A Deep Learning Based Approach towards Foreign Exchange Rate Prediction for USD/INR Pair

Bharatharaj Babu
Department of Computer Technology
Anna University, MIT Campus
Chennai, India
bharathbabu68@gmail.com

Suriyaa T
Department of Computer Technology
Anna University, MIT Campus
Chennai, India
tsuriyaa@gmail.com

Iraianbu T
Department of Computer Technology
Anna University, MIT Campus
Chennai, India
iraianbut20@gmail.com

Afzal MKT
Department of Computer Technology
Anna University, MIT Campus
Chennai, India
afzalmkt007@gmail.com

***Abstract*—Foreign exchange or FOREX is one of the most crucial trading markets in the world followed by the credit market. FOREX determines the exchange rate of different currencies in the world. It involves buying, and selling currencies in large volumes at current or pre-determined prices. Hence, predicting the exchange rate is a very valuable and important task in the field of currency trading and hence is required. This paper presents a deep learning based approach towards foreign exchange rate prediction using both deep learning models as well as sequence learning models such as ANN, RNN, etc. The trend in the exchange rates will also be identified using visual representation and carrying out some rudimentary exploratory analysis on the data before being fed into the model. The visual representation attempts to present a higher overview on the FOREX prices and establish its trend over the past 10 years. We will also carry out further optimization techniques such as ensemble learning in order to further improvise our performance metrics.**

***Index Terms*—FOREX, RNN, ANN**

I. INTRODUCTION

Today’s world is very much moved by Artificial Intelligence. The modelled systems provide greater prediction of results from training and applied in various fields. In business, the prediction of capital values between countries stands a significant factor over investment. The value of a Dollar in Indian Rupee has seen a gradual increase over the past few decades. But, in the view of smaller scale, the value fluctuates in a range that ends up raising or dropping at the end of the day. The prediction models for foreign currency has given a very much better results in a broder value. But, still the prediction of foreign exchange rate in a narrow scale is expected from many business minds. Our project is built to meet those expectation and analysis of several models. The Dataset for exchange rate between Indian rupee and US Dollar for last 10 years is extracted from inversting.com. The price value for each day besides the open price, high and low price are detailed in each column. Additionally, it has Volume and change % for comparison with multiple days. Based on this dataset, exploratory analysis is performed with Data visualization. The graph from the distplot is shown. The plot explains that, the value of Indian rupee to the US Dollar varies uncertainly for each day. And the value of Rs.60-70 has ruled the exchange rate for over the decade. It also explains how each other value shared with the total. The joint plot between the correlation between the high and low value for each day is plotted. This joint plot show how the value of currency exchange vary. Though, the evolution of exchange rate shows a gradual increase, the noise over a day in very small since the high and the low value remain with smaller difference and variations among them. As we already concluded that the value of range 60-70 has remained the maximum value of currency equivalence, this plot is another proof for our proposal. In the next plot, let us see what data we could infer from a violin plot. Though this plot looks simple, many statistical values like, complete flow of the values, percentile, median and the outliers could be extracted. Based on the graph, plotted by our dataset, it is seen that the variation in data follows a pattern of equilibrium and evolution and attains a maximum value of around 65. Thus, another proof for our hypothesis. The white dot, shows the median value of the data. The median value in very much related to the count of our total data than to the distribution. So, the medium also lies around the global maximum range.

In the description of our dataset, it is seen that we have taken the data of foreign exchange rate for a total of 2610 days. The mean value lies around 61 with a standard deviation of 8.5. We could see the value drops to minimum of 43.9 for a particular day and reaches a maximum of 76.9. The complete description also says about the percentile values of 25,50 and 75. We could infer that 55.3, 64.0 and 67.8 are the percentile score for 25%, 50% and the 75%. This percentile score is the comparison between each data to rest of the data for the given percentage. From the total data we fetched, we split the data into train and test in the ration of 80:20. Which comes as the beginning to the august, 2018 will be our training set and the remaining data will be the test set. In our study, we implemented seven different model and their performance over prediction of foreign exchange values are analyzed. So the train and test dataset are normalized before using them to train the model and predicting their output. In this phase, the inputs are normalized using tools from sklearn. The train and test data are normalized using the functions from these predefined packages.

