



# Comparison of Active COVID-19 Cases per Population Using Time-Series Models

Sakinat Oluwabukonla Folorunso, Olabisi Onabanjo University, Ago Iwoye, Nigeria

Joseph Bamidele Awotunde, University of Ilorin, Ilorin, Nigeria

 <https://orcid.org/0000-0002-1020-4432>

Oluwatobi Oluwaseyi Banjo, Olabisi Onabanjo University, Ago Iwoye, Nigeria

 <https://orcid.org/0000-0002-5440-7442>

Ezekiel Adebayo Ogundepo, Data Science Nigeria, Nigeria

Nureni Olawale Adeboye, Federal Polytechnic, Ilaro, Nigeria

## ABSTRACT

This research explored the precision of diverse time-series models for COVID-19 epidemic detection in all the 36 different states and the Federal Capital Territory (FCT) in Nigeria with the maximum count of daily cumulative of confirmed, recovered, and death cases as of 4 November 2020. A 14-step forecast system for active coronavirus cases was built, analyzed, and compared for six different deep learning-stimulated and statistical time-series models using two openly accessible datasets. The results obtained showed that based on RMSE metric, ARIMA model obtained the best values for four of the states (0.002537, 0.001969, 1.2E-058, 5.36E-05 values for Lagos, FCT, Edo and Delta states, respectively). While no method is all-encompassing for predicting daily active coronavirus cases for different states in Nigeria, ARIMA model obtains the highest-ranking prediction performance and attained a good position results in other states.

## KEYWORDS

ARIMA, COVID-19, Deep Learning, Epidemic, Exponential Smoothing, Forecasting

## 1. INTRODUCTION

The novel severe contagious respiratory syndrome coronavirus called (COVID-19) has cause greatest global challenge and public health, after the pandemic of influenza outbreak of 1918. The spread of CORonaVirus Disease 2019 (COVID-19) has caused significant changes in the style of living of communities around the world. The onset of the epidemic was discovered in Wuhan, China in December, 2019 and was triggered due to Severe Acute Respiratory Syndrome Coronavirus 2 (SARSCoV-2) virus (Ogundokun, Lukman, Kibria, Awotunde, & Aladeitan, 2020). As stated by World Health Organization (WHO) there have been 30, 949,804 confirmed cases and 959,116

DOI: 10.4018/IJEHMC.20220701.oa6

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

deaths of COVID-19 Globally, as of 4:30pm CEST, 21 September 2020 and affecting more than 180 countries, and due to the exponential level of spread, WHO stated COVID-19 as an epidemic. These make governments of different countries to take immediate control actions like separating profoundly influenced regions, stopping the cross-border traffic between nations, shutting down schools, working environments, regular spots, limiting resident's movements by encouraging them to stay at as much as possible. The social life and economics had significantly affected due to these measures put in place. The societies face different challenges ranges from education, healthcare, manufacturing, supply chain management travel, tourism and service delivery under the prevailing conditions and in a post COVID-19 world. As an example, the overcrowded of hospitals and other healthcare facilities due to exponential increase of COVID-19 patients and the inability to provide medical assistance to normal patients due to limited movement are significant barriers to the fight against COVID-19 in the healthcare sector. Similarly, delays and the increased resource demand for manual contact tracing, and unavailability of effective and automated contact tracing software impede the actions for controlling the spread. Hence, it is the duty of different parties to act with their fullest potential to control the prevailing COVID-19 situation, such parties are healthcare workers, government authorities, students, researchers, engineers, technology managers and the general public. In order to not only safeguard, but also handle the post-COVID-19 environment, digitalization and the implementation of information and communication technologies will be imperative. Technologies like Big data, 5G communication, Internet of Things (IoT), cloud computing, Artificial Intelligence (AI), and blockchain play vital role to assist the environment adopting to different protection and improvement of people and economies. The technologist and engineers will have to tackle important challenges to implement these promising solutions and realize their benefits, and carry out elaborate findings regard to risk management, resources, cost, scope, and quality.

In order to predict the prevalence of coronavirus in Nigeria, some researchers recently used the Linear Regression (LR) and curve estimation model (Ogundokun, Lukman, Kibria, Awotunde, & Aladeitan, 2020; Lukman, et al., 2020). In India (Pandey, Chaudhary, Gupta, & Pal, 2020) used the SEIR and Regression model to predict COVID 19 and LR to estimate the count of deaths in India due to coronavirus. (Ghosal, Sengupta, Majumder, & Sinha, 2020) proposed AutoRegressive Integrated Moving Average (ARIMA) model is among the numerous statistical methods that are used to forecast disease events. In multiple disease outbreaks, many researchers have used the ARIMA model for disease prevalence / incidence prediction. These include (Guan, Huang, & Zhou, 2004; Earnest, Chen, Ng, & Sin, 2005; Gaudart, et al., 2009; Liu, Liu, Jiang, & Yang, 2011; Nsoesie, Beckman, Shashaani, Nagaraj, & Marathe, 2013; Zheng, Zhang, Zhang, Wang, & Zheng, 2015; Fang, et al., 2020; Polwiang, 2020; Cao, et al., 2020)

Hence, this study is intended to compared six (6) statistical and deep learning-stimulated time-series models to evaluate the proportion of active coronavirus cases with reverence to the entire populace for the thirty-seven (37) states with the most active cases with multi step forecasting of 14 days (2 weeks). Specifically, six time-series models namely ARIMA (Box & Jenkins, 1990), the Holt–Winters Additive model (HWAAS) (Chatfield, The Holt-Winters Forecasting Procedure, 1978), TBAT (De Livera, Hyndman, & Snyder, 2011), Facebook's Prophet (Taylor & Letham, 2018), DeepAR (Salinas, Flunkert, Gasthaus, & Januschowski, 2020) but implemented as GluonTS (Alexandrov, et al., 2019) and N-BEATS (Oreshkin, Carpov, Chapados, & Bengio, 2019) were applied and their performances were accessed based on Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to find the optimum model that can predict future infections with least error for active cases in Nigerian states. These models were evaluated on publicly available COVID-19 dataset from NCDC website from 2/23/2020 to 11/4/2020. As far as we could possibly know, this is one of the few studies to compare time series models on COVID-19 active cases on all states in Nigeria. Expectantly, our addition to disease forecasting would be aidful to researchers, health practioners and the government.

Section 2 presents related work time-series prediction while section presents the methodology adopted for this study, the dataset used and the description of ARIMA, HWAAS, TBAT, Prophet,

GluonTS and N-BEATS models. Section 4 presents the result obtained from the methodology and discussion while conclusions are drawn in Section 5.

## 2. RELATED WORK

This section reports the prediction of coronavirus in the major countries, states, provinces and regions of the world as reported in the literature using statistical, machine and deep learning-stimulated time series models.

(Papastefanopoulos, Linardatos, & Kotsiantis, 2020) proposed and compared six (6) different time series model to evaluate the proportion of active coronavirus cases with reverence to the entire populace for the topmost-ten countries in the world. The models: ARIMA, NBEATS, HWAAS TBAT, GluonTS and Facebook's Prophet were compared for different countries like USA, Germany, UK, Iran, Italy, France, Spain, Turkey, Russia and Brazil and accessed their performances based on RMSE metric. The results obtained showed that ARIMA model performed best while TBAT performing second best out of six models.

(Zeroual, Harrou, Dairi, & Sun, 2020) proposed to compare five different deep learning models to forecast the count of new and recovered coronavirus cases based on daily confirmed and recovered cases in Australia, France, Spain, China, USA, and Italy. Specifically, simple Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Units (GRUs) and Variational AutoEncoder (VAE) models were applied evaluated based on RMSE, MAE, MAPE, Explained Variance (EV), and Root Mean Squared Log Error (RMSLE) metrics. VAE gave a superior performance to other models.

(Kumar, Chimmula, & Zhang, 2020) proposed a LSTM deep learning forecast model of coronavirus epidemic in Canada. They proposed to predict the trends and likely ending time of the recent coronavirus pandemic in Italy, Canada and USA. The proposed model achieved a value of 34.83 RMSE for short term predictions and 45.70 RMSE for long term predictions.

(Velásquez & Lara, 2020) proposed Reduced-Space Gaussian Process Regression model to forecast the daily accumulative count cases of coronavirus pandemic for all the states in USA for 82 days. This model achieved 98.91% correlation coefficient using correlation coefficient.

(Arora, Kumar, & Panigrahi, 2020) compared five (5) different Deep Learning-based models to predict the count of COVID-19 cases for all the states and regions in India. Convolutional LSTM, RNN, LSTM and Bi- directional LSTM were applied on Indian dataset to predict the count of coronavirus cases. LSTM model gave a superior performance with minimum error for short-term predicting day by day and week by week cases of 3% and 8% MAPE values respectively.

