Forecasting Financial Frontiers: Real-time Insights in Stock Price Prediction through LSTM, Linear Regression, and Sentiment Analysis

2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) | 979-8-3503-9463-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/HORA61326.2024.10550829

Ahmed A. Khalifa[1,2], Ramesh G[3], Deekshith R[3], Harsha K[3], Shreyas, J[4]

[1] Department of Networking and Security College of Information Technology, University of Fujairah Fujairah, UAE, akhalifa@uof.ac.ae
[2] Switching and Traffic Department, National Telecommunication Institute, Cairo, Egypt, akhalifa@nti.sci.eg.
[3] Dept. of Artificial Intelligence and Machine Learning, Alva's Institute of Engineering and Technology, Mangalore, India
[4]Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka, India-576104, shreyas.j@manipal.edu

Abstract— **The ever-changing world of financial markets makes it extremely difficult to predict stock values with any degree of accuracy. By combining real-time data analysis with cutting-edge methods including sentiment analysis, linear regression, and long short-term memory (LSTM), this study explores the field of financial frontier forecasting. Using LSTM's ability to recognize complicated patterns in time series data, the study makes its way through the challenges of stock price prediction. Sentiment analysis adds a qualitative element by estimating market sentiment from textual data, and linear regression, which supports LSTM, offers a strong foundation for modeling linear correlations. This research presents the efficacy of integrating these approaches, highlighting their capacity to reveal latent insights and improve predictive accuracy in stock price modeling through an empirical investigation. The findings point to the significance of real-time data integration and the synergy that can be attained by combining various analytical techniques, opening the door to more strategic financial forecasts and well-informed investment choices.**

Keywords*— *Stock Price Prediction, Real-time Data Analysis, LSTM (Long Short-Term Memory), Linear Regression, Sentiment Analysis, Financial Forecasting, Market Sentiment.

I. INTRODUCTION

Stock price prediction has long been a focal point in financial research, driven by the substantial impact accurate forecasts can have on investment decisions and risk management strategies. In the dynamic and often unpredictable realm of financial markets, the ability to anticipate future price movements with a high degree of accuracy is highly coveted yet inherently challenging [1]. This study delves into the intricate landscape of stock price prediction, employing a combination of advanced methodologies and real-time data analysis to enhance predictive accuracy and decision-making capabilities [2].

The inability of conventional statistical models to adequately represent the intricate, non-linear dynamics of stock price dynamics, which frequently result in differences between expected and actual values, is the driving force behind this study. To analyze stock price fluctuations and uncover underlying patterns and trends, machine learning techniques, in particular Long Short-Term Memory (LSTM) networks and Linear Regression models, offer promising methods [3]. Recurrent neural networks (RNNs) of the LSTM network type are excellent at identifying long-term patterns and temporal relationships in sequential data, which makes them a good fit for time series data such as historical stock prices. By examining market sentiment from textual sources like news stories and social media, sentiment analysis gives prediction models a qualitative element that can have a big impact on market dynamics. By including the most recent economic data, geopolitical events, and market circumstances into the models, real-time data analysis improves forecast accuracy even more and enables the models to respond swiftly to shifting market conditions. Comprehensive data collection from trustworthy sources, such as Yahoo Finance [4], is the first step in our methodology. The results and analysis section then follows a methodical approach. A comparative study of LSTM-based models, Linear Regression models, and hybrid techniques utilizing Sentiment Analysis and real-time data will be among the experimental results we will showcase. Performance indicators, advantages, disadvantages, and contributions of each methodology to forecast accuracy will all be included in this analysis.

Organization of the paper: The project's introduction and an outline of its algorithm are the main topics of Section I. Section II provides a review of the literature. Problems and its primary goals are the subject of Section III. Section V addresses the proposed methodology, whereas Section IV focuses on system design and architecture. Section VII finishes by summarizing important findings, talking about potential improvements, and considering the value of the study. Section VI details the experimental outcomes.

