furthermore, the effect of the cell's previous state under the identical conditions C_t at present time t is simultaneously incorporated. The updated version of deep LSTM enhances the prediction accuracy by an admirable training performance and improves the impact of the current input I_t and it is not assured by the hidden condition H_t, but also retains few attributes of the present input. Equations (10), (11), and (12) compute the entire final output O_r.

$$F_t = \sigma\left(W_f \cdot \left[I_t, O_{t-1}\right] + B_f\right) \tag{10}$$

$$C_t = \tanh\left(W_c \cdot \left[I_t, O_{t-1}\right] + B_c\right) \tag{11}$$

$$O_t = O_{t-1} \times \tanh\left(C_{t-1}\right) + \left(1 - O_{t-1}\right) \times C_t \tag{12}$$

The model accepts an input sequence $I = \{I^{(t-l_i)}, I^{(t-l_x+1)}, \ldots, I^{(t-1)}\}$ for every time instant t and produces the predicted result $P(t)$. Next, an error value $E_t = |P(t) - I(t)|$ is calculated by comparing the projected $P(t)$ and actual I_t values. The SPP method forecasts rising and falling stock values if the error rate exceeds a set threshold T_h, i.e., $E_t > T_h$, or when it is more than that number. The performance of SPP is therefore greatly influenced by the error threshold T_h, a crucial system parameter.

Table 2 illustrates the optimal settings of the BH-UDLSTM model. The deep LSTM network is configured with three hidden layers using UDLSTM architecture, each layer containing 120 units. The learning rate is set to 0.000625, and training occurs in batches of 16 samples. A dropout rate of 0.2 is applied to prevent overfitting, and input sequences have a length of 20-time steps. These parameter settings collectively define the architecture and training dynamics of the model.

Table 2 Optimal parameter settings

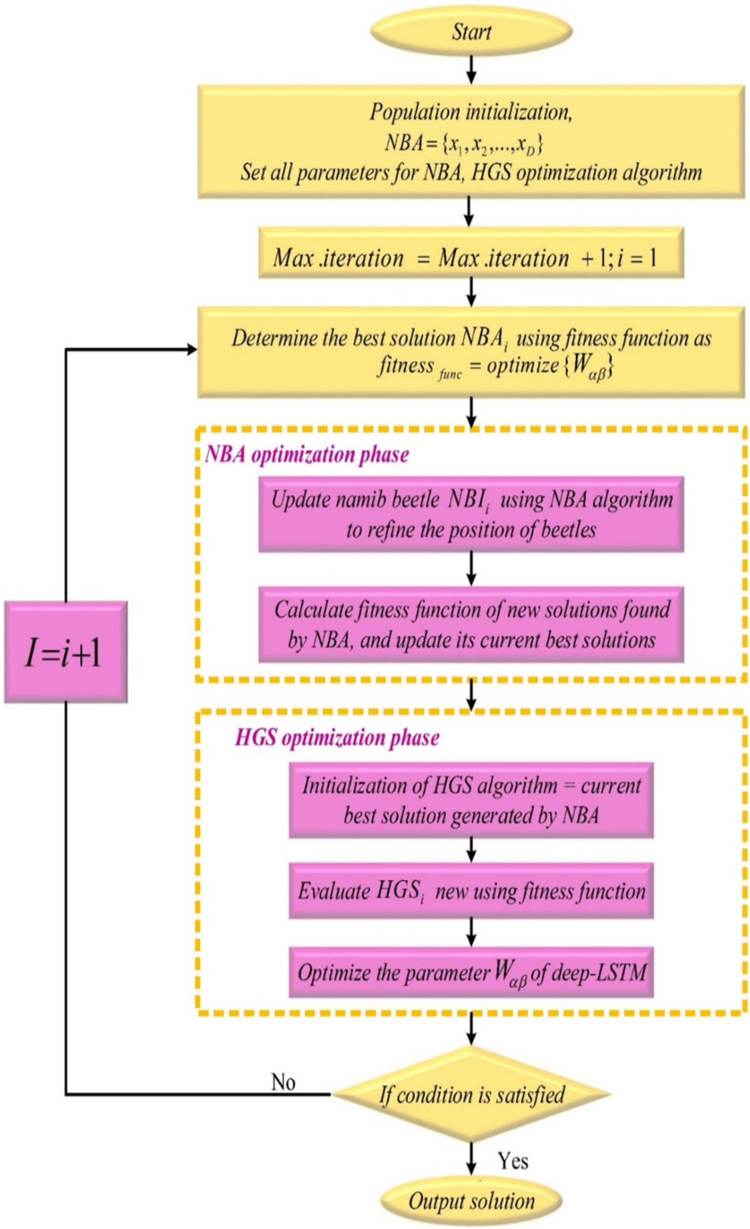
Parameters	*Values*
Number of UDLSTM hidden layers	3
Number of units per layer	120
Learning rate	0.000625
Batch sizes	16
Dropout rates	0.2
Sequence lengths	20

Backpropagation is the foundation of the UDLSTM training stage, which aims to minimise loss. Propagation occurs during training in the direction of the final concealed nodes and backward until the predetermined iterations have been attained. For prediction, this propagation moves to the SoftMax layer. The soft sign function predicts whether the stock price is rising (positive) or falling (negative) by producing a real number between –1 and 1 (Steven Eyobu and Han, 2018; Li et al., 2021).

The decision-making process of proposed model is intricately tied to the dynamic shifts in the current level of the index price. Choices and actions within this process are contingent upon the nuanced interplay of factors associated with the rise or fall of the index price, reflecting a sophisticated approach to interpreting and responding to market conditions. The dynamics of the securities market are intricately linked to the psychology and behaviour of investors. Within the proposed BH-DLSTM model, the pivotal element shaping the analysis of stock reviews lies in sentiment analysis, particularly focusing on the rising and falling sentiments. The innovation proposed here is the integration of a sentiment index, which serves to encapsulate the holistic emotional tendencies of investors. This approach goes beyond a mere examination of stock-related factors and delves into the broader emotional landscape influencing investment decisions. Consequently, the overarching objective of this proposed framework is to elevate our understanding of stock price trends, enabling more accurate predictions. By doing so, it empowers investors to make well-informed decisions, thereby reducing inherent risks and optimising potential returns in the market. The decision-making for the prediction occurs in the output gate, which determines what information from the cell state should be passed on as the final output. The output layer generates the prediction based on this decision.

By further reducing error threshold T_h, the hybrid NBHO method optimises the weight matrix w_β of UDLSTM to increase SPP accuracy.

Figure 3 Flow of proposed NBHO algorithm (see online version for colours)



3.5 Optimising the weight parameter of UDLSTM using hybrid NBHO

In this section, the NBA and HGS algorithm are hybridised and it is proposed for maximising the weight matrix and minimising the error threshold of UDLSTM in order and achieve accurate prediction results. A perfect stability between exploration and exploitation in the search space and good performance to avoid local optimum is the advantage of using this hybridised NBHO algorithm. The flowchart of hybrid NBHO is depicted in Figure 3.

- NBA

  This algorithm was developed using information about Namib beetle water search behaviour and the location of ideal mounds for collecting water. NBA discovers the solutions for exploitation and exploration. The NBA has weak exploitation ability. By hybrid the NBA with the Henry gas optimisation Algorithm, the ability of exploitation is enhanced (Chahardoli et al., 2022).

- HGS

  Henry's gas law's behaviour serves as the basis for this optimisation algorithm. In contrast to other rival metaheuristic algorithms, the control of the phases of exploration and exploitation is preserved by using HGS optimisation to prevent local optima (Hashim et al., 2019).

- A hybrid of Namib beetle with HGS optimisation algorithm

  Before starting the iteration, it is necessary to initialise the Namib beetle as $NBA_i = \{x_1, x_2, \ldots, x_D\}$ the entity that is encoded in the D dimensions. According to equation (13) below, the population of beetles is randomly initialised.

$$NBA_{i,j} = L_l + (U_l - L_l) \times RAND(0,1) \tag{13}$$

where $NBA_{i,j}$ is the j^{th} element related to the beetle i. L_l and U_l stands for the upper and lower bound, represents a random number.