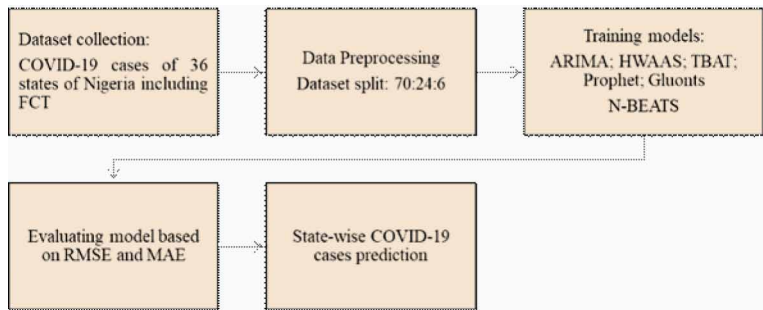
(Khan, Saeed, & Ali, 2020) proposed Vector Autoregressive (VAR) model with 5 lags to predict 10-days multi-step ahead of daily COVID-19 cases in the state and province of Pakistan. Firstly, the process of generating time series data is identified. A minimum and maximum forecasted were obtained with confidence interval of 95% for the next 10-days.

(Alzahrani, Aljamaan, & Al-Fakih, 2020) proposed the use of four (4) statistical time series models to predict the spread of coronavirus in Saudi Arabia. These models include ARIMA, Autoregressive (AR), Moving Average (MA) and ARMA. ARIMA model gave a superior performance over other models with values of 21.17, 14.93, 0.99, 2.16 and 0.024 for RMSE, MAE,  $R^2$ , MAPE and RMSRE measures respectively.

## 3. METHODOLOGY

This section describes the methodological process adopted for the study. The study models are supplied with state-based historic data of the daily count of COVID-19 cases in Nigeria. The description of the six different models were rendered and the metrics to assess the performances of each model. The layout of the study model is shown in Figure 1.

Figure 1. Layout of the study method

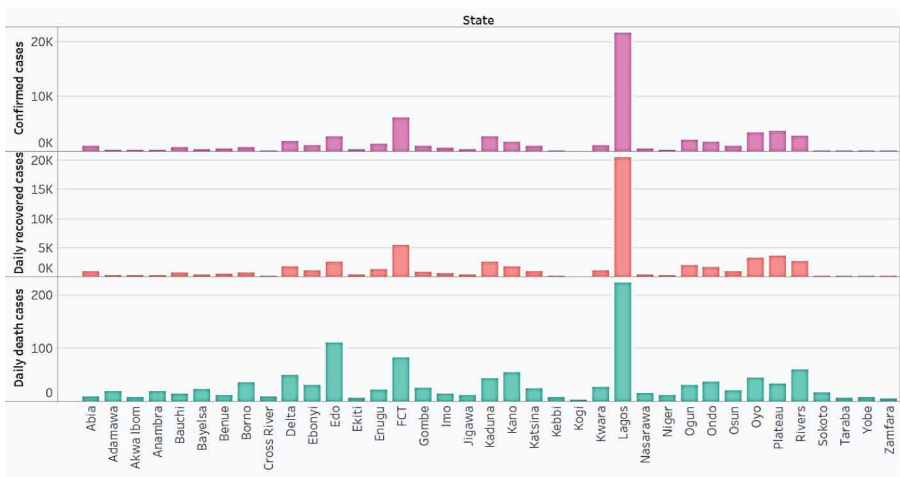


### 3.1 Dataset Description

The goal of the research is focused on forecasting active COVID-19 cases based on population for all states in Nigeria using daily cumulative count of confirmed, death and recovered cases. The second dataset consist of the health, social and economic COVID-19 data from (Ogundepo, et al., 2020). The variable constituents for the daily cases are presented in Table 1 while the health, social and economic COVID-19 data is presented in Table 2. The daily cases datasets considered for this study is obtained from NCDC website from the onset of the pandemic 2/23/2020 till 4/11/2020 as shown in Figure 2. The count of instances for each state is 256 totaling to 9472 instances for all the 37 states. The daily cases of COVID-19 from onset till November 4, 2020 for all states in Nigeria is shown Figure 2. It is observed that there is astronomical increase of count of 21483 for confirmed cases, 20511 for recovered cases and 218 for death cases in Lagos state in Nigeria only. As observed from Figure 3 that Lagos, Federal Capital Territory (FCT), Edo, Kaduna, Kano, Oyo, Plateau, Rivers, Delta and Kano were the ten (10) most affected states by the COVID-19, indicating a high epidemic propagation through a high count of confirmed, recovered and deaths as also shown by figure 3. Figure 2 also indicates that the rate of the three (3) cases were consistent for these ten states.

Table 3 gives the descriptive statistics of the top ten states with active cases in Nigeria. The distribution of the data for the considered states were all positively skewed with a flat Gaussian distribution. The spread of the COVID-19 dataset is evident from standard deviation (STD) and the

Figure 2. Daily COVID-19 cases for all states in Nigeria



**Table 1. variable of daily cases of COVID-19**

Variable name	Data type
State	String
Date	Date
Cumulative Confirmed	Numeric
Cumulative Recovered	Numeric
Cumulative Death	Numeric
Daily Confirmed	Numeric
Daily Recovered	Numeric
Daily Death	Numeric

**Table 2. Variable of health, social and economic COVID-19 data in Nigeria**

Variable Name	Data type
State	String
Population (2020)	Numeric
Fatality Rate	Numeric
Percentage Budget Reduction	Numeric
Health care	Numeric
COVID-19 Lab	Numeric
Density (P/Km <sup>2</sup> )	Numeric
Land Area (Km <sup>2</sup> )	Numeric

**Table 3. Summary of the top ten states with active cases**

State	Min	Max	Q-0.5	STD	Q-25	Q-75	Skew	Kurtosis
Delta	0	918	107	262.6	17	170.75	1.6	1.05
Edo	0	1,324	98.5	238	24.75	293.75	1.36	1.44
FCT	0	3,710	625.5	1,241	290.25	2,089.25	0.92	-0.67
Kaduna	0	506	97	108.3	57.00	222	0.69	-0.22
Kano	0	760	221	219.60	24	347	0.81	-0.34
Lagos	0	13,003	2,754.50	3,627	752.75	4,817.50	1.1	0.19
Ogun	0	627	142	112.4	78.00	235.5	0.69	0.54
Oyo	0	3,957	745	653.00	170	1,210.00	0.99	2.26
Plateau	0	2,034	348	430.00	71.75	819	0.86	0.47
Rivers	0	938	130	164.70	64.25	288	1.22	1.79

quartiles values (Q-25, Q-0.5 and Q-75). Hence, Table 3 exhibits a wide range of values for STD indicating a non-Gaussian distributed with positive support.

The count of active cases in each state of Nigeria on daily basis was determined by deducting both the recovered and deaths counts from the count of confirmed cases. Figure 3 shows the share of cumulative confirmed cases per state while Figure 4 displays the count of active coronavirus cases as at 03 November 2020 using the log scale. From the illustration displayed by Figure 3, the top ten states in Nigeria with the most count of cumulative confirmed cases as at 03 November 2020 accounted for more than 70% of all confirmed cases in Nigeria. Lagos state alone accounted for over

Figure 3. Share of daily accumulative confirmed coronavirus cases per state as at 03 November 2020

