A Deep Learning-Based Approach for Stock Price Prediction Using Bidirectional Gated Recurrent Unit and Bidirectional Long Short Term Memory Model

Conference Paper · October 2021

DOI: 10.1109/GCAT52182.2021.9587895

Md. Ebtidaul Karim
Khulna University of Engineering and Technology

A Deep Learning-Based Approach for Stock Price Prediction Using Bidirectional Gated Recurrent Unit and Bidirectional Long Short Term Memory Model

Md. Ebtidaul Karim
Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh
ebtikarim@ece.kuet.ac.bd

Sabrina Ahmed
Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh
sabrinaahmed.kuet@gmail.com

***Abstract*—Stock market investments make up a significant portion of any country's growing economy. The rise or decline in the share price has a significant impact on the investor's profit. Accurate prediction of the dynamic and volatile stock price movements can enhance the confidence of investors in the share market and hence can increase the volume of investment. Due to the chaotic, non-linear, and complex nature of the share price pattern, traditional time-series analysis exhibits poor performance. In this paper, we propose a new deep learning-based forecasting framework, Bidirectional Gated Recurrent Unit (BiGRU) incorporated with an external activation layer that utilizes both the forward and backward propagation feature of the Recurrent Neural Network (RNN). We compare our proposed model with the prevalent bidirectional RNN model, known as Bidirectional Long Short Term Memory (BiLSTM). We implemented both models on three different datasets collected from the NIFTY-50 index. We considered five different evaluation metrics to evaluate the model performance. For the different combinations of hidden layers and data, our model shows better performance in comparison to BiLSTM. The BiGRU model also shows better stability in comparison to BiLSTM, with smaller trainable parameters. It can also accurately forecast stock price for a greater prediction window, in comparison to BiLSTM. Our model can also predict all the sudden spikes on 1000 days ahead prediction precisely.**

***Index Terms*—stock prediction, time-series analysis, deep learning, recurrent neural network, bidirectional gated recurrent unit.**

I. Introduction

Anticipation of future occurrences or events by analyzing the series of historical data indexed in time order is commonly known as time series forecasting. It is a vast area for exploration. It encompasses a wide range of topics including industries, health care facilities, cloud workloads, finances, stock markets, to name a few. A time-series data analysis aids in detecting patterns, trends, and periods or cycles that exist in the data. It is very crucial in the field of stock market price prediction. Early knowledge of ever-shifting stock prices gives an upper hand to the investors in decision-making. There are at present 60 major stock exchanges all over the world. Over the years, the global market capitalization of the stock exchanges has risen from \$ 2.5 trillion in 1980 to \$ 68.65 trillion at the end of 2018 [1]. Even in the crisis period of the global pandemic, stock markets remain functional and are playing an active role in supporting the economy. By the end of 2020, global market capitalization has passed the \$ 100 trillion mark [2]. Strong confidence in venturing the market and a sustainable profit return over the investment keep the stock exchanges ticking. For this purpose, stockholders need a heads up on ever-changing stock price trends. It can be achieved by developing a reliable stock price prediction framework.

But forecasting of financial market is a challenging task due to the highly non-linear, mystifying, dynamic, and noisy in nature financial time series data. Earlier various statistical and classical computation methods were used in time-series analysis of stock market [3]–[5]. But it fails to achieve precise prediction accuracy due to the application of linear prediction procedure on non-linear data. Traditional methods based on linear regression struggles to accurately forecast the financial data due to the complex nature of the time series models. Auto-regression (AR), Auto-regressive Moving Average (ARMA), Auto-regressive Integrated Moving Average (ARIMA), and Simple Exponential Smoothing (SES) are an example of a few of the common traditional time series prediction methods.

Machine learning aims to uncover the underlying framework based on which data are produced. It realizes the linear and non-linear models that reside in these data by detecting new patterns in historical data. It provides robustness and boosts the ability of the prediction model to deal with the complex data [6], [7]. Subsequently, neural network-based models are used in time series forecasting. In comparison to statistical approaches, the neural network model performs better in processing the complicated observations seen in time series data [8].

Deep learning is one of the most efficient and popular tools prevalent in the field of time series analysis. Due to its capability of automatic learning, feature extraction, and ability to identify patterns in raw data that span a large number of sequences; Deep Neural Networks (DNNs) are gaining

978-1-7281-8402-9/21/\$31.00 ©2021 IEEE

substantial popularity in this field. Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU) are one of the most commonly used deep learning-based approaches. They are used in numerous time series applications, such as stock predictions, cloud workload predictions, weather forecasting, healthcare analysis, and so on [9], [10]. For stock market forecasting, various deep learning-based models have been proposed earlier. These models show superior performance in comparison to statistical and traditional machine learning models [11]. In the field of stock market price forecasting various deep learning models based on CNN, LSTM, GRU, and Bidirectional Long Short Term Memory (BiLSTM) have been proposed earlier. In some papers, hybrid models based on Recurrent Neural Networks (RNNs) are used to predict the stock price movements.