The objective function of hybrid NBHO is optimising the weight matrix and minimising the error threshold of UDLSTM which is described in equation (14).

$$fitness = \begin{Bmatrix} \max .weight\ matrix\left(W_\beta\right) \\ \min .error\ threshold\left(T_h\right) \end{Bmatrix} \tag{14}$$

Equations (15) and (16) are used to determine the humidity of vectors and the number of novel vectors C_i.

$$\rho = p_{MAX} - p_0 \times \left(1 - \frac{I_{teration}}{MAX\_I_{teration}}\right) \times RAND \tag{15}$$

$$C_i = C_{MAX} \times \sin\left(\frac{f\left(NBA_i\right) - f_{MIN}}{f_{MAX} - f_{MIN}} \times \frac{\pi}{2}\right) \tag{16}$$

where p_0 is the initial coefficient of humidity rise, p_{MAX} is the maximum coefficient of humidity rise, f_{MIN} and f_{MAX} signifies the population beetle's lowest and highest abilities respectively. C_{MAX} reflects the maximum capability of the number of beetles.

Finally, the following equations (17) and (18) yields the NBA's final solution.

$$NBA_j^{NEW} = NBA_j^{OLD} + H_{umidity} \times \left(NBA_i - NBA_j^{OLD}\right) + levy \tag{17}$$

A beetle seeking to move is presently in position NBA_j^{OLD} and further moves towards position NBA_j^{NEW} according to equation (17). NBA_i represents the location of the beetle that lures additional beetles.

$$NBA_i^{NEW} = NBA_i^{OLD} + RAND \times (NBA^* - NBA') + levy \quad (18)$$

where NBA^* is the location with the maximum humidity, position refers to the water's gravitational pull on the atmosphere and bodies of beetles and *levy* represents a random vector for beetle movement.

The current optimal solution discovered by NBA is used to determine the initialisation of the HGS algorithm which is obtained in equation (19). HGS directs the algorithm to discover the best solution more attentively in a local area to ensure exploitation and exploration.

$$NBA_i^{NEW} = NBA_{MIN} + RAND \times (NBA_{MAX} - NBA_{MIN}) \quad (19)$$

where *rand* the random number ranges from 0 to1. NBA_{MAX} and NBA_{MIN} denotes the maximum and the minimum bounds respectively.

3.6 Exploitation and exploration

Equation (20) is used to update the location.

$$\begin{aligned} NBA_i^{NEW}(t+1) = NBA_{i,j}(t) + F \cdot r \cdot \gamma \cdot \left(W_\beta \cdot NBA_{i,BEST}(t) - T_h \cdot NBA_{i,j}(t)\right) \\ + F \cdot r \cdot \gamma \cdot \left(W_\beta \cdot NBA_{BEST(t)} - T_h \cdot NBA_{i,j}(t)\right) \end{aligned} \quad (20)$$

where $NBA_{i,BEST}$ and NBA_{BEST} are used for controlling the exploration and exploitation, γ is the ability of gas, F changes the direction of individuals, r denotes a random number, t represents the iteration time. By using equation (17), the w_β and T_h of the UDLSTM is optimised.

3.7 Position updating of worst individuals

Equation (21) is utilised to update the ranking of the worst individuals.

$$W_{i,j} = W_{MIN(i,j)} + rand \cdot \left(W_{MAX(i,j)} - W_{MIN(i,j)}\right) \quad (21)$$

where $W_{i,j}$, $W_{MIN(i,j)}$, $W_{MAX(i,j)}$ represents the positions, minimum position and maximum position of gas. By using equation 18, the error rate is minimised. If this condition is satisfied, stop the process. Otherwise, repeat the step until the halting criteria is achieved (Agbaje et al., 2019).

Thus, the proposed BH-UDLSTM effectively predicts the stock price movement during the period of five years (January 1, 2018, to January 16, 2023) of various companies by optimising the weight matrix and also minimising the error rate with improved accuracy. UDLSTM networks with multiple layers learns complex representations and hierarchies in the data, allowing them to capture intricate patterns and make more accurate predictions. LSTM's memory cells and gate mechanisms enable them to remember important information over long periods, making them well-suited for tasks requiring context and temporal dependencies. The proposed hybrid optimisation algorithm efficiently updates the model's parameters during training. This helps in faster convergence and better generalisation. The UDLSTM networks often include dropout or other regularisation techniques, which prevent overfitting and lead to better generalisation on unseen data. UDLSTM networks are specifically designed to handle sequences with long-range dependencies, making them well-suited for tasks where past information is crucial for accurate SPPs. This capability allows them to outperform traditional feedforward neural networks in tasks involving time series, natural language processing, and other sequential data. The proposed neural networks' memory cells and gate mechanisms enable them to capture and retain relevant information from past time steps, providing the model with context and memory of the sequence history. This ability allows the model to make informed decisions based on past observations.

4 Outcomes of BH-UDLSTM

In this section, the outcomes of the proposed UDLSTM with NBHO are analysed. The proposed model is compared with several existing techniques like Deep generative adversarial network(D-GAN) (Kumar et al., 2022), bidirectional LSTM, and LSTM with AE (LSTM-AE) (Mndawe et al., 2022), multi-model GAN is a hybrid with prediction approach (MMGAN-HPA) (Polamuri et al., 2022) and DNN (Gupta et al., 2022). The PYTHON simulation of the overall process of stock prediction and sentiment analyses ML model is done using the TensorFlow library.

4.1 Evaluation metrics

The evaluation of the proposed technique is assessed using some metrics like MSE, RMSE, MAPE, accuracy (A), and correlation (C) in equations (22)–(26).

$$MSE = \frac{1}{t}\sum_{k=1}^{t}\left(y_k - \hat{y}_k\right)^2 \tag{22}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{t}\left(y_k - \hat{y}_k\right)^2}{t}} \tag{23}$$

$$MAPE = \frac{1}{t}\sum_{k=1}^{t}\frac{\left|y_k - \hat{y}_k\right|}{y_k} \tag{24}$$

$$C = \frac{\sum_{t=1}^{n}\left(y_k - \hat{y}_k\right)\left(y'_t - \hat{y}'_k\right)}{\sqrt{\sum_{t=1}^{n}\left(y_k - \hat{y}_k\right)^2\left(y'_t - \hat{y}'_k\right)^2}} \tag{25}$$

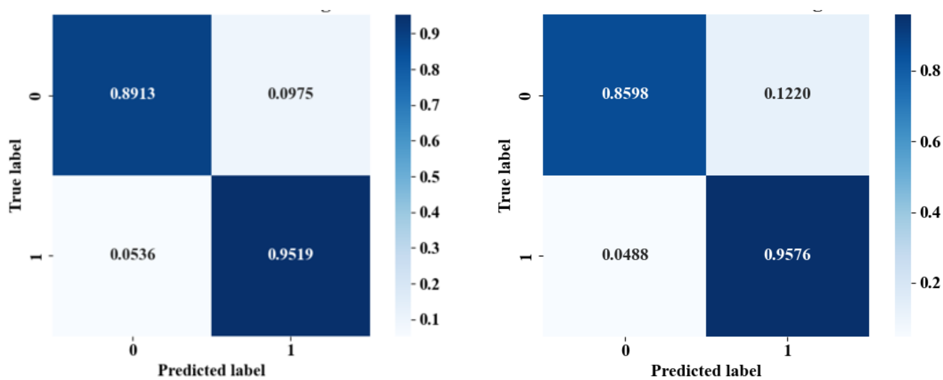
$$A = \frac{\text{No. of correct prediction rising falling}}{\text{Total no. of prediction}} \tag{26}$$

where t represents total samples, y_k and $\hat{y}_k$ denote the real and predicted price of the day k.

4.2 *Results of sentiment analysis*

The stock market prediction results by sentiment analysis of six companies including 'Bharti_airtel', 'Infosys', 'Adani_enterprises', 'Hcl_technologies', 'Tata_steel', 'SBI' using BH-UDLSTM is given below.