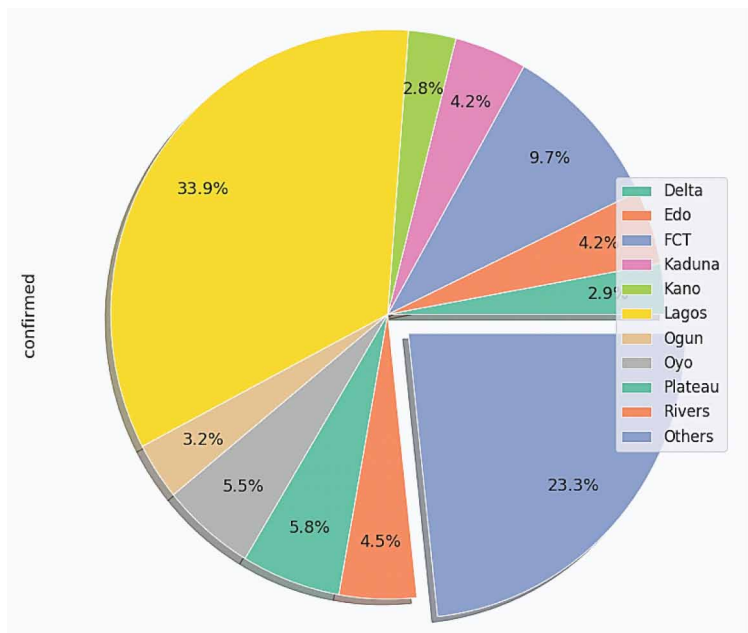
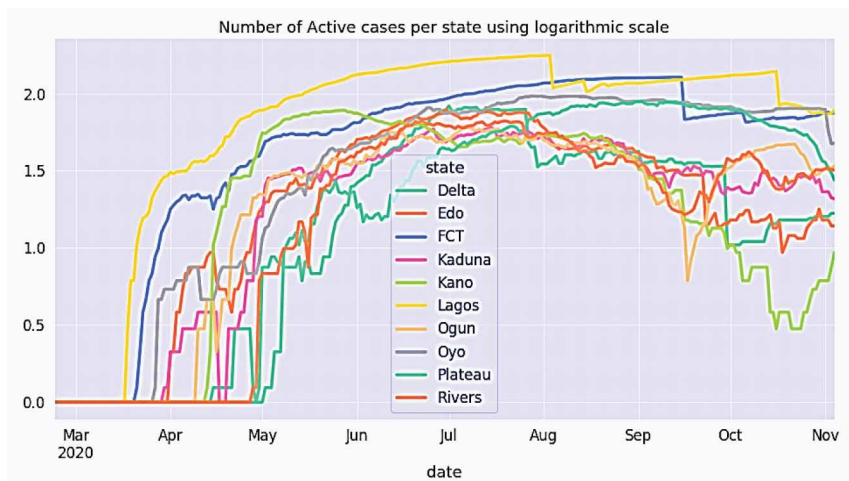


Figure 4. Count of daily active coronavirus cases per state with log scale



30%. These ten states are the most affected by the deadly coronavirus and maybe in later phases of the epidemic serve a point of reference and a model for newly-infected states. The count of instances for the top ten states is 2560.

Finally, the proportion of active coronavirus cases with reverence to the entire populace per state is displayed for every state in Figures 4 and 5. Figure 5 shows the spread of the epidemic based on the population for each state for the period under study using the linear scale, while Figure 6 displays the onset of active cases in each of the ten state till 4/11/2020.

Figure 5. Linear scale of proportion of active coronavirus cases with reverence to the entire populace per state

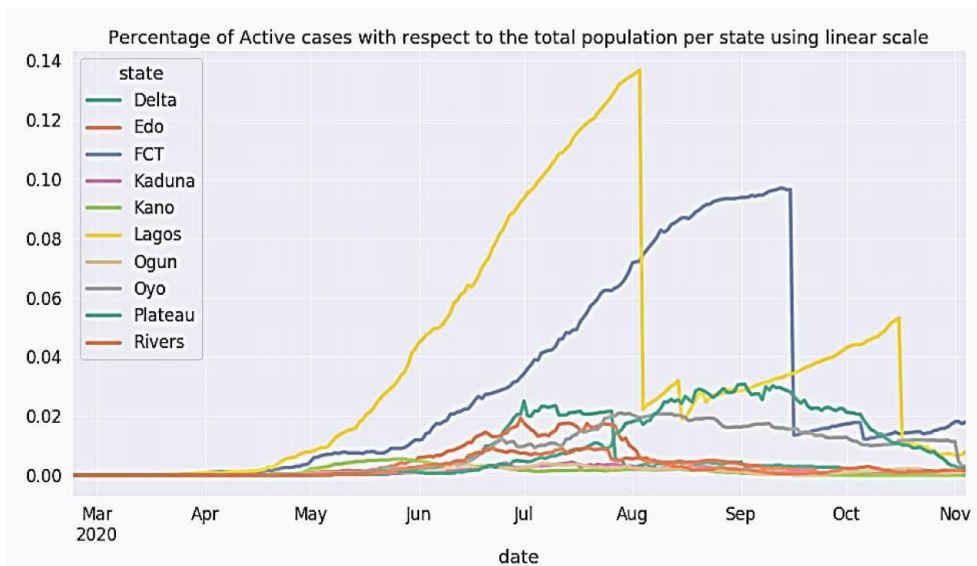
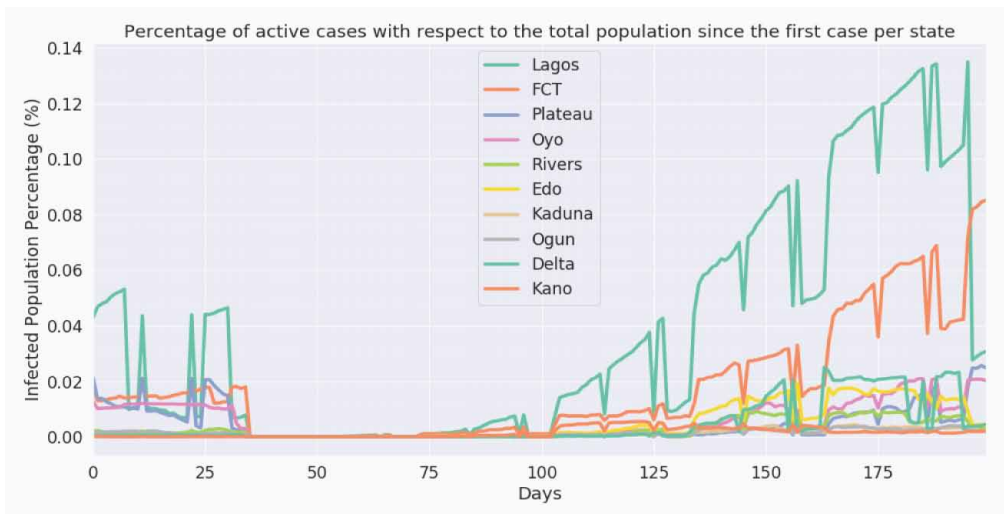


Figure 6. Proportion of active coronavirus cases with reverence to the entire populace per state from the onset of the pandemic



### 3.2 Time Series Models

Time series prediction focuses on investigating previous instances of a random variable towards building a model that finely catches the latent association, patterns and also forecast values of that variable. This strategy is invaluable when there is little to zero information accessible on the fundamental data-generating distribution/process or when there is no logical model that can sufficiently relate the random variable to other explanatory variables. There has also been a significant advancement in time series forecasting models due to great effort in research output in recent years. The following models: ARIMA, GluonTS, Prophet, HWAAS, N-BEATS and, TBAT will be discussed in this section.

#### 3.2.1 Auto-Regressive Integrated Moving Average (ARIMA)

ARIMA is the commonest and extensively used statistical time-series model (Box & Jenkins, 1990). Though, it was initially developed for applications in economic (Yule, 1926), its wide acceptance is based on its statistical properties, deployment of Box–Jenkins methods during training (Wold, 1939) coupled with its ability to apply different exponential smoothing models (Box & Jenkins, 1990). ARIMA is based on the principle that non-stationary time-series can be made stationary through differencing (Abdulmajeed, Adeleke, & Popoola, 2020). A time series  $X_t$  is said to follow an integrated autoregressive moving average model if the  $d^{\text{th}}$  difference  $W_t = \nabla^d X_t$  is a stationary ARIMA process. Regarding an ARIMA with components (p, 1, q) process where  $d = 1$ . For  $W_t = X_t - X_{t-1}$ , returns:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

or, based on the observed series:

$$\begin{aligned} X_t - X_{t-1} &= \phi_1 (X_{t-1} - X_{t-2}) + \phi_2 (X_{t-2} - X_{t-3}) + \dots + \phi_p (X_{t-p} - X_{t-p-1}) + \varepsilon_t \\ &\quad - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \end{aligned} \quad (2)$$

Equation (2) may be expressed as:

$$\begin{aligned} X_t &= (1 + \phi_1) X_{t-1} + (\phi_2 - \phi_1) X_{t-2} + (\phi_3 - \phi_2) X_{t-3} + \dots + (\phi_p - \phi_{p-1}) X_{t-p} \\ &\quad - \phi_p X_{t-p-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \end{aligned} \quad (3)$$

This is called the difference equation form of the model. Note this an ARMA (p + 1, q) process. This can be written in lag form as:

$$\left( 1 - \sum_{i=1}^p \phi_i B^i \right) X_t = \left( 1 + \sum_{i=1}^q \theta_i B^i \right) \varepsilon_t \quad (4)$$

Assuming the polynomial  $\left( 1 - \sum_{i=1}^p \phi_i B^i \right)$  as a unitary root of multiplicity  $d$ , it can be re-written as:



$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) = \left(1 + \sum_{i=1}^{p-d} \omega_i B^i\right) (1 - B)^d \quad (5)$$

A ARIMA (p, d, q) process expresses equation (4) as polynomial factorization property as shown by equation (5):

$$\left(1 - \sum_{i=1}^p \omega_i B^i\right) (1 - B)^d X_t = \left(1 + \sum_{i=1}^q \theta_i B^i\right) \varepsilon_t \quad (6)$$

### 3.2.2 Trigonometric Seasonal Formulation, Box–Cox Transformation, ARMA Errors and Trend Component (TBAT)

TBAT model consists of four (4) basic components: Trigonometric seasonal formulation (Harvey, Koopman, & Riani, 1997), Box–Cox transformation (Box & Cox, 1964), ARMA errors (Adhikari & Agrawal) and trend factor. The model depends on trigonometric functions to model non-integer seasonal frequencies and Box–Cox transformations to transform non-normal dependent variables into a normal shape and allow certain types of non-linearity (Papastefanopoulos, Linardatos, & Kotsiantis, 2020). This model is widely applied to different time-series problems.

