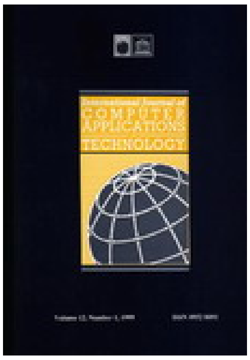


International Journal of Computer Applications in Technology

ISSN online: 1741-5047 - ISSN print: 0952-8091
https://www.inderscience.com/ijcat

Bi-LSTM GRU-based deep learning architecture for export trade forecasting

Vaishali Gupta

DOI: 10.1504/IJCAT.2024.10061555

Article History:

Received:	21 November 2022
Last revised:	28 February 2023
Accepted:	17 March 2023
Published online:	03 October 2024

Bi-LSTM GRU-based deep learning architecture for export trade forecasting

Vaishali Gupta

Department of Computer Science & Engineering,
Indira Gandhi Delhi Technical University For Women,
Delhi, India
Email: vaishaliagraw@gmail.com

Abstract: To assess a country's economic outlook and achieve higher economic growth, econometric models and prediction techniques are significant tools. Policymakers are always concerned with the correct future estimates of economic variables to take the right economic decisions, design better policies and effectively implement them. Therefore, there is a need to improve the predictive accuracy of the existing models and to use more sophisticated and superior algorithms for accurate forecasting. Deep learning models like recurrent neural networks are considered superior for forecasting as they provide better predictive results as compared to many of the econometric models. Against this backdrop, this paper presents the feasibility of using different deep-learning neural network architectures for trade forecasting. It predicts export trade using different recurrent neural architectures such as 'vanilla recurrent neural network (VRNN)', 'bi-directional long short-term memory network (Bi-LSTM)', 'bi-directional gated recurrent unit (Bi-GRU)' and a hybrid 'bi-directional LSTM and GRU neural network'. The performances of these models are evaluated and compared using different performance metrics such as Mean Square Error (MSE), Mean Absolute Error (MAE) Root Mean Squared Error (RMSE), Root Mean Squared Logarithmic Error (RMSLE) and coefficient of determination R-squared (R^2). The results validated the effective export prediction for India.

Keywords: Bi-LSTM; GRU; economic forecasting; international trade; RNN; recurrent neural network.

Reference to this paper should be made as follows: Gupta, V. (2024) 'Bi-LSTM GRU-based deep learning architecture for export trade forecasting', *Int. J. Computer Applications in Technology*, Vol. 74, No. 3, pp.235–245.

Biographical notes: Vaishali Gupta completed her BTech and MTech degree in the field of Computer Science and Engineering, and is currently a PhD Research Scholar in the Department of Computer Science and Engineering at Indira Gandhi Delhi Technical University for Women, Delhi, India. Her major research area is data analytics using machine learning techniques. The research mainly focuses on developing machine learning models for financial and economic forecasting.

1 Introduction

International trade denotes the economic transactions of any country with other countries. These economic transactions could include exports of goods and services from one country to other countries for international consumption or imports of goods and services from other countries for domestic consumption within the concerned country. Exports and imports of goods as well as services constitute the trade balance of any country. Therefore, both exports and imports are crucial economic variables utilised by policymakers for collaborating foreign trade policy and domestic policies. Policymakers focus more on increasing exports as the foreign currency obtained by way of increased exports provides financing cover for the imports of any country and acts as a cushion for such imports. If exports are less than imports, a country incurs a trade deficit, which may create other macroeconomic imbalances in the economy. Therefore, the prediction of exports is crucial as exports determine macroeconomic balance, economic stability and growth prospects.

Exports are an important component of the Gross Domestic Product (GDP). According to World Bank data, goods, and services exports constituted 18.7% of the GDP of India in the year 2020. It should also be noted that the COVID-19 pandemic had a major impact on India and other economies of the world, which in turn has affected India's exports. Mangla et al. (2021) argued that COVID-19 has adversely affected trade, travel and commerce and explored the impact of Covid-19 on financial sector. Rasheed et al. (2022) showed the impact of oil exports on consumer imports in Iraq during Covid-19.

A few studies have explored how exports and imports affect economic growth and improvements in the quality-of-life (Sirgy et al., 2007). A study (Omotor, 2008) on Nigeria has indicated that economic growth is dependent on exports, imports, labour, and exchange rate. The study also posited their co-integration and showed a positive relationship between exports and labour with the growth rate in the economy and the adverse impact of imports and exchange rates on various economic conditions. A research paper that used Ethiopian economic data (Chemeda, 2001) has demonstrated that real-export growth is positively associated with economic growth. Although the effect is insignificant in the near term, these two variables have a strong relationship in the long run. These studies indicated that changes in exports could lead to major impacts on the economic growth of any country. Apart from the contribution to economic growth, exports can also help to boost innovation and can support productive and innovative firms in increasing their revenues. Awokuse (2007) and Awokuse and Christopoulos (2009) emphasised the export-led growth models in which economic growth is mainly caused by exports. These researchers believed that domestic firms could be benefitted through the expansion of export activities by way of efficient resource allocation, increased economies of scale and imports of producer goods. These lead to an increase in investment in capital goods and thereby accelerate output and growth (Helpman and Krugman, 1985). According to Krugman (1994) and Ghartey (1993), economic growth that emanates from enhancement in skilled labour and technological innovation will lead to productivity gains, which eventually help in promoting exports.

Recently, machine learning and deep learning algorithms, a sub-field of Artificial Intelligence (AI), are recognised to perform far better than several econometric procedures applied for forecasting and predictions. The strong learning capability of these algorithms can be utilised across different fields including engineering, statistics and social sciences for anticipating the future event. Bengio (2009) highlighted shreds of evidence of the successful use of deep learning algorithms to showcase encouraging findings in many research areas and invited significant academic as well as industry attention. Bengio et al. (2013) showed the advantage of deep learning in terms of constructing a mix of various non-linear transformations and providing functional expressions, which may produce additional abstraction and larger benefits. However, these studies did not use hybrid deep learning models for trade forecasting. This study adds to the existing literature by showcasing that hybrid deep learning model can also be effective for the purpose of trade forecasting.

The machine-learning techniques can also be applied in the field of international trade to better analyse and predict the trade flows of any country. The export-led economic growth hypothesis and the numerous advantages of an artificial neural network form the basis of undertaking this research on deep learning. However, this work is novel in its approach as it develops an effective prediction model that synthesises LSTM and GRU neural networks with economic principles to show the significance of the deep learning model from a trade perspective.

This study aims to apply different recurrent neural network architectures such as 'Vanilla RNN', 'Bi-LSTM', 'Bi-GRU' and a hybrid 'Bi-LSTM-GRU' network for predicting export trade and shows the effectiveness of these models in trade forecasting.

