



Forecasting the Spread of COVID-19 Using Deep Learning and Big Data Analytics Methods

Cylas Kiganda¹ · Muhammet Ali Akcayol¹

Received: 2 June 2022 / Accepted: 22 March 2023 / Published online: 3 May 2023
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2023

Abstract

To contain the spread of the COVID-19 pandemic, there is a need for cutting-edge approaches that make use of existing technology capabilities. Forecasting its spread in a single or multiple countries ahead of time is a common strategy in most research. There is, however, a need for all-inclusive studies that capitalize on the entire regions on the African continent. This study closes this gap by conducting a wide-ranging investigation and analysis to forecast COVID-19 cases and identify the most critical countries in terms of the COVID-19 pandemic in all five major African regions. The proposed approach leveraged both statistical and deep learning models that included the autoregressive integrated moving average (ARIMA) model with a seasonal perspective, the long-term memory (LSTM), and Prophet models. In this approach, the forecasting problem was considered as a univariate time series problem using confirmed cumulative COVID-19 cases. The model performance was evaluated using seven performance metrics that included the mean-squared error, root mean-square error, mean absolute percentage error, symmetric mean absolute percentage error, peak signal-to-noise ratio, normalized root mean-square error, and the R2 score. The best-performing model was selected and used to make future predictions for the next 61 days. In this study, the long short-term memory model performed the best. Mali, Angola, Egypt, Somalia, and Gabon from the Western, Southern, Northern, Eastern, and Central African regions, with an expected increase of 22.77%, 18.97%, 11.83%, 10.72%, and 2.81%, respectively, were the most vulnerable countries with the highest expected increase in the number of cumulative positive cases.

Keywords Deep learning · COVID-19 · Artificial neural networks · Long short-term memory · Autoregressive integrated moving average · Prophet

Introduction

The coronavirus disease (COVID-19) is an epidemic that first appeared in Wuhan, Hubei Province, China, on December 31, 2019. It was initially reported as a cluster of pneumonia cases. After a thorough analysis of the severity of the spread, the World Health Organization (WHO) declared COVID-19 as a pandemic on March 11, 2020, as described in the research from WHO timeline to COVID-19 [4]. COVID-19 is caused by the SARS-CoV-2 virus and can infect anyone. In most cases, patients infected with

this disease will recover with no strong treatments applied. According to the findings by WHO [22], individuals may display symptoms ranging from low to average. Individuals with chronic medical illnesses, particularly the elderly, are more likely to experience severe symptoms of this.

The COVID-19 virus spreads from person to person via tiny fluid spread when an infected person coughs, sneezes, speaks, or breathes. When these fluids become stuck on surfaces such as door handles, the virus can be spread to others who come into contact with these surfaces without taking the necessary medical precautions. To prevent the spread of this virus, it is recommended that people keep a 1-m distance from other people, wash their hands frequently or use a disinfectant, wear a mask, and get vaccinated as recommended by WHO [22].

Various approaches have been deployed to prevent and control the spread of the COVID-19 pandemic. Among these strategies is the prediction of the spread of the

✉ Cylas Kiganda
kigandacylas@gmail.com; cylas.kiganda1@gazi.edu.tr
Muhammet Ali Akcayol
akcayol@gazi.edu.tr

¹ Computer Science Department, Institute of Informatics, Gazi University, Ankara, Turkey

COVID-19 virus. In this context, the spread of COVID-19 is considered a time series problem to which deep learning forecasting algorithms and big data statistical models are applied. Among the deep learning algorithms are the long short-term memory (LSTM) model as applied by Marzouk et al. [12], Hssayeni et al. [6], Yu et al. [24], Zeroual et al. [25], Pal et al. [14] and Shastri et al. [16]; the convolutional neural network (CNN) model as applied in research by Huang et al. [8], which performs well on image data such as X-ray images; the autoencoder model, which was applied by Hu [7]; gradient boosting, which provided the best results in research conducted by Zoabi et al. [27]; and the Prophet model, which was applied by P. Wang et al. [19] to perform epidemiological trend prediction. Big data statistical models include models such as the auto-regressive integrated moving average (ARIMA) model as applied by Gebretensae and Asmelash [5] and the susceptible-exposed infectious-removed (SEIR) model, which has been proven to be a robust model to predict the trend of COVID-19 as applied by Yang et al. [23]. Among the deep learning models used to perform time series prediction, the LSTM has been widely used due to its successful results in most research experiments. On the other hand, the ARIMA statistical model has also been widely applied in the health sector, for example, in a study by Y. W. Wang et al. [20] to predict the spread of hepatitis B disease, in the forecasting of medical service demand by Y. Huang et al. [9] and in the prediction of daily blood sampling room visits by Zhang et al. [26].