Furthermore, TBAT's proposed geometric expression which facilitates the breakdown of compound seasonal time series making it computationally efficient for maximum likelihood estimators. This decomposition enables the recognition and mining of seasonal components which may not be noticeable in the timeseries chart (De Livera, Hyndman, & Snyder, 2011; Papastefanopoulos, Linardatos, & Kotsiantis, 2020). Its structure comprises a sizeable specification set with the likelihood of superior predictions (Hyndman, Koehler, Ord, & Snyder, 2008; Papastefanopoulos, Linardatos, & Kotsiantis, 2020), management of the non-linear attributes that are found in everyday time-series data, auto-correlation considerations and entire estimation procedure (De Livera, Hyndman, & Snyder, 2011; Papastefanopoulos, Linardatos, & Kotsiantis, 2020).

In Box-Cox transformation, the transformation of the variable  $x$  is indexed by  $\lambda$ , and is defined as:

$$x'_\lambda = \frac{x^\lambda - 1}{\lambda} \quad (7)$$

Equation above is a scaled version of the Tukey transformation  $x^\lambda$ .

When  $\lambda = 0$ , the Box-Cox variable is the indeterminate form re-written as:

$$x'_\lambda = \frac{e^{\lambda \log(x)} - 1}{\lambda} \approx \frac{\left(1 + \lambda \log(x) + \frac{1}{2} \lambda^2 \log(x)^2 + \dots\right) - 1}{\lambda} \rightarrow \log(x) \quad (8)$$

as  $\lambda \rightarrow 0$ .

### 3.2.3 Probabilistic Forecasting With Auto-Regressive Recurrent Networks (DeepAR)

DeepAR forecast model is built on autoregressive RNN for probabilistic forecasting (Salinas, Flunkert, Gasthaus, & Januschowski, 2020) but implemented as GluonTS (Alexandrov, et al., 2019) in this study. Here, the forecasting problem is approached by including appropriate likelihoods and blending them

with deep neural network (Salinas, Flunkert, Gasthaus, & Januschowski, 2020; Papastefanopoulos, Linardatos, & Kotsiantis, 2020). Specifically, a LSTM built on RNN architecture as expressed in (Hochreiter & Schmidhuber, 1997), while building on past researches on deep learning model for time-series data (Graves, 2013; Zaremba, Sutskever, & Vinyals) towards addressing probabilistic forecasting problem. Deeper networks, as expressed in (Alexandrov, et al., 2019) is preferred to shallow and broad as it permits extra abstract data representations via complex transformations. One of the benefits of GluonTS are its ability to learn seasonal behaviors and composite dependencies. Another benefit is its support for wide-ranging probabilistic functions by not assuming Gaussian noise. Lastly, probabilistic predictions are produced in the form of Monte Carlo samples (Salinas, Flunkert, Gasthaus, & Januschowski, 2020; Papastefanopoulos, Linardatos, & Kotsiantis, 2020).

Given time series  $i$  at time  $t$  by  $z_{i,t}$ , with the aim to model the conditional distribution:

$$P\left(z_{i,t_0:T} \middle| z_{i,t_0-1}, X_i, 1 : T\right) \quad (9)$$

of the future of each time series  $\left[z_{i,t_0}, z_{i,t_0+1}, \dots, z_{i,T}\right] = z_{i,t_0:T}$  given its past  $\left[z_{i,1}, \dots, z_{i,t_0-2}, z_{i,t_0-1}\right] = z_{i,1:t_0-1}$  where  $t_0$  represents the time point from which we assume  $z_{i,t}$  to be unknown at prediction time, and  $X_{i,1:T}$  are covariates which are assumed to be known for all time points.

Based on autoregressive recurrent network architecture, the mode distribution:

$$Q_{\cdot} \left( z_{i,t_0:T} \middle| z_{i,t_0-1}, X_i, 1 : T \right)$$

consists of a product of likelihood factors:

$$Q_{\Theta} \left( z_{i,t_0:T} \middle| z_{i,t_0-1}, X_i, 1 : T \right) = \prod_{t=t_0}^T Q_{\Theta} \left( z_{i,t} \middle| z_{i,t-1}, X_i, 1 : T \right) = \prod_{t=t_0}^T \ell \left( z_{i,t} \middle| \theta \left( h_{i,t}, \Theta \right) \right) \quad (10)$$

Parameterize by the output  $h_{i,t}$  of an autoregressive recurrent network:

$$h_{i,t} = h \left( h_{i,t-1}, z_{i,t-1}, X_i, 1:T, \Theta \right) \quad (11)$$

where  $h$  is a function employed by a multi-layer RNN.

### 3.2.4 Holt–Winters Additive Model: Exponential Smoothing With Additive Trend and Seasonality (HWAAS)

HWAAS model extends Holt's exponential smoothing model by adding the seasonality and trend elements to forecasting, enabling data modelling with a seasonal or trend components. HWAAS is straight-forward and effective with an ability to handle trend and seasonal variation in time series data, serving as a substitute to Box–Jenkins methods. However, research had shown that HWAAS model is not more accurate than Box–Jenkins method (Chatfield, 1978; Papastefanopoulos, Linardatos, & Kotsiantis, 2020).

Holts exponential smoothing (ES) uses trend to forecast data. This technique includes a forecast and two smoothing equations with one for the level and trend respectively. The forecast equation is given as equation (12) – (14):

$$\hat{y}_{t+h|t} = \varepsilon_t + hb_t \quad (12)$$

The level equation is given as:

$$\varepsilon_t = \alpha y_t + (1 - \alpha)(\varepsilon_{t-1} + b_{t-1}) \quad (13)$$

and trend equation is given as:

$$b_t = \beta^* (\varepsilon_t - \varepsilon_{t-1}) + (1 - \beta^*) b_{t-1} \quad (14)$$

where:

- $\varepsilon_t$  denotes an estimate of the level of the series at time  $t$ ;
- $b_t$  denotes an estimate of the of the trend (slope) of the series at time  $t$ ;
- $\alpha$  is the smoothing parameter for the level,  $0 \leq \alpha \leq 1$ ;
- $\beta^*$  is the smoothing parameter for the trend,  $0 \leq \beta^* \leq 1$ ;
- $h$  is the step ahead forecast.

The level equation reveals that  $\varepsilon_t$  is a weighted average of observation  $y_t$  and the one-step-ahead training forecast for time  $t$  is given as  $\varepsilon_{t-1} + b_{t-1}$ . The trend equation illustrates that  $b_t$  is a weighted average of the estimated trend at time  $t$  based on  $\varepsilon_t - \varepsilon_{t-1}$  and  $b_{t-1}$ , the previous estimate of the trend.

### 3.2.5 Prophet: Automatic Forecasting Procedure

Prophet is an additive regression model for time-series forecasting problem developed by Facebook (Taylor & Letham, 2018). It is effective in treating time series data with outliers, shifts in trend making, strong seasonal effects and missing data. It has a Stan backend which introduced a fast solution to L-BFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shannon algorithm) for the forecasting problem (Abdulmajeed, Adeleke, & Popoola, 2020). *Prophet* includes a decomposable time series model defined by (15):

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \quad (15)$$

Here  $g(t)$  stands for the trend function,  $s(t)$  denotes periodic/seasonal changes (weekly and yearly seasonality), and  $h(t)$  is the irregular effects of holidays while error term is presented by  $\varepsilon(t)$  (Taylor & Letham, 2018; Abdulmajeed, Adeleke, & Popoola, 2020). The trend model in this work utilizes a piecewise saturated growth model with time-varying carrying capacity as shown in (16):

$$g(t) = \frac{C(t)}{1 + \exp\left[-\left(k + \alpha(t)^T \delta\right)\left(t - m + \alpha(t)^T \gamma\right)\right]} \quad (16)$$

$C(t)$  is the time-varying carrying capacity,  $k$  is the growth rate,  $m$  is an offset. The growth rate is not constant but piecewise with  $\alpha(t)$ ,  $\gamma$  and  $\delta$  defining its structure (Abdulmajeed, Adeleke, & Popoola, 2020).

