

# Deep-Learning Modelling of Dynamic Panel Data for African Economic Growth

Nureni Olawale Adeboye\*<sup>a</sup> and, Nurudeen Olawale Alabi<sup>b</sup>

<sup>a</sup> *Department of Mathematics & Statistics, The Federal Polytechnic, Ilaro, Nigeria*

<sup>b</sup> *Department of Mathematics & Statistics, The Federal Polytechnic, Ilaro, Nigeria*

## Article History

Received : 20 January 2022; Revised : 12 February 2022; Accepted : 09 March 2022; Published : 30 June 2022

## To cite this paper

Nureni Olawale Adeboye & Nurudeen Olawale Alabi (2021). Deep-Learning Modelling of Dynamic Panel Data for African Economic Growth. *Journal of Econometrics and Statistics*. 2(1), 47-60.

When modelling phenomena relating to the economy, dynamic panel models have proven to be a useful tool. Previous studies have modelled dynamic panel data using conventional methods of generalized method of moment, Instrumental variables and Maximum likelihood estimators among others. This study however focuses on modelling dynamic panel data using modern day approaches of deep-learning techniques. To this end, two macro-economic variables of Purchasing Power Parity (PPP) and Gross National Income (GNI) were employed to model the economic growth of twenty African countries. Dynamic panel information about these countries were sourced from UNESCO database between 1990 and 2019. Deep learning techniques of Long Term Short memory (LSTM), Bidirectional Long Short Term Memory (Bi-LSTM) and Gated Recurrent Units (GRU) were employed in the modelling process, and the findings revealed that LSTM having the least values of the adopted evaluation metrics, is the best and most suitable deep learning method for modelling dynamic panel data. Forecasts were also made for the next 20 years with the techniques, and the results show that LSTM gives the best predicting accuracy with its lowest Mean Absolute Error (MAE), MAPE, MSE and RMSE.

**keywords:** Africa Economic Growth, Deep-Learning, Dynamic Panel Data, Evaluation Metrics, Forecast.

\*Corresponding authors: [nureni.adeboye@federalpolyilaro.edu.ng](mailto:nureni.adeboye@federalpolyilaro.edu.ng)

©ARF India

ISSN: 2583-0473

<https://arfjournals.com/jes>

# 1 INTRODUCTION

In the last decade, dynamic panel models are increasingly gaining popularity amongst statisticians, economists and data scientist alike because of the increasing availability of widely known and feasible algorithms and software for their computations. Their use has provided new insights and understanding to econometric data (Xu, 2016). Some recent examples include investigations by Flannery and Hankins (2013), Foster (2008), Dutt et al. (2013), Pugh et al. (2014), in the areas of corporate finance, economic growth, foreign aids and education respectively.

Dynamic panel data like most times series data are often modelled by including time-lagged endogenous variable in the specified regression equation. This additional variable changes the interpretation of the variables, which now indicate correlations conditional on the history of the model. This situation shows that the lagged dependent variable violates strict exogeneity, and since common panel data estimators of fixed and first differences estimators both rely on the assumption of strict exogeneity, alternative techniques usually employed are the Instrumental variable and the GMM methods which are commonly used for solving problem of similar situation. Xu (2016) employed the use of particle filters to estimate a stochastic volatility induced dynamic panel data, and Monte Carlo studies revealed the goodness of the chosen estimator. However, this research tends to focus more on the estimation of a dynamic panel model with the use of Deep Learning techniques which has been less studied or adopted by Econometricians, as well as in major statistical field.

In literature, many researchers have studied the nexus existing amongst economic development and its predictors for developing countries which are mostly in exhaustive and not conclusive. These studies are that of Firebaugh (2002); Loizides and Vamvoukas (2005); Levine (2005); Ulku (2004) and Artelaris et al. (2005); Obadan (2006); Petrakos et al. (2007) Pritchett (2001); Mallik (2008); Arezki and Gylfason (2011); Collier and Goderis (2012); Yavari and Mohseni (2012). The main predictors of economic development included in these aforementioned studies are domestic investment, government expenditure, credit expansion, research and development, foreign aid, demographic trend, commodity prices, credit to private sector, trade openness. This list is not exhaustive.

There is a paucity of literature that relate directly to the modelling of dynamic panel data using deep learning techniques. Amongst the few existing literatures are that of Raissi (2018); Raissi et al. (2019); Reichstein (2019); Ye et al. (2019). While the first three authors only employed the application of deep learning techniques in the solution of nonlinear partial differential equations, solution of data-driven and data-driven discovery problems of partial differential equations, and solution of process understanding for data-driven earth system science respectively, only Ye et al. (2019) has employed deep learning in the analysis of dynamic panel data. The work of the authors was the pioneering research in the application of deep learning in the exploration of influential factors from dynamic panel data. The authors employed the technique of Long Short-Term Memory deep learning neural networks to a multi-dimensional, high-resolution spatiotemporal dataset which runs on a high-performance computing cluster.