The following questions will be addressed by this research:

1. What is the best-performing prediction model given the COVID-19 cumulative positive cases data from African countries in five key regions?
2. Is it possible to estimate the total number of cumulative positive cases 61 days ahead of time using the best prediction model?
3. After a 61-day forecasting period, which countries on the African continent are in the most vulnerable position in terms of the COVID-19 virus's spread?

In this study, a comparative and analytical approach were followed to predict the spread of the COVID-19 virus. This approach includes two deep learning models and a statistical model. The deep learning models include LSTM and Prophet. The statistical model comprises the ARIMA model. In most studies, the modeled ARIMA model does not include the seasonal component of the problem. However, in this study, it is modeled to take into consideration the seasonal component of the time series problem. The spread of COVID-19 was considered to be a univariate time series problem using the number of

COVID-19-positive cases. In “[Model Selection Criteria](#)”, the models used in this study are discussed in detail.

This study uses the African continent as a case study. In this comprehensive approach, the African continent was broken down into five major subregions, including Northern, Southern, Eastern, Western, and Central Africa. While most studies focus on a single or a few countries as a case study during the prediction of the spread of COVID-19, this study included and utilized all the African continent's regions. In this study, the successful prediction model was selected by using seven performance indicators. The performance indicators include mean-square error (MSE), root mean-square error (RMSE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE) R2 score, normalized root mean-square error (NRMSE), and peak signal-to-noise ratio (PSNR). In “[The Framework Of The Applied Approach](#)”, the performance metrics are provided in detail. The best-performing model was then used to perform the prediction of COVID-19 cases 61 days ahead of schedule. In “[Results and Discussion](#)”, the model results are provided and discussed in detail.

Related Work

In this section, prediction approaches and methods used in other research studies are addressed. These studies mainly concentrate on the prediction of the spread of COVID-19 using both statistical and deep learning tools.

In a research study by Gebretensae and Asmelash [5], the autoregressive integrated moving average (ARIMA) algorithm was used to forecast the spread of COVID-19 in Ethiopia. The autocorrelation function (ACF) and partial autocorrelation functions (PACF) were used to obtain the model's optimal terms. It was observed that the ARIMA models, ARIMA (0, 1, 5) and ARIMA (2, 1, 3), produced the best results. Ribeiro et al. [15] developed a stacking-ensemble learning algorithm that included ARIMA, cubist regression, random forest, and support vector regression. In this study, the Gaussian process was employed as a meta-learner, while the random forest, ridge regression, and other algorithms were utilized as foundational learners. In this study, it was observed that the support vector regression algorithm produced the best results.

Abdulmajeed et al. [1] applied a deep learning ensemble method to predict COVID-19 cases in Nigeria. The emphasis in this study was to create a prediction method that uses as little data as possible to give accurate predictions. This was because there was a problem with limited training data for models to learn the COVID-19 spread. This deep learning approach combined four prediction approaches, which included one statistical method called ARIMA. Among the other deep learning models in the ensemble approach were

the Prophet model (supported and provided by Facebook), the Holt–Winters exponential smoothing model, and the generalized autoregressive conditional heteroscedasticity (GARCH). While applying the ARIMA model, non-seasonal phenomena were used. To find the best ARIMA model, strategies such as brute search, autocorrelation function inspection, and partial autocorrelation function plots are used.