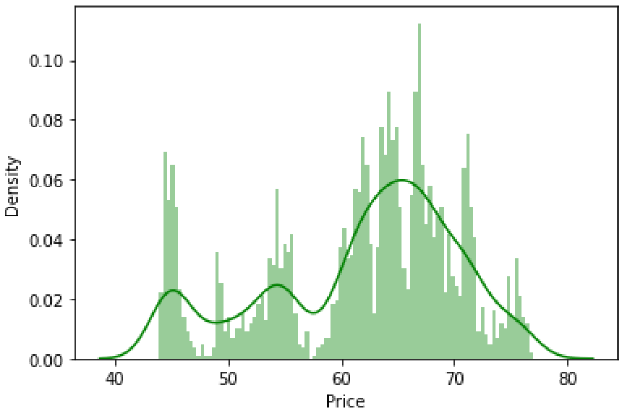

Figure 1: A distribution plot showing variation of probability density values

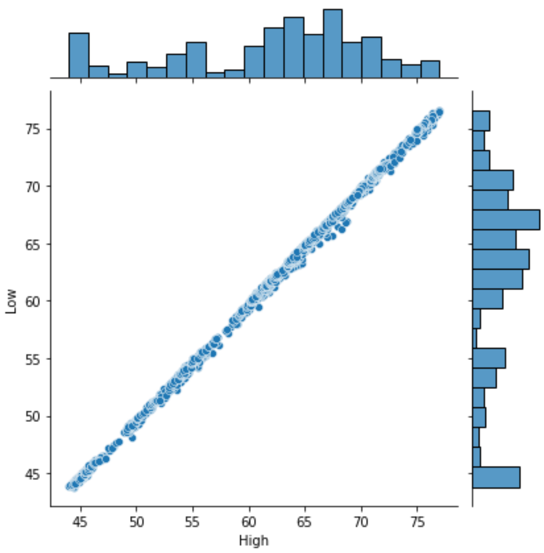

Figure 2: A joint plot showing the correlation between the highest and lowest price values.

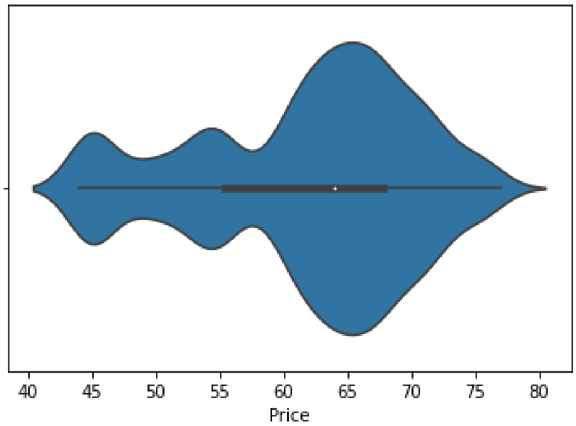

Figure 3: A violin plot showing the distribution of price along with its outliers.

II. RELATED WORKS

Vukovic et al. [1] had developed a neural network based model for the price prediction of the USD/EUR pair. This model does not correspond to the standard technical analysis and considers an analysis based on the autocorrelation and possible errors. Toroslu et al. [2] had used the LSTM deep learning tool which has been shown to be very effective in many time series forecasting problems to make direct predictions in forex. Two different datasets namely the macroeconomic data and the technical indicator data are used. The proposed hybrid model combines two LSTMs corresponding to the two datasets and was found to be quite successful in experiments using the real data.