II. RELATED WORK

Important discoveries have resulted from a number of developments in business management and forecasting, including ground fault prediction techniques that make use of the Hybrid Feature Selection (HFS) algorithm [1]. This approach is centered on identifying relevant variables from financial systems in order to improve forecast precision while reducing economic volatility. In order to better reliably predict stock prices, the Transformer Gated Recurrent Unit

(TEGRU) technique combines media theory with conventional indicators [2], offering analysts and investors easily understandable insights. In order to satisfy a variety of product quality requirements, a hybrid data module for comparison prediction was created [3], cleverly fusing stock and news data to enhance forecasting in volatile markets. To improve forecast accuracy and stability, a further price prediction model based on improved transformers and BiLSTM has been introduced [4]. The Saudi Arabian stock exchange has streamlined data collecting by using meta-heuristic search algorithms [5], which improve prediction accuracy while lowering manual labor. Deep neural network-based unsupervised learning technologies are becoming popular for recognizing odd business models that are suggestive of managerial approaches [7] [8] [18]. For predicting the stock market, long-term memory (LSTM) networks show promise, especially when applied to structural financial data and long-term study [12] [13], Cost estimate accuracy has significantly improved in hybrid models that include deep learning with optimization techniques [15] [16] [17] [19]. Together, these developments strengthen the capacity for technology management and business forecasting, enabling investors and financial analysts to adapt to the changing demands of a dynamic economy.

III. PROBLEM STATEMENT AND OBJECTIVES

Several variables, including news releases, market sentiment, geopolitical developments, and economic indicators, can have an impact on the stock market's volatility and dynamic nature. However, real-time insights are typically difficult for standard stock price prediction algorithms to incorporate properly, which results in less-than-ideal predictions and investment choices [4]. Accurately capturing the most recent market dynamics and trends requires using real-time data streams and smoothly integrating them into prediction models.

Objectives:

- Analyze how well Sentiment Analysis, Linear Regression, and LSTM work together to predict stock prices in real time.
- To demonstrate the benefits of the suggested methodology, compare it with conventional techniques.
- Examine how sentiment analysis affects the accuracy of stock price predictions.
- Develop strategies for dynamic model adaption in response to shifting market circumstances.
- Employing an integrated approach gives traders and investors better decision support.

IV. SYSTEM ARCHITECTURE

The stock price prediction system is designed as a web application using the Django framework, integrating LSTM, Linear Regression, and real-time data integration for accurate and timely stock price forecasts. The architecture comprises multiple layers and components, facilitating data collection, preprocessing, machine learning modeling, real-time data updates, and user interaction through a web interface as shown in Figure 1.

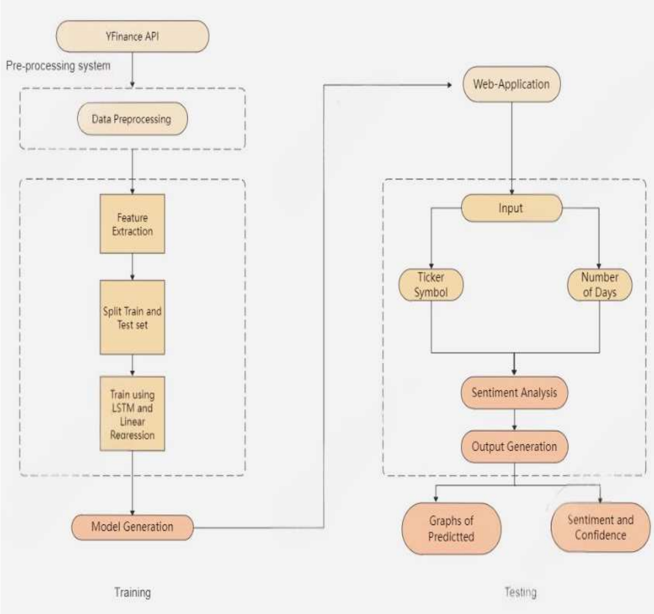

Figure 1 LSTM-LR-SA Framework

The Django framework, which provides scalability, security, and extensibility for data-driven applications, is used by the web application layer. Users may easily input parameters, view predictions, and receive recommendations thanks to its user interface. Sources such as "StockTwits," a trading social media platform, are where data collection starts. Preprocessing is the next step, in which stop words, symbols, extraneous text, and punctuation are eliminated from the data. Words must be stemmed for simplicity and converted to lowercase as part of the standardization process. The prediction phase forecasts future stock prices based on sentiment analysis findings. An upward trend in stock prices, for instance, may be indicated by positive sentiment on StockTwits, which helps investors make decisions. This methodical methodology guarantees effective data processing, analysis, and well-informed forecasts, augmenting the application's usability and user value.