It was observed that Deep learning techniques greatly outperformed a cutting-edge spatial-econometric model at continental, state, and grid-cell scale levels. Thus, this research adopted dynamic panel model to aims at studying the pattern and extent of the cause-effect relationship of the considered macro-economic indicators on the economic development across some African countries, in order to explore some policy implications. PPP uses market "basket of goods" approach to make comparisons between currencies of different countries. This concept establishes that two currencies are in equilibrium or at par whenever prices of a market basket of goods is the same in both countries, taking into account the exchange rate. GNI, on the other hand, is an estimate of the value of all goods and services produced by the nationals of the country whether currently resident in the country or living aboard. The primary focus of government policies in African countries is to have high and sustainable economic growth which are necessary and sufficient conditions for extensive development. Generally, the presence of sustainable economic growth in a country is a manifestation of advancement in the socio-economic endowment of the citizenry. Becsi and Wang (2002) asserted that a decline in growth rates, as seen in most developing countries, is an indicator of a drop in people's standard of life, leading to poverty.

Africa, the second largest continent (in terms of size) is the world's poorest region with numerous economic, social and political crisis. It is the richest continent in the world in terms of mineral deposit such as cobalt, platinum, gold, chromium, diamond, tantalite, manganese, iron ore and uranium etc. Agriculture accounts for about 67 per cent of the labour force, about 35 per cent and 40 per cent of GNP and foreign exchange earnings respectively. In spite of these abundant human and natural resources, African countries remain the world's poorest and most underdeveloped. Consequently, there is the needs for the continuous modelling of her economic growth.

Due to rising data of Gross Domestic Product (GDP), the African economy has been growing every year in recent years. Despite this fact, nearly half of the nations in Africa are currently battling with poverty, poor living standard and unemployment. Thus, an efficient modelling of African economic growth is highly needed, which will be capable of providing a true measure of economic growth for the populace using appropriate machine learning techniques. It is pertinent to note that understanding the future evolution of the Economic Growth of a country is very important as it serves as a great basis for proper planning, allocation of funds, and decision making.

## **2 MATERIALS AND METHODS**

Due to rising data of Gross Domestic Product (GDP), the African economy has been growing every year in recent years. Despite this fact, nearly half of the nations in Africa are currently battling with poverty, poor living standard and unemployment. Thus, an efficient modelling of African economic growth is highly needed, which will be capable of providing a true measure of economic growth for the populace using appropriate machine learning techniques. It is pertinent

to note that understanding the future evolution of the Economic Growth of a country is very important as it serves as a great basis for proper planning, allocation of funds, and decision making.

## 2.1 Model Specification

The functional dynamic panel data model for this research is specified as

$$GDP_{it} = \alpha_{it} + \beta (PPP)_{it} + \gamma (GNI)_{it} + e_{it} \quad (1)$$

$$GDP_{it} = \alpha_{it} + \beta (PPP)_{it} + \varphi GDP_{it-1} + e_{it} \quad (2)$$

Where  $i$  and  $t$  denote different African countries and the sample time dimension respectively. GDP is the measure of growth for the African countries;  $\alpha, \beta, \gamma$  and  $\varphi$  are the estimable parameters while  $e_{it}$  is the error term. Considering the fact that the countries are heterogeneous, a panel unit root test was carried out on the variables through the adoption of Lm *et al.* (2003) (LPS) test for individual unit root process given as

$$\Delta y_{it} = \rho_{it} y_{i,t-1} + \sum_{L=1}^{pi} \phi_{iL} \Delta y_{i,t-L} + z'_{it} y + u_{it} \quad (3)$$

The LPS test assumes the unit root can differ across the cross sections in the model.

## 2.2 Estimation Techniques

Three different deep-learning techniques of Long Term Short memory (LSTM), Bidirectional Long Short Term Memory (Bi-LSTM) and Gated Recurrent Units (GRU) as employed in Edmond Zhang *et al.* (2018) were adopted for modelling and forecasting for comparison.