Islam et al. [3] presented a new model that combines two powerful neural networks used for time series prediction namely the GRU and the LSTM for predicting the future closing prices of forex currencies. The first layer of the proposed model is a GRU layer with 20 hidden neurons and the second is the LSTM layer with 256 hidden neurons. The prediction was done for a 10 minutes timeframe and the performance of the model is validated using MSE, RMSE, MAE and R2 score. The performance of this hybrid model is also compared against the standalone GRU and LSTM models and was found to be outperforming.

Aggarwal et al. [4] had incorporated a comparison between LSTM, SRNN and GRU to predict the forex rate of 22 currencies against the USD simultaneously. The model has predicted the forex rates for 30 consecutive days by taking last year's data as input. The models have been worked with the same number of neural network layers, input, targeted output, and optimizer and learning rate. Ahmed et al. [5] showed that a significant enhancement in the prediction of forex rate can be achieved by incorporating domain knowledge in the process of training machine learning models. The proposed system has integrated the forex loss function into the LSTM that minimizes the difference between the actual and predictive average of forex candles. It is shown that compared to the classic LSTM model, the hybrid model has achieved a decrease in overall MAE rate.
Zhelev et al. [6] have tested whether the LSTM neural network is suitable for high frequency forex trading. The fact

that the major world currencies often correlate and affect each other is taken as an advantage and the time series data are fed and the correlation analysis is done. Qian et al. [7] investigated the predictability of forex spot rates of USD against the British pound to show that not all periods are equally random. The Hurst exponent is used in order to select a period with great predictability. The parameters for generating the training patterns were determined heuristically by auto-mutual information and false nearest-neighbor methods. Inductive machine learning classifiers like the ANN, KNN, Decision Tree and the Naive Bayes classifiers are trained with the generated patterns.

Rout et al. [8] have proposed an evolutionary algorithm based hybrid parallel model to enhance the prediction of the Asia forex rates. The model is made up of trained adaptive linear combiner as linear model and a functional link ANN as a nonlinear model in parallel. To set the parameters of the nonlinear model, the differential evolution learning algorithm is used whereas the linear model has already been trained using the LMS algorithm. Pradeepkumar et al. [9] have presented a review of 82 soft computing hybrid forex rate prediction models. According to them, the ANN based hybrids turned out to be more prevalent, pervasive and powerful. This observation is corroborated by the fact that both evolutionary computation based hybrids as well as the fuzzy logic based hybrids contained some architecture of neural networks as a predominant constituent.

Memon et al. [10] have depicted the combination of the ANNs with different strategies and report the correlation between the exhibitions of the ANNs also those of other anticipating techniques and finding blended outcomes. They have presented the forecast of top exchange monetary utilizing diverse machine learning models which incorporate top forex monetary standards utilizing a hybrid comparison of SVR and ANN, STM and neural networks with hidden layers. Wang et al. [11] have proposed a prediction model combined with ARIMA and a three layer ANN. The performance measure is quantified in terms of MAE, MSE and RMSE. The results and comparisons have shown that the proposed model outperforms the global modelling techniques in terms of profit returns.

Ranjit et al. [12] have presented different machine learning techniques like the ANN, RNN to develop prediction models between Nepalese rupees against Euro, Pound sterling and USD. The prediction model is based on different RNN architectures and feeds forward ANN with a back propagation algorithm. Different ANN architecture models like the Feed forward neural network, SRNN, GRU and LSTM were used. They have also shown that the LSTM networks provide better results than SRNN and GRU. Tenti et al. [13] have shown the use of RNN in forecasting the forex rates. They have shown that the recurrent networks in which the activity patterns pass through the network more than once before they generate an output pattern can learn extremely complex temporal sequences. Three recurrent architectures are compared in terms of prediction accuracy of future forecast for Deutsche mark currency.
Galeshchuk et al. [14] have described the exploration of ANNs for economic purposes. The input selection in which the processing steps to prepare the raw data to a suitable input for the models are investigated. Visual graphs on the experiment's dataset are presented after the processing steps to illustrate the results. The out-of-sample results are compared with training ones. Ince et al. [15] have proposed a two stage forecasting models which incorporates parametric techniques such as ARIMA, VAR and co-integration techniques and non-parametric techniques such as the SVR and ANN. Comparison of these models have shown that the input selection is very important. The SVR technique is found to be outperforming the ANN for two input selection methods.

