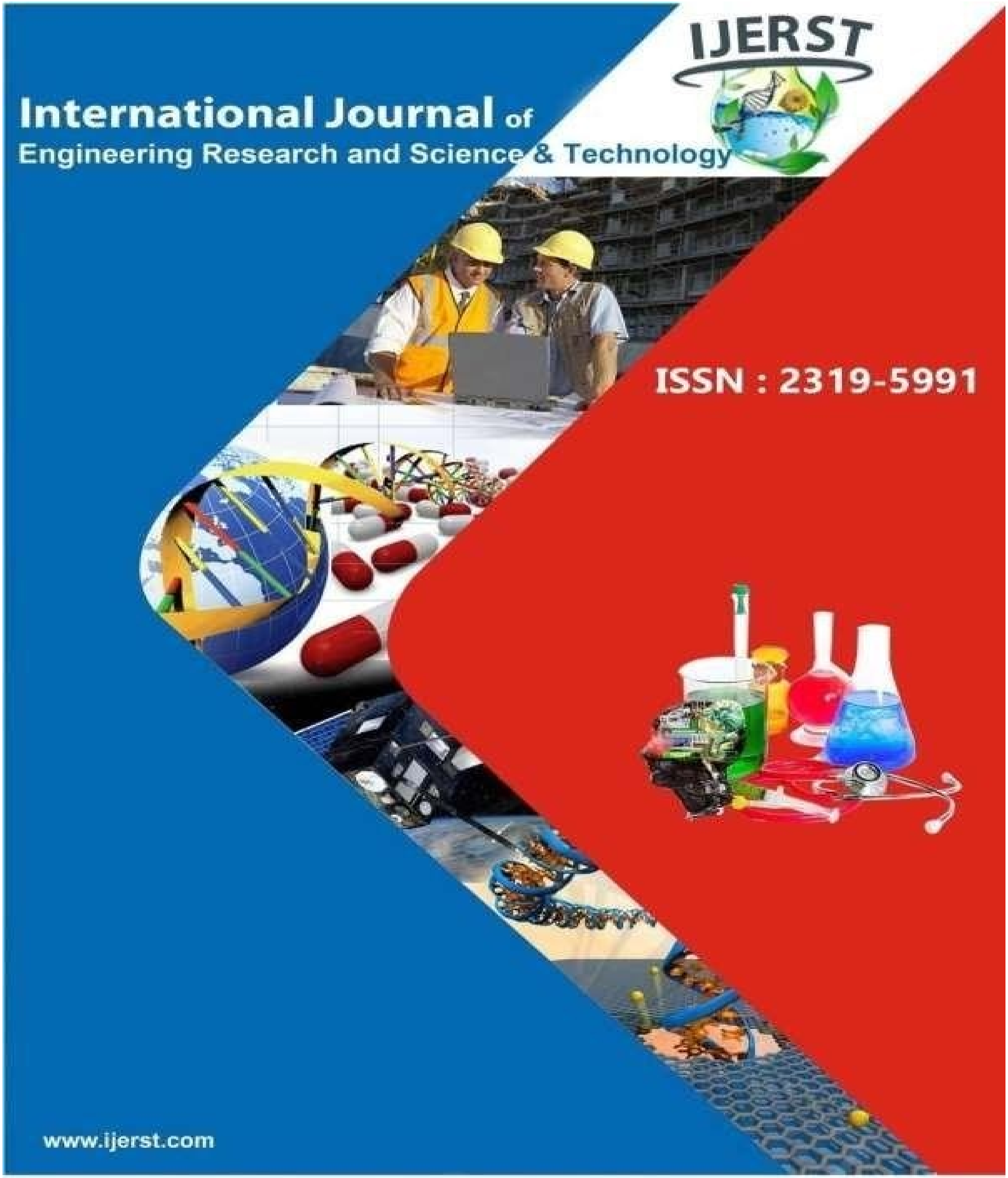


Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp884-891 Vol. 21, Issue 2, 2025