Wang et al. [19] used a hybrid prediction strategy to predict the COVID-19 cumulative cases in their study. This included the logistic and Prophet models. With the Prophet model, the primary focus was on modeling non-periodic changes. The model included the date and the total number of COVID-19 cases obtained from a specific country. The logistic model was used to identify the quickest rising point in the data in this hybrid method. The output of this model is then fed into the Prophet model, which is used to make the final forecast. Marzouk et al. [12] used three deep learning models to forecast the spread of COVID-19 in Egypt: the LSTM, convolutional neural network, and multilayer perceptron neural network. In this study, the COVID-19 data was modeled as a time series data. In this study, the LSTM outperformed the other two models.

Hssayeni et al. [6] used mobility data to predict the COVID-19 risk spread using the LSTM model and the gradient tree boosting model in their study. In this study, it was discovered that the number of daily cases decreased in the retiree context, while it increased in the youth context. Yang et al. [23], on the other hand, used the susceptible-exposed-infectious-removed (SEIR) and the LSTM models to forecast the spread of the COVID-19 pandemic in China. The SIER algorithm was used to model epidemiological and mobility data by specifying parameters, and the parameter was defined as the product of the daily number of people in contact with COVID-19 patients and the likelihood of transmission. σ was the amount of time it took for a COVID-19 patient to develop infection symptoms. Finally, γ was determined to be the average mortality or recovery rate. The rate of pandemic spread in Hubei province was determined using these parameters. These parameters were then fed into the LSTM model as input.

Zeroual et al. [25] used five models to predict new and recovered COVID-19 cases. The recurrent neural network, long short-term memory, bidirectional LSTM, gated recurrent units, and variational autoencoder were among the models used. The study was carried out in six different countries: Italy, Spain, France, China, the USA, and Australia. The variational autoencoder model produced the best results. The best model was used to forecast cases for the next 2 weeks. To forecast the positive COVID-19 outcome in a PCR test, Zoabi et al. [27] used the gradient-boosting algorithm in conjunction with the Shapley additive explanations (SHAP) bee-swarm plot. Sex, contact with COVID-19 patients, and the presence of the five most notable COVID-19 symptoms

were all model input features. Techniques such as early stopping were used to improve the results.

Pal et al. [14] used the LSTM model and Bayesian optimization to determine COVID-19 risk categories. To obtain the hyperparameters, the search space had to be defined. The optimal hyperparameters were obtained and used by the model in the local trend prediction phase to perform country-specific predictions. Finally, a fuzzy rule-based risk categorization process was carried out, in which the data obtained from the previous module was used to determine each country's risk status. This study concluded that weather had no significant impact on the spread of COVID-19.

Shastri et al. [16] conducted research on COVID-19 time series prediction and comparative analysis using variants of long short-term memory neural network models. Among them were models such as bidirectional long short-term memory, convolutional long short-term memory, and stacked long short-term memory. Two countries were used as case studies. Among these are the USA and India. Because models are sensitive to the size of data input values, tools like MinMaxScaler were used to perform data normalization. Various regions of the USA and India were divided into groups based on the severity of the COVID-19 situation. These were the initial, moderate, and severe groups. Regions with a high number of COVID-19 cases were classified as severe. When compared to the other two models, the convolutional LSTM model produced the best results.

In the related literature, several models have been used to forecast the spread of COVID-19 in a couple of countries. However, the African continent has not been extensively studied in this regard. This study aimed to close this gap by applying the most successful model (LSTM) among the rest of the forecasting models to conduct an extensive investigation and analysis of African states from the five major regions of the continent. In addition, the most critical states with the highest expected COVID-19 increase rate from each region were identified for immediate action in the region.