Figure 4 Confusion matrix of 3(a) training and 3(b) testing data (see online version for colours)



(a) (b)

Table 3 Reset indices and drop unnecessary columns for likes, retweets, quotes, negative, positive and label

Tweets	*Likes*	*Retweets*	*Quotes*	*Negative*	*Positive*	*Label*
0	0.0	0.0	0.0	0.043852	0.048182	0
1	0.0	0.0	0.0	0.078235	0.062213	0
2	2.0	1.0	0.0	0.093096	0.049736	0
…	…	…	…	..	…	…
7050	36.0	11.0	2.0	0.102514	0.099030	1
7051	23.0	16.0	3.0	0.094142	0.128177	0

Figures 4(a) and 4(b) show the training and validation confusion matrix. The true label and predicted label are represented in columns and rows. The stock prediction results of the sentiment analysis and stock market data are based on the confusion matrix's output.

The process of resetting indices and removing extra columns before the feature selection process, is depicted in Tables 3, 4 and 5. The pair plot of Figure 5 shows that the characteristics of positive tweets have a significant influence on the rising and falling of stock prices.

Table 4 Reset indices and drop unnecessary columns for data time, open, high, low, close and volume

Tweets	*Datetime*	*Open*	*High*	*Low*	*Close*	*Volume*
0	2018-01-01	473.682852	480.016471	468.687340	470.917480	4719333
1	2018-01-08	476.492854	476.492854	458.696295	460.614227	12061390
2	2018-01-09	462.086138	464.806917	450.221775	454.771271	7402279
…	…	…	…	…	…	…
7050	2018-01-11	596.450012	599.500000	592.250000	596.700012	7419630
7051	2018-01-12	597.900024	597.900024	590.099976	594.900024	8183715

Figure 5 Features of positivity of tweets (see online version for colours)

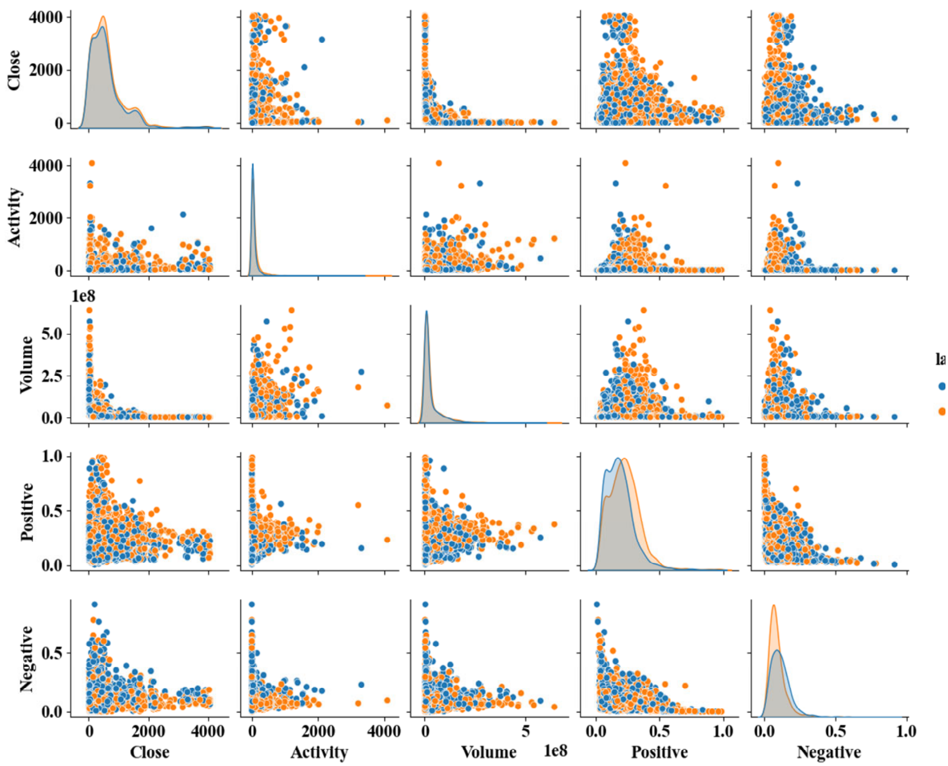
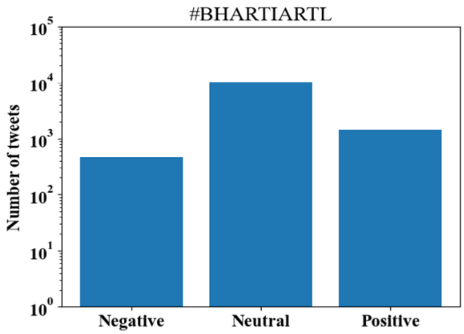


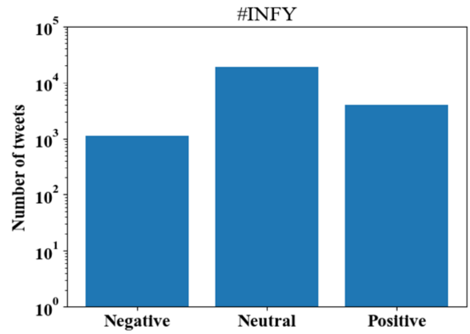
Table 5 Reset indices and drop unnecessary columns for difference and activity

Tweets	*Difference*	*Activity*
0	Nan	0.0
1	–10.303253	0.0
2	–5.842957	3.0
…	…	…
7,050	1.549988	49.0
7,051	–1.799988	42.0

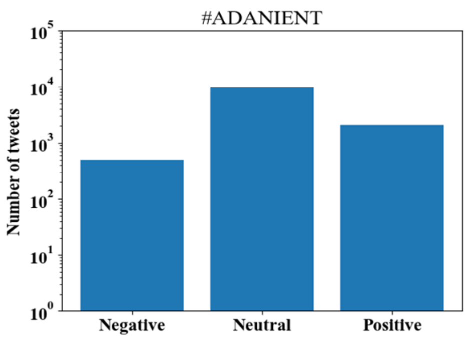
Figure 6 SPP of (a) Bharti_airtel (b) Infosys (c) Adani_enterprises (d) Hcl_technologies (e) Tata_steel (f) SBI using sentiment analysis (see online version for colours)



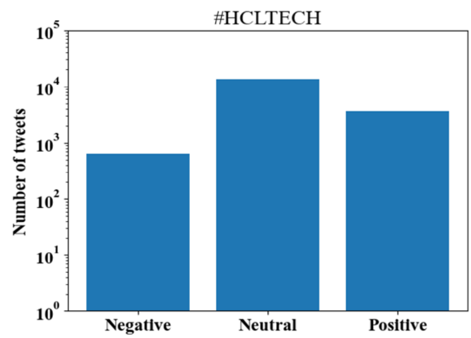
(a)



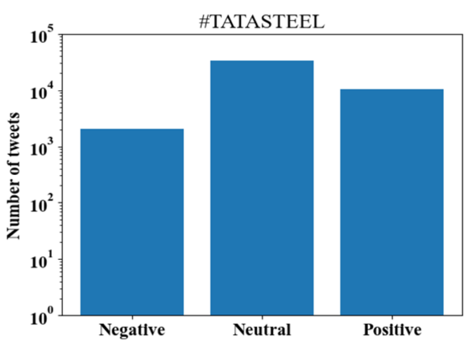
(b)



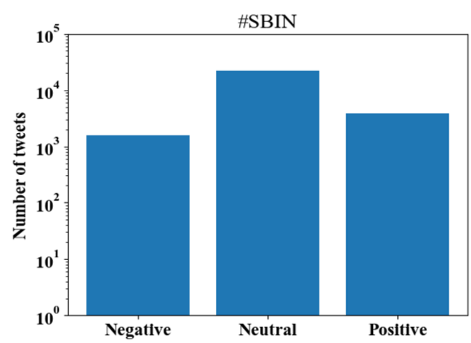
(c)



(d)



(e)