**Long Term Short Memory (LSTM):** Previous studies have employed Recurrent Neural Networks (RNN, *hereon* in addressing temporal dependencies in sequential time series applications. Despite its benefits, according to Farah *et al.* (2020), RNN has the problem associated with training the long term data dependencies. Schmidhuber & Hochreiter (1997) developed a procedure for solving this problem in Long Term Short Memory (LSTM, *hereon* which is a variant of RNN. This procedure uses memory cells, which are hidden layer units with self-connections that store the network's temporal state and are controlled by three gates: input gate, output gate, and forget gate. Memory cell inputs and outputs are controlled by the input and output gates in the rest of the network. The forget gate sends high-weighted output information from one neuron to the next. The information saved in memory is determined by the high activation results of the input unit; for example, if the input unit has a high activation, the information is kept in memory cell. However, if there exists a high activation in the output unit, information is passed to the next neuron. Otherwise, high-weighted input data is stored in memory cells. The LSTM network is created by mapping the input and output sequences denoted as

$$x = (x_1, x_2, \dots, x_n) \text{ and } y = (y_1, y_2, \dots, y_n)$$

Calculated by the following equations:

$$\text{forget gate} = \text{sigmoid}(W_{fg}X_t + W_{hfg}h_{t-1} + b_{fg}) \quad (4)$$

$$\text{input gate} = \text{sigmoid}(W_{ig}X_t + W_{hig}h_{t-1} + b_{ig}) \quad (5)$$

$$\text{output gate} = \text{sigmoid}(W_{og}X_t + W_{hog}h_{t-1} + b_{og}) \quad (6)$$

$$(C)_t = (C)_{t-1} \text{forget gate} + \text{input gate}_t (\tanh((W_cX_t + W_{hc}h_{t-1} + bC)) \quad (7)$$

$$h_t = \text{output gate} + \tanh((C)_{t-1}) \quad (8)$$

In the above equations,  $W_{fg}$ ,  $W_{ig}$ ,  $W_{ag}$ ,  $W_c$ , and  $b_{fg}$ ,  $b_{ig}$ ,  $b_{og}$ ,  $b$  and  $C$  denote the weights and bias variables of three gates and a memory cell, respectively. Also,  $h_{t-1}$  represents the prior hidden layer's units that added element-wise to the weights of three gates. After the execution of model (7),  $(C)_t$  metamorphosed into current memory cell unit and the element-wise multiplication of prior hidden unit outputs and previous memory cell unit is shown in Model (8). By adding the non-linearity on top of the three gates the tanh and sigmoid activation functions are generated. Further,  $t - 1$  and  $t$  are previous and current time steps.

**Bidirectional Long Short Term Memory:** Schuster & Paliwal (2012) proposed a bi-directional recurrent neural network (BRNN) that consists of two independent LSTM hidden layers with similar output in opposite directions to address the restrictions of an LSTM cell that can work on previous content but not future content. Previous and future information are used in the output layer with this design. Thus, both activations (forward, backward) would be considered to calculate the output  $\hat{y}$  at time  $t$  given as

$$\hat{y}_t = g(W_y [a_t^+, a_t^-] + b_y) \quad (9)$$

**Gated Recurrent Unit (GRU):** Gated Recurrent Unit (GRU, *hereon*) is a two gates LSTM variant comprising an "update gate" which is composed of input, forget gates and a "reset gate". Due to inexistence of an additional memory cell to store data, GRU can only control data within the unit.

$$\text{Up gate} = \text{sigmoid}(W_{ug}X_t + W_{hug}h_{t-1} + b_{fg}) \quad (10)$$

$$\text{Rest gate} = \text{sigmoid}(W_{rg}X_t + W_{hrg}h_{t-1} + b_{ig}) \quad (11)$$

$$h_t = \tanh W(\text{rest gate})_t (Wh_{t-1}, X_t) \quad (12)$$

$$h_t = 1 - \text{update gate}_t h_{t-1} + (\text{update gate}, X_t) \quad (13)$$

The *update gate* in Equation (10) controls the amount of information updated. If the gate is set to zero, the *rest gate* in Equation (11) behaves like the update gate by reading input sequences and forgetting the previously calculated state. Furthermore,  $h_t$  shows the same functionality as in recurrent unit and  $h_t$  of GRU at time  $t$  represents the linear interpolation among the current  $h_t$  and previous  $h_{t-1}$  activation states in Equations (11) and (12) respectively.