### 3.2.6 Neural Basis Expansion Analysis for Interpretable Time Series Forecasting (NBEATS)

N-BEATS model deploys a deep neural architecture comprising a forward and backward residual links with a deep heap of completely-joined layers (Papastefanopoulos, Linardatos, & Kotsiantis, 2020; Oreshkin, Carpov, Chapados, & Bengio, 2019). NBEATS function similarly to a conventional decomposition approach (Cleveland, Cleveland, McRae, & Terpenning, 1990). The architecture is made up of the trend and seasonality stacks with each consisting many blocks connected through their residuals. With a mixture of forecast/backcast rule, this residual heap brings about trend element being separated from the input window prior to being passed into seasonality pile. The incomplete estimates of trend and seasonality are isolated opening up was independent yield, carrying a layer of interpretability to the model (Oreshkin, Carpov, Chapados, & Bengio, 2019). The model is accurate, easy to interpret, quick to train and easy to deploy to many target domains (Oreshkin, Carpov, Chapados, & Bengio, 2019; Papastefanopoulos, Linardatos, & Kotsiantis, 2020).

### 3.3 Evaluation Metric

The study prediction models were assessed based on RMSE and MAE as presented by (17) and (18) respectively. A low value of RMSE and MAE for a model signify more accurate forecasting performances:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (17)$$

$$MAE = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n} \quad (18)$$

where  $y_t$  is the real value,  $\hat{y}_t$  are the matching projected value and  $n$  is the count of instances.

## 4. EXPERIMENTS AND RESULTS

This section describes the modelling, evaluation and compare the results of time-series forecasting models namely ARIMA, HWAAS, TBAT, Prophet, GluonTS and N-BEATS time series models. This study focused on the daily cumulative active cases from the top ten states namely for Rivers, Plateau, Oyo, Ogun, Lagos, Kano, Kaduna, FCT, Edo and Delta in Nigeria. Firstly, each of the six models is trained and validated with training measurements of 70:24 ratio. Secondly, each variable is predicted with the trained models for the previously unseen test dataset of 14 multi-step ahead. The train data contain univariate time series data of active cases from 2/23/2020 to 11/4/2020. This study is aimed at examining the forecast performances of these six time series models despite small data size.

## 4.1 Modelling

The study is aimed building a comparative study of six time-series models to forecasting the proportion of active coronavirus cases with reverence to the entire populace of top ten states in Nigeria. Hence, the ten states with the highest count of active cases were picked as the basis of experiment and comparisons having accounted over 70% of the COVID-19 pandemic cases nationally. For each of the Nigerian states, 256 samples were created, each representing the proportion of daily active coronavirus cases with reverence to entire populace per state, confirmed, recovered and death cases. The 256 instances for each of the state is split into train-validation-train to approximately ratio 70:24:6 (181:61: 14) percentage. Consequently, 242 out of 256 instances were applied, while a window of fourteen days (14 instances) was used to assess the performance of the predictive models with RMSE and MAE metrics.

## 4.2 Predicting New Active Cases

All the six time-series models in this study were trained, tested and assessed with proportions of active coronavirus cases with respects to the entire populace per the ten states. Notably, 256 samples were created for each state where each sample is representing the proportion of daily active coronavirus cases with reverence to the entire populace in the corresponding state. It is worthy to note that the top ten states were used as a reference point for active cases. The count of daily cumulative active cases in each of the top ten states was estimated by deducting from both the recovered and deaths from the confirmed cases as shown in (19), while the proportion of the entire population is presented by (20):

$$\text{Active case} = \text{confirmed} - \text{recovered} - \text{death} \quad (19)$$

$$\text{Percentage of total population} = \frac{\text{Active case}}{\text{population}} \quad (20)$$

Tables 4 and 5 shows the prediction performances of ARIMA, HWAAS, TBAT, Prophet, GluonTS and N-Beats models in terms of RMSE and MAE respectively for active COVID-19 cases from top ten states in Nigeria. The results established that no model is all encompassing for forecasting the proportion of active coronavirus cases with reverence to the entire populace in each state. The results obtained also shows the superiority performance of ARIMA model in four of the states compared to the other models. The ARIMA model captured nearly all variances in active cases and gives precise predictions than other time series models.

This study attempted to compare and give highlights the potentials between statistical methods like ARIMA, TBAT and HWAAS and DL models like N-BEATS and GluonTS in predicting the count of active COVID-19 cases recorded from top ten different states in Nigeria. Overall, as demonstrated both by RMSE and MAE measured performances as shown in Tables 4 and 5, statistical methods like ARIMA, HWAAS and Prophet gave a superior performance than the deep learning methods respectively. One reason for the poor performance could be the small sample size of the dataset as DL methods thrive more with huge data sample size to capture the dynamics of the COVID-19 data. Surprisingly, Prophet also performed well despite its origin of solving problems related to business.

Six time-series models were applied to the active cases to predict the trend of newly active cases. Figure 7 (a)-(j) displays the prediction outcomes of the active coronavirus cases in the top ten states using ARIMA, HWAAS, TBAT, Prophet, GluonTS and N-BEATS models from 22<sup>nd</sup> October to 4<sup>th</sup> November. The prediction results of these six models following the general trend of the logged active COVID-19 cases, showing the superior performance of ARIMA model based on RMSE in four of the states.

**Table 4. Models performance based on RMSE metric for top-ten states with the highest active coronavirus cases in Nigeria**

Models	Lagos	FCT	Plateau	Oyo	Rivers	Edo	Kaduna	Ogun	Delta	Kano
ARIMA	<b>0.002537</b>	<b>0.001968</b>	0.000963	0.004135	0.000381	<b>9.12E-05</b>	0.000163	0.001023	<b>5.36E-05</b>	8.75E-05
Prophet	0.014757	0.021863	0.002944	<b>0.003185</b>	0.001564	0.003687	0.000805	<b>0.000533</b>	0.001167	0.000606
HWAAS	0.002548	0.002281	0.002668	0.003994	<b>0.000322</b>	0.000731	<b>0.000161</b>	0.00106	0.001809	0.000169
NBEATS	0.038309	0.001972	0.00624	0.004199	0.00142	0.000302	0.000214	0.000931	0.000204	<b>3.02E-05</b>
GluonTS	0.024467	0.005443	0.004174	0.004484	0.002911	0.000608	0.000248	0.001531	0.000751	0.000127
TBAT	0.004998	0.007709	<b>0.00083</b>	0.003967	0.000397	0.000122	0.000229	0.001061	0.00129	4.15E-05

**Table 5. Models performance based on MAE for ten states with the highest active COVID-19 cases in Nigeria**

Model	Lagos	FCT	Plateau	Oyo	Rivers	Edo	Kaduna	Ogun	Delta	Kano
ARIMA	<b>6.44E-06</b>	<b>3.87E-06</b>	9.27E-07	1.71E-05	1.45E-07	<b>8.32E-09</b>	2.65E-08	1.05E-06	<b>2.87E-09</b>	7.66E-09
Prophet	0.000218	0.000478	8.67E-06	<b>1.01E-05</b>	2.45E-06	1.36E-05	6.47E-07	<b>2.85E-07</b>	1.36E-06	3.67E-07
HWAAS	6.49E-06	5.2E-06	7.12E-06	1.6E-05	<b>1.04E-07</b>	5.35E-07	<b>2.6E-08</b>	1.12E-06	3.27E-06	2.85E-08
NBEATS	0.001468	3.89E-06	3.89E-05	1.76E-05	2.02E-06	9.14E-08	4.57E-08	8.67E-07	4.17E-08	<b>9.09E-10</b>
GluonTS	0.000599	2.96E-05	1.74E-05	2.01E-05	8.47E-06	3.7E-07	6.13E-08	2.34E-06	5.64E-07	1.62E-08
TBAT	2.5E-05	5.94E-05	<b>6.89E-07</b>	1.57E-05	1.57E-07	1.5E-08	5.25E-08	1.13E-06	1.66E-06	1.73E-09

Precisely, ARIMA model gave the best performance based on low RMSE values in Lagos, Edo FCT and Delta states as shown in Figures 7 (e), (i), (h) and (j), while Prophet achieved the best results in Oyo and Ogun as shown by Figure 7 (c) and (d). HWAAS achieved the best results for Rivers and Kaduna as displayed by Figures 7 (a) and (g), N-BEATS performed best for Kano as displayed by Figures 7 (f) while TBAT performed best for Plateau state (Figure 7 (b)). GluonTS is least performing model and did not performed best in any state.