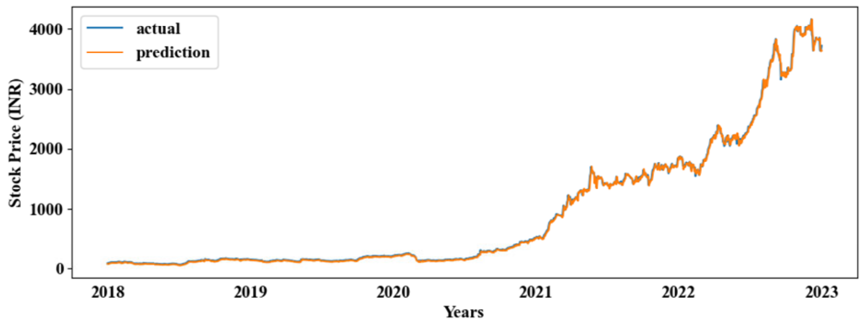
(f)

Figures 6(a), 6(b), 6(c), 6(d), 6(e) and 6(f) depict the sentiment analysis of Bharti_airtel, Infosys, Adani_enterprises, Hcl_technologies, Tata_steel and SBI using real-time twitter data. The positive, negative, and normal prices of stocks are predicted using information from tweets. 12,367 tweets from Bharti_airtel, 24,698 tweets from Infosys, 12,336 tweets from Adani_enterprises, 17,906 tweets from Hcl_technologies, 47,163 tweets from Tata_steel and 27,717 tweets from SBI are taken for sentiment analysis. From Figure 6, prediction results show that neutral tweets are more than positive and negative tweets.

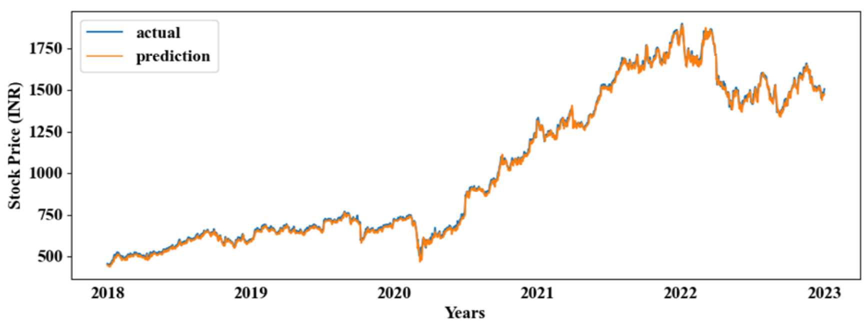
4.3 Results of stock market data

The stock prediction results based on the stock market data of six companies are presented below.

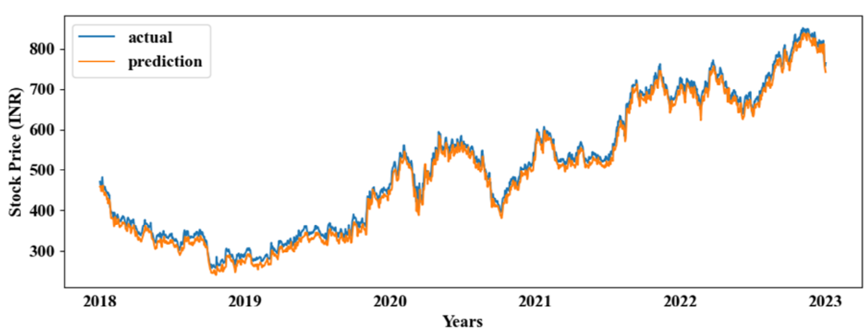
Figure 7 SPP of (a) Bharti_airtel (b) Infosys (c) Adani_enterprises (d) Hcl_technologies (e) Tata_steel (f) SBI using stock market data (see online version for colours)



(a)

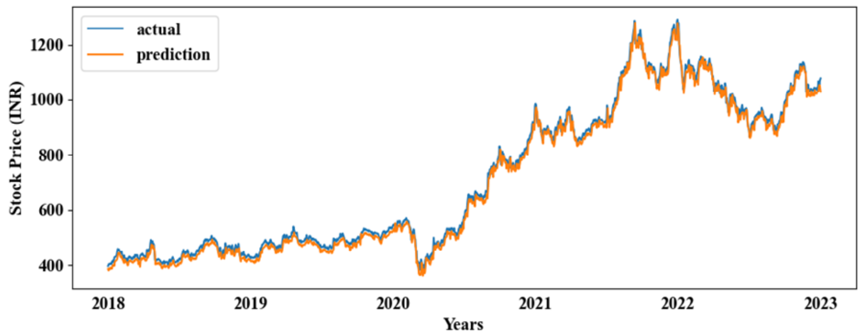


(b)



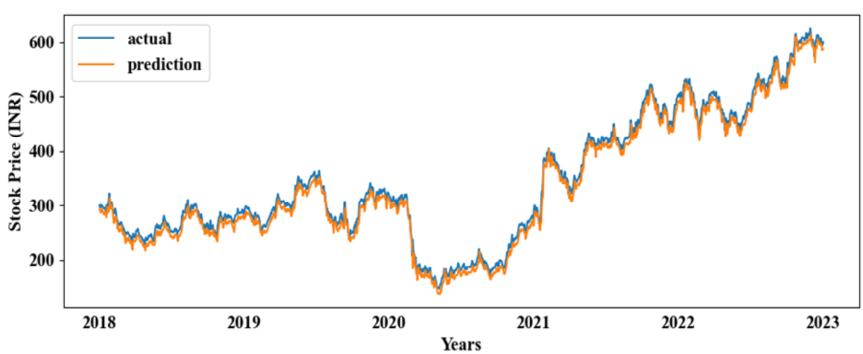
(c)

Figure 7 SPP of (a) Bharti_airtel (b) Infosys (c) Adani_enterprises (d) Hcl_technologies (e) Tata_steel (f) SBI using stock market data (continued) (see online version for colours)



(d)

(e)

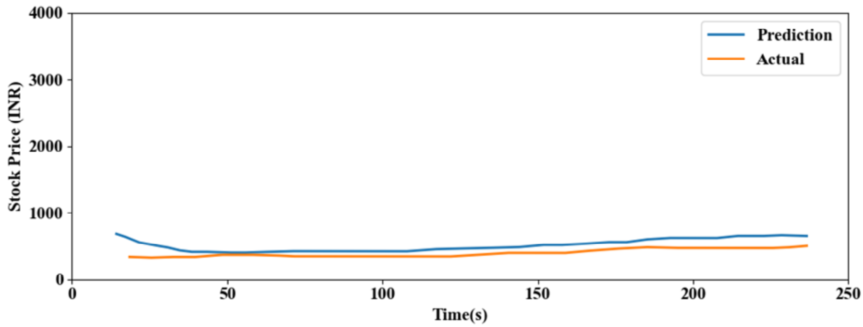


(f)

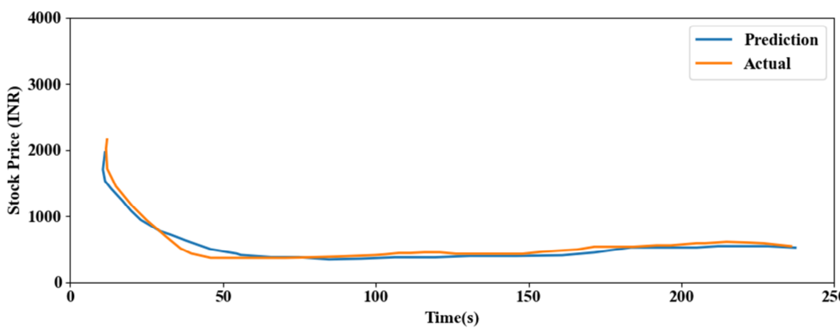
Figures 7(a), 7(b), 7(c), 7(d), 7(e) and 7(f) display the analysis by stock market data of Bharti_airtel, Infosys, Adani_enterprises, Hcl_technologies, Tata_steel and SBI using Yahoo Finance. Actual and predicted values are compared in this instance for proving the better performance of BH-UDLSTM. The blue curve displays the actual price for a given input index, while the orange curve displays the predicted price. Year-wise prediction rate is depicted for finding the rising and falling of the stock market. As demonstrated in

Figure 7, the prediction error is fairly minimum and impressive as the projected values are frequently near actual values for all six companies from 2018 to 2023 and move in the direction of the expected values.