### 2.3 Evaluation Metrics

The four performance measures used to evaluate the performance of the proposed models are Mean absolute error (MAE), Mean square error (MSE), Root mean square error (RMSE) and Mean absolute percentage error (MAPE) where  $C$  denotes the actual value,  $C^*$  for estimated value and  $N$  is the number of times the summation iteration is executed.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^M |C - C^*| \quad (14)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^M (C - C^*)^2 \quad (15)$$

RMSE is well defined as:

$$\text{RSME} = \sqrt{\frac{1}{M} \sum_{i=1}^M (C - C^*)^2} \quad (16)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^M \left| \frac{c - c^*}{c} \right| \quad (17)$$

## 3 RESULTS

Data preprocessing becomes a necessary step in compliance with the requirement for modelling with deep learning techniques. The original dataset is a panel data with 600 observations from which 16 missing values were traced to GNI. Thus, one cannot simply use simple descriptive measures of mean, median or mode to fill out the missing values in the dataset as required by a normal machine learning or time series modelling. Therefore, the research applied a backward filling method in the imputation of the missing values. It is pertinent to note that there is no evidence of outliers in the dataset, which was converted to dynamic panel data by taking the lag of the dependent variable (GDP) and included the lagged variable as part of the explanatory variables.

The dataset was well prepared, and then transformed into a supervised learning problem using the *"series-to-supervised"* function which was developed specifically for this research. The inputs were normalized for robust prediction of the selected countries' economic growth. Normalized data leads to higher performance in Neural Networks, according to a good rule of thumb, hence a data normalization technique called MinMaxScaler from Scikit-learn was employed and dataset converted into a 3D input num-samples, num-timesteps, num-features as a necessary condition for deep-learning modelling. A time steps of 10 was used which implies that the model makes projections based on the last ten years of the data.

Table 1 below shows the descriptive statistics of the dataset after dropping the duplicates. The dataset cannot be randomly split into 80% and 20% train

and test data basically because this data was from different cross-sections at different time frame. Thus, the data splitting was carried out for each country in the first instance, and each country's training and test data were combined as final train data and final test data respectively. It was observed that the minimum height is 55cm and that of the weight is 10kg. It is observed that the minimum GDP and GNI for all the counters respectively is 8,003,491 and 2,176,091 dollars while the maximum is 11,754,730 and 4,454,235 dollars respectively. These results show that there are estranged values in the dataset and It is thus imperative to drop the values that are above 75% percentile and those below 25% percentile distributions.

The parity in the countries purchasing power is estimated at a negative average rate of 2,000,000 dollars, which implies that the purchasing power among the chosen countries is largely being eroded rather than improving.

**Table 1: Descriptive Statistics Table**

	<b>GDP</b>	<b>PPP</b>	<b>GNI</b>
<b>Count</b>	600	600	584
<b>Mean</b>	9.996470	1.530795	3.247477
<b>Standard deviation</b>	0.710921	1.155709	0.454293
<b>Min</b>	8.003491	-2.000000	2.176091
<b>25%</b>	9.479272	0.451000	2.892095
<b>50%</b>	9.973737	2.144775	3.244275
<b>75%</b>	10.398536	2.363880	3.507179
<b>Max</b>	11.754730	3.563451	4.454235

All the modelling and preprocessing approaches were carried out using Python 3.4 which was computed on Google Colab, and this provides a free computational power for the training of the model. Scikit Learn, Pandas and Numpy library were used for the preprocessing, the visualization was carried out using Matplotlib and Seaborn Library; and the modelling was carried out using Tensorflow and Keras.

Specifically, the performances of LSTM, GRU and BiLSTM models are compared in forecasting the rate of Economic growth in the selected African countries. To achieve a robust specification, the research model was trained with the 80% of the partitioned dataset using LSTM, BiLSTM and GRU techniques after which prediction was made on the test set and the output of the prediction made on the remaining 20% of the partition dataset as shown in Fig 1.