III. PROPOSED WORK

The proposed work for our project basically involves the implementation of various models for the prediction of exchange rates between the USD/INR pair, along with the preliminary analysis of the exchange rates using various graphs and plots to gather insights from the data. The dataset used has been downloaded from a popular FOREX website for a period of 10 years. A total of seven models have been implemented and compared for their performance on the dataset, and the models have been listed below in their order of implementation.

1. Linear Regression
2. Neural Network
3. LSTM
4. Bidirectional LSTM
5. GRU
6. Ensemble Model
7. Concatenated Model

The first model implemented is the linear regression model. The model is trained using the above data in order to predict the future rates of the USD/INR pair. Linear Regression follows a linear function to calculate the output. Here we are supposed to deduce two different coefficients to get the final target value .The Linear function is as given in equation 1. Here the different coefficients are b and a which can be calculated from the training set.

$$Y = bX + a \tag{1}$$

$$B = \frac{cov(X < Y)}{var(X)} \tag{2}$$

$$A = var(y) - mean(y) * X \tag{3}$$

Equations 2 and 3 will give us the slope and the intercept value. The supporting values like Cov(X,Y),Var(x),mean(y) can be calculated from the training set. The main goal is to fit a line through all the points such that the distance between the predicted point and original point is minimized. The error function used here is the sum-of-square error which will scale up negative errors. The final R2 score obtained from the linear regression model is 0.839.

An artificial neural network model has been implemented in order to perform the task of predicting the prices of the USD/INR pair having an architecture of one hidden layer consisting of 6 nodes. The model performs slightly better when compared to linear regression but however still has a significant error rate. We aim to solve this problem in the upcoming models by making use of sequence models which can better interpret the data, hence giving better results. The ANN model performs the task of forward propagation in order to compute the activations of the nodes. The activation function used here is the linear activation function, and the error values are back propagated from the output layer back to the other nodes in order to update the weights. We have used stochastic gradient descent as the optimizer for this model, and have trained for about 100 epochs.

In general RNN, we may not considering the prominent features from very beginning of the input. These features might be important in deciding the output. This is where the sequential model like LSTM and GRU resolves these problem by using cell state. In LSTM, forget gate, input and output gate perform activations to decide the flow of data through each layer. In our LSTM model, the architecture of the model includes a LSTM layer followed by a dense layer. The model is compiled with mean square error loss function with stochastic gradient descent of Adam optimizer. We follow the Root Mean Square Error as our performance metric. The model is trained with early stopping, and achieves an appreciable value of rmse value = 0.048. From the graph plotted, it is seen that loss and rmse values drops to a minimum and does not show much variations in further epochs. In model estimation, the prediction is done for train and test set and their performance metrics are analyzed. The r2-score values are calculated for each set. And due the correlation between the features, the r2-score values is trimmed and the adjusted r2-score value is calculated. Finally, this model predicts the output with r2-score value of 0.977 and 0.047 for train and test set respectively.