However, to the best of our knowledge, none of the papers so far has shown a comparative study between two bidirectional RNN prediction models, namely Bidirectional Gated Recurrent Unit (BiGRU) and BiLSTM for stock prediction. In this paper we propose a novel deep learning-based forecasting model that uses BiGRU along with an immediate activation layer and a fully connected neural network at the end. For this purpose, across the full training data set, we use an overlapping sliding window approach strategy. A systematic comparison between both the bidirectional RNNs, BiLSTM and BiGRU has been shown in stock prediction for the first time. Our model shows better performance in comparison to BiLSTM and yields precise forecasting with minimum error generation and trainable parameters. The integrity of our model is also preserved in multi-step ahead forecasting.

We have sorted the rest of the paper as follows. Researches conducted so far related to our work are discussed in section II. A detailed discussion about the methodology implemented for this work along with the evaluation tools is done in section III. Section IV compares the results obtained from both the prediction models. Conclusion of the whole work along with possible future extensions are discussed in section V.

II. Related Works

For successfully predicting the stock future prices, in [6] an optimized Least Square Support Vector Machine (LS-SVM) algorithm is proposed which integrates Particle Swarm Optimization (PSO) algorithm. PSO algorithm helps to solve the over-fitting and local minima problem by selecting a combination of best free parameters and helps to improve the prediction accuracy. A comparative study between deep learning algorithms with various machine learning algorithms is shown in [12]. Here, iShares MSCI United Kingdom exchange-traded fund data for the daily close stock price is investigated. They compared the forecasting accuracy of the deep learning-based algorithm LSTM to three machine learning algorithms: Artificial Neural Network (ANN), Support Vector Regression (SVR), and Random Forest (RF). Study results show that LSTM outperforms other machine learning algorithms in terms of prediction accuracy.

DNN architectures are viewed as being able to capture hidden dynamics that lie within data and can anticipate future events precisely [13]. Here, instead of fitting the data to a certain model, deep learning architectures such as RNN, CNN, and LSTM are used to uncover the underlying dynamics in the data. It also shows that deep learning-based approaches can predict the future stock price values more accurately than the classical traditional time series forecasting model, ARIMA. A 2-Directional 2-Dimensional Principal Component Analysis $(2D)^2$PCA is implemented, along with DNN in [14]. Here, $(2D)^2$PCA is utilized for reducing the dimensions and DNN is used for forecasting. This proposed model is compared with state of the art 2-Directional 2-Dimensional Principal Component Analysis $(2D)^2$PCA method accompanied by Radial Basis Function Neural Network (RBFNN). Although the proposed model performs better, still there is a lot of scopes for improvement in predicting the upcoming stock prices. Here, the error generated while forecasting in both the model is relatively high.

Apart from the models that rely on individual RNNs, different hybrid models are also proposed to predict the stock prices in [11], [15]. These models combine different RNN models to optimize and improve forecasting performance. In [15], LSTM and GRU layers are stacked together along with dropout layers in between to form a hybrid prediction model. This hybrid model illustrates better performance than individual RNN models. In [11], a hybrid model is introduced which combines BiLSTM and GRU layers on top of a 1-D CNN layer. This model performs better than individual LSTM, GRU, BiLSTM, and Neural Network with 3 hidden layers in both single and multi-step ahead predictions. The hybrid layers with proper optimization perform better than the individual models, but it adds more complexity along with large trainable parameters. So prediction cost is high. A systemic comparison between LSTM and BiLSTM is conducted in [1]. It shows that BiLSTM performs better than LSTM on predicting the Google Stock Market data by using the same tuning parameters.

Nonetheless, to the best of our knowledge, none of the works reported thus far have shown a comparison between BiGRU and BI-LSTM in terms of stock price prediction performance improvement. Both of the models utilize the forward and backward propagation of the RNN and have the ability to retain both past and future information, making it suitable for long-term ahead prediction. Thus a systemic comparative study between both the models may help the researchers in the future in selecting a suitable candidate for accurate long-time ahead stock prediction.

III. Methodology

A. Gated Recurrent Unit

A BiGRU Layer is a combination of two GRU layers placed in such a manner so that one contributes to forward propagation and the another in backward propagation. GRU is a sophisticated variant of RNN that includes a gating mechanism. It can avoid the vanishing gradient problem that occurs with standard RNNs. The internal architecture of GRU is shown in Fig. 1.

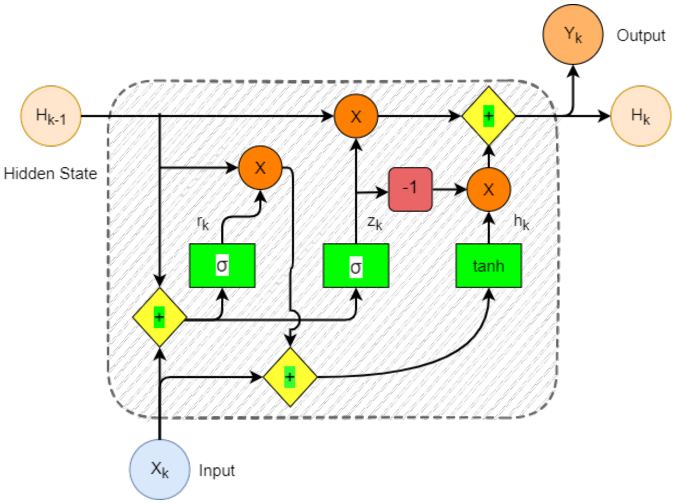


Fig. 1: The internal cell architecture of GRU

A GRU cell consists of two gates namely, reset gate and update gate. Gate is a method of determining whether or not data can enter the cell state. A sigmoid function σ and a point-wise multiplication algorithm are combined to form Gate. The sigmoid function can produce any value between 0 and 1. This number governs data passage in such a way that an estimated zero restricts data entry and an estimated one allows the entry of data. Determination of the quantity of prior data (from previous time steps) that can be carried over to the next time step is assisted by the update gate. The reset gate, on the other hand, is used to define how much data from the past can be deleted. Apart from two sigmoid gates, GRU has a candidate activation vector h_k, generated from the $tanh$ activation function. This vector along with the update gate vector, and previous hidden state value generates the output and the value of the next hidden state. Here, both the next hidden state value and the value of output are the same. The internal operations that take place inside a GRU cell can be expressed as:

$$\text{Update Gate:} \quad z_k = \sigma\left(W_z X_k + U_z H_{k-1} + b_z\right) \tag{1}$$
$$\text{Reset Gate:} \quad r_k = \sigma\left(W_r X_k + U_r H_{k-1} + b_r\right) \tag{2}$$

Here, X_k is the input data at time k and H_{k-1} is the previous hidden state value. The candidate activation value and the final output of the GRU are given below.

$$h_k = tanh\left(W_h X_k + U_h\left(r_k \otimes h_{k-1}\right) + b_h\right) \tag{3}$$
$$H_k = z_k \otimes h_k + (1 - z_k) \otimes H_{k-1} \tag{4}$$

B. Bidirectional Gated Recurrent Unit

The internal structure of a BiGRU layer is shown in Fig. 2. Analyzing Fig. 2, we can clearly say that BiGRU is a

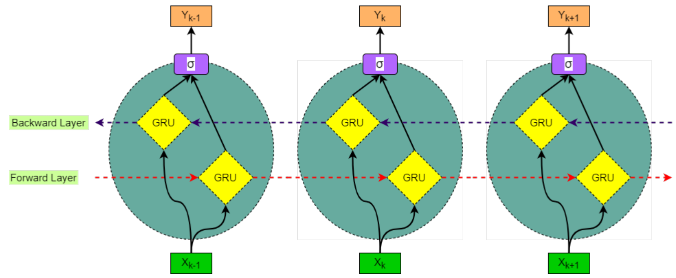


Fig. 2: The internal structure of a BiGRU Layer

modified version of two GRU layers placed together. In this case, one GRU unit handles inputs in the forward direction, while the other unit handles inputs in the reverse direction. Even though the number of trainable parameters grows in BiGRU, there is a significant rise in the amount of information available to the network. As a result, the network has a greater understanding of the context and can predict accurately the long-term sequence of data.

The forward layer GRU present in BiGRU for input X_k at time k follows the Eqn. 1, 2, 3, and 4 sequentially and hidden state value is passed from left to right. Backward layer GRU in BiGRU also follows the same pattern as the forward layer GRU, but here hidden state value is carried from right to left. As a result, for input X_k at time k, we obtain the two outputs. One uses the past value and another uses the future value. Concatenation of these two different outputs gives rise to the final output Y_k.

C. System Architecture

Multiple hidden layers and a significant number of hidden units per layer are used in a deep learning model to retrieve higher-level features from input raw data progressively. It facilitates the opportunity to learn the non-linear patterns embedded inside the raw data. Here, in this work, we propose an optimized BiGRU model equipped with a subsequent external activation layer to accurately predict future stock prices. Fig. 3 illustrates the proposed model with 03 hidden layers. This model initially takes the raw input data and activates a MinMaxScaler function for normalizing the data in the range of 0 to 1. It helps in faster convergence and learning process. After preprocessing the data, refurbished data is split into two parts for training and testing the model. 80 percent of the total data is used for training the model. The rest of the data is used for testing the model performance. Data is passed into the BiGRU layer. We used one, two, and three successive BiGRU layers stacked together for analyzing the model performance. Each BiGRU layer has 64 hidden units in both the forward and backward direction, resulting in a total of 128 hidden units per layer.

Each BiGRU layer is followed by an external activation layer 'tanh'. This hyperbolic tangent activation function takes any real value and converts the BiGRU layer normalized output data in the range of 0 to +1 into a stretched half S-shaped output data in the range of 0 to +1. Here, the sensitivity