The paper is structured as follows. Section 2 represents a review of research in this field and provides a brief overview of different approaches for trade forecasting. Section 3 represents data, data pre-processing, and the methodology. In Section 4 performances of different approaches are mentioned and results are compared. Finally, Section 5 concludes the paper with the identification of the future scope related to this research.

2 Literature review

The main objective of the trade forecasting procedures is to get correct future information about exports and imports. Many studies have examined different forecasting approaches and made a comparison between them for assessing the future values of exports or imports or both. This section presents a review of some related papers.

2.1 Conventional statistical models for trade prediction

Quimba and Barral (2018) made a comparison of gravity model estimation methods including 'Ordinary Least Squares (OLS)', 'Poisson Pseudo Maximum Likelihood (PPML)' and 'Gaussian Processes (GPML)' with neural network methods for analysing and predicting trade flows. Veenstra and Haralambides (2001) applied 'multivariate autoregressive' time-series-based models for predicting seaborne trade. They used four commodities, namely, crude oil, iron ores, grains and coal as independent variables in their study. They found that their models were able to achieve relatively low forecast errors while estimating long-term seaborne trade flows. Urrutia et al. (2019) made use of 'Autoregressive Integrated Moving Average (ARIMA)' as well as 'Bayesian Artificial Neural Network (BANN)' techniques on various economic variables in the case of the Philippines to forecast its exports and imports. The authors compared these models using statistical tests and used forecast accuracy measures (error measures) for comparing the performances across these models.

2.2 Machine learning models for trade prediction

Sun et al. (2018) predicted import as well as export flows through the evaluation of trade statistics correlations between these flows for all countries between 1960 and 2017. They also examined trade flows and compared various ML algorithms to forecast GDP using the country's import and export data. They considered five ML algorithms, namely, 'Linear Regression (LR)', 'RBF Regressor (RBF)', 'Support Vector Machine (SVM)', 'Regression by Discretisation (RD)' and ‘Reduced Error Pruning Tress (REP)’ to identify the maximum performing ML algorithm for superior analysis of trade data set and forecasting the GDP. Gopinath et al. (2020) compared supervised and unsupervised ML techniques with the gravity model to decrypt trade flow patterns at a bilateral level. The ML models were trained on the trade data set of agriculture products for predicting future trade flows. They found that ML techniques are more relevant as compared to traditional econometric methods to predict trade patterns. Nummelin and Hanninen (2016) made a comparison of ‘Support Vector Machine’, ‘Neural Network’ and ‘Random Forest (ensemble method)’ techniques for predicting trends in sawn wood trade at the bilateral level between various countries. Kuo and Li (2016) done to predict export in the case of Taiwan. In this study, the authors combined ‘wavelet transforms’, ‘firefly algorithm-based *K*-means algorithms’ and ‘firefly algorithm-based SVR’ to do the forecasting. They found that the ML algorithm having both ‘wavelet transforms’ and ‘clustering’ was able to perform highly and the forecasting accuracy of ‘firefly algorithm-based SVR’ was much better than other algorithms. Gupta and Kumar (2021) gave a snapshot of the relevant literature on ML approaches that were applied to international trade forecasting. The literature review reveals an increasing attraction towards the use of ML techniques for trade forecasting as compared to conventional statistical techniques.

2.3 Artificial neural network (ANN) for trade prediction

Batarseh et al. (2019) presented the application of neural networks to predict exports and imports in Saudi Arabia. The author analysed the performance of two models, namely, ‘Artificial Neural Network (ANN)’ and ‘Autoregressive Integrated Moving Average (ARIMA)’ models for this purpose. As mentioned earlier, Nummelin and Hanninen (2016) also used ‘Autoregressive Integrated Moving Average (ARIMA)’ and ‘Bayesian Artificial Neural Network (BANN)’ algorithms for the Philippines’ economic data set for forecasting its exports and imports. Junoh (2004) proposed neural network models for predicting GDP growth in Malaysia and they compared the accuracy of the models with the economic statistical methods. Their findings reveal that neural network offers great potential in trade prediction compared to traditional econometric techniques.

2.4 Deep learning models for trade prediction

The idea behind deep neural networks is to use Multilayer Perceptron (MLP) with more hidden layers to model more complicated functions (Alpaydin, 2009). Jung et al. (2018) pursued a deep learning approach and employed three deep learning algorithms, namely, ‘Elastic Net’, ‘Super-Learner’ and ‘Recurring Neural Network’, to a common economic forecasting problem and concluded that deep learning algorithms can outperform conventional statistical forecasting models. Shen et al. (2021) proposed an effective method to forecast international trade based on a neural network with Long Short-Term Memory (LSTM) and their findings validated that the chosen method was effective.

2.5 Research contributions

The main research contributions of our study are outlined below:

- The trade data of 219 countries including India for the period of 50 years from 1970 to 2019 is investigated.
- Unlike the previous studies that have focused on applying classical statistical prediction models, such as ARIMA and OLS and ML techniques, such as SVM, Decision Tree and Random Forest, we apply advanced deep learning techniques using Recurrent Neural Networks, namely, ‘Vanilla Recurrent Neural Network’, ‘Bi-directional Long Short Term Memory (LSTM)’, ‘Bi-directional Gated Recurrent Unit (GRU)’ and one ensemble model by combining ‘Bi-LSTM’ and ‘Bi-GRU’ for predicting India’s exports.
- The results are compared using different performance metrics, such as RMSE, MSE, RMSLE, MAE and R^2.
- Results show the effectiveness of advanced deep learning architectures in trade forecasting.

3 Methodology

This study predicts India’s annual export trade by applying different deep learning models, namely, ‘Vanilla Recurrent Neural Network’, ‘Bi-directional LSTM’, ‘Bi-directional GRU’ and a hybrid approach of Bidirectional LSTM and GRU. In this work, the best model is chosen based on more accurate prediction results for India’s export.

3.1 Data set

The trade data for this study is collected from the UNCTAD Statistics database and Federal Reserve Economic Data (UNCATD, n.d.). The data consisted of variables such as the country's yearly imports, exports, and GDP for 219 countries including India for the period from 1970 to 2019. Data is presented in billion USD. There were some missing data entries in the data set as there was insignificant import, export or GDP for a few countries in some years.

Export and import statistics are critical to a country’s GDP as they contribute the most to the economy, and various algorithms have been developed to analyse trade data. The focus of our research was on export data analysis. For a better understanding, GDP, Exports and Imports data statistics are given in Tables 1, 2 and 3, respectively. Figures 2 and 3 show the graphical representations of data variables to visualise the data variations.