Methods and Materials

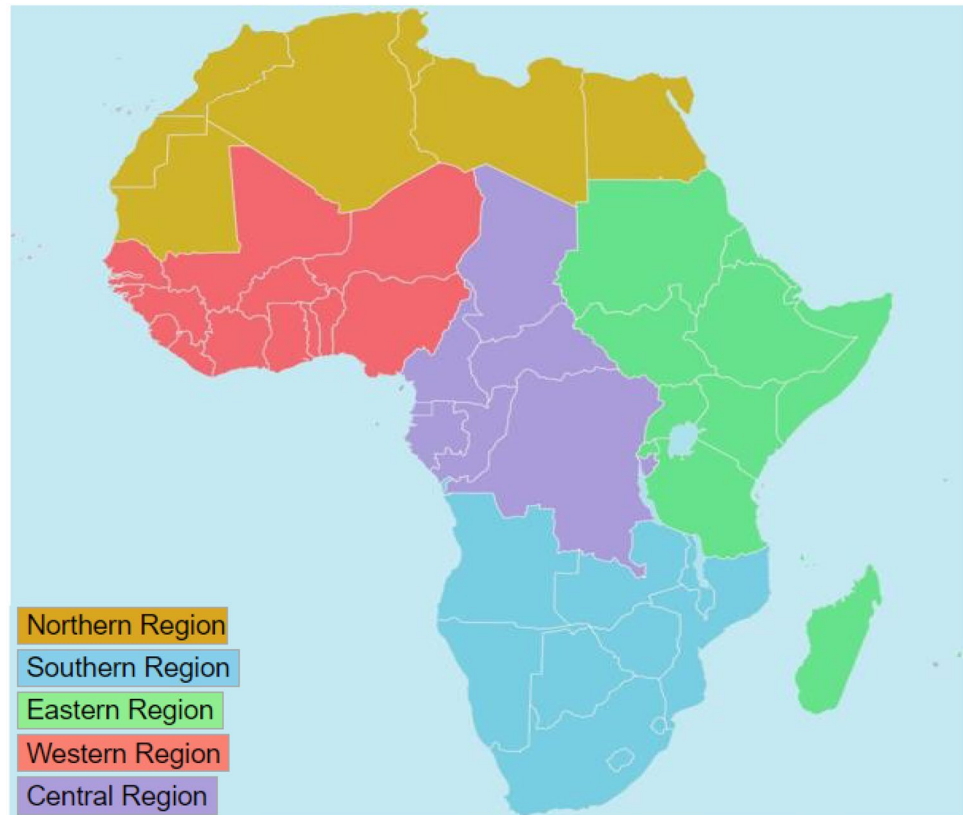
Data Gathering

Africa's Geographical Regions and Populations

The case studies used in this study included countries from the five major regions of the African continent. These regions, as depicted in Figure 1, include the Northern, Eastern, Southern, Central, and Western regions.

Much work on the COVID-19 pandemic has been done in the literature. In some research, several or individual African countries have been used as case studies, for example, research done by Abdulmajeed et al. [1]. In this study, the

Fig. 1 Africa's five major regions



African continent is considered from a broader perspective, including countries from each of the major regions that make up the continent. This study performs a comparative analysis of the COVID-19 pandemic spread.

COVID-19 Data

A humanitarian data exchange [2] source provided the COVID-19 dataset used in this research. This information was gathered by first splitting the data of each country into distinct groups based on the country's geography. The Northern, Southern, Central, Eastern, and Western regions of Africa were used in the study. Model fitting was then done for each country separately. This data was split into training and testing datasets, with the former accounting for 80% of the total prediction models.

ARIMA Model

The ARIMA model is made up of three main parts: the terms "AR," "I," and "MA" are among these elements. As mentioned by Noureen et al. [13], the "AR" term refers to the autoregression parameter. This shows that the variable under consideration in this context has a linear relationship between its present and prior values. That is to say, an AR(1) of order one implies that the current data point in the series

is based directly on the immediate past data point, while an AR(2) implies that it is based on two past data points in the series by Kirbaş et al. [10]. The "I" component stands for the integrated element, which shows the amount of difference between the current data points and their preceding values. This is part of the ARIMA model that handles the data stationarity requirement for better results in ARIMA time series processing, which is attained by the differencing process as explained in the research by the Noureen et al. [13]. Stationarity in ARIMA processing refers to the condition when the mean and variance statistical parameters in the time series data are constant with respect to the time factor. The last part in the basic ARIMA structure is the "MA" part, which represents the moving average. This component displays the linear combination that exists between the error values at past intervals in the time series as denoted by Ribeiro et al. [15]. The standard notation of the basic ARIMA model is denoted as ARIMA (p, d, q). The p , d , and q terms represent the autoregressive, differencing, and moving average terms as described in the research by Abdulmajeed et al. [1]. The mathematical notation for the AR (p) term can be represented as shown in Eq. 1.

$$Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t. \quad (1)$$

In the above equation, Y_t denotes the time series value at a given time point t . The p , δ , and ϵ_t denote the autoregression term, fixed value, and the error value, respectively. The moving average component can be defined mathematically in Eq. 2.