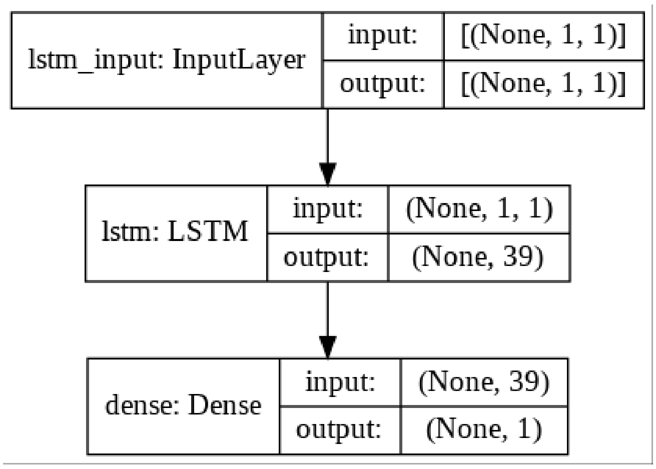

Figure 4: Architecture of implemented LSTM model

Bidirectional LSTMs are an updated version of traditional LSTMs that can improve the performance of sequence prediction models. Bidirectional LSTMs train two LSTMs, one in the forward direction and the other in the reverse direction of the sequence. This results in more complete learning and comparatively accurate predictions. The python keras libraries for bidirectional LSTM model are imported. The sequential model layer is added with an output shape of 20 each for the two LSTMs. A dense hidden layer is also added. The model is then trained using the compile function with the RMSE as the metric and also with the adam optimizer. The model is initially coded to fit for 100 epochs but then early stopping is also implemented. The performance metrics for the training and the testing set are calculated. The R2 score obtained after applying the methods to the dataset for the training and the testing set are 0.969 and 0.875 respectively. The RMSE score obtained for the training and the testing set are 0.175 and 0.163 respectively. The loss plot obtained after 8 epochs of fitting is shown below.

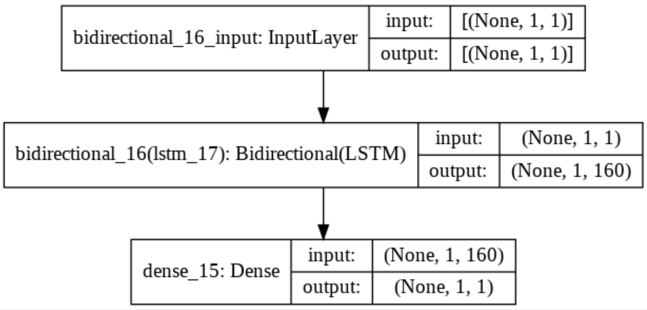

Figure 5: Architecture of Bidirectional LSTM model

GRU is an updated standard recurrent neural network model which overcomes the vanishing gradient problem. GRU basically uses update and reset gates in order to decide what information should be passed on to the output. The python keras libraries for the GRU model are imported. The sequential model is added with an output shape of 7 and a linear activation function. A dense layer is also added to the existing GRU model. The model is then trained using the compile function and the metric used is the RMSE. The adam optimizer is also used.

The model is initially coded to fit 100 epochs. But the early stopping method is implemented in order to prevent overfitting. The final loss plot for 22 epochs is shown below. The performance metrics for the training and the testing set are calculated. The R2 scores obtained for the training and the testing set of the considered dataset are 0.986 and 0.919 respectively. The RMSE scores obtained for the training and testing set are 0.119 and 0.084 respectively. The final prediction plot for the dataset obtained from the GRU model is shown below.

In order to further improve the accuracy of our models, we have also tried out ensemble learning models which make use of the bootstrap aggregation technique. The bootstrap aggregation technique randomly samples values from the dataset, and totally 5 GRU models have been trained. Each of the GRU models make predictions and the final prediction is computed for each training example by making use of the median rule rather than using the mean. Using the median value for the predictions obtained by each of the models aims to improve the accuracy by making the overall ensemble model less susceptible to outliers. The ensemble model basically applies bagging of 5 GRU models, and the optimizer used for each of the models is the Adam optimizer. The final R2 score obtained by this method is 0.960.