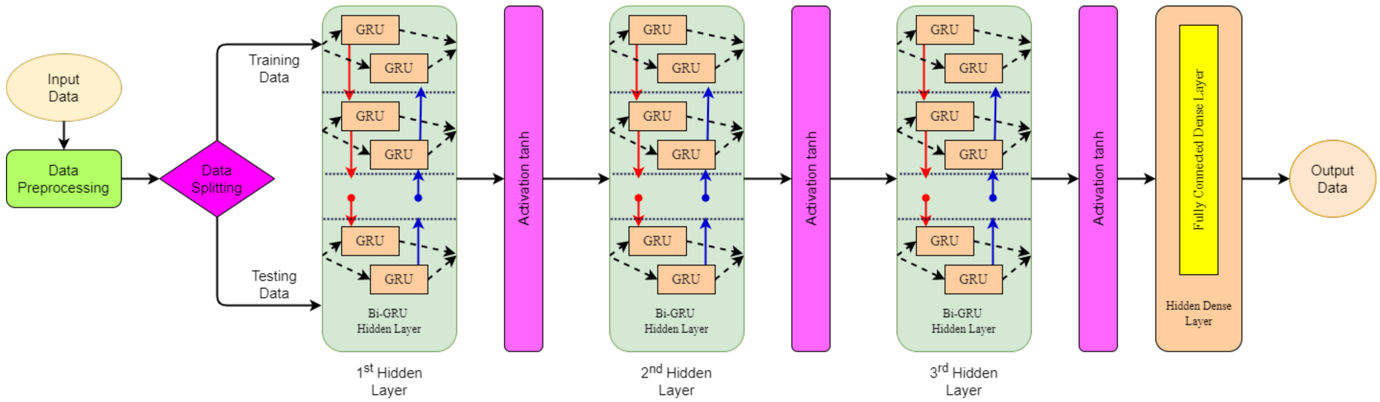


Fig. 3: The proposed BiGRU model with 3 hidden layers

of the smaller data points is enhanced to a great extent. As a result, the model can predict even the smallest data spikes. The BiGRU layer along with the layer of activation forms the basic core of the prediction model. The resulting output data from the BiGRU-activation layer is passed into a fully connected dense neural network. It has one hidden unit. Hyperparameters of our proposed model are set after proper tuning. For this purpose, we employ the grid search technique.

D. Evaluation Metrics

To precisely estimate the forecasting accuracy of our model, we used various evaluation metrics. System performance of our proposed model in comparison to BiLSTM is conducted by analyzing Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Logarithmic Error (MSLE), and coefficient of determination (R^2). The equations used for calculating the evaluation metrics are as follows:

$$\text{MSE} = \frac{1}{k}\sum_{j=1}^{k}(y_j - \overline{y_j})^2 \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{k}\sum_{j=1}^{k}(y_j - \overline{y_j})^2} \tag{6}$$

$$\text{MAE} = \frac{1}{k}\sum_{j=1}^{k}|y_j - \overline{y_j}| \tag{7}$$

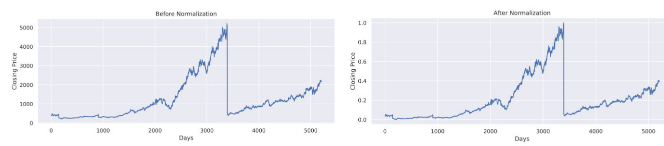
$$\text{MSLE} = \frac{1}{k}\sum_{j=1}^{k}\left(\log\left(y_j+1\right) - \log\left(\overline{y_j}+1\right)\right)^2 \tag{8}$$

$$\text{R}^2 = 1 - \frac{\sum_{j=1}^{k}(y_j - \overline{y_j})^2}{\sum_{j=1}^{k}\left(y_j - \frac{1}{k}\sum_{j=1}^{k} y_j\right)^2} \tag{9}$$

Here, in Eqn. 5, 6, 7, 8, and 9; y_j is the actual closing stock price and $\overline{y_j}$ is the corresponding predicted value for the j^{th} day respectively. Total number of the sample is k. Smaller MSE, RMSE, MAE, and MSLE values indicate better forecasting accuracy. Again, the closer the R^2 value is to 1, better the data fitted, and higher the model performance.

TABLE I: Hyperparameters of our proposed Model

Hyper parameters	Values
History Window Size	60
Batch Size	64
Hidden BiGRU Layer Neurons	128
Number of Hidden BiGRU Layers	{ 1,2, and 3}
Learning Rate	0.001
1^{st} Moment Exponential Decay Rate, β_1	0.9
2^{nd} Moment Exponential Decay Rate, β_2	0.999
Numerical Stability, ϵ	10^{-7}
Iteration	100

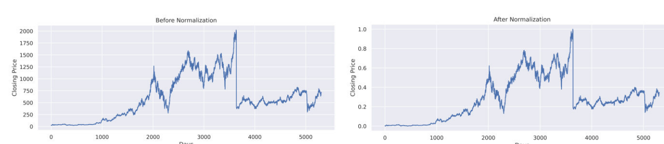


(a) Before Normalization (b) After Normalization

Fig. 4: Closing Stock Price of ASIAN PAINT

E. Description of Dataset and Preprocessing

In this work, we used real datasets containing stock price data collected from the Kaggle dataset repository. The information comes from the National Stock Exchange (NSE) of India, and it includes the price history and trading volumes of the fifty equities that make up the NIFTY 50 index [16]. All datasets are on a daily basis, with pricing and trading values spread from 1st January 2000 to 30th April 2021. From this dataset, we used stock price data of ASIAN PAINT, AXIS BANK, and BRITANNIA for training and validating our proposed model. Same data are also used to evaluate the



(a) Before Normalization (b) After Normalization

Fig. 5: Closing Stock Price of AXIS BANK

TABLE II: Comparison of evaluation metrics of BiGRU and BiLSTM for forecasting of closing ASIAN PAINT stock values