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_2 \epsilon_{t-2} + \theta_q \epsilon_{t-q}. \tag{2}$$

In Eq. 2, q depicts the order of the moving average term. The difference term d can be obtained from Eq. 3.

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - LY_t. \tag{3}$$

In Eq. 3, ΔY_t denotes the stationary time series value at a time interval t .

$$(1 - \varphi_1 L - \varphi_1 L^2 - \dots - \varphi_p L^p) \Delta Y_t = \delta + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \tag{4}$$

Equation 4 is a combination of all the equations for the basic ARIMA model terms. This denoted the full ARIMA (p, d, q) model equation with the complete set of terms computed and represented.

The partial autocorrelation function (PACF) and autocorrelation function (ACF) graphs, as shown in Fig. 2, can also be used to obtain the ARIMA model's p and q terms. The ACF plot is a graphical representation of the average correlation between data and prior values in a time series over different lag intervals. The only difference between the two exists in the fact that PACF reveals correlations within a shorter lag interval, as explained in the research by Noureen et al. [13].

Prophet Model

The Prophet model is a deep learning model for time series forecasting. The Facebook group created and maintains this model as an open-source initiative. According to Taylor and Letham [18], it is based on the generic specification of a

generative additive model (GAM), which is a linear regression model whose linear variable is reliant on smoothing functions. GAMs can be quantitatively represented using Eq. 5.

$$g(E(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_m(x_m). \tag{5}$$

In Eq. 6, Y represents the univariate response variable, x_1 represents the predictor variable, and f_1 represents the smoothing functions. Due to its use of GAM model formulation, the Prophet model has a variety of benefits, including flexibility and quick fitting times, and evaluates a time series problem from three perspectives, including trend, seasonality, and holiday components, as discussed in research carried out by Taylor and Letham [18]. The trend component takes into account the likelihood of time series data increasing or decreasing over time. Seasonality, on the other hand, looks at data changes that happen over a short time period.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t. \tag{6}$$

The final predicted value $y(t)$ is obtained from a combination of the trend, seasonal and holiday component functions as shown in Eq. 6 above, where ϵ_t represents the changes that are not captured by the model [18].

LSTM Model

The LSTM model is composed of three main core components. These include the forget gate, input gate, and output gate [16]. The forget gate identifies the degree to which past data is obliterated. The input gate receives the data that is taken into the cell's internal state, while the output gate is used to create the next hidden state or output that is obtained from the existing internal state value.

The above figure displays the major building blocks of the LSTM model. It is evident that the main building blocks of the LSTM model consist of the forget gate, input gate

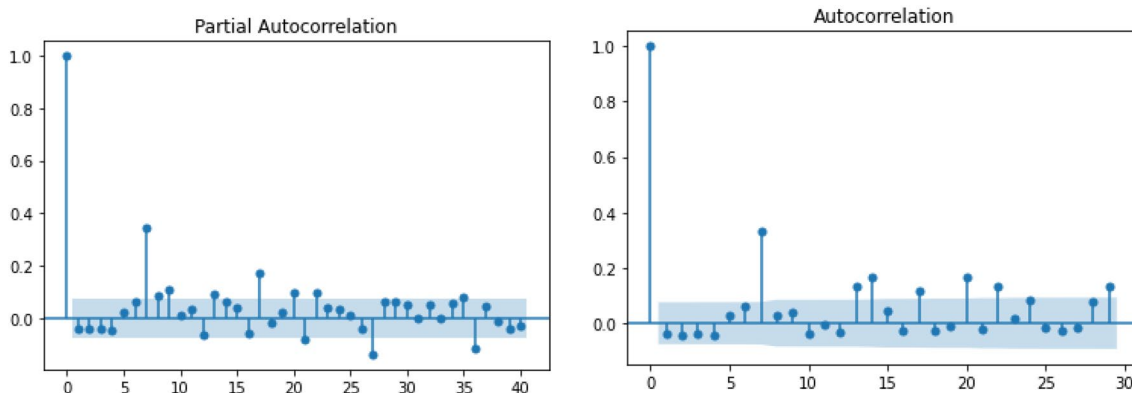


Fig. 2 Representation of PACF and ACF plots

and output gates as described by Le and Lee [11]. Several activation functions are used such as the tanh and sigmoid functions for extracting the optimal model weight values.