A HYBRID MODEL FOR STOCK PRICE FORECASTING BASED ON MARKET SENTIMENT AND DEEP LEARNING OPTIMIZATION

[1]T. Chandana, MCA Student, Department of MCA
[2] P. Rohini Bai, M.Tech, Assistant Professor, Department of MCA
[12]Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool

ABSTRACT

If you can accurately predict the stock price, you may increase your returns while decreasing your investment risks. Combining deep learning, swarm intelligence techniques, sentiment analysis, and multi-source data affecting stock prices, this research constructs the MS-SSA-LSTM model. The first step in developing a unique emotion dictionary and calculating the sentiment index is to peruse the posts made on the East Money forum. After that, the Sparrow Search Algorithm is used to optimise the hyperparameters of the Long and Short-Term Memory networks (LSTM). Last but not least, LSTM is used to forecast stock values in the future by combining fundamental trade data with the sentiment index. Results demonstrate that the MS-SSA-LSTM model outperforms the competition and is very generalisable. Compared to ordinary LSTM, MS-SSA-LSTM improves R2 by an average of 10.74%. We found that: 1) Adding the sentiment index to the model may increase its prediction ability. 2) SSA is used to fine-tune the LSTM's hyperparameters, which enhances the prediction's effect and offers a neutral justification for the model's parameter values. Thirdly, short-term forecasts work better in the very volatile Chinese financial market.

I. INTRODUCTION

More and more individuals are interested in working in the financial sector due to factors such as the development of China's stock market and the fast expansion of online banking. Yet, the stock market is well-known for its enormous data sets and for its infamous volatility. Improving data-mining abilities is crucial for many individual investors. Companies and investors may be able to lessen investment risks and increase investment returns with the aid of reliable stock price forecasts.

The first study's authors built a linear model using statistical techniques to mimic the historical trend of stock prices. Traditional approaches may be shown by several instances, such as GARCH, ARIMA, ARMA, and many more. For the purpose of analysing stocks using time series data, the ARMA was developed. Based on the ARMA, the ARIMA model was created to forecast the overall trend of stock price changes [2]. To further enhance the accuracy of the Shanghai Composite Index's fitting, one might use wavelet analysis to the ARIMA model [3]. The GARCH model offers novel concepts for long-term stock time series forecasting [4]. Concurrently, a number of writers offered theoretical backing for multivariate stock volumetric price research by developing a novel prediction model that fused ARMA and GARCH [5]. A lot of the time, these older systems can only handle data that is regular and organised. On the other hand, conventional approaches to predicting depend on assumptions that are just not practical. As a result, defining nonlinear financial data using statistical methods is rather complex.

As a result, many academics rely on machine learning techniques like SVMs and neural networks to try to predict stock values. The core principle of machine learning is that computers have the ability to learn and make predictions

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp884-891
Vol. 21, Issue 2, 2025

based on existing data. Academics often use the SVM for stock forecasting because of its exceptional capabilities in handling low-sample-size data, high-dimensional data, and nonlinear scenarios. Hossain and Nasser [6] discovered that support vector machine (SVM) approaches generated more accurate stock forecasts compared to statistical methods. Using a mix of least squares SVM and Genetic Algorithm (GA), Chai et al. [7] discovered that a hybrid SVM model was the most effective in predicting changes to the HS300 index. Since SVM requires a lot of memory and processing power when trained on large-scale datasets, it may not be able to predict a lot of stock data. We next go on to address financial time series difficulties using ANN and multi-layer ANN. There is a lot of evidence from experiments indicating ANNs are beneficial, and one of them is that they are quite accurate and converge quickly [8, 9, 10]. To find out how well various feed forward artificial neural networks predicted stock prices, Moghaddam and Esfandyari [11] ran tests on them. By using a Bayesian regularisation strategy, Liu and Hou [12] enhanced the BP (Back Propagation) neural network. However, the time-tested neural network approach may need improvement in the following areas. A lack of generalisability is soon followed by overfitting and optimisation at the local level. To address these obstacles, a large number of samples must be trained, so better models are required.