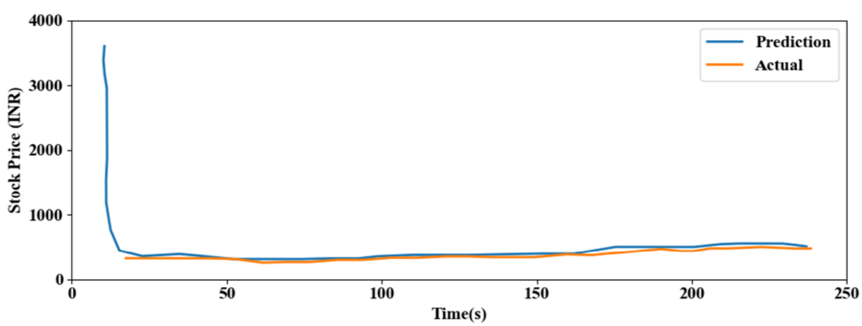
Figure 8 SPP of (a) Bharti_airtel (b) Infosys (c) Adani_enterprises (d) Hcl_technologies (e) Tata_steel (f) SBI with respect to time (see online version for colours)



(a)

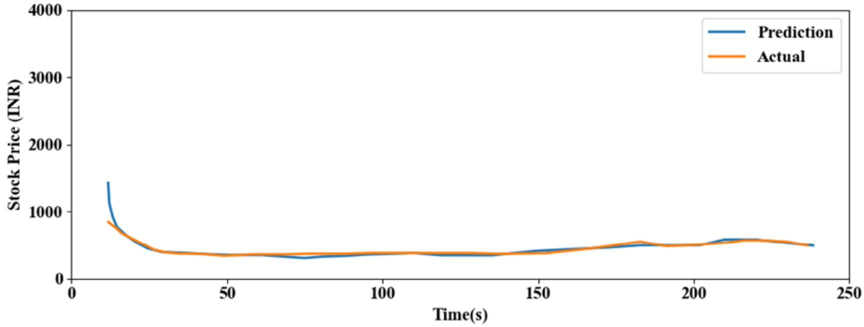


(b)

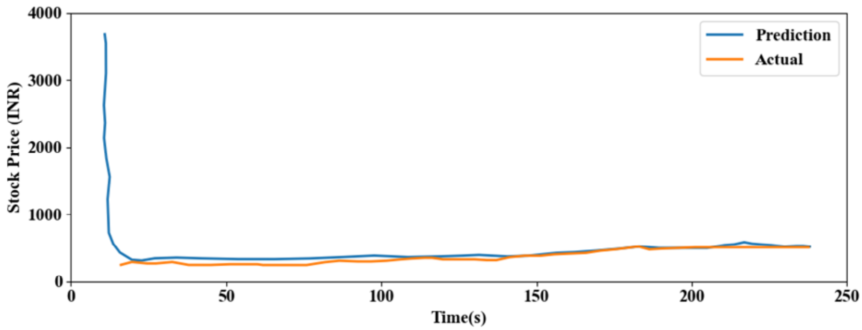


(c)

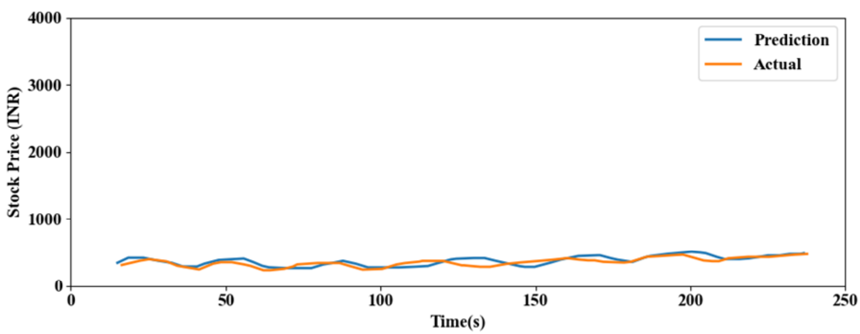
Figure 8 SPP of (a) Bharti_airtel (b) Infosys (c) Adani_enterprises (d) Hcl_technologies (e) Tata_steel (f) SBI with respect to time (continued) (see online version for colours)



(d)



(e)



(f)

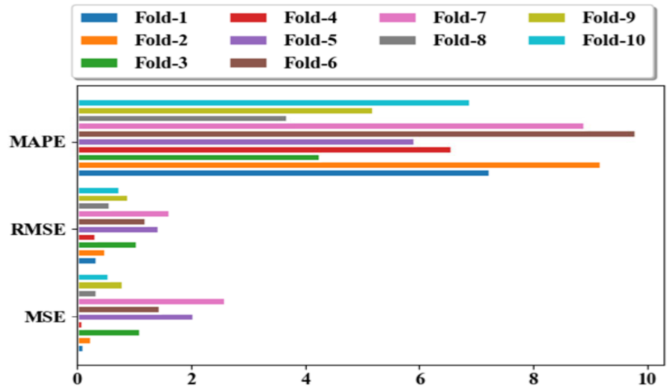
This research initiative involves the integration of BH-UDLSTM with sentiment analysis for the purpose of predicting whether stock prices will rise or fall. The outcomes of this predictive model are visually represented in Figure 8, offering a clear illustration of the forecasted stock prices over time. The utilisation of BH-UDLSTM, coupled with sentiment analysis, enhances the accuracy of these predictions, contributing valuable insights for investors and stakeholders in navigating the dynamic landscape of stock markets. In the given Figure 8, the horizontal axis represents the timeline, and the vertical axis represents the corresponding stock prices. Remarkably, the predicted stock prices demonstrate a strong resemblance to the actual stock prices, signifying a minimal error rate in the predictions. These predictions are derived from the stock performance of 8(a) Bharti_airtel, 8(b) Infosys, 8(c) Adani_enterprises, 8(d) Hcl_technologies, 8(e) Tata_steel, and 8(f) SBI. The depiction of stock prices is an outcome of the model's predictive capabilities. The process involves a transformation or post-processing step wherein sentiment scores, derived from the sentiment analysis component of the proposed model, are converted into predictions of stock price movements. The integration of stock prices in Figure 8 is justified as it provides a holistic representation of the model's performance, offering insights into both sentiment analysis and the subsequent impact on stock prices. The BH-UDLSTM model is designed, allowing it to capture intricate patterns and dependencies in sentiment data, ultimately influencing SPPs. The goal of enhancing the accuracy of SPP is achieved by the hybrid NBHO algorithm. This hybridisation maximises the weight matrix and minimises the error threshold from the input to each gate of UDLSTM, leading to improved accuracy in the SPP task with a minimum computational duration. Incorporating BH-UDLSTM with a sentiment index in the proposed approach consistently yields noteworthy results, achieving superior accuracy, minimal time offset, and closely aligned predictive values in SPPs over time.

4.4 Cross validation and statistical analysis

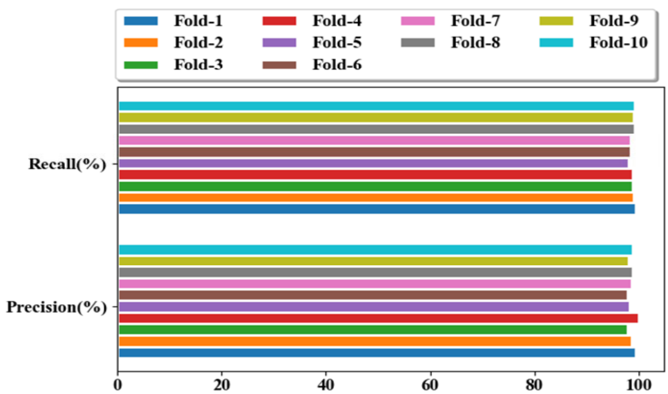
Nested cross-validation (Shen et al., 2021) is employed for the meticulous tuning of hyperparameters and a thorough evaluation of a model's performance. The approach includes an outer loop with ten folds, utilising the training subset for hyperparameter optimisation and the testing subset for estimating the prediction model's performance. BH-UDLSTM underwent evaluation using a nested cross-validation approach. This involved ten outer folds and ten inner folds for hyperparameter tuning in each iteration. By integrating ten-fold cross-validation into both the outer and inner loops, this nested approach ensures a comprehensive evaluation of the model's performance and effective hyperparameter tuning across diverse subsets of the data. This strategy is instrumental in preventing the model from overfitting to specific datasets or hyperparameter settings, thereby enhancing the reliability of the model's generalisation performance on unseen data. The careful integration of nested cross-validation provides a robust framework for model development and evaluation in ML.