Model Selection Criteria

In this study, seven metrics were adopted to assess the predictive performance of the models. These metrics include, the peak signal-to-noise ratio (PSNR), mean-squared error (MSE), root mean-square error (RMSE), symmetric mean absolute percentage error (SMAPE), mean absolute percentage error (MAPE), normalized root mean-square error (NRMSE), and R2 score.

Mean-Square Error

The mean-squared error can be calculated numerically as below.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i^{\wedge} - Y_i)^2 \quad (7)$$

The overall number of observations n , the exact value Y , and the anticipated value Y^{\wedge} are all represented in Eq. 7.

Root Mean-Square Error

The RMSE can be calculated using Eq. 8.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i^{\wedge} - Y_i)^2} \quad (8)$$

The overall number of observations n , the actual value Y , and the anticipated value Y^{\wedge} are all represented in Eq. 8.

Mean Absolute Percentage Error

Equation 12 can be used to represent this performance measure numerically.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{A_t - F_t}{A_t} \right|. \quad (9)$$

The observed vector of numbers is represented by A_t , the projected value is expressed by F_t , and the total number of data points is represented by n in Eq. 9.

Symmetric Mean Absolute Percentage Error

Equation 10 can be used to represent this measurement numerically.

$$SMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}. \quad (10)$$

The observed vector numbers are represented by A_t , the forecasted value is represented by F_t , and the overall number of observations is represented by n in Eq. 10.

Peak Signal-to-Noise Ratio

$$PSNR = 20 \log_{10} \left(\frac{MAX_f}{\sqrt{MSE}} \right). \quad (11)$$

The highest signal value is expressed by MAX_f in Eq. 11. MSE stands for mean-square error.

Normalized Root Mean-Square Error

$$NRMSE = \frac{RMSE}{Y_{max} - Y_{min}}. \quad (12)$$

The root mean-square deviation (RMSD) is defined in Eq. 12. The RMSD measure is also known as the RMSE statistic (Fig. 3).

R2 Score

$$R_2 = \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - Y^i)^2}. \quad (13)$$

In Eq. 13, the projected values are represented by f_i , whereas the original values are represented by y_i , and the mean is represented by Y^i .

The Framework of Applied Approach

In Fig. 4, the major stages of this study include splitting the preprocessed positive COVID-19 cumulative cases data into 80% training and 20% testing datasets, fitting the models, validating the model performance using the performance, and then selecting the best-performing model to use it to forecast the future positive COVID-19 cases for the next 61 days.

Rationale for the Selected Models

This section aims to address the reasons for choosing the LSTM, ARIMA, and Prophet models to perform the prediction and forecasting of the COVID-19 cumulative positive cases data for the various African countries in this study.