This research introduces a novel model (MS-SSA-LSTM) for stock price forecasting that integrates LSTM neural networks with data from several sources using the Sparrow Search Algorithm. By projecting the future value of a company, the MS-SSA-LSTM model may assist traders and investors in making more informed investment choices. The MS-SSA-LSTM model is used by traders and investors to enter data pertaining to specific equities, including shareholder comments and transaction history. The algorithm not only predicts the stock price for the next day, but it also automatically creates a trend chart for the stock price.

This is where the concepts presented in the article first appeared.

Improving the accuracy of predictions is possible by include sentiment indicators in the model's parameters. But a large vocabulary won't get you far in the financial industry. The development of an emotion language that is particular to stocks is therefore imperative.

Some of the most challenging features of stock price series are high time-variability, nonlinearity, and excessive noise. Large time series data sets can be handled using long short-term memory. This is where the long short-term memory (LSTM) network, a deep learning technique, comes in handy.

Thirdly, the LSTM network's prediction performance is affected by adjusting the hyperparameters. The network model's hyperparameters may be expensively and artificially selected. Improving LSTM may so include using the 2020 Sparrow Search Algorithm.

These are the primary findings from the research.

(1) The SSA-LSTM model uses data from several sources, including stock forum posts and historical transaction data, to influence stock prices. Predictions made by the MS-SSA-LSTM model are more precise than those of the model that uses just one data source.

Scroll down to the textual comments (2) on the stock forum. For the purpose of calculating sentiment indicators and doing individual stock sentiment analysis, our next move is to create a sentiment language by compiling reliable financial investing dictionaries.

With the help of the Sparrow Search Algorithm, LSTM models can achieve better results than individual LSTM networks when it comes to predicting stock prices and hyperparameter values.

II. LITERATURE SURVEY

Earlier studies on stock market prediction were based on the Random Walk and the Efficient

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp884-891 Vol. 21, Issue 2, 2025

Market Hypothesis (EMH). Multiple studies show that stock market prices are not completely unpredictable and can be predicted, according to Gallagher, Kavussanos, and Butler, among others [3][7][10]. Another theory under investigation is whether or whether early indications derived from online sources (e.g., blogs, Twitter feeds, etc.) may be used to predict future changes in economic and commercial indicators. Similar analyses have been conducted in several fields of study; for instance, Gruhl et al. [8] found a correlation between book sales and online debate participation. Blog sentiment analysis has been used by Mishne, Glance, and colleagues to predict movie office revenues [15]. Schumaker et al. [18] investigated the correlation between sudden financial news and stock price changes. Investigating the connection between the DJIA and public opinion, Bollen, Mao, et al. (2011) produced a landmark study in the field of stock prediction. We found out how people were feeling in general (happy, calm, nervous) by looking at their Twitter feeds [2]. Chen and Lazer came up with investment techniques by sorting through Twitter feeds [4]. Bing et al. concluded from their examination of the Twitter stream that stock prediction was industry-specific [1]. According to Zhang et al. [20], there was a substantial negative connection between unpleasant emotions experienced on social networks and the DJIA index. Pagolu et al. discovered that there is a strong correlation between the emotion on Twitter and the movement of a stock price. Their research focused on developing a sentiment analyser that, instead of using the conventional word embedding method, could sort tweets into three distinct mood states: positive, negative, and neutral. [16]. Mittal et al. sought to develop a portfolio management tool using sentiment analysis from Twitter in their study. Their model was reviewed and validated using DJIA. The greedy strategy-based model used social media sentiment data to predict the DJIA holdings' Buy/Sell decisions a day in advance. Based on their study, Chen et al. contrasted the Long Short-Term Memory (LSTM) model with the Random Estimation (RA) model, proving that the LSTM model was more accurate. The method served as the foundation for the Long Short-Term Memory (LSTM) model, which was used to predict the movement of equities on the Chinese Stock Exchange [4]. In their analysis of data from 25 of the largest companies, Tekin et al. [19] employed a number of different forecasting systems. According to their findings, the Random Forest method is more important. When allocating funds for a portfolio, Malandri et al. employ a classifier from random forests and an MLP from long short-term memory (LSTM). An examination of NYSE data indicates that LSTM yields better experimental results [13]. In their presentation to the Turkish Stock Exchange (BIST 100), Kilimci et al. detailed their work on developing a deep learning-based model that effectively uses word embedding to predict the market direction of the stock exchange by analysing financial news sites and Twitter [11]. To determine which combination of word embedding and deep learning models may provide the most accurate stock forecasts, their study comprised a variety of such models. Data labelling allows them to classify the information as either good or negative. After that, we feed the data into word embedding models like Word2Vec, FastText, and GloVe. These models generate several word embeddings, which we then test using CNN, RNN, and LSTM, three separate deep learning methods. The Word2Vec embedding model and LSTM combination had the highest average accuracy out of the nine equities that were evaluated, with Twitter data serving as the basis.