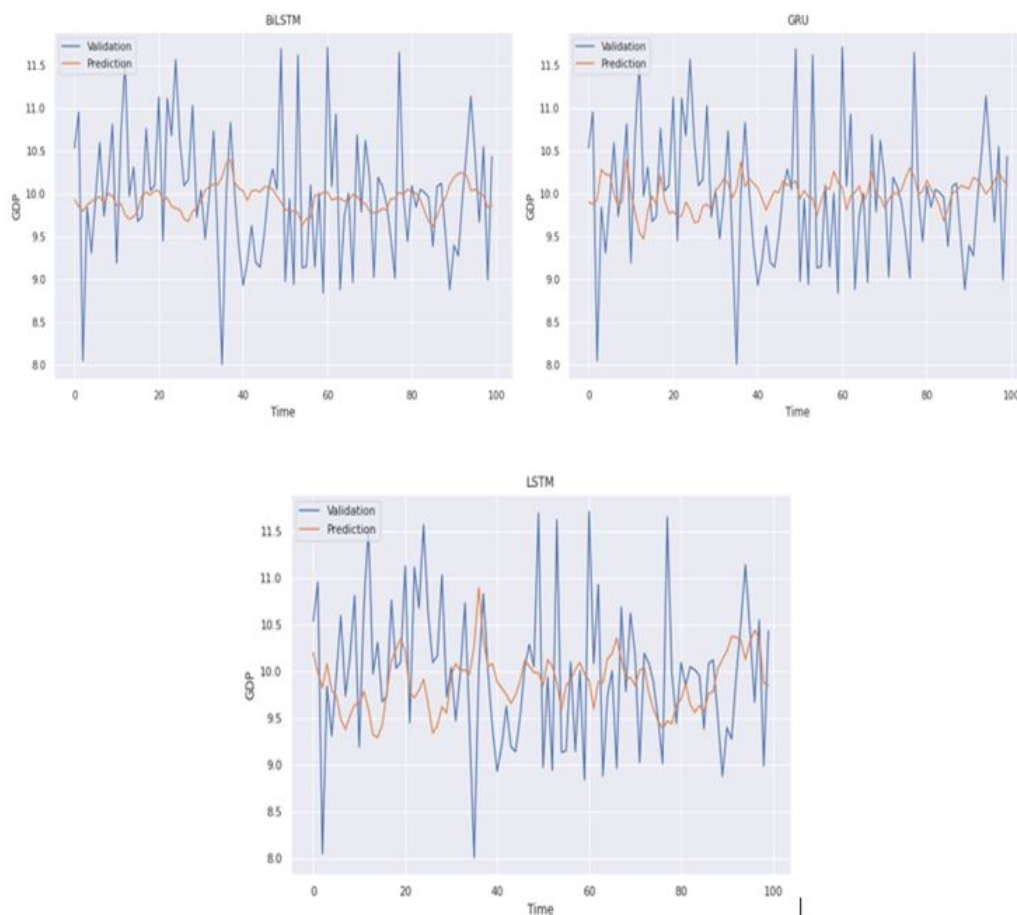


Figure 1: Validation and Prediction of each forecasting approach

Listed in Table 2 below are the parameters of the constructed Neural Networks models using training data.

**Table 2: Neural Networks Model Parameters**

Model	Hidden Unit	Epochs	Learning rate	Optimizer	Batch size
<b>LSTM</b>	64	100	0.001	Adam	32
<b>BiLSTM</b>	64	100	0.001	Adam	32
<b>GRU</b>	64	100	0.001	Adam	32

These parameters are obtained by minimizing the cross-entropy of the reconstructed error during the training. For each of the LSTM, Bi-LSTM and GRU models, we created the number of neurons in hidden layers, number of units in hidden layers and model name. The adopted techniques all have 64 neurons in the input layer, one hidden layer including 64 neurons and 1 neuron in the output layer. To make the LSTM and GRU model robust to changes, **Dropout** function “**Dropout (0.2)**” which randomly drops 20% of units from the network was used.



The model was trained with a Batch Size of 32 while the remaining 20% of the training data was used for validation. However, in order to avoid over-fitting, an *early stop* process was used to truncate training whenever *validation loss* did not improve after 10 epochs just as illustrated in Fig. 2.