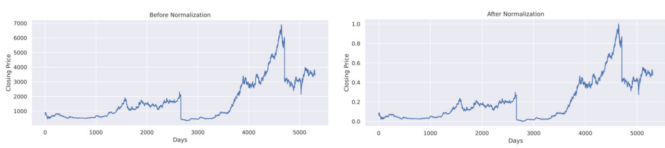
Hidden Layers	**MSE**		**RMSE**		**MAE**		**MSLE**		**R^2**	
	BiGRU	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***
01	**0.0000244**	0.0000259	**0.00494**	0.00509	**0.00325**	0.00337	**0.000015**	0.0000159	**0.9936**	0.9932
02	**0.0000247**	0.0000257	**0.00497**	0.00507	**0.00329**	0.0033	**0.0000152**	0.0000158	**0.9935**	0.9933
03	**0.0000251**	0.000026	**0.00501**	0.0051	**0.00331**	0.00341	**0.0000154**	0.0000159	**0.9934**	0.9932

TABLE III: Comparison of evaluation metrics of BiGRU and BiLSTM for forecasting of closing AXIS BANK stock values

Hidden Layers	**MSE**		**RMSE**		**MAE**		**MSLE**		**R^2**	
	BiGRU	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***
01	**0.0000453**	0.0000478	**0.00673**	0.00691	**0.0047**	0.00479	**0.0000282**	0.0000299	**0.9862**	0.9854
02	**0.000045**	0.0000476	**0.00671**	0.0069	**0.00468**	0.0048	**0.0000281**	0.0000297	**0.9863**	0.9855
03	**0.0000461**	0.0000476	**0.00679**	0.0069	**0.00472**	0.00476	**0.0000288**	0.0000297	**0.9859**	0.9855

TABLE IV: Comparison of evaluation metrics of BiGRU and BiLSTM for forecasting of closing BRITANNIA stock values

Hidden Layers	**MSE**		**RMSE**		**MAE**		**MSLE**		**R^2**	
	BiGRU	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***	***BiGRU***	***BiLSTM***
01	**0.0002866**	0.0003278	**0.01693**	0.0181	**0.0072**	0.0091	**0.0001118**	0.0001268	**0.9892**	0.9877
02	**0.0002859**	0.000296	**0.01691**	0.0172	**0.00715**	0.00789	**0.0001116**	0.000116	**0.9892**	0.9889
03	**0.0002888**	0.0003017	**0.01699**	0.01737	**0.00743**	0.00771	**0.0001126**	0.0001186	**0.9891**	0.9886



(a) Before Normalization (b) After Normalization

Fig. 6: Closing Stock Price of BRITANNIA

performance of our model in comparison to the widely used BiLSTM model. In this work, we predicted the daily closing price of the stock in each case.

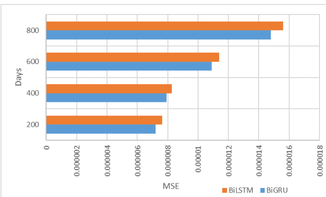
For normalizing the raw closing price data in the range of 0 to +1 by using the MinMaxScaler function. The preprocessed data along with the original raw data for the three different companies are shown in Fig. 4, Fig. 5, and Fig. 6. Analyzing the closing prices of the stock on daily basis we observe the non-linear, periodic, and non-periodic nature of the data. For extracting the hidden features from this data, we will employ a deep learning approach.

IV. Result Analysis and Discussions

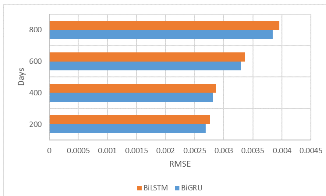
For performing this experiment, we used the Google Colaboratory platform. We used Python Keras and TensorFlow library to build and evaluate our model. Our proposed model, after proper hyperparameter tuning was trained and validated for three different closing stock prices from the NIFTY 50 index. Using the same set of hyperparameters and the same dataset with 80:20 training and testing data split we tuned a BiLSTM model. This model is used as a baseline model in this work. A comparison between our BiGRU model and the BiLSTM model in terms of different evaluation metrics with the change of hidden layers for Asian Paint stock price data is shown in Table II. Here, we observe that all the forecasting error is less in BiGRU in comparison to BiLSTM Again R^2 value is higher in BiGRU in comparison to BiLSTM indicating better regression line fit. Change of hidden RNN layer changes the forecasting performance, but BiGRU still performs better in comparison to BiLSTM with the change of hidden layers.

A similar comparative study in also conducted for comparing both the BiGRU and BiLSTM models with two other closing stock prices data. Table III and Table IV points out the comparison of these two models for AXIS BANK and BRITANNIA stock prices respectively. Analyzing Table II, Table III, and Table IV we can clearly see that, error values and R^2 value is quite different for predicting set of data. More complex data pattern of BRITANNIA stock prices shown in Fig. 6 results in more error while predicting. Meanwhile, the forecasting error is lowest for ASIAN PAINT closing stock price data. The addition of hidden RNN layers changes the forecasting performance of both models. But for each of the datasets and each of the hidden layer combinations, the BiGRU model shows better performance than the BiLSTM model. In all the cases, in comparison to BiLSTM, BiGRU yields smaller MSE, RMSE, MAE, MSLE, and higher R^2 scores.