V. PROPOSED METHODOLOGY

The suggested approach to stock price prediction combines several innovative techniques such as sentiment analysis, linear regression, long short-term memory (LSTM) networks, and real-time data processing. The methodical methodology used in data collecting, preprocessing, model creation, assessment metrics, and validation procedures is described in this section [5].

A. Long Short-Term Memory (LSTM) Network

LSTM networks are a type of recurrent neural network (RNN) designed to overcome the limitations of traditional RNNs in capturing long-term dependencies in sequence data. They achieve this by offering specialized storage facilities that can retain data for long periods. The design of an LSTM network is shown in Figure 2 which consists of an input layer, an LSTM layer with buffers, and an output layer.

Algorithm:

Step 1: Gather and prepare historical stock data.
Step 2: Select relevant features and establish a historical data time frame.

Step 3: Design the architecture and initialize the LSTM model.
Step 4: Divide the data into training and testing sets.
Step 5: Train the LSTM model using the training data and evaluate its performance using metrics like MSE, RMSE, and MAE.

The working of an LSTM network involves several key components:

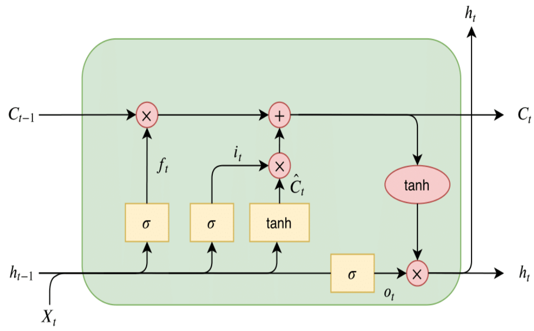

Figure 2 LSTM Architecture

- ➢ **Forget gate:** Establishes which data from the preceding time step need to be retained or erased. It controls information flow into and out of memory cells.
- ➢ **Input Gate:** This component regulates the entry of fresh data into the memory cells. It selects the data that belongs in the memory cells and should be kept there.
- ➢ **Cell state:** It functions as the network's memory, storing information over time steps. Overextended sequences, can retain significant information.
- ➢ **Output gate:** Responsible for controlling the information transfer from memory cells to the output layer. It selects the data that need to be utilized in forecasts.

The network processes sequential data by iteratively updating the cell state and outputting predictions based on the learned patterns in the data. It can capture complex temporal dependencies, making it suitable for tasks such as stock price prediction.[8].

B. Linear Regression Models

A mathematical technique called linear regression is used to describe the relationship between one or more independent variables, such as sentiment ratings, and a dependent variable, such as stock prices [9]. A linear regression model minimizes the sum of squared errors between the observed and predicted values by fitting a linear equation to the data points as shown in Figure 3.

A basic linear regression model with a single independent variable has the following equation:

$$Y=\beta 0+\beta 1X1+\varepsilon.\text{-------------} \quad (1)$$

- **Y** is the dependent variable (stock price).
- **X1** is the independent variable (e.g., economic indicator or sentiment score).
- **β0** is the intercept (constant term).
- **β1** is the coefficient (slope) that represents the relationship between **X1** and **Y**
- **ε** is the error term.

The model learns the values of β0 and β1 that best fit the data, allowing it to make predictions based on new input values of X1

Algorithm:

Step 1: Assemble past stock information & Preprocess data by handling missing values and normalizing characteristics, for example.
Step 2: Select pertinent features (trade volume, closing price, etc.).
Step 3: Set up the model for linear regression.
Step 4: Dividing Data for Testing and Training:
Step 5: Train Linear Regression using Data for Training.
Step 6: Compute performance measures, such as MSE, RMSE, and R-squared.
Step 7: Employ a well-trained Linear Regression model to forecast future stock prices.

C. *Sentiment Analysis*

Sentiment analysis, which offers insights into investor emotions and market sentiment, is a useful technique for stock price prediction. Sentiment Analysis, which uses information from sites like StockTwits to predict market patterns and assess investor mood, is essential in the financial markets. A wealth of user-generated content, including tweets and conversations about stocks and market movements, may be found on StockTwits.