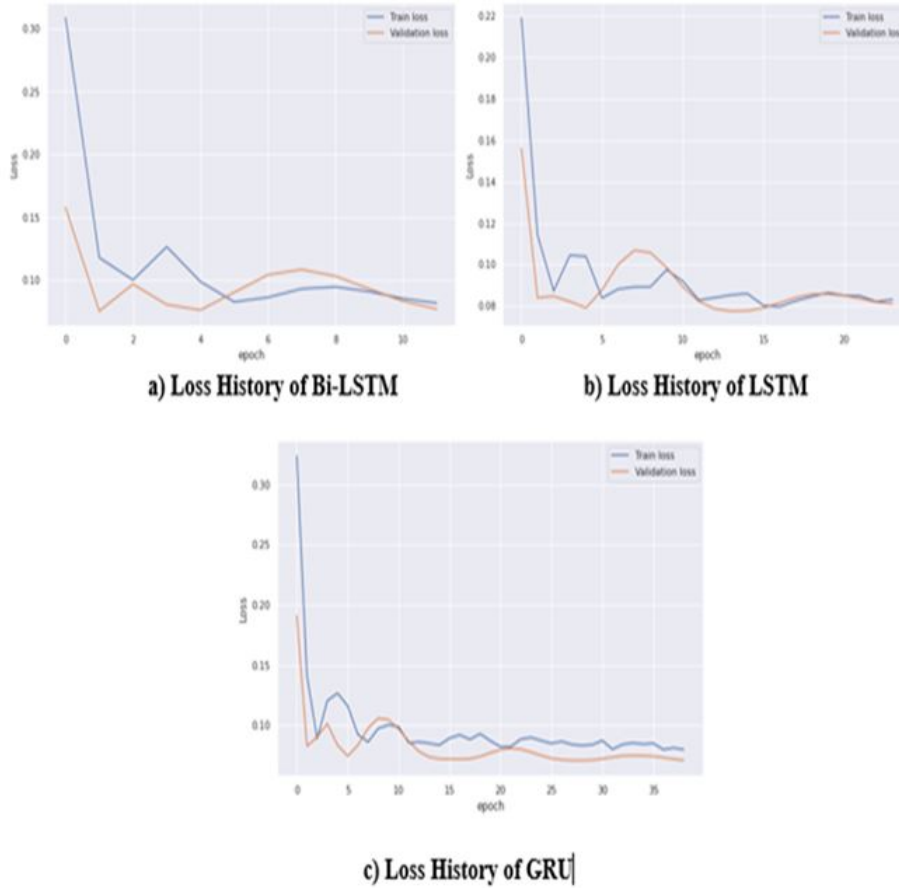


Figure 2: Loss History of the Fitted Models

Table 2: Neural Networks Model Parameters

Model	MAE	RMSE	MSE	MAPE
Bidirectional-LSTM	0.566	0.7187	0.5165	5.6221
LSTM	0.5519	0.6978	0.4869	5.5217
GRU	0.569	0.7157	0.5123	5.6606

The mean absolute percentage error (*MAPE*) which is the mean or average of the absolute percentage errors of forecasts was shown in Table 3 alongside other measures of error adopted, and a corresponding visual description displayed in Fig.3.

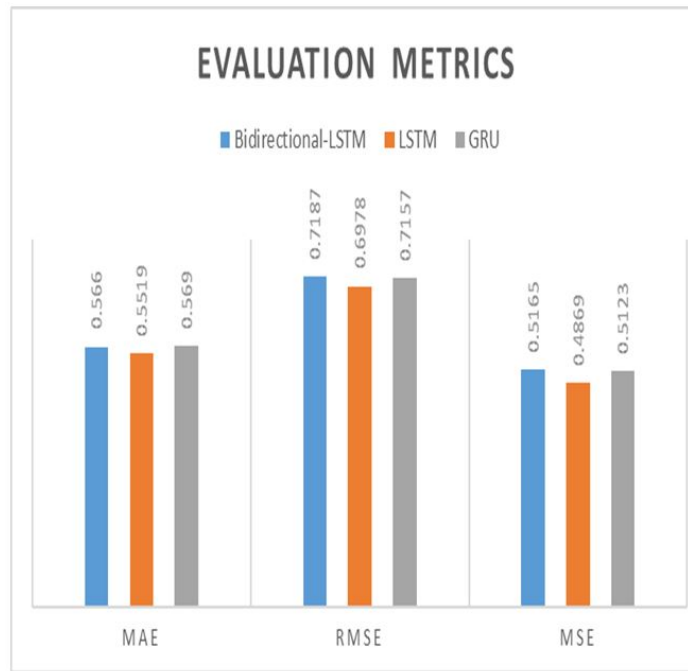


Figure 3: Performance of Fitted Models

The results of MAE, RMSE, MSE and MAPE of the three models as tabulated in **Table 3** showed that the technique with the least measures of error is LSTM with respective values of 0.5519, 0.6978, 0.4869 and 5.5217. Thus, LSTM performed better than the remaining two techniques in the modeling of our specified dynamic panel model.

Furthermore, the predicted and actual plots of each of the utilized models is shown in Fig. 4.