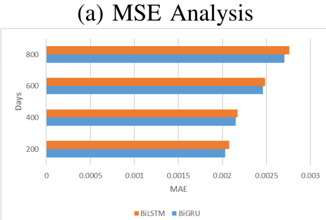
A key performance indicator of a prediction model is the analysis of the effect of the change of prediction length. Larger the size of the prediction window, the greater the probability of degradation of prediction accuracy. The model that upholds its forecasting performance even with the increase of prediction length is more suitable in multi-step prediction. Fig. 7, 8, and 9 illustrates the effect of increase of prediction window size for both the BiGRU and BiLSTM model with two hidden layers. Here, we observe that, with the increase of forecasting days, there is a gradual upward shift of error values. In Fig. 7, with the increase of the number of prediction days, the forecasting error of ASIAN PAINT closing stock prices increases for both the models. But for BiLSTM upward shift of the error curve is more prominent. In Fig. 8, we can observe the same nature of upward shift for AXIS BANK closing stock price. But, in the case of Fig. 9a, 9b, and 9c, we can see that prediction length of 600 days yields more MSE, RMSE, and MAE respectively, in comparison to the prediction length of 800 days for both the models. It happens due to the sudden fall of closing stock prices of BRITANNIA during the prediction window of 400 to 500 days. For each of the datasets and each of the prediction lengths, BiGRU generates smaller forecasting error in comparison to BiLSTM.



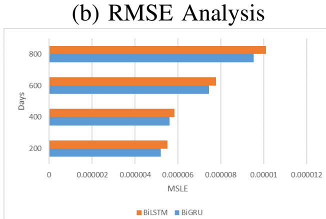
(a) MSE Analysis



(b) RMSE Analysis

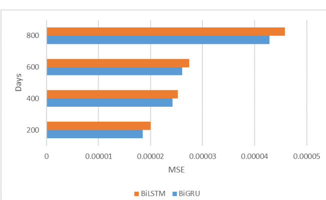


(c) MAE Analysis

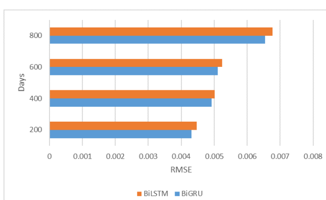


(d) MSLE Analysis

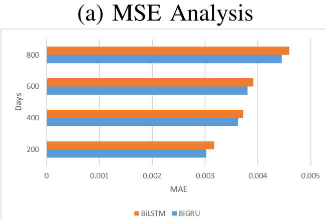
Fig. 7: Analysis of change of prediction length of BiGRU and BiLSTM models with two hidden layers for closing stock prices of ASIAN PAINT



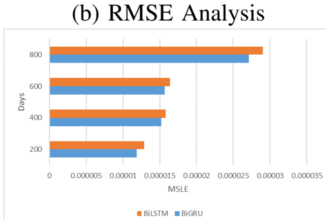
(a) MSE Analysis



(b) RMSE Analysis



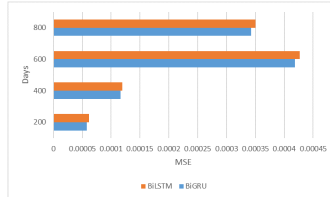
(c) MAE Analysis



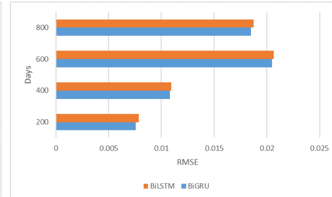
(d) MSLE Analysis

Fig. 8: Analysis of change of prediction length of BiGRU and BiLSTM models with two hidden layers for closing stock prices of AXIS BANK

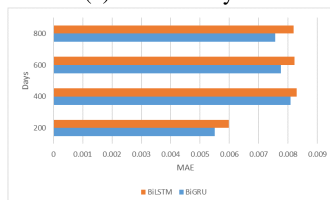
Cumulative Distribution Function (CDF) analysis of different forecasting errors can accurately scrutiny the model harmony and stability. CDF analysis of both BiGRU and BiLSTM model for closing stock price of ASIAN PAINT, AXIS BANK, and BRITANNIA is illustrated in Fig. 10, 11,



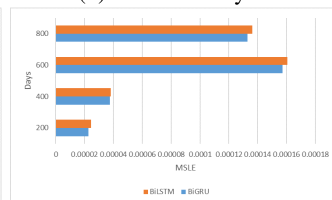
(a) MSE Analysis



(b) RMSE Analysis



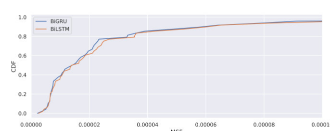
(c) MAE Analysis



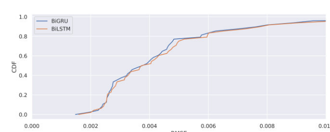
(d) MSLE Analysis

Fig. 9: Analysis of change of prediction length of BiGRU and BiLSTM models with two hidden layers for closing stock prices of AXIS BANK

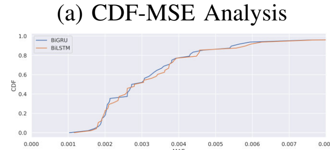
and 12 respectively. The model which generates the 'CDF' curve more closer to the y-axis is more suitable and stable in the forecasting task. In Fig. 10, 11, and 12, we can see that 'blue' CDF curve obtained from BiGRU model is more closer to y-axis in comparison to 'orange' BiLSTM curve. It demonstrates the superior performance of BiGRU in comparison to BiLSTM. Elements of the networks that are affected due to backpropagation are known as trainable parameters. Table V shows the comparison of trainable parameters of both models. Here, we can observe that with the introduction of new hidden layers, the total trainable parameter of the network increases. But, in each scenario, BiGRU has less trainable parameters in comparison to the BiLSTM model. It further solidifies the suitability of BiGRU in real-time stock prediction.