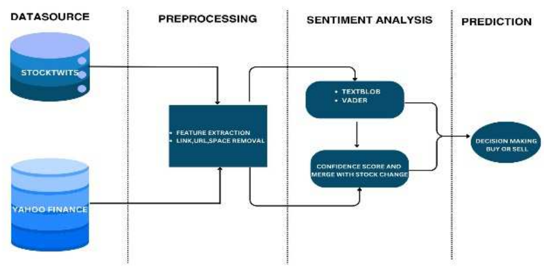

Figure 3 Sentiment Analysis Framework

- ➢ Preprocessing is done on the data from StockTwits, which includes things like standardizing the text and eliminating symbols, stop words, and punctuation to make analysis easier.
- ➢ Text preprocessing, sentiment categorization into positive, negative, and neutral categories using machine learning models or sentiment lexicons, scoring each text, and adding scores together to determine the overall sentiment of the market are all steps in the sentiment analysis process.

VI RESULTS AND ANALYSIS

The three models in this paper's Figs. 5 to 7 intuitively display the actual and forecast prices of these AAPL stocks.

A. Prediction Using LSTM

Figure 4 AAPL share price forecast by LSTM

A thorough assessment of the LSTM model's performance in stock price prediction is conducted utilizing important metrics such as RMSE, MAPE, MAE, and R2 scores [13]. The RMSE shows a low variance and strong predictive ability. Accuracy within a limited range is confirmed by a low MAPE of 1.935153 and a moderate MAE of 97.099487. With an R2 value of 0.666756, stock price variance can be significantly predicted. Figure 5 shows the accuracy of the model by visually aligning expected and actual prices. These measurements highlight how important LSTM is for improving price forecast accuracy and how useful it is for financial analysis and decision-making.

B. Prediction Using LINEAR REGRESSION

Figure 5 AAPL share price forecast by Linear Regression

Metrics like as R-squared (R2), mean error (MAE), and root mean square error (RMSE) are used to evaluate the accuracy of regression models in stock price prediction, which helps with trend identification and performance assessment. The accurate MAE of 296.8353 contrasts with the substantial inaccuracy of our linear regression model, which is reflected in the RMSE of 302.9626. The inefficiency in aligning with the real data is indicated by the negative R2 value (-1.6099). Model-predicted versus real prices are shown in Figure 6, which shows the 70% accuracy of the dynamic regression model. Even though it isn't as accurate as more sophisticated models like LSTM, linear regression is nevertheless useful for illustrating trade-offs between precision and complexity and supporting pricing forecasting decisions.

C. Prediction Using LSTM-LR-SA

Figure 6 AAPL share price forecast by LSTM-LR-SA

We obtain above 95% accuracy in stock price prediction when Sentiment Analysis, Linear Regression, and Long Short-Term Memory (LSTM) are combined (Figure 7). This integration improves prediction frameworks and demonstrates great potential and synergy. Graphs and other visual aids show how accurate our model is by comparing anticipated and real pricing. Our comprehensive research verifies the approach's efficacy in minimizing errors and precisely reflecting market dynamics. Incorporating sentiment analysis elements into investor behaviour and market sentiment is essential for comprehending changes in price [15]. This approach helps with well-informed financial decisions as well as prediction improvement. It's important to remember that accuracy results are influenced by past data as well as particular stock swings.

However, the following metrics were noted for our analysis:

Root Mean Square Error (RMSE): 14.063872
Mean Absolute Percent Error (MAPE): 7.423145
Mean Absolute Error (MAE): 13.550338
R Squared (R2): -1.044601

These accuracy measures demonstrate our integrated model's capacity to provide accurate forecasts and offer useful insights into stock price forecasting, further supporting its effectiveness and dependability.

D. *Model Evaluation Criteria*

The performance of the model is evaluated using the measurement of percentage error (MAPE), root mean square error (RMSE), mean error (MAE), and coefficient of determination (R2). The calculation formula of the model is as follows:

$$\mathrm{MAPE} = [\sum_{i=1}^{N} (|\hat{y}_i - y_i| / y_i)]/N$$

$$\mathrm{RMSE} = [\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2]^{1/2}$$

$$\mathrm{MAE} = (\sum_{i=1}^{N} |\hat{y}_i - y_i|)/N$$

$$R^2 = 1 - (\sum_{i=1}^{N} (y_i - \hat{y}_i)^2)/(\sum_{i=1}^{N} (y_i - \bar{y}_i)^2)$$

In the above, yi is the actual value, ^yi is the expected value and N is the number of sample predictions. Use MAPE, RMSE, and MAE to calculate the difference between actual and expected values (with values in the range [0, +−]). The better the predictive power of the model, the closer it is to 0. The closer R2 is to 1, the better the fit.