For statistical time-series models, ARIMA gave the best performance followed by HWAAS. ARIMA achieved the topmost-four (4) performance from ten states with active cases while HWAAS model achieving the topmost result in Kaduna state.

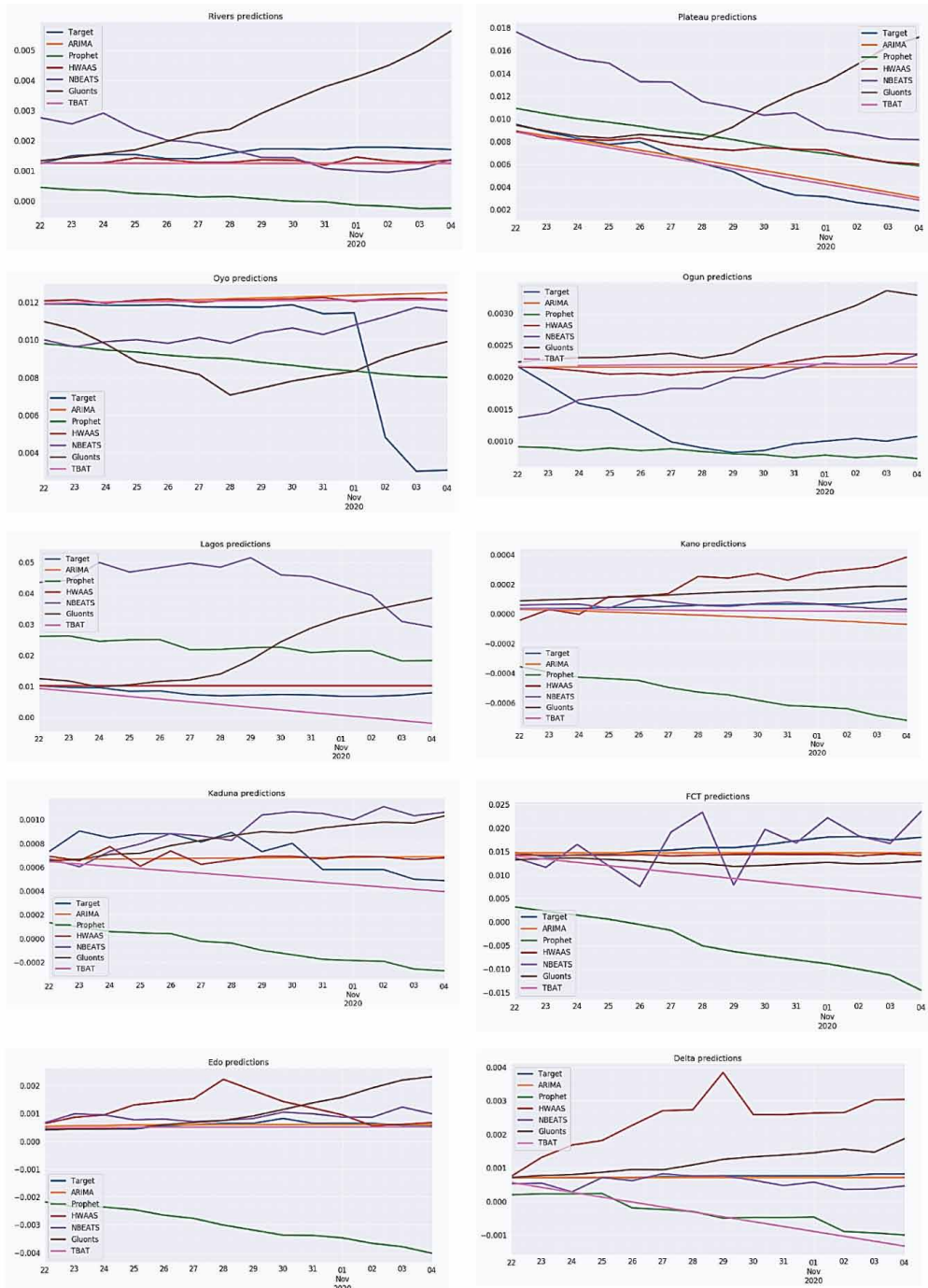
Finally, for deep learning inspired models, though N-BEATS and GluonTS obtained poor performance overall, N-BEATS gave a superior performance in Plateau state.

### 4.3 Week Number Analysis of the States

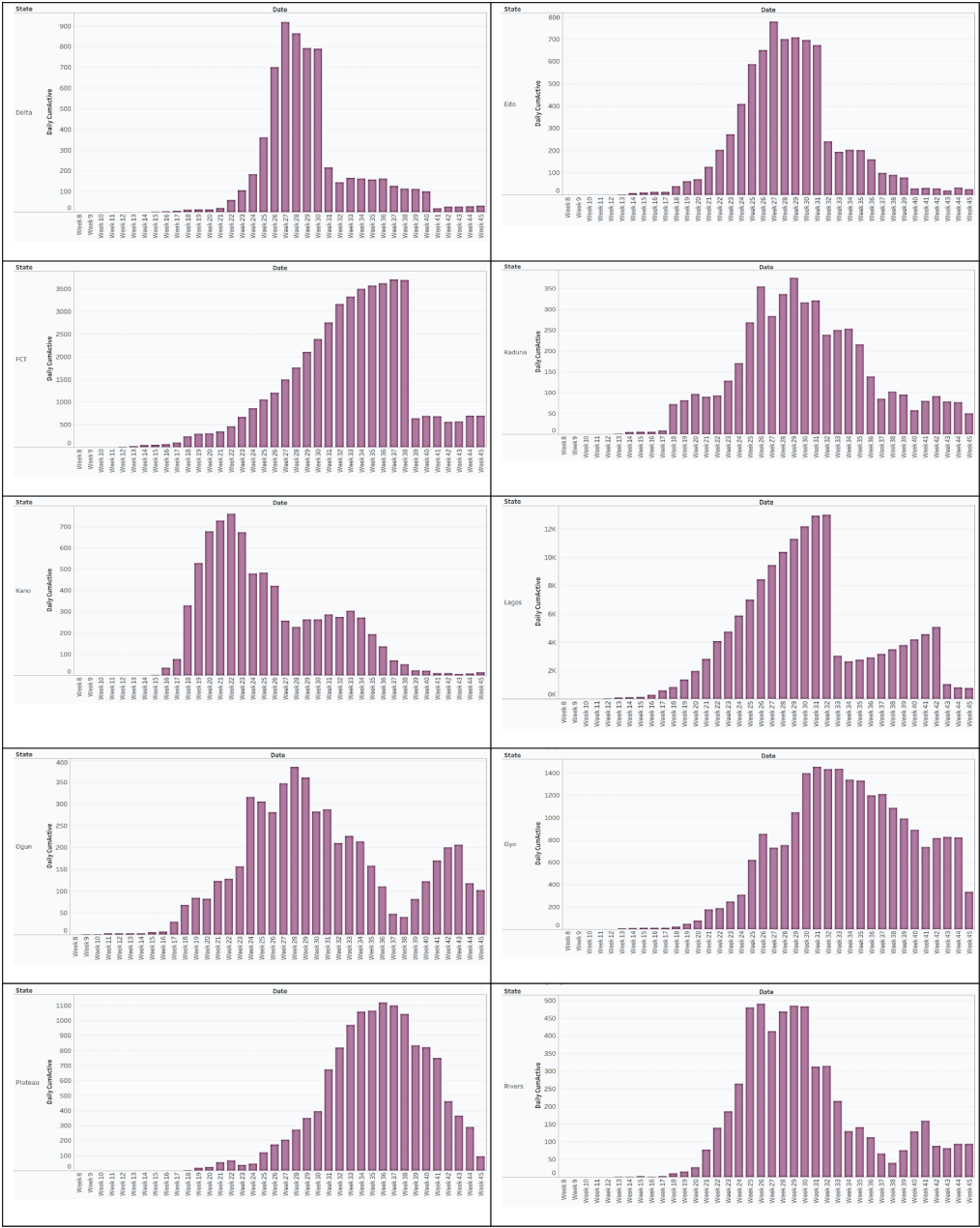
This section presents the trend in the active cases in the top ten states in Nigeria as presented in Figure 8. For Lagos state which has over 30% percent active COVID-19 cases, a heap of 22 active cases is noticed in week 12. Since then, a steady steep increase in the active cases 182, 4879 is noticed in week 16 and 24 respectively. The count of active cases (13,003) reached its peak for Lagos state at week 32. An immediate decrease surge in active cases of 2,923 and 2,634 is noticed in week 33 and 34 respectively. Then, a new upward trend started by week 35 and got peaked to 5,061 new active cases by week 42. A surge decrease is noticed to 1,019 active cases by week 43 till 45.