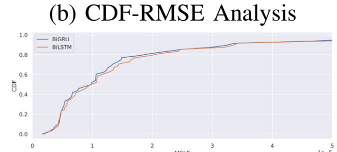
(a) CDF-MSE Analysis



(b) CDF-RMSE Analysis



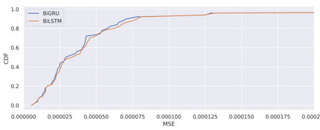
(c) CDF-MAE Analysis



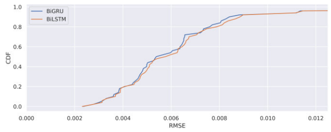
(d) CDF-MSLE Analysiss

Fig. 10: CDF analysis of BiGRU and BiLSTM models with two hidden layers for closing stock prices of ASIAN PAINT

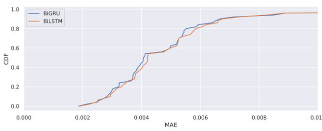
Another important feature of BiGRU is the capability to handle long-term ahead data. Knowledge of accurate fluctuation of stock prices in the future facilitates the opportunity to reshape the investment policy efficiently. The graph of actual and predicted closing stock price values of ASIAN PAINT, AXIS BANK, and BRITANNIA is shown in Fig. 13, 14, and



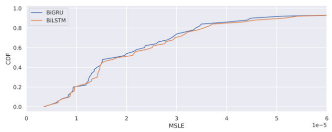
(a) CDF-MSE Analysis



(b) CDF-RMSE Analysis

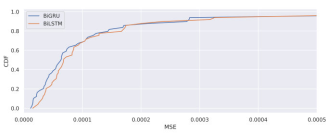


(c) CDF-MAE Analysis

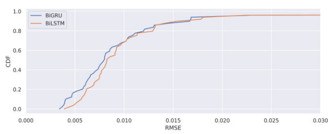


(d) CDF-MSLE Analysis

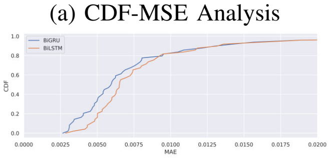
Fig. 11: CDF analysis of BiGRU and BiLSTM models with two hidden layers for closing stock prices of AXIS BANK



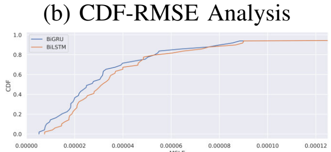
(a) CDF-MSE Analysis



(b) CDF-RMSE Analysis



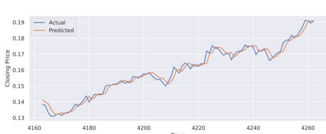
(c) CDF-MAE Analysis



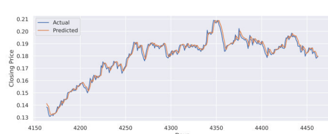
(d) CDF-MSLE Analysis

Fig. 12: CDF analysis of BiGRU and BiLSTM models with two hidden layers for closing stock prices of BRITANNIA

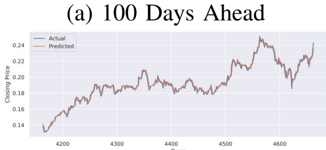
15 respectively. Predicted values are obtained by using the BiGRU model having 2 hidden BiGRU layers incorporated with 2 activation layers. Here, the 'blue' curve indicates the actual value and the 'orange' curve indicates the predicted value over time. Analyzing Fig. 13, 14, and 15, we can say that proposed BiGRU model can accurately predict closing stock prices of next 100, 300, 500, and 1000 days with almost zero deviation. Our model can predict and replicate all the sudden peaks and troughs present on the data almost accurately.



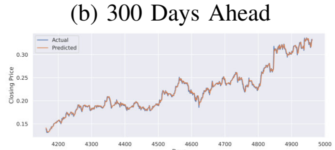
(a) 100 Days Ahead



(b) 300 Days Ahead



(c) 500 Days Ahead

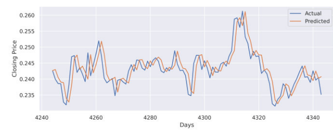


(d) 1000 Days Ahead

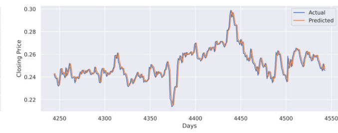
Fig. 13: Predicted versus actual closing stock price of ASIAN PAINT by BiGRU with 2 Hidden Layers

TABLE V: Trainable parameter comparison

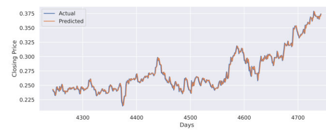
Number of Hidden Layers	BiGRU	BiLSTM
01	**25,857**	33,921
02	**100,353**	132,737
03	**174,849**	231,553



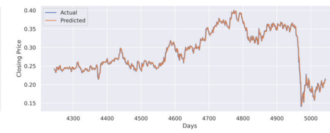
(a) 100 Days Ahead



(b) 300 Days Ahead



(c) 500 Days Ahead

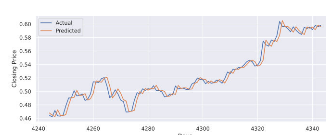


(d) 1000 Days Ahead

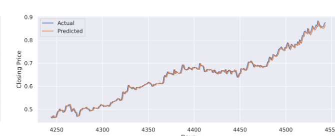
Fig. 14: Predicted versus actual closing stock price of AXIS BANK by BiGRU with 2 Hidden Layers