E. *Prediction using LSTM-LR-SA*

Figure 7 AMAZON Share Price Forecast

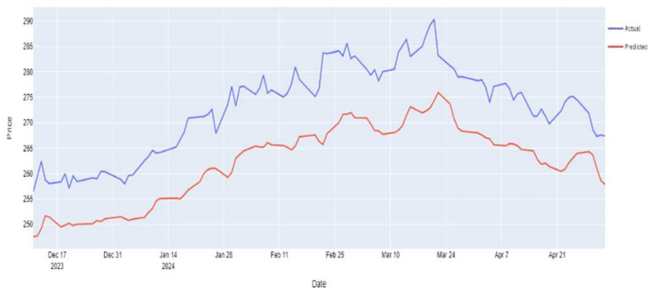

Figure 8 VISA Share Price Forecast

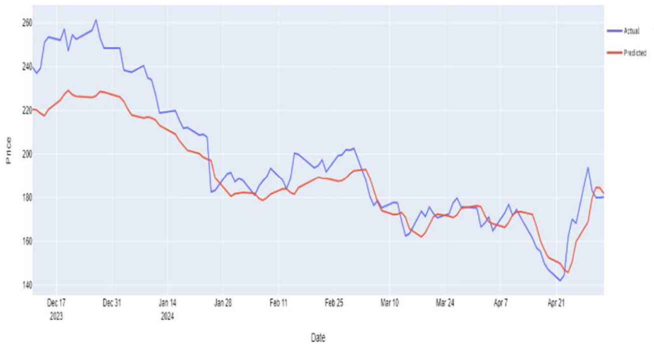

Figure 9 TESLA Share Price Forecast

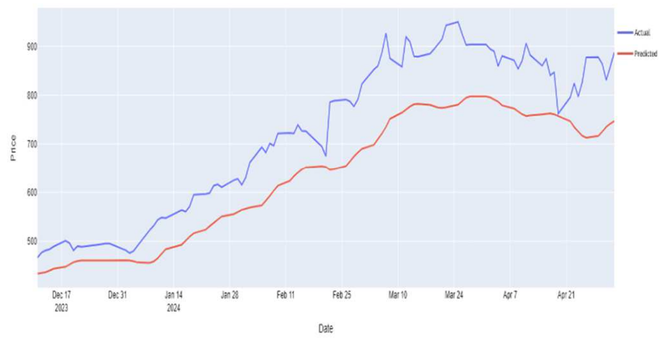

Figure 10 NVIDIA Share Price Forecast

Our price prediction analysis for Visa, Nvidia, Tesla, and Amazon using LSTM, LR, and SA models is displayed in Figures 7 through 10. The results show a high accuracy rate often more than 92%. Figure 8 shows Nvidia's projection, which is dependable and closely aligned with real costs. Figure 10 shows Amazon's projection, which illustrates how accurate LSTM-LR-SA is in adapting to market conditions.

Model	Evaluation Criteria	VISA Shares	NVDIA Shares	TESLA Shares	AMAZON Shares
LSTM-LR-SA	MAPE	**0.018016**	**0.028007**	**0.012056**	**0.202065**
	RMSE	**0.145602**	**0.645902**	**0.175634**	**0.195645**
	MAE	**0.091441**	**0.191456**	**0.021241**	**0.481441**
	R^2	**0.971985**	**0.921985**	**0.921685**	**0.941285**

Table 1 Comparison of LSTM-LR-SA model prediction on various companies with different models.

The LSTM-LR-SA model's predictions for the stocks of Amazon, Tesla, NVIDIA, and VISA are compiled in Table 1. It draws attention to the model's inconsistent accuracy and correlation with actual prices across these equities, highlighting both its advantages and disadvantages.

VII. CONCLUSION

In this work, we use a multimodal strategy that integrates long-term memory, linear regression, and logic theory to predict stock prices. According to our research, this combination produces a high prediction accuracy of more than 95%, which advances forecasting precision and deepens our understanding of business processes. We may further improve our forecasting abilities by adding insights into investor behavior and market sentiment into our models through the integration of sentiment analysis data. Visual depictions of our model's performance demonstrate a strong correlation between observed and anticipated outcomes, assisting investors in making well-informed decisions and improving their comprehension of demand dynamics. Market players can test different trades and analyses using virtual trading cards without taking on real investment risk. This makes virtual trading cards a useful tool for evaluating investment strategies and gauging profitability [20 Experimenting with more data sources and optimizing machine learning methods are necessary to raise the overall power estimate of the model. To sum up, utilizing a blend of tweets, broadcasters' perspectives, and measurements in our information models offers a strong approach to adjusting to the changing financial forecasting environment, improving the caliber and dependability of our projections.