Table 6 provides performance metrics for the proposed model evaluated using the average ten-fold cross-validation on both outer and inner loops. The 'mean value' row at the bottom of Table 6 provides the average of each metric over all ten-fold cross-validation iterations, offering a consolidated view of the model's overall performance. Figure 9 depicts the (a) MSE, RMSE, MAPE (b) recall, precision (c) correlation and accuracy evaluation of the proposed model with ten-folds and the mean value is determined. This methodology ensures robust hyperparameter tuning and a comprehensive assessment of the model's predictive capabilities across diverse subsets of the data. Table 7 provides statistical measures like the mean, minimum, maximum, and standard deviation for six different companies (Adani_enterprises, Bharti_airtel, Hcl_technologies, Infosys, SBI and Tata_steel) based on a dataset containing 1,246 data points for each company. These statistical measures can be used to gain insights into the distribution and characteristics of the data for each company.

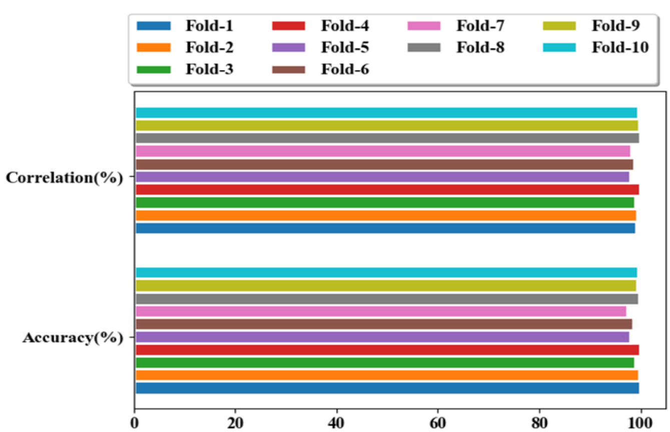
Figure 9 Evaluation of (a) MSE, RMSE, MAPE (b) recall, precision and (c) correlation and accuracy for average ten-fold validation on both outer and inner loops (see online version for colours)



(a)



(b)



(c)

Tables 8 and 9 provide the results of chi-square statistical test and Pearson's correlation coefficient statistical test comparison performed on different companies (Adani_enterprises, Bharti_airtel, Hcl_technologies, Infosys, SBI, and Tata_steel) to test for independence among two categorical variables. The chi-square test statistic quantifies the difference between the observed and expected frequencies in a contingency Table 8. The degree of freedom depends on the number of categories or levels of the variables involved in the test.

The p-value indicates the level of statistical significance, and it is below the chosen significance level (0.05), therefore reject the null hypothesis of independence is rejected. This means that there is a statistically significant association between the variables being tested for the six companies. The correlation coefficients in Table 9 measure the direction and strength of a linear relationship among two variables, and p-values indicate the statistical significance of the correlation. For example, for Adani_enterprises and Bharti_airtel, the correlation coefficient is 0.99956, indicating a very strong positive linear relationship between the stock price movements of these two companies. The high correlation coefficients close to 1 indicate strong positive linear relationships, while the low p-values suggest that these correlations are statistically significant.

Table 6 Performance metrices comparison for average 10-fold cross-validation on both outer and inner loops

K-fold	*Precision (%)*	*Accuracy (%)*	*Recall (%)*	*MSE*	*RMSE*	*MAPE*	*Correlation accuracy (%)*
1-fold	99.4	99.9	99.4	0.12	0.332	7.2	99.12
2-fold	98.7	99.8	99.03	0.23	0.5	9.19	99.2
3-fold	97.9	99	98.8	1.1	1.05	4.245	99.02
4-fold	99.9	99.9	98.8	0.1	0.32	6.57	99.8
5-fold	98.2	97.96	98	2.04	1.43	5.92	98.03
6-fold	98	98.6	98.5	1.4	1.2	9.79	98.8
7-fold	98.6	97.	98.4	2.6	1.6125	8.89125	98.1
8-fold	98.8	99.7	99.3	0.3	0.6	3.68	99.9
9-fold	98	99.2	99	0.8	0.89443	5.2	99.8
10-fold	99	99.5	99.2	0.6	0.742	6.9	99.5
Average mean value	98.6	99.1	98.83	0.9	0.864	6.76	99.14

Table 7 Statistical analysis outcomes of six companies' stock price data

Index	*Count*	*Mean*	*Min*	*Max*	*Standard deviation*
Adani_enterprises	1,246	917.814594	66.427170	4165.299805	1,092.758047
Bharti_airtel	1,246	508.132894	257.815002	851.200012	163.748634
Hcl_technologies	1,246	718.609771	372.515594	1,291.272949	270.284820
Infosys	1,246	1,029.905001	445.934540	1,898.427979	443.552055
SBI	1,246	344.935913	147.173218	625.500000	118.752087
Tata_steel	1,246	38.374763	10.753931	120.449997	28.032194

Table 8 Chi-square statistical test comparison on six companies

	Adani_enterprises	*Bharti_airtel*	*Hcl_technologies*	*Infosys*	*SBI*	*Tata_steel*
Chi_square	1,513,889.99	1,456,573.99	1,522,611.99	1,520,120	1,402,061.50	1,472,858.52
p value	0.0242	0.0382	0.0263	0.0250	0.0420	0.0310
Degrees of freedom	1,512,675	1,456,065	1,521,510	1,518,946	1,401,725	1,466,325

Table 9 Pearson's correlation coefficient statistical test comparison on six companies

	Adani_enterprises	*Bharti_airtel*	*Hcl_technologies*	*Infosys*	*SBI*	*Tata_steel*
Correlation	0.99956	0.9982	0.99901	0.99945	0.998554	0.99941681
p value	0.00043	0.00174	0.00098	0.00054	0.0014453	0.00058