Table 1 Descriptive statistics for GDP

		GDP		
Level	*Percentiles*	*Smallest*		
1%	110	1		
5%	544	2		
10%	1090.5	3	**Obs.**	10,950
25%	2732	4	**Sum of Wgt.**	10,950
50%	5455.5		**Mean**	5388.684
		Largest	**Std. Dev.**	3047.726
75%	8180	9703		
90%	9703	9703	**Variance**	9,288,632
95%	9703	9703	**Skewness**	–.0883126
99%	9703	9703	**Kurtosis**	1.726643

Table 2 Descriptive statistics for exports

		Exports		
Level	*Percentiles*	*Smallest*		
1%	2	1		
5%	266	1		
10%	778.5	1	**Obs.**	10,950
25%	2317	1	**Sum of Wgt.**	10,950
50%	4835.5		**Mean**	4778.827
		Largest	**Std. Dev.**	2830.612
75%	7357	8810		
90%	8810	8810	**Variance**	8,012,364
95%	8810	8810	**Skewness**	–.0782255
99%	8810	8810	**Kurtosis**	1.726181

Table 3 Descriptive statistics for imports

		Imports		
Level	*Percentiles*	*Smallest*		
1%	2	1		
5%	286	1		
10%	797.5	1	**Obs.**	10,950
25%	2368	1	**Sum of Wgt.**	10,950
50%	4966.5		**Mean**	4910.988
		Largest	**Std. Dev.**	2906.033
75%	7571	9032		
90%	9032	9032	**Variance**	8,445,026
95%	9032	9032	**Skewness**	–.0829963
99%	9032	9032	**Kurtosis**	1.719305

Figure 1 Vertical box plot of data variables (see online version for colours)

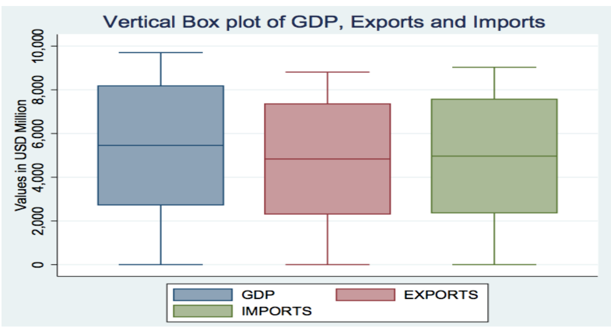
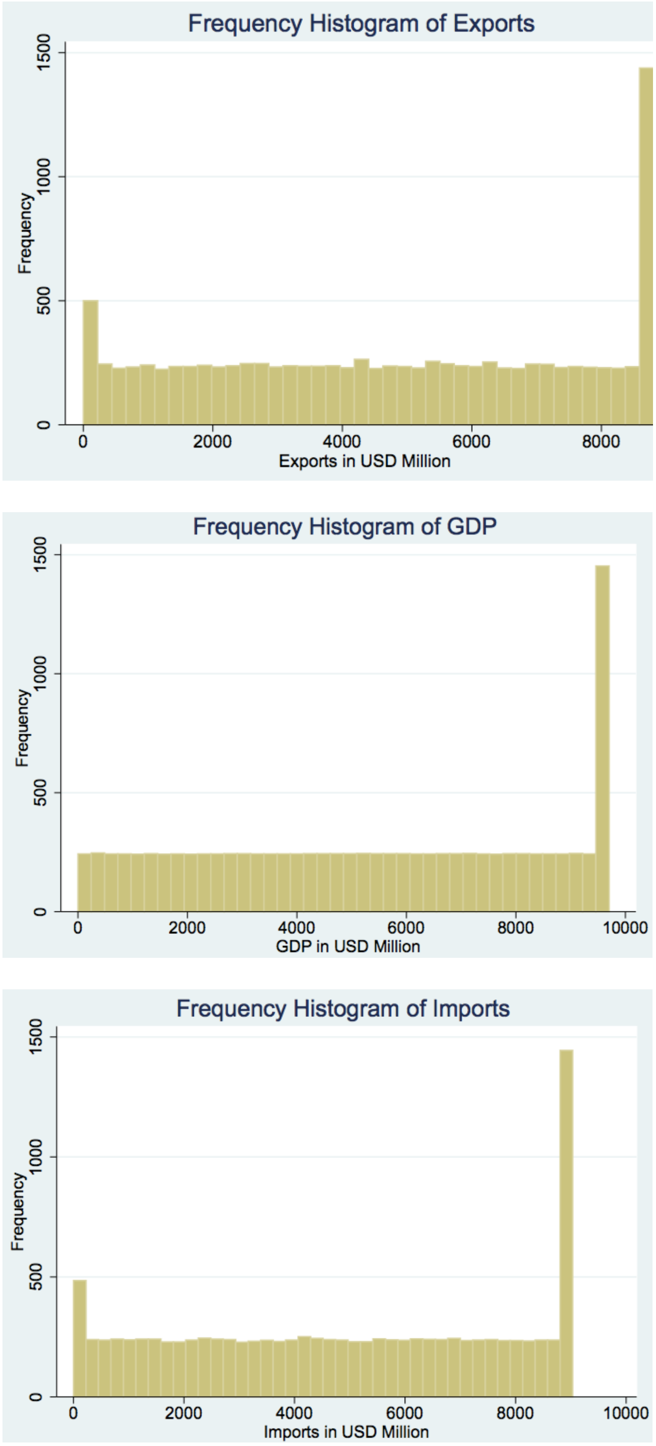


Figure 2 Frequency histograms of data variables



3.2 *Data preparation*

Data pre-processing is necessary to avoid irregularities and missing values in data and potential obstacles to forecasting. Incomplete data has been removed from the data set. It implies that the data set used for this study does not include data points that have missing feature values.

As the data set was in time series format and there was a great chance of it having a non-stationarity problem, this needed to be removed before modelling. The stationarity in time series implies that statistical properties of data, i.e., mean, variance, etc. remain constant over time. These statistical assumptions in a time series can be easily violated because of trends and other time-dependent formats. This creates the problem of non-stationarity in time series data. Since time series problems are distinctive from conventional classification and regression predictive problems, stationary becomes an important factor in time series forecasting analysis. Therefore, before ML analysis, an Augmented Dickey–Fuller test (ADF) (Dickey and Fuller, 1979) has been performed to assess stationarity in the time series data set.