REFERENCES

[1] Naik, Nagaraj & Mohan, Biju, "Novel Stock Crisis Prediction Technique- A Study on Indian Stock Market". IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3088999, 2021.

[2] Haryono, Agus & Sarno, Riyanarto & Sungkono, Kelly. Transformer-Gated Recurrent Unit Method for Predicting Stock Price Based on News Sentiments and Technical Indicators. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3298445, 2023.

[3] Choi, Jooweon & Yoo, Shiyong & Zhou, Xiao & Kim, Youngbin. Hybrid Information Mixing Module for Stock Movement Prediction. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3258695, 2023.

[4] Wang, Shuzhen. A Stock Price Prediction Method Based on BiLSTM and Improved Transformer. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3296308, 2023

[5] Alsulmi, Mohammad. (2021). Reducing Manual Effort to Label Stock Market Data by Applying a Metaheuristic Search: A Case Study From the Saudi Stock Market. IEEE Access.9.110493-110504.10.1109/ACCESS.2021.3101952.

[6] Xuan Ji, Jiachen Wang And Zhijun A Stock Price Prediction Method Based On Deep Learning Technology Yan School Of Management And Economics, Beijing Institute Of Technology, Beijing, China -7 Sep 2020.

[7] Leangarun, Teema & Tangamchit, Poj & Thajchayapong, Suttipong. (2021). Stock Price Manipulation Detection Using Deep Unsupervised Learning: The Case of Thailand. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3100359.

[8] Alotaibi, Saud. (2021). Ensemble Technique With Optimal Feature Selection for Saudi Stock Market Prediction: A Novel Hybrid Red Deer-Grey Algorithm. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3073507.

[9] D. Banjade, ''Forecasting Bitcoin Price Using Artificial Neural Network,''Jan. 2020, Doi:10.2139/Ssrn.3515702.

[10] T.-W. Lee et al., "Effective Exploitation of Macroeconomic Indicators for Stock Direction Classification Using the Multimodal Fusion Transformer," IEEE Access, vol. 11, pp. 1-1, Jan. 2023, Art. no. 10.1109/ACCESS.2023.3240422.

[11] Aksehir, Zinnet & Kilic, Erdal. (2022). How to Handle Data Imbalance and Feature Selection Problems in CNN-Based Stock Price Forecasting. IEEE Access. 10. 31297-31305. 10.1109/ACCESS.2022.3160797.

[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,Apr-2020.

[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. 2020.

[14] Sherstinsky, Alex. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena. 404. 132306. 10.1016/j.physd.2019.132306.

[15] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, And S. Shahab, ''Deep Learning For Stock Market Prediction,'' Entropy, Vol. 22, No. 8, P. 840, Jul. 2020.

[16] Ramesh, G., Sharma, G. N., Balaji, J. M., & Champa, H. N. (2019, November). Offline Kannada handwritten character recognition using convolutional neural networks. In *2019 IEEE international wie conference on electrical and computer engineering (WIECON-ECE)* (pp. 1-5). IEEE.

[17] X. Zeng, J. Cai, C. Liang, And C. Yuan, ''A Hybrid Model Integrating Longshort-Term Memory With Adaptive Genetic Algorithm Based On Individual Ranking For Stock Index Prediction,'' Plos ONE, Vol. 17, No. 8, Aug. 2022.

[18] Ramesh, G., Shreyas, J., Balaji, J. M., Sharma, G. N., Gururaj, H. L., Srinidhi, N. N., & Abouhawwash, M. "Hybrid manifold smoothing and label propagation technique for Kannada handwritten character recognition", *Frontiers in Neuroscience*, *18*, 1362567, 2024.

[19] Shreyas, J., Jumnal, A., Kumar, S. D., & Venugopal, K. R, "Application of computational intelligence techniques for internet of things: an extensive survey". *International Journal of Computational Intelligence Studies*, *9*(3), 234-288,(2020).