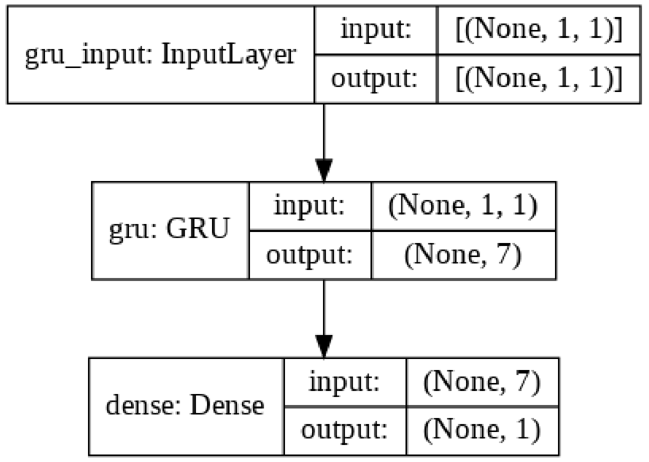

Figure 6: Architecture of implemented GRU model

Recurrent neural networks train well on time-series data. We have already developed two recurrent neural networks namely LSTM and GRU. In the concatenated model, we will be merging both the models when they attain a similar output shape. After concatenating them, a bunch of dense layers are added for further training and the final activation is produced.

Here the dataset is given two both the RNN's individually and parallel training is carried out. Both models train on their neurons and update their gates individually. When both the models complete training, they produce an activation which is merged into a single layer and is passed on into a set of dense layers to compute the errors and activations. In Backward propagation, the weights of the final dense layers are first updated using Gradient descent with momentum and the weights of both RNN's are updated individually. LSTM uses Adam optimization to update its weights and GRU uses Gradient descent with momentum to update its weights. The final weights on both the models are stored individually and the weights of the dense layers are stored separately. The Main advantage of this concatenated model is that we get to train two different Recurrent Neural Networks simultaneously on the same training set and use both the results together to compute the final activation. Further Dense and dropout layers are added to avoid regularization and prevent overfitting of the model. The main drawback is that the training complexity increases rapidly. This model gives the best performance as it combines two recurrent neural networks to predict the final price.

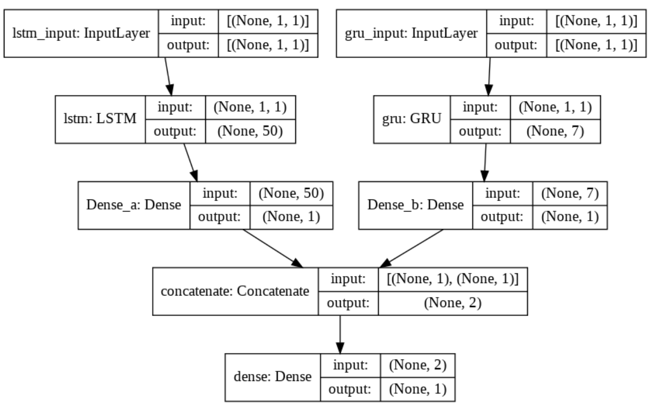

Figure 7: Architecture of concatenated model

IV. RESULTS AND DISCUSSION

We have successfully been able to predict the exchange rate for an upcoming day based on the data available from the previous days. We have first preprocessed the data by converting the values of tuples into the required format so that it can be fed into the model for the purpose of prediction. Seven machine learning and deep learning models have been built and compared using various performance metrics. The Bar graph indicates that the concatenated model consisting of a GRU and LSTM concatenated together performs better than the other models with an R-2 score of 96.7%. We have further used performance metrics such as MAE and RMSE score to get a better analysis of our models. The various models implemented have been compared in Table 1 on the basis of their performance metrics. We have successfully been able to predict the exchange rate for an upcoming day based on the data available from the previous days. We have first preprocessed the data by converting the values of tuples into the required format so that it can be fed into the model for the purpose of prediction. Seven machine learning and deep learning models have been built and compared using various performance metrics. The bar graph indicates that the concatenated model consisting of a GRU and LSTM concatenated together performs better than the other models with an R-2 score of 96.7%. Linear Regression performs less compared to other models with R-2 score of 0.83. Here the training R-2 score for linear regression is 0.96 and test R-2 score is 0.83 which indicates that the model is over fitted to the training set. The ensemble model stands as the second best model with an R-2 score 0.96 and is very close to the concatenated model. We have further used performance metrics such as MAE and RMSE score to get a better analysis of our models. The second bar graph is a comparison of all models with their Mean Absolute Error (MAE) value. The bar graph indicates that the neural network has the least Mean Absolute Error compared to the other models. The ensemble model and Concatenated model get nearly same error values and they both outperform other models. Linear Regression gets the highest error value of 0.456. Any good model should have error values less than 0.2, but Linear Regression has a relatively high error. The Final graph is used to compare the models and Root Mean Square (RMSE) values. Here the Concatenated model gives the lowest RMSE value. It gives a RMSE value of 0.054 which is the least value among all the models. Highest RMSE value of 0.13 is given by Linear Regression. The Ensemble model gives higher RMSE than Concatenated model but still less than the other models. From all the graphs, it is evident that the Concatenated model performs as the best model with highest R-2 score and least RMSE value.