ADF is a statistical 'Unit Root Test' for checking the statistical significance. It includes null and alternate hypothesis testing to compute a test measure and generate *p*-values. The computed statistics and *p*-value help in making inferences about the stationarity or non-stationarity in the time series. The Dickey-Fuller test assesses the null hypothesis (α = 1 which implies the presence of unit root) in our model equation as below

$$y_t = a + \beta t + \alpha y_{t-1} + \varphi \Delta Y_{t-1} + e_t \tag{1}$$

Here,

y_{t-1} = lag *1* of time series

ΔY_{t-1} = First difference of the series at the time $(t-1)$

For the ADF test, if we add more differencing terms for including higher-order regressive processes in our model, it will expand the Dickey-Fuller equation as below:

$$\begin{aligned} y_t &= a + \beta t + \alpha y_{t-1} + \varphi_1 \Delta Y_{t-1} \\ &+ \varphi_2 \Delta Y_{t-2} . + \varphi_p \Delta Y_{t-p} + e_t \end{aligned} \tag{2}$$

For rejecting the null hypothesis and to infer the stationarity of the time series, the *p*-value computed should be less than the significance level (0.05). In this study, after applying the ADF test, it is found that computed statistic values for ADF and *p*-values are greater than critical values and significant values (0.05) for all economic variables.

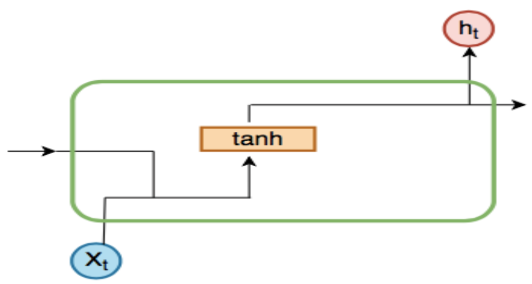
To remove the problem of non-stationarity from the time series data, the data-shuffling technique is used.

3.3 Predictive models

3.3.1 Vanilla recurrent neural network (V-RNN)

It is the simplest of the recurrent neural network models. The vanilla RNN has a quite straightforward form, which can be represented as a simple unit as shown in Figure 3.

Figure 3 Unit architecture of vanilla RNN



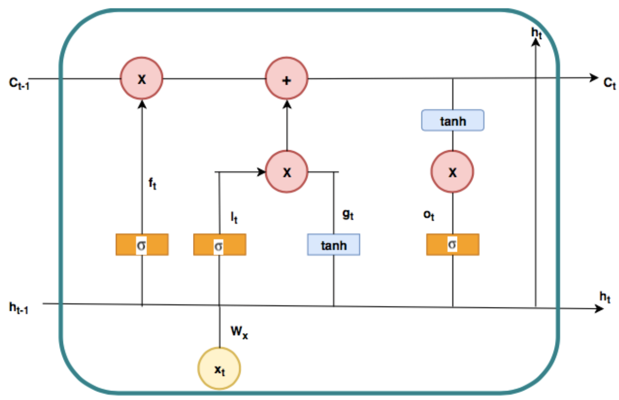
In this architecture, the hidden state h_t is directly dependent on the present input x_t and the previous hidden state h_{t-1} at each time point *t*. Generally, the tanh activation function is employed in the hidden recurrent layer, while the output layer activation function is chosen to be dependent on the problem to be solved. In this study, ReLU activation function is considered in the context of the output layer. The mathematical representation for vanilla RNN is as follows:

$$h_t = tanh\left(Ux_t + Wh_{t-1}\right) \tag{3}$$

3.3.2 Bi-directional long short-term memory (Bi-LSTM)

The LSTM is a kind of RNN model that can learn long-term dependencies, especially in time series forecasting problems. The model is specifically made to do away with the problem of vanishing and exploding gradients. The LSTM has a more complicated recurrence relation than the standard RNN. The LSTM units have several architectures. An LSTM unit common architecture comprises four things – memory cell (c_t), input gate (i_t), output gate (o_t) and forget gate (f_t) (Hochreiter and Schmidhuber, 1997), as shown in Figure 4.

Figure 4 Unit architecture of LSTM



The information can be added to or removed from memory blocks and is regulated by gates using the activation function for a long duration of time. In LSTM, the forget gate is responsible for remembering or discarding the information received from the memory blocks while the input gate is responsible for storing or updating newly received information in the memory block. The output gate is used to decide the new hidden state by considering the newly updated cell state, the previous hidden state and the newly received input data.

Every LSTM cell also computes new hidden state values (g_t) and cell states. The LSTM cell concerning each of the gates and states can be represented mathematically as follows:

$$f_t = \sigma\left(x_t W_f + h_{t-1} U_f\right) \tag{4}$$

$$i_t = \sigma\left(x_t W_i + h_{t-1} U_i\right) \tag{5}$$

$$o_t = \sigma\left(x_t Wo + h_{t-1} Uo\right) \tag{6}$$

$$h'_t = \tanh\left(x_t Wg + h_{t-1} Ug\right) \tag{7}$$

$$c_t = \sigma\left(c_{t-1} \times f_t + i_t \times h'_t\right) \tag{8}$$

$$h_t = tanh\left(c_t\right) \times o_t \tag{9}$$

where the terms h and c are hidden state and memory state respectively, σ is the sigmoid function and W and U are weights. This LSTM network works only in one direction to preserve only the previous state. To preserve both the previous and next state, Bi-directional Long Short-Term Memory (Bi-LSTM) network has been applied. The network operates in both backward and forward directions having two layers that are hidden and separated. It computes $\overrightarrow{h_t}$ as the forward hidden state and $\overleftarrow{h_t}$ as the backward hidden state. The architecture of a bidirectional network is presented in Figure 5.

The given structure has two layers: one is the forward LSTM layer and the other one is backward. The forward layer output order, $\overrightarrow{h_t}$, is recurrently calculated through inputs in a positive order from time T–n to time T–1. On the other hand, the computation of the backward layer outcome order, $\overleftarrow{h_t}$, is done by taking into consideration the inverted inputs from time T-n to time T-1 (Cui et al., 2017). Both these outputs are computed in the same way as the unidirectional LSTM. The output Y_t in LSTM layer is computed as follows:

$$Y_t = \sigma\left(\overrightarrow{h_t}, \overleftarrow{h_t}\right) \tag{10}$$

Here, σ denotes a function for merging $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ outputs.