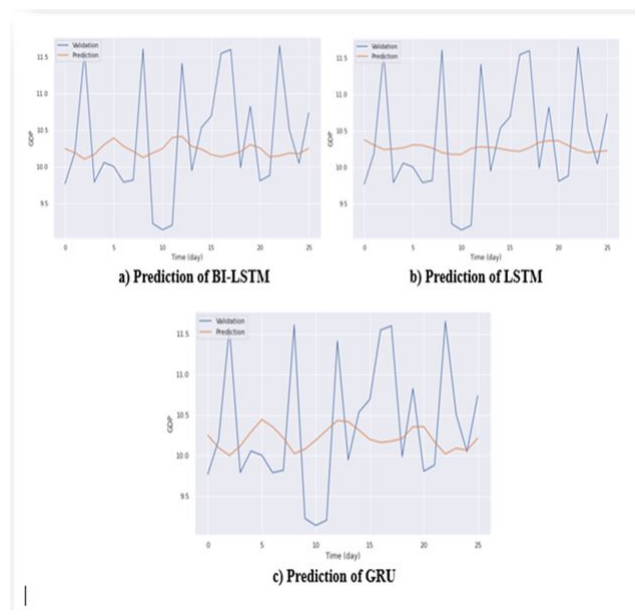


Figure 4: Performance of Fitted Models

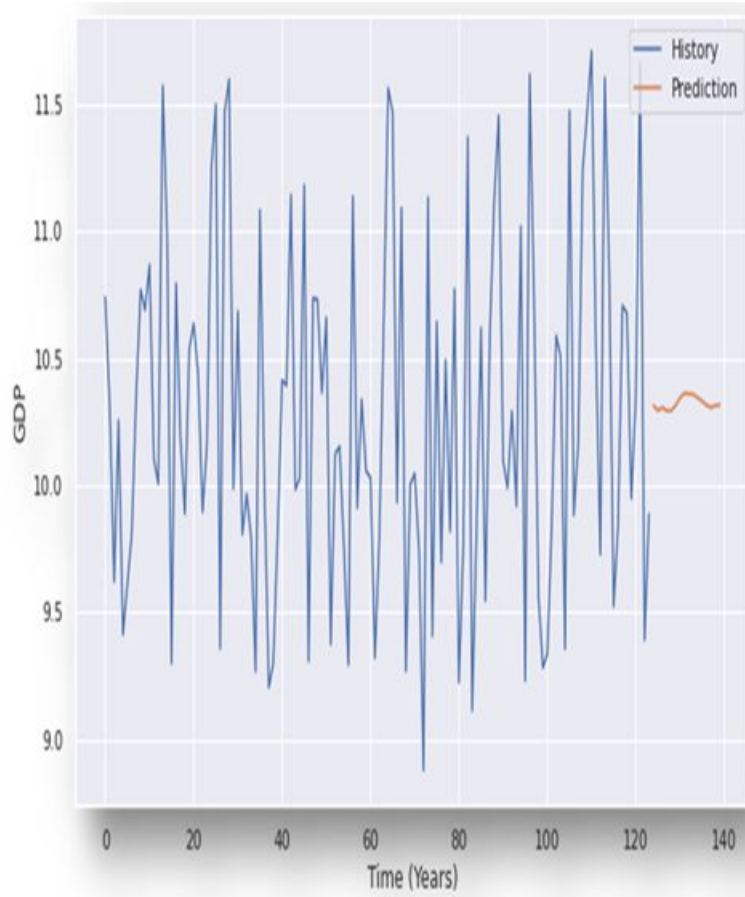


Figure 5: Economic Growth Forecasts

Considering the results of the models' performances, we safely concluded that after parameter training, LSTM outperforms the other methods in predicting the Economic growth of African countries. Thus, Fig. 5 depicts the prediction of the next 20 years for all the countries using LSTM technique.

The forecasts showed that the economic growth of the African countries will maintain a steady growth with a constant trend for most part of the forecasts years.

## 4 DISCUSSIONS AND CONCLUSIONS

Considering the results obtained after parameter training, LSTM performs better in predicting the Economic growth of African countries giving its lowest MAE, MSE, MAPE and RMSE amidst other methods used in the research. Also, forecasts showed that the economic growth of the African countries will maintain a steady growth with a constant trend for most part of the forecasts years. It is thus recommended that LSTM should be adopted over Bi-LSTM & GRU when modelling dynamic panel data with deep learning techniques.

It will be of interest to note that this research is not targeted at discrediting the application of convectional modeling techniques for dynamic panel data,

otherwise it is targeted at exploring the suitability of modern techniques of deep learning in modeling dynamic panel data. While deep learning targets accurate prediction of the models based on the splitting of dataset into training and test data, the conventional methods are more superior in terms of model interpretability of which major concerns are about the parameter estimates, standard error, P-value and confidence interval.

## Conflict of Interest

The authors confirm that this article content has no conflict of interest.

## References

- Adeboye, N. O., Osuolale, P. P. & Ogunnusi, O. N. (2020). Data Science Skills: Building Partnership for Efficient School Curriculum Delivery in Africa. *Statistical Journal of the IAOS*, 36 (S1), 49-62, DOI 10.3233/SJI-200693.
- Arezki, B. & Gylfason, T (2011). A Dynamic Panel Data Analysis for Relationship between Structural Policy and Economic Growth. *International Journal of Econometrics and Financial Management*. 4 (1), 1-10. <http://pubs.sciepub.com/ijefm/4/1/1>.
- Artellaris, A., Calamitsis E. A. & Ghura D. (2005). Adjustment and Growth in Sub-Saharan Africa. Economic Issue is based on IMF Working Paper 99/51.
- Becsi, Z., & Wang, P. (2002). Financial development and growth. *Economic Review*, 82, 46–62.
- Collier & Goderis (2012). Estimation of Nonlinear Dynamic Panel Data Models with Individual Effects. *Mathematical Problems in Engineering*, Article ID 672610.
- Dutt, N., Lahiri, K., & Montiel, P. (1990). A Rational-Expectations Macro Econometric Model for Developing countries. *International Monetary Fund*, 37 (3), 537.
- Edmond Zhang, Reece Robinson, Bernhard P. (2018). Deep Holistic Representation Learning from EHR, 2018. 12th International Symposium on Medical Information and Communication Technology (ISMICT).
- Farah Shahid, Aneela Zameer, Muhammad M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*, 2020.
- Firebaugh, C. (1992). Macroeconomic Policy Rules in Theory and Practice. *Policy Rules—The Next Steps*, 1-31.