This pattern is noticed for other states as displayed by Figure 8. The upward and downward trend might occur at different weeks for different states. FCT got peaked at week 37 with 3,709 while Oyo reached its peak at week 31 with 1, 447 new active cases. Plateau reached its peak at week 36 with 1,119 while Delta reach its peak by week 27 with 918 new active cases. Edo state attained its peak by week 27 with 778 while Rivers got peaked at week 26 with 490 new active cases. Kaduna reached

**Figure 7. Predicted active COVID19 cases from 22<sup>nd</sup> October to 4<sup>th</sup> November 2020 using ARIMA, HWAAS, TBAT, Prophet, GluonTS and N-Beats using RMSE for states: (a) Rivers, (b) Plateau, (c) Oyo, (d) Ogun, (e) Lagos, (f) Kano, (g) Kaduna, (h) FCT, (i) Edo, (j) Delta**



**Figure 8.** Week number for active COVID19 cases from 2<sup>nd</sup> February to 4<sup>th</sup> November 2020 for the top ten states (a) Delta, (b) Edo, (c) FCT, (d) Kaduna, (e) Kano, (f) Lagos, (g) Ogun, (h) Oyo, (i) Plateau, (j) Rivers



its peak by week 29 with 375, Kano reached its peak by week 22 with 760 while Ogun reached its peak by week 28 with 385 new active cases.

#### 4.4 Discussion

The methodology presented in this study is proficient for producing significant and well-grounded multi-step ahead (14-days ahead) predictions of COVID-19 activity, at the state level in Nigeria. This



**Table 6. Comparison of model, data and results of related work**

Author and Year	Type of model	Model	Location	Metric	Best model
(Papastefanopoulos, Linardatos, & Kotsiantis, 2020)	Statistical + ML	ARIMA, HWAAS, TBAT, Prophet, GluonTS and N-Beats	Brazil, USA, France, UK, Italy, Germany, Spain, Russia and Turkey	RMSE	ARIMA and TBAT
(Zeroual, Harrou, Dairi, & Sun, 2020)	Deep Learning	VAE, RNN, GRUs, LSTM and Bi-LSTM,	Australia, Italy, China, Spain, France and USA.	RMSE, MAPE, MAE, EV and RMSLE	VAE
(Folorunso, Taiwo, & Olatayo, 2019)	ML	SVR, MLP, RFR, <i>k</i> -NNR and LR	Nigeria	RMSE, MAE and MAPE	SVR
This research	ML + Statistical	ARIMA, HWAAS, TBAT, Prophet, GluonTS and N-Beats	Top ten active cases in Nigeria: Lagos, Kano, Ogun, FCT, Rivers, Oyo, Edo, Delta, Kaduna, Plateau	RMSE and MAE	ARIMA

method is capable of getting over many challenges typical of evolving epidemics caused by novel virus. It is hard to explain precisely why some models performed better than the others in one state than the other. Some of contributing factors could be the climate and geographic features of each state. Another reason could be the population size, the land mass area and the population densities of different countries. Some of the states have different trade and market activities. Furthermore, diversity in quarantine, social distancing measures, duration, severity and strict adherence to these measures. More so, the method of data collection among different states regarding the discrepancies in testing and procedures. This could be based on the count of COVID-19 laboratories and medical care facilities available in the different states. Table 6 presents similar work in the literature where similar models and data but in different countries were compared. This shows that this study is relevant and essential to the pandemic in Nigeria.

## 5. CONCLUSION

The COVID-19 epidemic is exponentially spreading over the world, and the healthcare systems in some high impacted states in Nigeria like Lagos, FCT, Delta, Oyo etc are already overcrowded. Precise prediction of the count of active cases provides pertinent information to governments and decision-makers about the expected situation and the needed measures to impose. Also, prediction of information can be useful for motivating the wider public to consider the imposed measures for down slowing the spread of this virus. In this study, statistical and ML-stimulated time-series models like ARIMA, HWAAS, TBAT, Prophet, GluonTS and N-BEATS have been applied to the real-time predictions of the daily cumulative confirmed, recovered and death COVID-19 cases in all states in Nigeria especially active cases in top ten states in the country. The choice of models is highly motivated by their extended capacity in capturing, processing non-linearity and their flexibility in modelling time-dependent data. Fourteen days-ahead predictions are provided based on historical data of 256 days since February 23, 2020, for all states in Nigeria especially top ten states namely Rivers, Plateau, Oyo, Ogun, Lagos, Kano, Kaduna, FCT, Edo and Delta. The performances of each model are evaluated based on RMSE and MAE. The results obtained established that ARIMA achieved a superior prediction performance than other study models. These results will guide the public health care givers and government of Nigeria to plan for the catastrophe against coronavirus.

## REFERENCES

- Abdulmajeed, K., Adeleke, M., & Popoola, L. (2020). Online forecasting of COVID-19 cases in Nigeria using limited data. *Data in Brief*, 20(105683), 105683. Advance online publication. doi:10.1016/j.dib.2020.105683 PMID:32391409
- Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *ArXiv 2013*.
- Adhikari, R., & Agrawal, R. K. (n.d.). *An introductory study on time series modeling and forecasting*. arXiv 2013:arXiv:1302.6613.
- Alexandrov, A., Benidis, K., Bohlke-Schneider, M., Flunkert, V., Gasthaus, J., Januschowski, T., . . . Wang, Y. (2019). *Gluonts: Probabilistic time series models in python*. arXiv:1906.05264.
- Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A. (2020). Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *Journal of Infection and Public Health*, 13(7), 914–919. doi:10.1016/j.jiph.2020.06.001 PMID:32546438
- Arora, P., Kumar, H., & Panigrahi, K. B. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons, and Fractals*, 139(110017), 1–9. doi:10.1016/j.chaos.2020.110017 PMID:32572310
- Box, G., & Jenkins, G. (1990). *Time Series Analysis Forecasting and Control*. Holden-Day, Inc.
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *J. R. Stat. Soc. Ser.*, 26, 211–243.
- Cao, L. T., Liu, H. H., Li, J., Yin, X. D., Duan, Y., & Wang, J. (2020). Relationship of meteorological factors and human brucellosis in Hebei province, China. *The Science of the Total Environment*, 703(135491), 135491. doi:10.1016/j.scitotenv.2019.135491 PMID:31740063
- Chatfield, C. (1978). The Holt-Winters Forecasting Procedure. *Applied Statistics*, 27(3), 264–279. doi:10.2307/2347162
- Chatfield, C., & Yar, M. (1988). Holt-Winters forecasting: Some practical issues. *J. R. Stat. Soc. Ser.* 1988, 37, 129–140. *J. R. Stat. Soc. Ser.*, 37, 129–140.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *Journal of Official Statistics*, 6, 3–73.
- De Livera, A. M., Hyndman, R., & Snyder, R. D. (2011). Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing. *Journal of the American Statistical Association*, 106(496), 1513–1527. doi:10.1198/jasa.2011.tm09771
- Earnest, A., Chen, M. I., Ng, D., & Sin, L. Y. (2005). Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. *BMC Health Services Research*, 5(1), 36. doi:10.1186/1472-6963-5-36 PMID:15885149
- Fang, X., Liu, W., Ai, J., He, M., Wu, Y., Shi, Y., Shen, W., & Bao, C. (2020). Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC Infectious Diseases*, 20(1), 1–8. doi:10.1186/s12879-020-4930-2 PMID:32171261
- Folorunso, S. O., Taiwo, A. I., & Olatayo, T. O. (2019). Empirical Comparison of Time Series Data Mining Algorithms for Electrical Power prediction. *Ind. Journal of Computing*, 4(2), 109–118. doi:10.21108/indojc.2019.4.2.329
- Gaudart, J., Touré, O., Dessay, N., Dicko, A. L., Ranque, S., Forest, L., Demongeot, J., & Doumbo, O. K. (2009). Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area. *Mali. Malaria Journal*, 8(1), 61. doi:10.1186/1475-2875-8-61 PMID:19361335
- Gelper, S., Fried, R., & Croux, C. (2010). Robust forecasting with exponential and Holt–Winters smoothing. *Journal of Forecasting*, 29, 285–300.
- Ghosal, S., Sengupta, S., Majumder, M., & Sinha, B. (2020). Prediction of the number of deaths in India due to SARS-CoV-2 at 5–6 weeks. *Diabetes & Metabolic Syndrome*. Advance online publication. doi:10.1016/j.dsx.2020.03.017