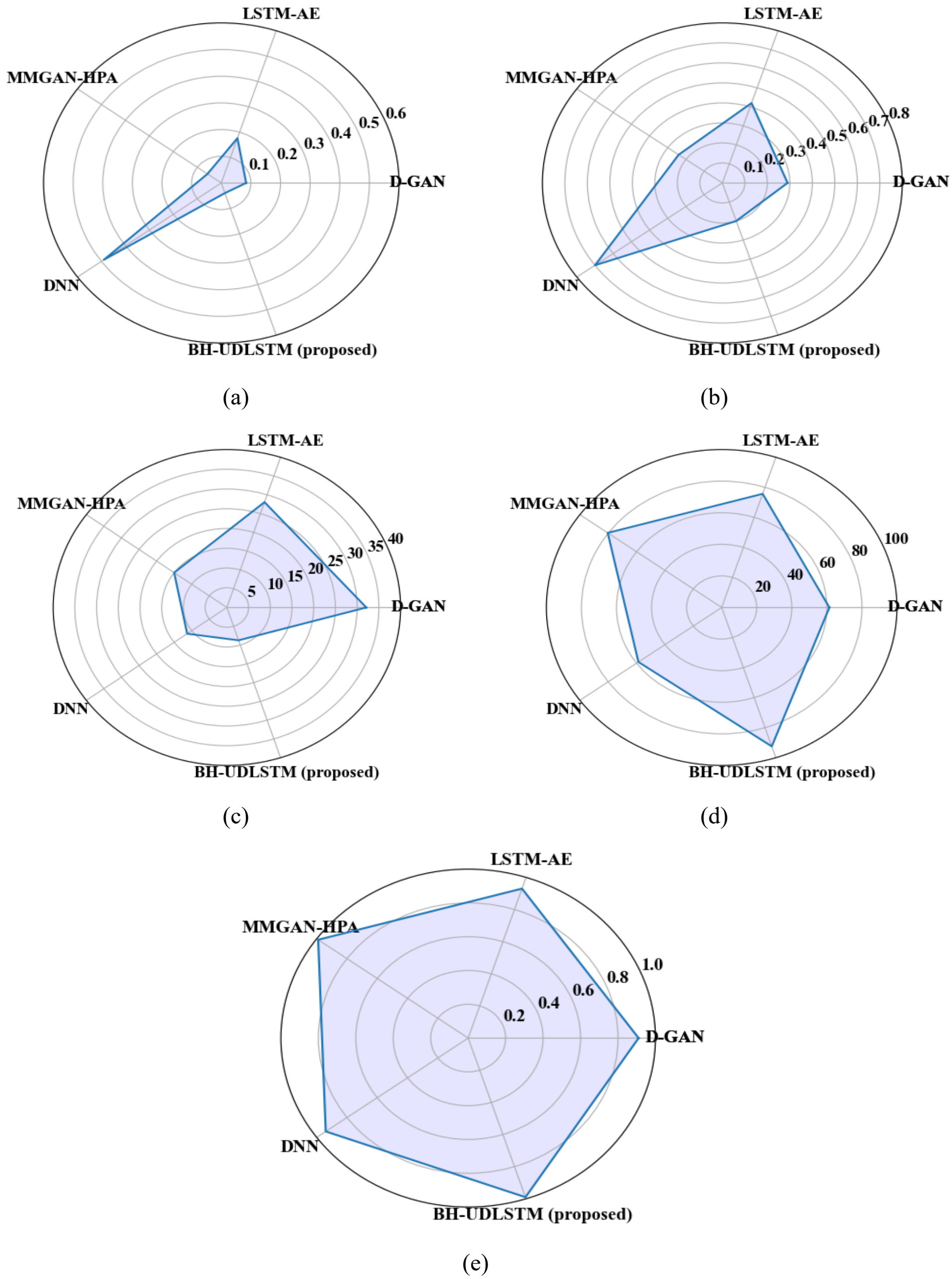
4.5 Performance evaluation

Table 10 depicts the presents the average running times per sample for each of the mentioned DL techniques (Nabipour et al., 2020). The running times are measured in seconds and represent the time taken by the models to process individual samples of input data. The average time taken by the BH-UDLSTM (proposed) model to process one sample of sequential input data is 0.019923 seconds which is comparatively lower than other deep-learning techniques. Lower running times are generally preferred as they indicate faster processing speed, making the model more efficient and suitable for real-time or large-scale applications. The hybrid optimisation algorithm used in this research work efficiently update the model's weights during the training process. This enables the LSTM network to converge to an optimal solution faster, reducing the number of iterations needed for training.

Table 10 Overall running time comparison with DL techniques

	Deep learning-based techniques			
Techniques	*ANN (Nabipour et al., 2020)*	*RNN (Nabipour et al., 2020)*	*LSTM (Nabipour et al., 2020)*	*BH-UDLSTM (proposed)*
Average running time per sample	0.020088 s	0.02063 s	0.080902 s	0.019923 s

Figure 10 Comparison of (a) MSE (b) RMSE (c) MAPE (d) overall accuracy and (e) correlation (see online version for colours)



The proposed BH-UDLSTM method of SPP is compared with several existing techniques such as DNN, MMGAN, LSTM-AE, and D-GAN. Figures 10(a), 10(b), 10(c), 10(d) and 10(e) show the evaluation of MSE, RMSE, MAPE, accuracy, and correlation metrics. The MSE, MAPE, and RMSE assessment metrics are used, as researchers actively investigated these to depict the prediction performance of several techniques. The proposed BH-UDLSTM method proved more reliable than other recent approaches in predicting future stock market patterns, as evidenced by the results, which exhibit lower MAPE (8.74%), RMSE (0.2%), and MSE (0.04%) values. A perfect correlation coefficient between the prices of stocks of the proposed technique (0.99%) is higher than other existing approaches. The primary factor in the proposed solution's improved performance is due to the optimisation of the weight and error threshold parameters of UDLSTM by the hybrid optimisation algorithm. Also, the proposed updated version accomplishes accurate time-series prediction with various dispersals, which makes the proposed scheme more adaptable. Accuracy of a technique in prediction is greater when compared to other points if it has a smaller standard deviation and a higher correlation coefficient. Thus, the BH-UDLSTM obtains an overall higher accuracy of 92.45%.

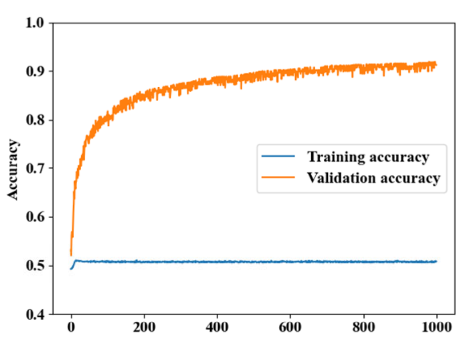
Table 11 Overall accuracy comparison with existing researches which takes input data from datasets

Techniques	*Overall accuracy*
S-I-LSTM (Wu et al., 2022)	52.72%
HiSA-SMFM (Srivinay et al., 2022)	91.99%
Ensemble learning (Li and Pan, 2022)	66.67%
LSTM (Alsayat, 2021)	90.25%
BH-UDLSTM (proposed)	92.45%

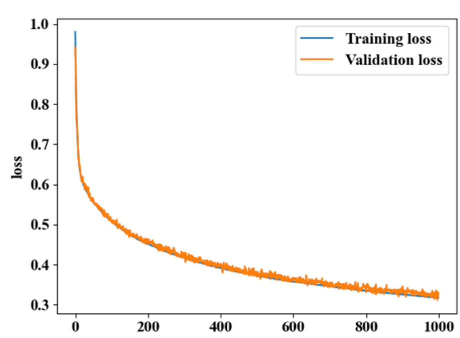
The proposed model is also evaluated with the traditional approaches which uses several datasets for input acquisition as displayed in Table 11. BH-UDLSTM uses real time stock and sentiment values as input for prediction, this proposed model achieves a superior performance than the existing techniques.

The accuracy and loss curves shown in Figures 11(a) and 10(b) on training and validation data of BH-UDLSTM provide performance variations over the number of epochs. It is configured to have 1000 epochs. The training accuracy remains the same at every epoch and the testing accuracy is increased until 200 epochs and maintained almost constant in 400 epochs. The training and testing loss of BH-UDLSTM decreased until 500 epochs and then, the loss curves slightly decreased.

Figure 11 (a) Accuracy and (b) loss curves (see online version for colours)

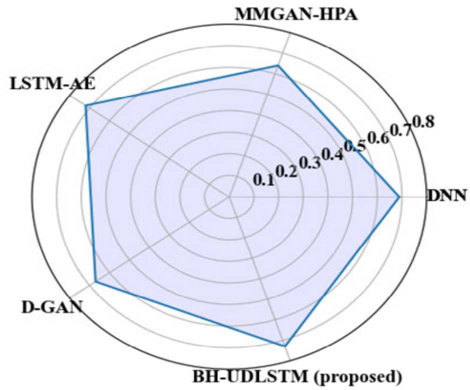


(a)



(b)

Figure 12 Comparison of training performance (see online version for colours)



The training stability of the proposed neural network is evaluated using Figure 12 in comparison with several traditional techniques. It is observed that the training process is more excellent in case of BH-UDLSTM by coupling the input gate of UDLSTM with forget gate and also updating the cell conditions. The evaluation is done by comparing with traditional approaches like DGAN, LSTM-AE, MMGAN-HPA and DNN and proven that the proposed technique achieved maximum training stability of 0.77.