3.3.3 *Bi-directional gated recurrent unit (bi-GRU)*

GRU is another of the RNN architectures. It resembles the LSTM network but is simple in architecture. There are two gates in GRU unit: an update gate and a reset gate (r_t) as shown in Figurev6. In the GRU architecture, the update gate makes a combination of both the input and forget-gates and forms a single gate. It also has a current memory cell $\left(\widetilde{h_t}\right)$, with the output (h_t) being saved in the GRU's final memory. In GRU, the input value makes interaction with the information arriving from the previous state for calculating various values for intermediate gates, which will eventually be used for deciding the value for the output. GRU has a simplified gate mechanism and thus may train a bit faster than LSTM. The mathematical representation of GRU unit is as follows:

$$z_t = \sigma\left(x_t.Wz + h_{t-1}Uz\right) \tag{11}$$

$$r_t = \sigma\left(x_t.Wr + h_{t-1}Ur\right) \tag{12}$$

$$\widetilde{\mathrm{h}}_t = \tanh\left(x_t W_h + \left(r_t * h_{t-1}\right) W_h\right) \tag{13}$$

$$h_t = \left(1 - z_t\right) * h_{t-1} + z_t * \widetilde{\mathrm{h}}_t \tag{14}$$

Here, σ is the sigmoid function, z_t is the update gate, r_t is the reset gate, $\widetilde{h_t}$ is the candidate vector and h_t is the output vector of GRU unit.

Figure 5 Architecture of Bi-LSTM

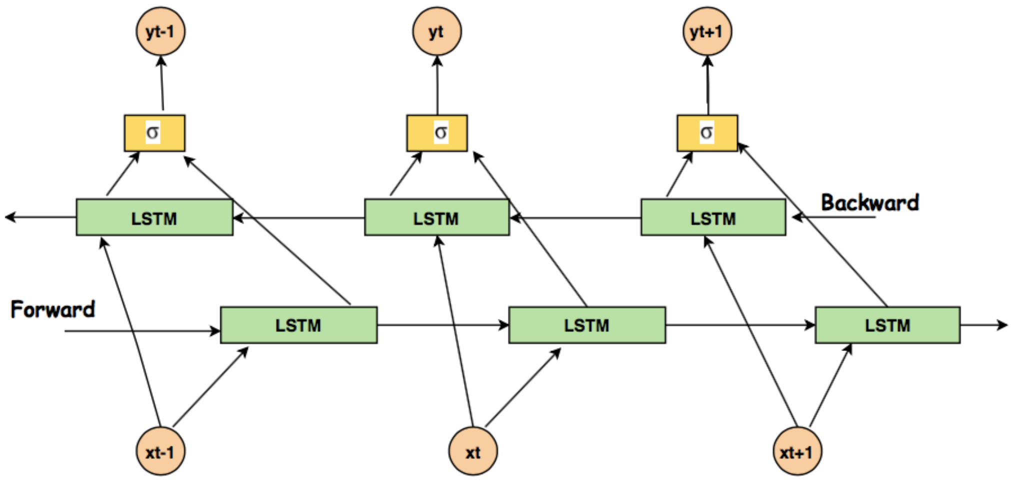
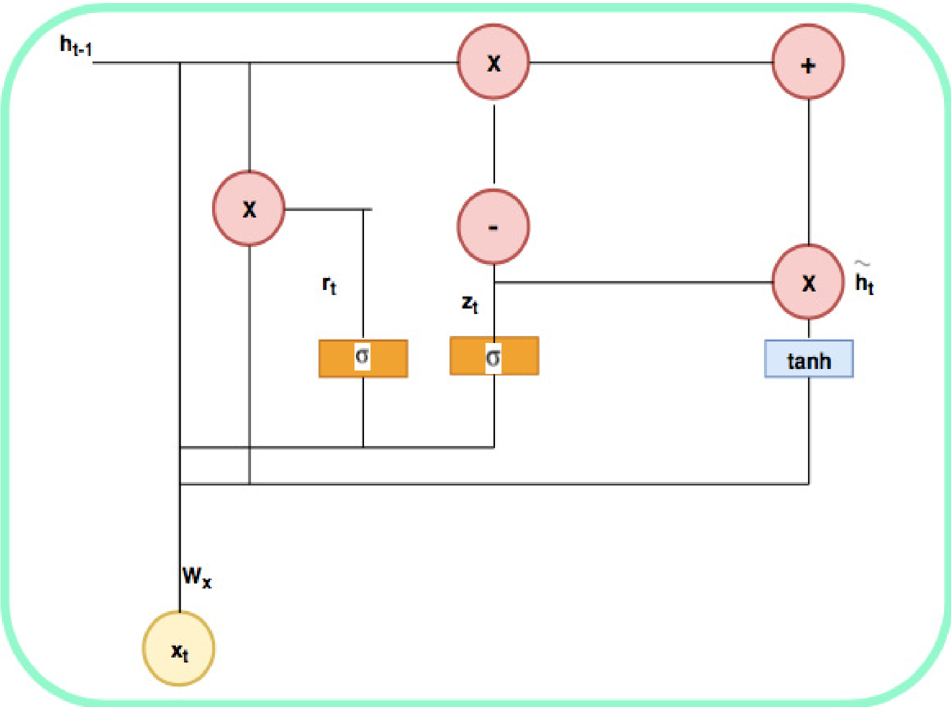


Figure 6 Unit architecture of GRU



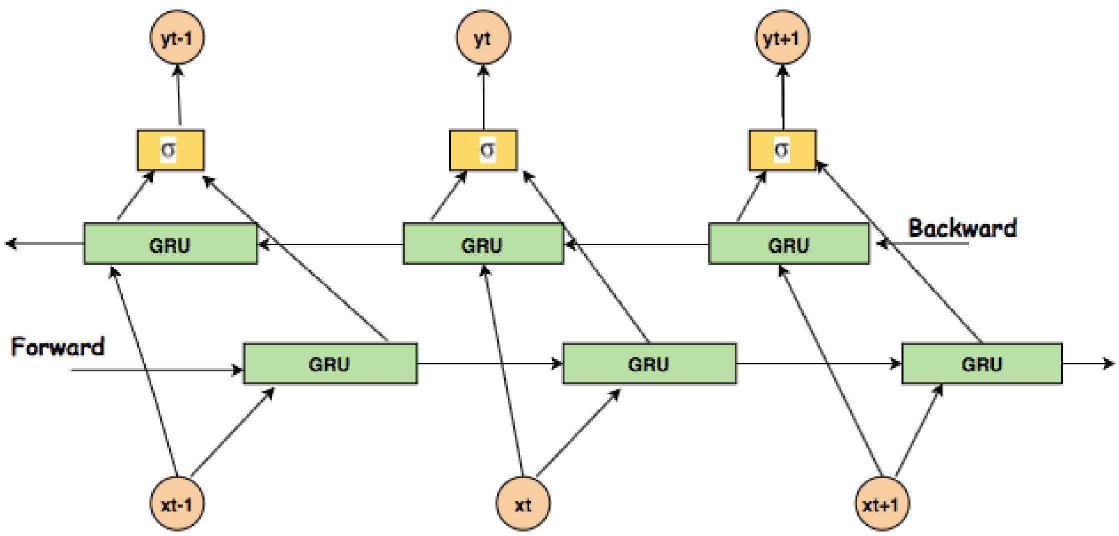
A bi-directional GRU is a neural network model based on a sequential process and consists of two GRUs. The model works based on the state of two GRUs that are uni-directional in opposite directions (Xiong et al., 2016) as shown in Figure 7. One of the GRUs makes forward movement, beginning from the starting of the data sequence, while the other GRU does the backward movement as it begins from the last of the data sequence. This makes it possible for future as well as past information to affect the current state. The bi-GRU is presented below:

$$\overrightarrow{h_t} = GRU_{fwd}\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{15}$$

$$\overleftarrow{h_t} = GRU_{bwd}\left(x_t, \overleftarrow{h_{t+1}}\right) \tag{16}$$

$$h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t} \tag{17}$$

Figure 7 Architecture of Bi-GRU



3.4 Performance measures

The performance measures indicate the prediction capability of the models and determine the extent to which the objective of the prediction models is achieved. In this study, for evaluating the model prediction performance, Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Root Mean Squared Logarithmic Error (RMSLE) and co-efficient of determination *R*-squared (R^2) are considered as performance measures. They are defined as below:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - p_t\right)^2 \quad (18)$$

$$RMSE = \sqrt{MSE} \quad (19)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t - p_t\right| \quad (20)$$

$$RMSLE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\log\left(p_t + 1\right) - log\left(a_t + 1\right)\right)^2} \quad (21)$$

where y_t is the actual value and p_t is the predicted value of period *t* and *n* is the total number of observations.

4 Results and discussion

In this study, the annual export of India is predicted using a different variant of recurrent neural networks. For the experiments, the trade data set is divided into train and test data subsets. The test subset is used for validation during the training process and for final evaluation. The training subset comprised 75% of the data points and the test subset has the rest 25% of the data. For deep learning models, Keras and Tensor flow are employed. All the recurrent neural network architectures applied in this study have four-layer structures; the number of neurons in each hidden layer is set as 64, 32, 16 and 8, respectively and the activation function is ReLU. It links the last layer output of the network at the final time step to a dense layer with a single output neuron. To train the models, the repetition step works with batch size, which is the number of data points or samples that are fed into the model at a time. The batch size is a hyperparameter that is obtained using the trial and error method. Our study used a batch size of 64 in all models. During training, in each repetition step, error metrics (MSE, RMSE, MAE, RMLSE and R^2) are computed for 64 data points observed and predicted export. In all models, the ReLU activation function is used for the hidden layer. The key benefit of applying ReLU is that, for any input more than 0, we have a constant derivative that augments network learning. Each model is run using epoch numbers 100, 200 and 500. In model performance evaluation, MSE, RMSE, MAE, RMLSE and R^2 are utilised to compare the prediction performance of these models. Smaller error values represent a high degree of predictive accuracy. The results are presented in Table 4. Among the RNN models, based on error metrics the performance sequence is Vanilla RNN, Bi-LSTM-GRU, Bi-GRU and Bi-LSTM.

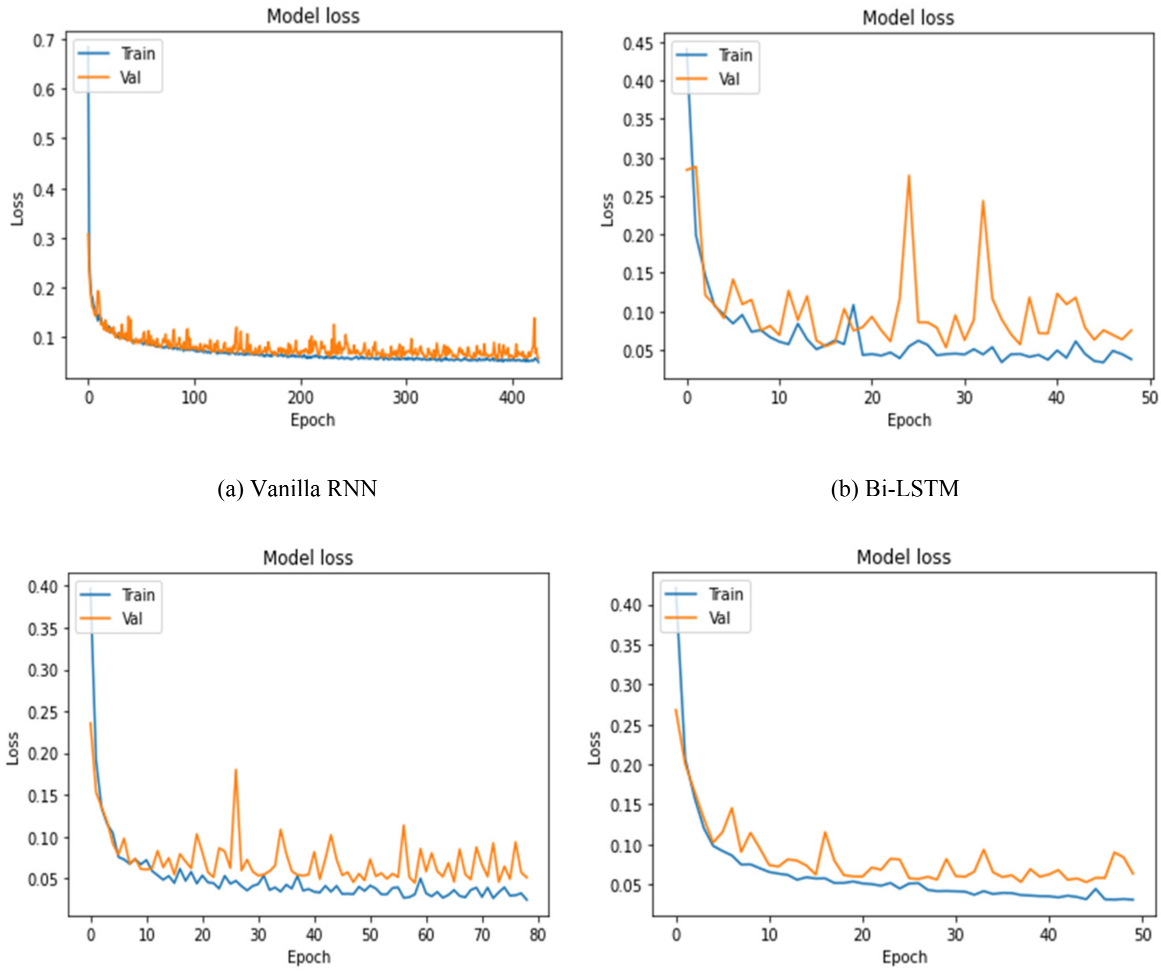
Table 4 Performance analysis of different RNN architectures