III. SYSTEM ANALYSIS

EXISTING SYSTEM

Not only that, but Yan et al. [16] developed an LSTM deep neural network–based, very accurate prediction model for a short-term financial market. The LSTM neural network can reliably forecast market values, but the BP and regular RNN are not as good at making predictions. For

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp884-891
Vol. 21, Issue 2, 2025

their prediction model, Nabipour et al. [17] settled on ten technical markers. In comparison to other algorithms like ANN, RNN, XGBoost, Adaboost, decision tree, and random forest, LSTM performed better in the experiment when it came to fitting models. In order to predict the trading activity of the Dow Jones 30 Index equities the next day, Aksehir and Kilic [18] suggested a CNN-based model. The program is fed data on technical indicators, oil and gold prices. The results show that the new CNN-based algorithms are 3-22% more accurate than the old ones.

The very subjective values of LSTM network hyperparameters, however, typically need researchers to deliberately set them. A model's prediction performance is affected by the hyperparameter setting, which directly controls the topology of the network model. Optimal hyperparameter selection for the network model requires a large investment of time and computing power. Using Swarm Intelligence (SI) algorithms, researchers try to adjust the hyperparameters of neural networks in an effort to fix the problems outlined before. Using its global search capabilities, SI can, to an extent, evaluate the best answer's value. The LSTM model outperformed the LSTM optimised using the Improved Particle Swarm Optimisation (IPSO) approach, as shown by Ji et al. [19]. Zeng et al. [20] improved prediction accuracy by optimising LSTM using an Adaptive Genetic Algorithm (AGA). Optimisation using neural networks often also employs additional advanced methods. The creation of the Sparrow Search Algorithm (SSA) [21] was inspired by the sparrows' abilities to both hunt and defend themselves from predators. When compared to more traditional optimisation methods like particle swarm and grey wolf, the novel SSA technique outperforms them in terms of price prediction. In addition, the approach demonstrates great global search capabilities by considering the stochastic nature of humans throughout the search process. Furthermore, the approach may adaptively change the search parameters to fit different problems' needs.

The data collected from the stock forum's real-time content updates more accurately portrays the sentiment of investors than other sources of information. However, it will greatly influence investors' investment preferences, which in turn will greatly affect stock price fluctuations. This is why our analysis cannot be complete without analysing the textual data collected from stock forums. Researchers mostly use machine learning and semantic analysis methods for sentiment classification. Since the machine learning method necessitates the human categorisation of training sets by finance professionals, the time expenditure is increased despite the high classification accuracy. Applying a standard dictionary to a business setting could be challenging, but the semantic analysis method makes financial and economic analysis much easier to understand and implement [34]. Making your own unique financial language is the key.

Disadvantages

• Difficulty with large and complex datasets: most existing machine learning techniques need such datasets to successfully forecast stock prices.

• Data accessibility: Machine learning algorithms sometimes want enormous datasets to provide trustworthy predictions. Without sufficient data, the model's reliability might be affected.

• False labels: The capacity of existing ML models to learn is limited by the quality of the training data. It is the responsibility of the data labels to ensure that the model produces accurate predictions.

PROPOSED SYSTEM

This research introduces a novel model (MS-SSA-LSTM) for stock price forecasting that integrates LSTM neural networks with data from several sources using the Sparrow Search Algorithm. By projecting the future value of a company, the MS-SSA-LSTM model may assist traders and investors in making more informed investment choices. The MS-SSA-LSTM model

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp884-891
Vol. 21, Issue 2, 2025

is used by traders and investors to enter data pertaining to specific equities, including shareholder comments and transaction history. The algorithm not only predicts the stock price for the next day, but it also automatically creates a trend chart for the stock price.