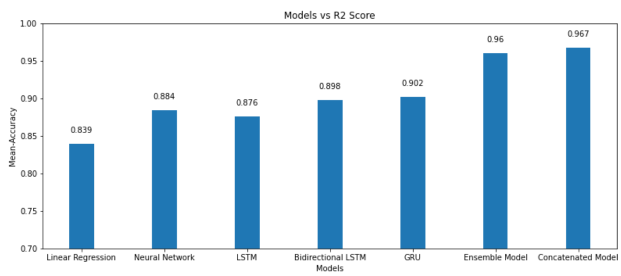

Figure 8: Models vs R2 Score

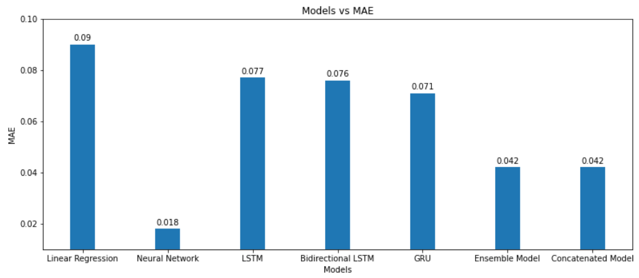

Figure 9: Models vs MAE

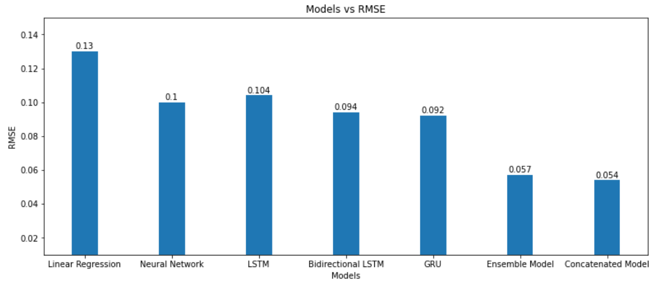

Figure 10: Models vs RMSE

Model	R2 Score	MAE	RMSE
Linear Regression	0.839	0.090	0.030
Neural Network	0.884	0.018	0.100
LSTM	0.876	0.077	0.104
Bidirectional LSTM	0.898	0.076	0.094
GRU	0.902	0.071	0.092
Ensemble Model	0.960	0.042	0.057
Concatenated Model	0.967	0.042	0.054

Table 2: Comparison of various models

V. CONCLUSION

In this project we have covered 7 models in which 5 of them are sequential models. Every model is trained with the same splitting composition of train and validation data and contains 3605 training examples. The best trained model is the model which contains concatenated layers of LSTM and GRU. This model gave a test r2 value of 0.96. This model can be deployed and can be used to predict the Foreign Exchange rate for the upcoming years. The main drawback of this model is that it does not consider factors related to the country's growth. Financial markets cannot be predicted only with the help of market data but also need data about the asset. In Foreign-Exchange rate prediction we have taken features like high-price, low-price, change, opening-price etc but should also take features like GDP of country, capital of the country, Gain value of the country etc which can affect the Forex value.