Performance metrics	*Vanilla RNN*	*Bi –LSTM*	*Bi –GRU*	*Bi –LSTM –GRU*
MSE	0.0241465	0.03169948	0.02473511	0.02433474
MAE	0.0872746	0.07290424	0.09099498	0.10132296
RMSE	0.1553915	0.1780435	0.15727402	0.15599597
RMLSE	–1.861807	–1.7257274	–1.8497657	–1.8579252
R-squared	0.975853	0.96830051	0.97526488	0.97566525

Here, 10% of the training data set is used as the validation data set to validate all RNN architectures using the early stopping technique. Various training epochs are considered for Vanilla RNN, Bi-LSTM, Bi-GRU and Bi-LSTM-GRU. The *x*-and *y*-axis of the figures are showing the epoch and loss, respectively. The curve shows the lower training loss in all models. The model's loss improved as the number of epochs increased. The Bi-LSTM and Bi-GRU models have the closest loss learning curve. Various models are run using different iterations. The number of epochs is one of the influential parameters so in fewer iterations, the models have lower accuracy with larger errors. When we increase the number of iterations, the model gradually converges. Therefore, a maximum of 500 epochs are considered. The number of neurons in hidden layers also affects the model's accuracy. In case of a low value of the number of neurons, the model will have incorrect simulation, and if the value is very high, it may result in over-fitting. This issue is resolved by L2 regularisation technique.

Table 5 shows the predicted values of India's export using Vanilla RNN, bi-LSTM, Bi-GRU and Bi-LSTM-GRU architectures. Predictions results show that RNN model' predicted values are quite close to the actual values for India's exports. In Figure 9, graphs of the predicted values and the actual values show similar trends as the predicted series has an ascending trend with some fluctuations, as also shown by the actual series. By examining the loss function and prediction graphs, we found that the highest successful methods are Vanilla RNN and Bi-LSTM-GRU. The Bi-GRU and Bi-LSTM-GRU models have the closest results while Bi-LSTM model performs worst. The results indicate that all RNN architectures perform well in predicting future exports of India except Bi-LSTM model.

Figure 8 Learning curves of different RNN architectures (see online version for colours)



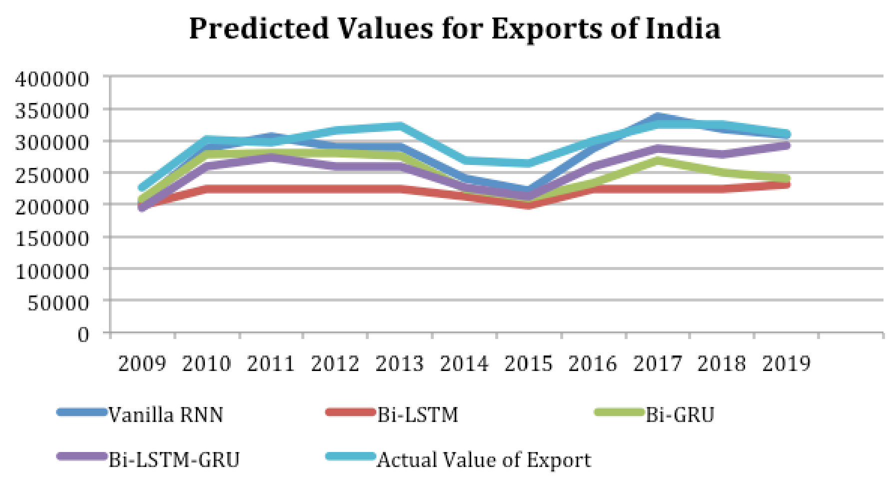
(a) Vanilla RNN (b) Bi-LSTM

(c) Bi-GRU (d) Bi-LSTM-GRU

Table 5 Prediction results of exports using different RNN architectures

Year	*Actual value of export*	*Predicted value for export (in billion USD)*			
		Vanilla RNN	*Bi-LSTM*	*Bi-GRU*	*Bi-LSTM-GRU*
2009	226,351.4	203,733.712	198,732.581	208,794.759	194,813.565
2010	302,905.38	287,871.542	225,342.155	279,103.942	260,447.561
2011	296,828.2	305,706.938	225,371.135	281,878.616	274,816.535
2012	314,847.75	289,294.632	225,363.142	279,872.625	260,176.517
2013	322,693.7	289,378.64	225,370.295	276,291.398	259,225.260
2014	267,951.38	242,027.298	213,172.360	225,541.231	225,701.783
2015	264,542.2	221,045.904	199,607.029	210,831.995	213,509.23
2016	299,241.4	288,627.897	225,373.435	232,981.053	259,535.484
2017	324,778.38	336,671.838	225,373.642	268,793.812	287,199.022
2018	324,339.62	318,219.26	225,373.607	251,219.662	277,637.623
2019	312,339.62	309,371.340	231,111.540	241,232.550	293,222.430

Figure 9 Prediction results of different RNN architectures (see online version for colours)



5 Contributions to literature

As per available literature, this is the only study that has compared different deep learning models for predicting export trade. The study also proposes one hybrid RNN architecture (Bi-LSTM-GRU) for export trade prediction and results also validate the effectiveness of the model.

6 Conclusion

This study employed an export trade prediction framework featuring an analysis of deep learning architectures. To achieve the effectiveness of the framework and to increase the prediction capabilities, deep learning (RNN) models such as Vanilla RNN (VRNN), Bi-directional Long Short-Term Memory (Bi-LSTM), Bi-Gated Recurrent Unit (Bi-GRU) and embedded Bi-LSTM-GRU, are trained with trade data from 219 countries, including India from 1970 to 2019 period and tested for predicting exports of India. To assess the predictive power of models, these models are evaluated and compared using different performance metrics. The results indicate that all RNN architectures can predict the exports of India except the Bi-LSTM model. The performance sequence of the RNN architectures is as follows: Vanilla RNN, Bi-LSTM-GRU, Bi-GRU and Bi-LSTM with lower MSE values.

This study explores the impact of a variety of deep learning (RNN) architectures on the prediction for exports in India. The results of this study depend on the selected data set. There is the possibility to apply the same models to predict other countries' exports and other economic variables as well. However, it is still required to tune the parameters or choice of hyperparameters and find the most optimal prediction model. This study suggests the use of more advanced deep learning models in various scenarios to model economic variables for providing further important insights for trade forecasting. This work has significant implications for advanced strategic decision-making to improve any country's economic outlook.