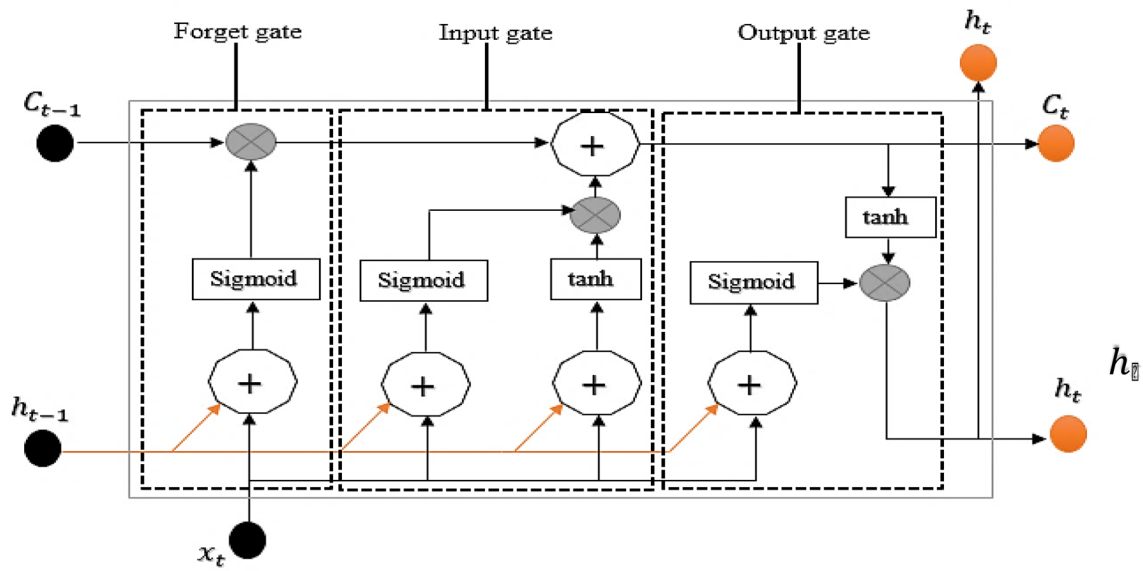
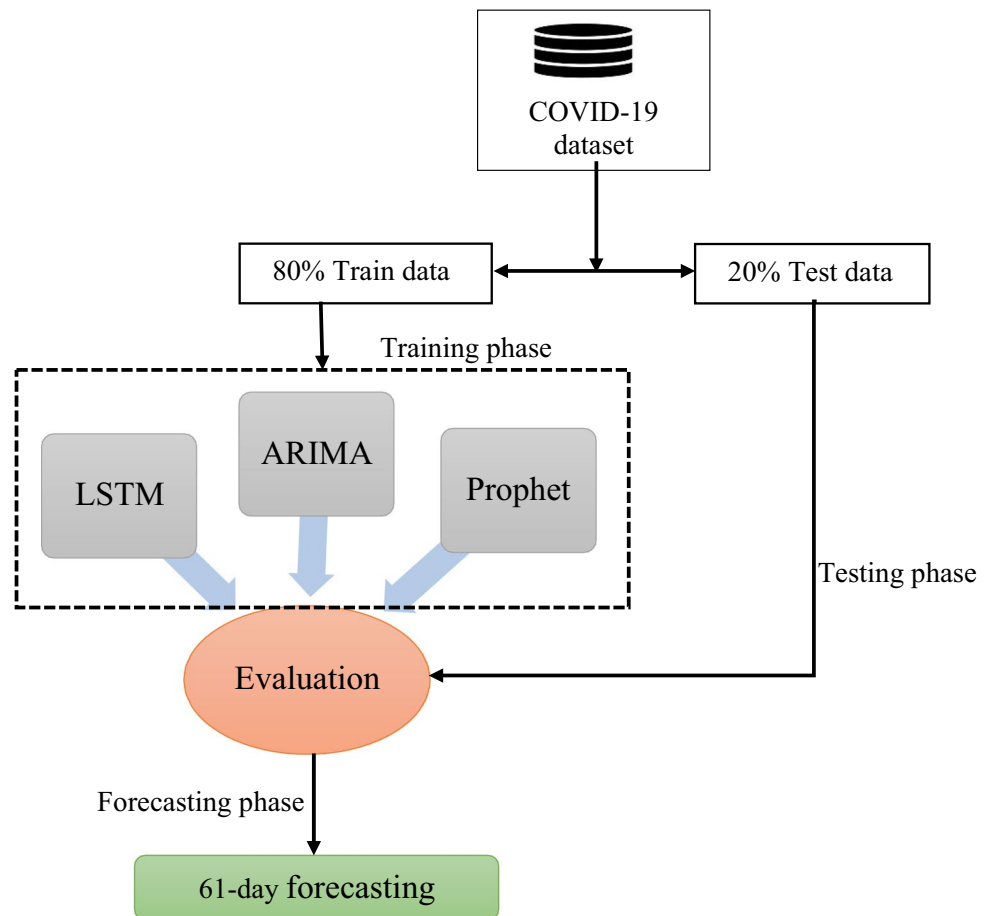