In the future Works other necessary features which defines the country's growth will also be taken into consideration and the model will be trained for the data consisting training examples for the past 20 years. The neural network can contain residual layers so that the current layer can compute activations for the next layer which in turn will increase the performance of the model. We can also add a RBF(Radial Basis Function) layer with n RBF neuros which will increase the performance.
The splitting composition can follow k-Cross Validation technique where the dataset is split into K parts and each part is chosen at random and trained for k times. Here the performance is taken as an average for all the k times.

VI. REFERENCES

[1] Lo Bosco G., Di Gangi M.A. (2017) Deep Learning Architectures for DNA Sequence Classification. In: Petrosino A., Loia V., Pedrycz W. (eds) Fuzzy Logic and Soft Computing Applications.

[2] Vyklyuk, Y., Vukovic, D. and Jovanovic, A., 2013. Forex prediction with neural network: USD/EUR currency pair. *Актуальні проблеми економіки*, (10), pp.261-273.

[3] Yıldırım, D.C., Toroslu, I.H. and Fiore, U., 2021. Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators. *Financial Innovation*, *7*(1), pp.1-36.

[4] Ince, H. and Trafalis, T.B., 2006. A hybrid model for exchange rate prediction. Decision Support Systems, 42(2), pp.1054-1062.

[5] Galeshchuk, S., 2016. Neural networks performance in exchange rate prediction. Neurocomputing, 172, pp.446-452.

[6] Ranjit, S., Shrestha, S., Subedi, S. and Shakya, S., 2018, October. Comparison of algorithms in foreign exchange rate prediction. In 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS) (pp. 9-13). IEEE.

[7] Wang, S., Tang, Z. and Chai, B., 2016, October. Exchange rate prediction model analysis based on improved artificial neural network algorithm. In 2016 International Conference on Communication and Electronics Systems (ICCES) (pp. 1-5). IEEE.

[8] Zhelev, S. and Avresky, D.R., 2019, September. Using LSTM Neural Network for Time Series Predictions in Financial Markets. In 2019 IEEE 18th International Symposium on Network Computing and Applications (NCA) (pp. 1-5). IEEE.

[9] Qian, B. and Rasheed, K., 2010. Foreign exchange market prediction with multiple classifiers. Journal of Forecasting, 29(3), pp.271-284.

[10] Tenti, P., 1996. Forecasting foreign exchange rates using recurrent neural networks. Applied Artificial Intelligence, 10(6), pp.567-582.

[11] Islam, M.S. and Hossain, E., 2020. Foreign Exchange Currency Rate Prediction using a GRU-

LSTM Hybrid Network. Soft Computing Letters, p.100009.

[12] Aggarwal, P. and Sahani, A.K., 2020, November. Comparison of Neural Networks for Foreign Exchange Rate Prediction. In 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS) (pp. 415-419). IEEE.

[13] Ahmed, S., Hassan, S.U., Aljohani, N.R. and Nawaz, R., 2020. FLF-LSTM: A novel prediction system using Forex Loss Function. Applied Soft Computing, 97, p.106780.

[14] Rout, M. and Koudjonou, K.M., 2020. An evolutionary algorithm based hybrid parallel framework for Asia foreign exchange rate prediction. In Nature Inspired Computing for Data Science (pp. 279-295). Springer, Cham.

[15] Pradeepkumar, D. and Ravi, V., 2018. Soft computing hybrids for FOREX rate prediction: A comprehensive review. Computers & Operations Research, 99, pp.262-284.

[16] Memon, I.Z., Talpur, S., Narejo, S., Junejo, A.Z. and Hassan, E.F., 2020, March. Short-Term Prediction Model for Multi-Currency Exchange Using Artificial Neural Network. In 2020 3rd International Conference on Information and Computer Technologies (ICICT) (pp. 102-106). IEEE.