7 Future work

In future work, other hybrid RNN models can be developed and tested to improve predictive accuracy. Future work focusing on bilateral trade flow prediction, trade prediction in specific commodities, e.g., agriculture trade, multivariate response variables and prescriptive deep learning models to compare with conventional econometric models would greatly assist in public and private decision making.

References

Alpaydin, E. (2009) *Introduction to Machine Learning*, MIT Press.

Awokuse, T.O. (2007) 'Causality between exports, imports, and economic growth: evidence from transition economies', *Economics Letters*, Vol. 94, No. 3, pp.389–395. Doi: 10.1016/j.econlet.2006.08.025

Awokuse, T.O. and Christopoulos, D.K. (2009) 'Nonlinear dynamics and the exports–output growth nexus', *Economic Modelling*, Vol. 26, No. 1, pp.184–190. Doi: 10.1016/j.econmod.2008.06.009.

Batarseh, F. et al. (2019) 'Application of machine learning in forecasting international trade trends', *arXiv:1910.03112v1*.

Bengio, Y. (2009) 'Learning deep architectures for AI', *Foundations and Trends in Machine Learning*, Vol. 2, No. 1, pp.1–127. Doi: 10.1561/2200000006.

Bengio, Y., Courville, A. and Vincent, P. (2013) 'Representation learning: a review and new perspectives', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp.1798–1828. Doi: 10.1109/TPAMI.2013.50.

Chemeda, F.E. (2001) 'The role of exports in economic growth concerning Ethiopian country', *Conference on Annual Meeting of American Agricultural Economics Association in Chicago*, Chicago, USA.

Cui, Z., Member, S., Ke, R., Member, S. and Wang, Y. (2017) 'Deep stacked bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction', *Proceedings of the 6th International Workshop on Urban Computing*, Halifax, NS, Canada, pp.1–12.

Dickey, D.A. and Fuller, W.A. (1979) 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, Vol. 74, pp.427–431. Doi: 10.2307/2286348.

Ghartey, E.E. (1993) 'The causal relationship between exports and economic growth: some empirical evidence in Taiwan, Japan and us', *Applied Economics*, Vol. 25, No. 9, pp.1145–1152. Doi: 10.1080/00036849300000175.

Gopinath, M., Batarseh, F.A. and Beckman, J. (2020) *Machine Learning in Gravity Models: an Application to Agricultural Trade*, National Bureau of Economic Research, Cambridge Working Paper ID-27151.

Gupta, V. and Kumar, E. (2021) 'Review on machine learning techniques for international trade trends prediction', *3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp.544-547, Doi: 10.1109/ICAC3N53548.2021.9725585.

Helpman, E. and Krugman, P.R. (1985) *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, MIT Press, New York, NY, USA.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural Computation*, Vol. 9, No. 8, pp.1735–1780. Doi: 10.1162/neco.1997.9.8.1735.

Jung, J-K., Patnam, M. and Ter-Martirosyan, M. (2018) *An Algorithmic Crystal Ball: Forecasts-based on Machine Learning*, IMF Working Papers 18/230, International Monetary Fund.

Junoh, M.Z.M. (2004) 'Predicting GDP growth in Malaysia using knowledge-based economy indicators: a comparison between neural network and econometric approaches', *Sunway College Journal*, Vol. 1, pp.39–50.

Krugman, P. (1994) *Import Protection as Export Promotion: International Competition in the Presence of Oligopoly and Economies of Scale*, MIT Press, pp.185–197.

Kuo, R.J. and Li, P.S. (2016) 'Taiwanese export trade forecasting using re y algorithm based k-means algorithm and SVR with wavelet transform', *Computers and Industrial Engineering*, Vol. 99, pp.153–161. Doi: 10.1016/j.cie.2016.07.012.

Mangla, M. et al. (2021) 'Multivariate economic analysis of the government policies and COVID-19 on financial sector', *International Journal of Computer Applications in Technology (IJCAT)*, Vol. 66, No. 3/4, pp.294–302.

Nummelin, T. and Hanninen, R. (2016) 'Model for international trade of sawn- wood using machine learning models', *Natural Resources and Bioeconomy Studies*. Doi: 10.13140/RG.2.2.26478.41284.

Nummelin, T. and Hanninen, R. (2016) 'Model for international trade of sawn-wood using machine learning models', *Natural Resources and Bioeconomy Studies*, Vol. 74.

Omotor, D. (2008) 'The role of exports in the economic growth of Nigeria: the bounds test analysis', *International Journal of Economic Perspectives*, Vol. 2, No. 3, pp.222–235.

Quimba, F.M.A. and Barral, M.A.A. (2018) *Exploring neural Network Models in understanding Bilateral Trade in APEC: A Review of History and Concepts*, Discussion Papers DP 2018-33, Philippine Institute for Development Studies.

Rasheed, M.K. et al. (2022) 'The impact of oil exports on consumer imports in the Iraqi economy and COVID-19 period, a theoretical study', *International Journal of Computer Applications in Technology (IJCAT)*, Vol. 66, No. 3/4, pp.246–253.

Shen, M-L., Lee, C-F., Liu, H-H., Chang, P-Y. and Yang, C-H. (2021) 'Effective multinational trade forecasting using LSTM recurrent neural network', *Expert Systems with Applications*, Vol. 182. Doi: 10.1016/j.eswa.2021.115199.

Sirgy, M.J., Lee, D.J., Miller, C., Littlefield, J.E. and Atay, E.G. (2007) 'The impact of imports and exports on a country's quality of life', *Social Indicators Research*, Vol. 83, No. 2, pp.245–281.

Stlouisfed (n.d.) *Federal reserve economic data*. Available online at: https://fred.stlouisfed.org

Sun, J. et al. (2018) 'Analysis of bilateral trade flow and machine learning algorithms for GDP forecasting', *Engineering, Technology and Applied Science Research Journal,* Vol. 8, No. 5, pp.3432–3438.

UNCATD (n.d.) *Statistics*. Available online at: https://unctadstat.unctad.org/EN/

Urrutia, J.D., Abdul, A.M. and Atienza, J.B.E. (2019) 'Forecasting philippines imports and exports using Bayesian artificial neural network and autoregressive integrated moving average', *AIP Conference Proceedings.* Doi: 10.1063/1.5139185.

Veenstra, A.W. and Haralambides, H.E. (2001) 'Multivariate autoregressive models for forecasting seaborne trade OWS', *Transportation Research Part E: Logistics and Transportation Review*, Vol. 37, No. 4, pp.311–319. Doi: 10.1016/S1366-5545(00) 00020-X.

Xiong, C., Merity, S. and Socher, R. (2016) 'Dynamic memory networks for visual and textual question answering', *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, Vol. 48, pp.2397–2406.