- Cambridge University, Cambridge UK: Monetary Policy Committee, Bank of England.
- Flannery, R. and S. Hankins (2013). Initial conditions and moment restrictions in dynamic panel-data models. *Journal of Econometrics*, 87, 115–143.
- Foster, E. (2008). *Econometric Methodology at the Cowles Commission: Rise and Maturity*. *Econometric Theory*; 4(02), 187-209.
- Levine O. (2005). Another look at the instrumental variable estimation of error component models. *Journal of Econometrics*, 68, 5–27.
- Lm, K., Pesaran, H., & Shin, Y. (2003). Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, 115, 53-74.
- Loizides, D. & Vamvoukas, F.G (2005) Assessing Simple Policy Rules: A view from a Complete Macroeconomic Model. *Federal Reserve Bank of St. Louis review*, 83(4), 83-110.
- Mallik, R. (2008). Health and economic growth: Evidence from dynamic panel data of 143 years. *PLoS ONE* 13(10): e0204940. <https://doi.org/10.1371/journal.pone.0204940> unpublished MA, Department of Economics, Central European University.
- Obadan, M. (2006). *Prospects for Diversification in Nigeria's Export Trade in Annual Conference of the Nigerian Economic Society*, Heinemann press, Ibadan.
- Pugh, R. (2017) Universities and economic development in lagging regions: ‘triple helix’ policy in Wales, *Regional Studies*, 51:7, 982-993, DOI: 10.1080/00343404.2016.1171306.
- Petrakos, G., Paschalis, A., & Sotris, P. (2007). Determinant of Economic Growth: The Experts' View. EU Working Paper.
- Pritchett, L. (2018). Alleviating Global Poverty: Labor Mobility, Direct Assistance, and Economic Growth. Working Paper 479. Washington, DC: Center for Global Development <https://www.cgdev.org/sites/default/files/alleviating-global-poverty-labor-mobility-direct-assistance-and-economic-growth>.
- Raissi, M., Perdikaris, B., Karniadakis, G, E. (2018). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*. 378, 686-707. <https://doi.org/10.1016/j.jcp.2018.10.045>.
- Raissi, M. (2018). Deep Hidden Physics Models: Deep Learning of Nonlinear Partial Differential Equations, *Journal of Machine Learning Research* 19 (2018) 1-24.

- Reichstein, M., Camps-Valls, G., Stevens, B. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature* 566 (2019), 195–204 Science. *Nature* 566, 195–204 (2019). <https://doi.org/10.1038/s41586-019-0912-1>.
- Schmidhuber, J. & Hochreiter. S. (1997). Long short-term memory. *Neural Computation*, 9,1735–1780, 1997.
- Schuster M. & Paliwal K.K (2012). Bidirectional recurrent neural networks, *IEEE Transactions on Signal Processing*, 45(11), 2673-2681.
- Ulku, F (2004). Foreign Direct Investment: Flows, Volatility and the Impact on Growth, *Review of International Economics*, 14(3), 478-493.
- Xu W. (2016). Estimation of Dynamic Panel Data Models with Stochastic Volatility Using Particle Filters. *Econometrics*4(4),39;<https://doi.org/10.3390/econometrics4040039>.
- Yavari, C. C., Moseni B.V (2012). Macroeconomic modelling and economic policy making: A survey of experience in Africa.
- Ye L, Gao, L., Marcos-Martinez, R., Mallants, D., Bryan, A. B. (2019). Projecting Australia’s forest cover dynamics and exploring influential factors using deep learning. *Environmental Modeling and solution*, 119 (2019), 407-417. <https://doi.org/10.1016/j.envsoft.2019.07.013>.
- Yi Hu, Dongmei G., Ying D., and Shouyang W. (2014). Estimation of Nonlinear Dynamic Panel Data Models with Individual Effects. *Mathematical Problems in Engineering*, Article ID 672610, <http://dx.doi.org/10.1155/2014/672610>.