This is where the concepts presented in the article first appeared.

Improving the accuracy of predictions is possible by include sentiment indicators in the model's parameters. But a large vocabulary won't get you far in the financial industry. The development of an emotion language that is particular to stocks is therefore imperative.

Some of the most challenging features of stock price series are high time-variability, nonlinearity, and excessive noise. Large time series data sets can be handled using long short-term memory.

This is where the long short-term memory (LSTM) network, a deep learning technique, comes in handy. the third Altering the hyperparameters of the LSTM network directly impacts the accuracy of the model's predictions. The network model's hyperparameters may be expensively and artificially selected. Consequently, the 2020 Sparrow Search Algorithm might be useful for LSTM..

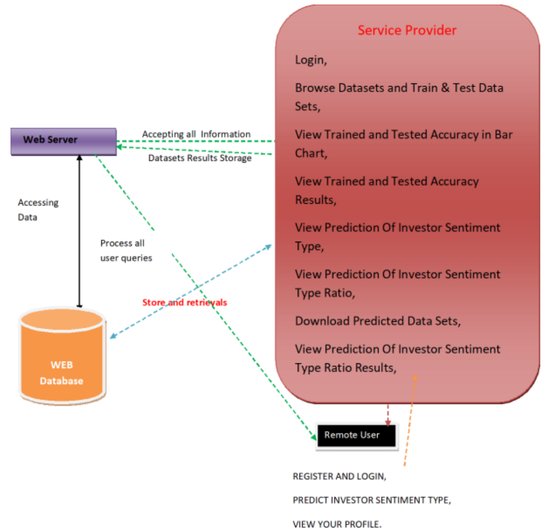
Advantages

(1) The SSA-LSTM model uses a variety of data sources to determine stock values, including past transactions and user comments on stock forums. The MS-SSA-LSTM model outperforms the model that uses a single data source in terms of prediction accuracy.

Second, read the stock forum posts for any relevant content. The next step is to compile trustworthy financial investment dictionaries into a sentiment dictionary suitable for use in determining sentiment indicators and evaluating individual stocks.

(3) Tuning an LSTM model using the Sparrow Search Algorithm may lead to more accurate stock price predictions and better hyper parameter values than a single LSTM network.

IV. SYSTEM ARCHITECTURE



V. SYSTEM IMPLEMENTATION MODULES

Service Provider

The Service Provider can't access this module without a valid login and password. Once logged in, he may do tasks like as viewing datasets, executing tests, and training on them. Verify the Trained and Tested Accuracy Bar Chart, Verify the Trained and Tested Accuracy Results, Verify All Remote Users, Download Forecasted Data Sets, Verify Investor Sentiment Type Prediction Results, and Verify Investor Sentiment Type Ratio Prediction.

View and Authorize Users

Here the administrator may get a comprehensive list of all users who have signed up. A user's name, email, and physical address are shown here so the administrator can approve their request for access.

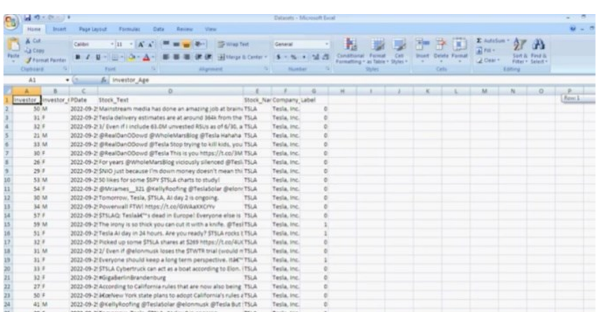
Remote User

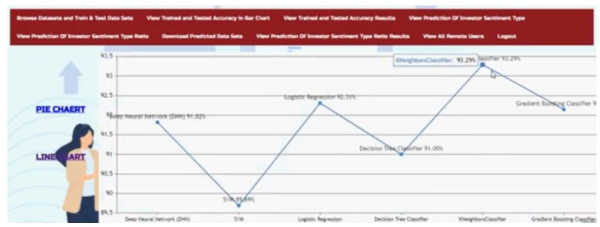
In this module, n users are total. Before any procedures may be carried out, registration must be completed. Once a person has registered, their details will be added to the database. After the registration is finished, he will be prompted to log

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp884-891
Vol. 21, Issue 2, 2025

in using the authorised username and password. Users may examine their profiles, forecast the sentiment type of investors, and register and log in when they log in.