5 Conclusions

The stock market prediction i.e. rising (positive) and falling (negative) based on sentiment analysis and stock market data from January 1, 2018, to January 16, 2023, using the proposed BH-UDLSTM is successfully implemented using python. Six companies such as Bharti_airtel, Infosys, Adani_enterprises, Hcl_technologies, Tata_steel, and SBI are selected for prediction. The performance of the proposed method is evaluated with the existing procedures and thus proved that BH-UDLSTM performs better in MSE (0.04%), RMSE (0.2%), MAPE (8.74%), accuracy (92.45%), and correlation (0.99%).The UDLSTM model offered a more consistent training performance, which improved the data's prediction accuracy. Optimising the parameter of the UDLSTM network with the hybrid of the NBA and HGS algorithm makes the network predict the stock value accurately with a minimum error rate.

Future research will examine the proposed technique's applicability to other time series prediction application domains like climate, disaster, etc. To broaden the applicability of the proposed technique to diverse time series prediction domains like climate and disaster, the approach involves expanding the dataset, incorporating domain-specific features, exploring transfer learning, collaborating with domain experts, rigorous benchmarking, iterative testing, and transparent documentation of methodology and results. These measures aim to enhance the model's adaptability beyond its initial focus on stock market and sentiment analysis data.

Disclaimer

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Agbaje, M.B., Ezugwu, A.E. and Els, R. (2019) 'Automatic data clustering using hybrid firefly particle swarm optimization algorithm', *IEEE Access*, Vol. 7, pp.184963–184984.

Agrawal, M., Khan, D.A. and Shukla, D.P. (2019) 'Stock price prediction using technical indicators: a predictive model using optimal deep learning', *International Journal of Recent Technology and Engineering (IJRTE)*, Vol. 8, No. 2, pp.2297–2305.

Albahli, S., Nazir, T., Mehmood, A., Irtaza, A., Alkhalifah, A. and Albattah, W. (2022) 'AEI-DNET: a novel DenseNet model with an autoencoder for the stock market predictions using stock technical indicators', *Electronics*, Vol. 11, No. 4, p.611.

Alsayat, A. (2021) 'Improving sentiment analysis for social media applications using an ensemble deep learning language model', *Arabian Journal for Science and Engineering*, Vol. 47, No. 2, pp.2499–2511.

Asgarian, S., Ghasemi, R. and Momtazi, S. (2023) 'Generative adversarial network for sentiment-based stock prediction', *Concurrency and Computation: Practice and Experience*, Vol. 35, No. 2, p.e7467.

Chahardoli, M., Eraghi, N.O. and Nazari, S. (2022) 'Namib beetle optimization algorithm: a new meta-heuristic method for feature selection and dimension reduction', *Concurrency and Computation: Practice and Experience*, Vol. 34, No. 1, p.e6524.

Chen, W., Zhang, H., Mehlawat, M.K. and Jia, L. (2021) 'Mean-variance portfolio optimization using machine learning-based stock price prediction', *Applied Soft Computing*, Vol. 100, p.106943.

Chun, S-H. and Jang, J-W. (2022) 'A new trend pattern-matching method of interactive case-based reasoning for stock price predictions', *Sustainability*, Vol. 14, No. 3, p.1366.

Gülmez, B. (2023) 'Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm', *Expert Systems with Applications*, Vol. 227, p.120346.

Gupta, I., Singh, A., Singh, S. and Madan, T. (2022) *HiSA-SMFM: Historical and Sentiment Analysis based Stock Market Forecasting Model*, arXiv preprint arXiv:2203.08143.

Hashim, F.A., Houssein, E.H., Mabrouk, M.S., Al-Atabany, W. and Mirjalili, S. (2019) 'Henry gas solubility optimization: a novel physics-based algorithm', *Future Generation Computer Systems*, Vol. 101, pp.646–667.

Karimi, S., Amiri, B.J. and Malekian, A. (2019) 'Similarity metrics-based uncertainty analysis of river water quality models', *Water Resources Management*, Vol. 33, No. 6, pp.1927–1945.

Kumar, A., Alsadoon, A., Prasad, P.W., Abdullah, S., Rashid, T.A., Pham, D.T. and Nguyen, T.Q. (2022) 'Generative adversarial network (GAN) and enhanced root mean square error (ERMSE): deep learning for stock price movement prediction', *Multimedia Tools and Applications*, Vol. 81, No. 3, pp.3995–4013.

Li, W., Sengupta, N., Dechent, P., Howey, D., Annaswamy, A. and Sauer, D.U. (2021) 'Online capacity estimation of lithium-ion batteries with deep long short-term memory networks', *Journal of Power Sources*, Vol. 482, p.228863.

Li, Y. and 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*, Vol. 13, No. 2, pp.139–149.

Li, Y., Bu, H., Li, J. and Wu, J. (2021) 'The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning', *International Journal of Forecasting*, Vol. 36, No. 4, pp.1541–1562.

Lu, W., Li, J., Wang, J. and Qin, L. (2020) 'A CNN-BILSTM-AM method for stock price prediction', *Neural Computing and Applications*, Vol. 33, No. 10, pp.4741–4753.

Ma, Y., Mao, R., Lin, Q., Wu, P. and Cambria, E. (2023) 'Multi-source aggregated classification for stock price movement prediction', *Information Fusion*, Vol. 91, pp.515–528.

Mehtab, S. and Sen, J. (2019) *A Robust Predictive Model for Stock Price Prediction using Deep Learning and Natural Language Processing*, arXiv preprint arXiv:1912.07700.

Mndawe, S.T., Paul, B.S. and Doorsamy, W. (2022) 'Development of a stock price prediction framework for intelligent media and technical analysis', Applied Sciences, Vol. 12, No. 2, p.719.

Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E. and Shahab, S. (2020) 'Deep learning for stock market prediction', *Entropy*, Vol. 22, No. 8, p.840.

Polamuri, S.R., Srinivas, K. and Krishna Mohan, A. (2022) 'Multi-model generative adversarial network hybrid prediction algorithm (MMGAN-HPA) for stock market prices prediction', *Journal of King Saud University – Computer and Information Sciences*, Vol. 34, No. 9, pp.7433–7444.

Rezaei, H., Faaljou, H. and Mansourfar, G. (2021) 'Stock price prediction using deep learning and frequency decomposition', *Expert Systems with Applications*, Vol. 169, p.114332.

Shen, F., Zhao, X., Kou, G. and Alsaadi, F.E. (2021) 'A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique', *Applied Soft Computing*, Vol. 98, p.106852.

Soemartojo, S.M., Siswantining, T., Fernando, Y., Sarwinda, D., Al-Ash, H.S., Syarofina, S. and Saputra, N. (2022) 'Iterative bicluster-based Bayesian principal component analysis and least squares for missing-value imputation in microarray and RNA-sequencing data', *Mathematical Biosciences and Engineering*, Vol. 19, No. 9, pp.8741–8759.

Srivinay, Manujakshi, B., Kabadi, M. and Naik, N. (2022) 'A hybrid stock price prediction model based on pre and deep neural network', *Data*, Vol. 7, No. 5, p.51.

Steven Eyobu, O. and Han, D. (2018) 'Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network', *Sensors*, Vol. 18, No. 9, p.2892.

Stoean, C., Paja, W., Stoean, R. and Sandita, A. (2019) 'Deep architectures for long-term stock price prediction with a heuristic-based strategy for trading simulations', *PLOS ONE*, Vol. 14, No. 10.

Vijh, M., Chandola, D., Tikkiwal, V.A. and Kumar, A. (2020) 'Stock closing price prediction using machine learning techniques', *Procedia Computer Science*, Vol. 167, pp.599–606.

Wang, Y., Du, X., Lu, Z., Duan, Q. and Wu, J. (2022) 'Improved LSTM-based time-series anomaly detection in rail transit operation environments', *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 12, pp.9027–9036.

Wu, J.M-T., Li, Z., Herencsar, N., Vo, B. and Lin, J.C-W. (2021) 'A graph-based CNN-LSTM stock price prediction algorithm with leading indicators', *Multimedia Systems*, Vol. 29, No. 3, pp.1751–1770.

Wu, S., Liu, Y., Zou, Z. and Weng, T-H. (2022) 'S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis', *Connection Science*, Vol. 34, No. 1, pp.44–62.

Yu, P. and Yan, X. (2019) 'Stock price prediction based on deep neural networks', *Neural Computing and Applications*, Vol. 32, No. 6, pp.1609–1628.