V. Conclusion and Future Work

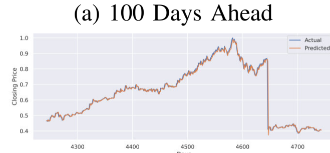
The demand for a reliable and precise forecasting financial market tool among investors is increasing rapidly. In this work, we see that deep learning-based models can successfully extract the long-term dependencies present on time series data. We propose a BiGRU model with an external activation layer for predicting the closing stock price of three different datasets. Using the same set of hyperparameters and the same combination of activation layers we compare the BiGRU model with the state-of-the-art BiLSTM model. For each of the stock datasets and each of the combinations of hidden layers, the result analysis shows that BiGRU performs better than the state-of-the-art BiLSTM model, in terms of MSE, RMSE, MAE, MSLE, and R^2 metrics. The BiGRU model also performs better in comparison to BiLSTM in each step of the change of prediction length analysis. We also can see that the stability of BiGRU is far superior to the BiLSTM model. In comparison to BiLSTM, BiGRU also has lesser trainable



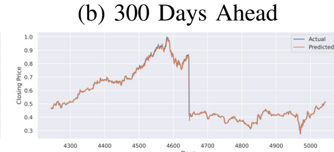
(a) 100 Days Ahead



(b) 300 Days Ahead



(c) 500 Days Ahead



(d) 1000 Days Ahead

Fig. 15: Predicted versus actual closing stock price of BRITANNIA by BiGRU with 2 Hidden Layers

parameters. Our suggested model can also accurately predict long-time-ahead stock prices.

In the future, we want to develop a more robust prediction model which can explore and extract hidden long-term dependencies from a range of datasets, ensemble them, and can predict any time-series data. We are also interested to develop a multi-variant time-series model for stock prediction.

REFERENCES

[1] M. A. I. Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep learning-based stock price prediction using lstm and bi-directional lstm model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*. IEEE, 2020, pp. 87–92.

[2] "Global market capitalisation passed $100 trillion," accessed: 2021-06-21. [Online]. Available: https://www.marketsmedia.com/global-market-capitalisation-passed-100-trillion/

[3] G. S. Atsalakis and K. P. Valavanis, "Surveying stock market forecasting techniques–part ii: Soft computing methods," *Expert systems with applications*, vol. 36, no. 3, pp. 5932–5941, 2009.

[4] P. Li, C. Jing, T. Liang, M. Liu, Z. Chen, and L. Guo, "Autoregressive moving average modeling in the financial sector," in *2015 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. IEEE, 2015, pp. 68–71.

[5] G. Zhang, X. Zhang, and H. Feng, "Forecasting financial time series using a methodology based on autoregressive integrated moving average and taylor expansion," *Expert Systems*, vol. 33, no. 5, pp. 501–516, 2016.

[6] O. Hegazy, O. S. Soliman, and M. A. Salam, "A machine learning model for stock market prediction," *arXiv preprint arXiv:1402.7351*, 2014.

[7] P. D. Yoo, M. H. Kim, and T. Jan, "Machine learning techniques and use of event information for stock market prediction: A survey and evaluation," in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, vol. 2. IEEE, 2005, pp. 835–841.

[8] J. Xue, F. Yan, R. Birke, L. Y. Chen, T. Scherer, and E. Smirni, "Practise: Robust prediction of data center time series," in *2015 11th International Conference on Network and Service Management (CNSM)*. IEEE, 2015, pp. 126–134.

[9] M. N. H. Shuvo, M. M. S. Maswood, and A. G. Alharbi, "Lsru: A novel deep learning based hybrid method to predict the workload of virtual machines in cloud data center," in *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 2020, pp. 1604–1607.

[10] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, "Single layer & multi-layer long short-term memory (lstm) model with intermediate variables for weather forecasting," *Procedia Computer Science*, vol. 135, pp. 89–98, 2018.

[11] M. E. Karim, M. Foysal, and S. Das, "Stock price prediction using bi-lstm and gru based hybrid deep learning approach," in *2021 International Conference on Computing and Technological Solutions with Artificial Intelligence*, (in press).

[12] M. Nikou, G. Mansourfar, and J. Bagherzadeh, "Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms," *Intelligent Systems in Accounting, Finance and Management*, vol. 26, no. 4, pp. 164–174, 2019.

[13] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman, "Stock price prediction using lstm, rnn and cnn-sliding window model," in *2017 international conference on advances in computing, communications and informatics (icacci)*. IEEE, 2017, pp. 1643–1647.

[14] R. Singh and S. Srivastava, "Stock prediction using deep learning," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 18 569–18 584, 2017.

[15] M. A. Hossain, R. Karim, R. Thulasiram, N. D. Bruce, and Y. Wang, "Hybrid deep learning model for stock price prediction," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2018, pp. 1837–1844.

[16] Vopani, "Nifty-50 stock market data (2000 - 2021)," May 2021, accessed: 2021-07-02. [Online]. Available: https://www.kaggle.com/rohanrao/nifty50-stock-market-data