VI. RESULTS



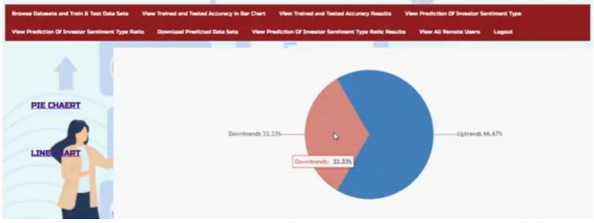


International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2025.v21.i2.pp884-891

Vol. 21, Issue 2, 2025



VII. CONCLUSION

The hyper parameters in the LSTM network are often chosen based on subjective experience, and the stock price is affected by the emotions of shareholders. Consequently, this study introduces the MS-SSA-LSTM stock price prediction model. In order to train and test the model, six datasets including individual stocks from the Chinese financial market were used. We also utilise four assessment markers to measure how well the model predicts. It is possible to draw these findings via the use of comparative analysis.

Before anything else, the MS-SSA-LSTM model considers a wide range of data sources. The first source of information is the characteristics of past transaction data, which include things like opening price, closing price, prior closing price, etc. On the flip side, the data is derived from the sentiment of market shareholders. The model's prediction performance is enhanced when basic stock trading indicators are inputted with a mood indicator. Therefore, the stock bar might be a tool for manipulating the emotions of investors. The platform administrators may control the flow of information, provide warnings before widespread unrest breaks out, encourage shareholders to make prudent investments, and more. A constructive and orderly stock market environment is what we aim for at the end of the day.

Secondly, it is not easy to produce the best prediction results since the LSTM parameters need to be altered on purpose. Consequently, the SSA is chosen in order to optimise the LSTM model's hyperparameters. In addition to giving an unbiased description of the model's parameter values and network architecture, this strategy improves the model's predicting abilities and makes it more flexible. Even in the most competitive financial markets, the SSA-designed LSTM model can provide investors with low-risk price predictions, quick and precise feature understanding, and increased efficiency.

Thirdly, comparison studies performed on six separate stocks in different models validate the MS-SSA-LSTM model's outstanding accuracy, reliability, and stock market adaptability. Based on the results of the experiment, short-term forecasts are more appropriate in light of the tremendous volatility of China's financial markets, and the model's prediction effect needs 5–10 time steps to attain its optimal level. While we wait, this method could fix other time series problems.

So, to summarise, the MS-SSA-LSTM model is found everywhere. Although we limited our testing to the Chinese financial sector, the model holds up well when applied to other global stock markets. Among the several data sets used as inputs by the algorithm are technical stock data and the shareholder sentiment index. The Chinese and international financial markets have access to this multi-source data. Debates on stock prices may take place on a variety of easily accessible social media platforms. Finding the stock comments will allow us to calculate the sentiment index, which can then be used as an input feature in the model. In addition, we can gather input more quickly and easily by establishing a stock market commenting platform.

The MS-SSA-LSTM model still has limitations when it comes to our multi-source data. When it comes to sentimental analysis, this research only differentiates between positive and negative feelings. Alterations to policy and macroeconomic conditions are two other potential variables that could impact the stock price forecast. Moving on, sentiment analysis must be

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp884-891
Vol. 21, Issue 2, 2025

enhanced. It would be beneficial for our research to include several emotional indicators, such as sadness, anxiety, anger, and contempt. Meanwhile, official WeChat and Weibo accounts are only two examples of the additional data sources that may be included to estimate the market sentiment. It would be wise to use mining techniques in order to discover more factors that impact the forecasting of stock prices.