- Graves, A. (2013). *Generating sequences with recurrent neural networks*. arXiv 2013, arXiv:1308.0850.
- Guan, P., Huang, D. S., & Zhou, B. S. (2004). Forecasting model for the incidence of hepatitis A based on artificial neural network. *World journal of gastroenterology*. *WJG*, 10(24), 3579. doi:10.3748/wjg.v10.i24.3579 PMID:15534910
- Harvey, A., Koopman, S. J., & Riani, M. (1997). The modeling and seasonal adjustment of weekly observations. *Journal of Business & Economic Statistics*, 15, 354–368.
- Harvey, A. C., & Peters, S. (1990). Estimation procedures for structural time series models. *Journal of Forecasting*, 9(2), 89–108. doi:10.1002/for.3980090203
- Harvey, A. C., & Shephard, N. (1993). Estimation and Testing of Stochastic Variance Models. Technical report; Suntory and Toyota International Centres for Economics and Related: London, UK.
- Hochreiter, S., & Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. In *In Advances in Neural Information Processing Systems* (pp. 473–479). MIT Press.
- Hutchinson, G. E. (1978). *An Introduction to Population Ecology*. John Wiley & Sons.
- Hyndman, G., & Athanasopoulos, R. J. (2018). *Forecasting: principles and practice*. Academic Press.
- Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer Science & Business Media. doi:10.1007/978-3-540-71918-2
- Kalekar, P. S. (2004). Time series forecasting using holt-winters exponential smoothing. *Kanwal Rekhi Sch. Inf. Technol*, 4329008, 1–13.
- Kane, M., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics*, 15(276), 276. Advance online publication. doi:10.1186/1471-2105-15-276 PMID:25123979
- Khan, F., Saeed, A., & Ali, S. (2020). Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan. *Chaos, Solitons, and Fractals*, 140(110189). Advance online publication. doi:10.1016/j.chaos.2020.110189 PMID:32834659
- Kumar, V., Chimmula, R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons, and Fractals*, 135, 109864. doi:10.1016/j.chaos.2020.109864 PMID:32390691
- Liu, Q., Liu, X., Jiang, B., & Yang, W. (2011). Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. *BMC Infectious Diseases*, 11(1), 218. doi:10.1186/1471-2334-11-218 PMID:21838933
- Lukman, A. F., Rauf, R. I., Abiodun, O., Oludoun, O., Ayinde, K., & Ogundokun, R. O. (2020). COVID-19 prevalence estimation: Four most affected African countries. *Infectious Disease Modelling*, 5, 827–838. doi:10.1016/j.idm.2020.10.002 PMID:33073068
- Nsoesie, E. O., Beckman, R. J., Shashaani, S., Nagaraj, K. S., & Marathe, M. V. (2013). A simulation optimization approach to epidemic forecasting. *PLoS One*, 8(6), e67164. doi:10.1371/journal.pone.0067164 PMID:23826222
- Ogundepo, E., Folorunso, S., Adekanmbi, O., Akinsande, O., Banjo, O., Ogbuju, E., Oladipo, F., Abimbola, O., Oseghale, E., & Babajide, O. (2020). An Exploratory Assessment of a Multidimensional Healthcare and Economic Data on COVID-19 in Nigeria. *Data in Brief*, 33, 106424. doi:10.1016/j.dib.2020.106424 PMID:33102643
- Ogundokun, R. O., Lukman, A. F., Kibria, G. B., Awotunde, J. B., & Aladeitan, B. B. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infectious Disease Modelling*, 5, 543–548. doi:10.1016/j.idm.2020.08.003 PMID:32835145
- Oreshkin, B. N., Carpov, D., Chapados, N., & Bengio, Y. (2019). *N-BEATS: Neural basis expansion analysis for interpretable time series forecasting*. arXiv 2019, arXiv:1905.10437.
- Pandey, G., Chaudhary, P., Gupta, R., & Pal, S. (2020). *SEIR and Regression Model based COVID-19 outbreak predictions in India*. arXiv:2004.00958.
- Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S. (2020). COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population. *Applied Sciences*, 10(3880), 1-15. 10.3390/app10113880

- Polwiang, S. (2020). The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003-2017). *BMC Infectious Diseases*, 20(1), 1–10. doi:10.1186/s12879-020-4902-6 PMID:32164548
- Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 181–1191. doi:10.1016/j.ijforecast.2019.07.001
- Sato, R. C. (n.d.). Disease management with ARIMA model in time series. *Einstein (Sao Paulo, Brazil)*, 11(1), 128. doi:10.1590/S1679-45082013000100024 PMID:23579758
- Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, 72(1), 37–45. doi:10.1080/00031305.2017.1380080
- Velásquez, R. M., & Lara, J. V. (2020). Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression. *Chaos, Solitons, and Fractals*, 136, 109924. doi:10.1016/j.chaos.2020.109924 PMID:32501372
- Wold, H. (1939). A Study in Analysis of Stationary Time Series. *Journal of the Royal Statistical Society*, 102(2), 295. doi:10.2307/2980009
- Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between Time-Series?—A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89(1), 1–63. doi:10.2307/2341482
- Zaremba, W., Sutskever, I., & Vinyals, O. (n.d.). *Recurrent neural network regularization*. arXiv 2014, arXiv:1409.2329.
- Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos, Solitons and Fractals*, 140, 110121. 10.1016/j.chaos.2020.110121
- Zhang, P. G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. doi:10.1016/S0925-2312(01)00702-0
- Zheng, Y. L., Zhang, L. P., Zhang, X. L., Wang, K., & Zheng, Y. J. (2015). Forecast model analysis for the morbidity of tuberculosis in Xinjiang, China. *PloS one*, 10(3), e0116832.

*Sakinat Folorunso is a computer science lecturer and researcher at Olabisi Onabanjo University, Ago Iwoye, Ogun State, Nigeria. She had a B. Tech in Computer Science from the Federal University of Technology, Akure, Ondo State, Nigeria. She also had an M. Sc and PhD in Computer science for the University of Ibadan, Ibadan, Oyo State, Nigeria. She has authored many journals and conference articles in Data mining, Machine Learning and Data Science.*

*Joseph Bamidele Awotunde was born in Ayetoro-Ile Village, Ilorin, Kwara State, Nigeria in 1982. He received a B.Sc. degree in Mathematics/Computer Science from the Federal University of Technology, Minna, Nigeria, in 2007. M.Sc. and Ph.D. degrees in Computer Science from the University of Ilorin, Ilorin, Nigeria, in 2014 and 2019 respectively. From 2012 to 2015, and 2018, he was a Computer Science Instructor with the University School, University of Ilorin, Ilorin, Nigeria. From 2017 to 2018, he was a Lecturer II with the McPherson University, Ijebo, Seriki-Sotayo, Nigeria. Since 2019, he has been a Lecturer II with the Computer Science Department, University, of Ilorin, Ilorin, Nigeria. He is the author of more than 40 articles, and more than 15 Conference Proceedings. His research interests include information security, Cybersecurity, Bioinformatics Artificial Intelligence, Information Security, Software Engineering, and Biometrics. He is a member of the International Association of Engineers and Computer Scientist (MIAENG), Computer Professional Registration Council of Nigeria (MCPN), and Nigeria Computer Society (MNCS). Internet Society.*

*Banjo Oluwatobi O. is a seasoned researcher and programmer with several research works and information systems to his credit. He obtained his first degree from TASUED thereafter he proceeded to UNN for his M. Sc. also in Computer Science. He's currently undergoing his PhD at FUNAAB. His research interest is cognitive Computing, exploring the detection and prediction of human emotion from both text and images. He's currently a Lecturer of Computer Science in the Department of Mathematical Sciences, Olabisi Onabanjo University, Ago Iwoye, Ogun State, Nigeria.*

*Ezekiel Adebayo Ogundepo is a smart and innovative professional in the field of Mathematics and Statistics. He is a first-class graduate of Statistics from the University of Ilorin, Nigeria. He holds a Master's degree from the African Institute for Mathematical Sciences (AIMS), Rwanda with a MasterCard scholarship. He is an expert in advanced statistical modelling with professional experience at the Rwanda Revenue Authority (RRA) where he worked to build dynamic taxpayers' data portal and automated key statistical reports required by internal and external users. Ezekiel currently works at Data Science Nigeria (DSN). He is an active member of the Africa R programming community and a co-editor of Beginners' Artificial Intelligence & Python Programming Book and a certified Rstudio Tidyverse instructor.*

*Wale Adeboye, Ph., is a Statistician and Data Analysis specialist. He parades National Diplomas (ND & HND), B.Sc., M.Sc., and PhD degrees in Statistics as well as Post Graduate Diploma in Planning, Research and Statistics. He is currently a Lecturer at the Department of Mathematics & Statistics, Federal Polytechnic, Ilaro, Nigeria and his research activities are mainly in Computational Statistics, Econometrics / Time Series, Operations Research, Data Management, Information Analysis, Machine Learning, Probability Statistics, and Statistical Inference. He is a member of both National and International Statistical bodies and a Research fellow, strategic institute for natural resources and human development.*