REFERENCES

[1] M. M. Rounaghi and F. N. Zadeh, ‘‘Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: Monthly and yearly forecasting of time series stock returns using ARMA model,’’ *Phys. A, Stat. Mech. Appl.*, vol. 456, pp. 10–21, Aug. 2016, doi: 10.1016/j.physa.2016.03.006.

[2] G. Bandyopadhyay, ‘‘Gold price forecasting using ARIMA model,’’ *J. Adv. Manage. Sci.*, vol. 4, no. 2, pp. 117–121, 2016, doi: 10.12720/joams.4.2.117-121.

[3] H. Shi, Z. You, and Z. Chen, ‘‘Analysis and prediction of Shanghai composite index by ARIMA model based on wavelet analysis,’’ *J. Math. Pract. Theory*, vol. 44, no. 23, pp. 66–72, 2014.

[4] H. Herwartz, ‘‘Stock return prediction under GARCH—An empirical assessment,’’ *Int. J. Forecasting*, vol. 33, no. 3, pp. 569–580, Jul. 2017, doi: 10.1016/j.ijforecast.2017.01.002.

[5] H. Mohammadi and L. Su, ‘‘International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models,’’ *Energy Econ.*, vol. 32, no. 5, pp. 1001–1008, Sep. 2010, doi: 10.1016/j.eneco.2010.04.009.

[6] A. Hossain and M. Nasser, ‘‘Recurrent support and relevance vector machines based model with application to forecasting volatility of financial returns,’’ *J. Intell. Learn. Syst. Appl.*, vol. 3, no. 4, pp. 230–241, 2011, doi: 10.4236/jilsa.2011.34026.

[7] J. Chai, J. Du, K. K. Lai, and Y. P. Lee, ‘‘A hybrid least square support vector machine model with parameters optimization for stock forecasting,’’ *Math. Problems Eng.*, vol. 2015, pp. 1–7, Jan. 2015, doi: 10.1155/2015/231394.

[8] A. Murkute and T. Sarode, ‘‘Forecasting market price of stock using artificial neural network,’’ *Int. J. Comput. Appl.*, vol. 124, no. 12, pp. 11–15, Aug. 2015, doi: 10.5120/ijca2015905681.

[9] D. Banjade, ‘‘Forecasting Bitcoin price using artificial neural network,’’ Jan. 2020, doi: 10.2139/ssrn.3515702.

[10] J. Zahedi and M. M. Rounaghi, ‘‘Application of artificial neural network models and principal component analysis method in predicting stock prices on Tehran stock exchange,’’ *Phys. A, Stat. Mech. Appl.*, vol. 438, pp. 178–187, Nov. 2015, doi: 10.1016/j.physa.2015.06.033.

[11] A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, ‘‘Stock market index prediction using artificial neural network,’’ *J. Econ., Finance Administ. Sci.*, vol. 21, no. 41, pp. 89–93, Dec. 2016, doi: 10.1016/j.jefas.2016.07.002.

[12] H. Liu and Y. Hou, ‘‘Application of Bayesian neural network in prediction of stock time series,’’ *Comput. Eng. Appl.*, vol. 55, no. 12, pp. 225–229,2019.

[13] A. M. Rather, A. Agarwal, andV. N. Sastry, ‘‘Recurrent neural network and a hybrid model for prediction of stock returns,’’ *Expert Syst. Appl.*, vol. 42,no. 6, pp. 3234–3241, Apr. 2015, doi: 10.1016/j.eswa.2014.12.003.

[14] A. Sherstinsky, ‘‘Fundamentals of recurrent neural network(RNN) and long short-term memory (LSTM) network,’’ *Phys. D,Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306, doi:10.1016/j.physd.2019.132306.

[15] G. Ding and L. Qin, ‘‘Study on the prediction of stock price based on the associated network model of LSTM,’’ *Int. J. Mach. Learn. Cybern.*, vol. 11,no. 6, pp. 1307–1317, Nov. 2019, doi: 10.1007/s13042-019-01041-1.

e0272637, doi: 10.1371/journal.pone.0272637.