Fig. 3 General structure of an LSTM model

Fig. 4 Structural depiction of the methodology used in this study



LSTM

This model is a special class of recurrent neural network deep learning models with the capability to identify and learn the relationship that exists within a given series of data observations, as described in the research by Yu et al. [24]. This is possible because the LSTM has memory modules that act as a connection between past and current data points. Important data points with strong desired insights are retained, while those with weaker weights are disposed of in the forget module of the LSTM model. This both optimizes the model to concentrate on extracting the dependence that exists within a given input sequence and also minimizes the error by eliminating noise points from the learned data at this stage. As described by Zeroual et al. [25], the LSTM model eliminates the problem of vanishing gradients that is faced with traditional recurrent neural networks, whereby the computed gradient fluctuates within peak ranges, that is to say, either too big or too small. According to Zeroual et al. [25], this issue arises during the training phase. The LSTM model solves the vanishing gradient problem with the help of activation vectors used in the forget gate to determine the gradient values. It is at this point that the LSTM model, by using a summative strategy, identifies the optimal terms to adjust at a given time step, which improves accuracy and overall performance. The LSTM model implementation provides several hyperparameters, such as the batch and epoch numbers, which can be easily adjusted to obtain better results. This makes it easy to fit and use the LSTM model to achieve accurate results. These robust qualities of the LSTM model make it ideal for performing the time series prediction task.

ARIMA

This is a statistical method that uses regression in which past data points and errors are connected using weight factors, which improves the overall prediction results, as described in the research by Singh et al. [17]. This model also amalgamates the strengths of both the autoregression and moving average models, which further makes it a robust choice that extracts the inherent statistical relationship between the dependent and independent variables. It is a flexible model to use, since it incorporates the difference between data points both in the past and present context, which makes it able to handle and process data which is not stationary using a few parameters as described by Abdulmajeed et al. [1]. Another factor lies in the fact that it is easier to obtain the optimal parameter terms of this model using simple methods like the PACF and ACF plots, as described in the research by Gebretensae and Asmelash [5]. Also, metrics such as the Akaike information criteria and Bayesian information criteria make it possible to measure how good the ARIMA

model is for a given combination of hyperparameter terms, which further makes it easier to streamline the prediction results. This model has the ability to process data with seasonal trends by further increasing the hyperparameter terms to include the seasonal factors, as explained by Y. Wang et al. [21]. This makes it possible to capture any seasonal relationship within the COVID-19 dataset at any given time.

Prophet

According to Abdulmajeed et al. [1], this is an additive regression model supported by Facebook with a robust architecture that takes into account seasonal dynamics within a given data sequence, such as yearly, weekly, and daily trends. It also handles data with missing data points and extreme values well, since it has the ability to identify data anomalies as described by Y. Wang et al. [21]. This makes it an ideal solution to process and predict the COVID-19 datasets in some countries with data of this nature, such as data that has sharp spikes from the normal trend in the general data. According to research by Letham and Taylor [18], the Prophet model has built-in computational support that handles non-linear growth curves when the natural boundary is reached and also offers flexibility in tuning, such as smoothing features that capture and model seasonality constraints in the data to make a good fit regarding historical cycles. It is also easy to capture and model the effects of events such as holidays in the time series data with the Prophet model using limited data [18]. These qualities make this model appropriate to perform the prediction of the COVID-19 spread.

Results and Discussion

In this study, countries from the African continent were grouped into the five groups named in “Data Gathering”. Three forecasting models were used, including the ARIMA, LSTM, and Prophet. In this section, the performance results obtained from these models are given for each region of Africa.

Model Training and Testing

Northern Africa

In the Northern region of Africa, of the six countries studied, the most densely populated country is Egypt, as shown in Fig. 2, with a population of 102334404, while the least populated country is Mauritania, with a population of 4649658 as observed in the work by Worldometer [3].

In Fig. 5, it can be seen that Morocco has maintained the highest number of COVID-19 cases over time. This was

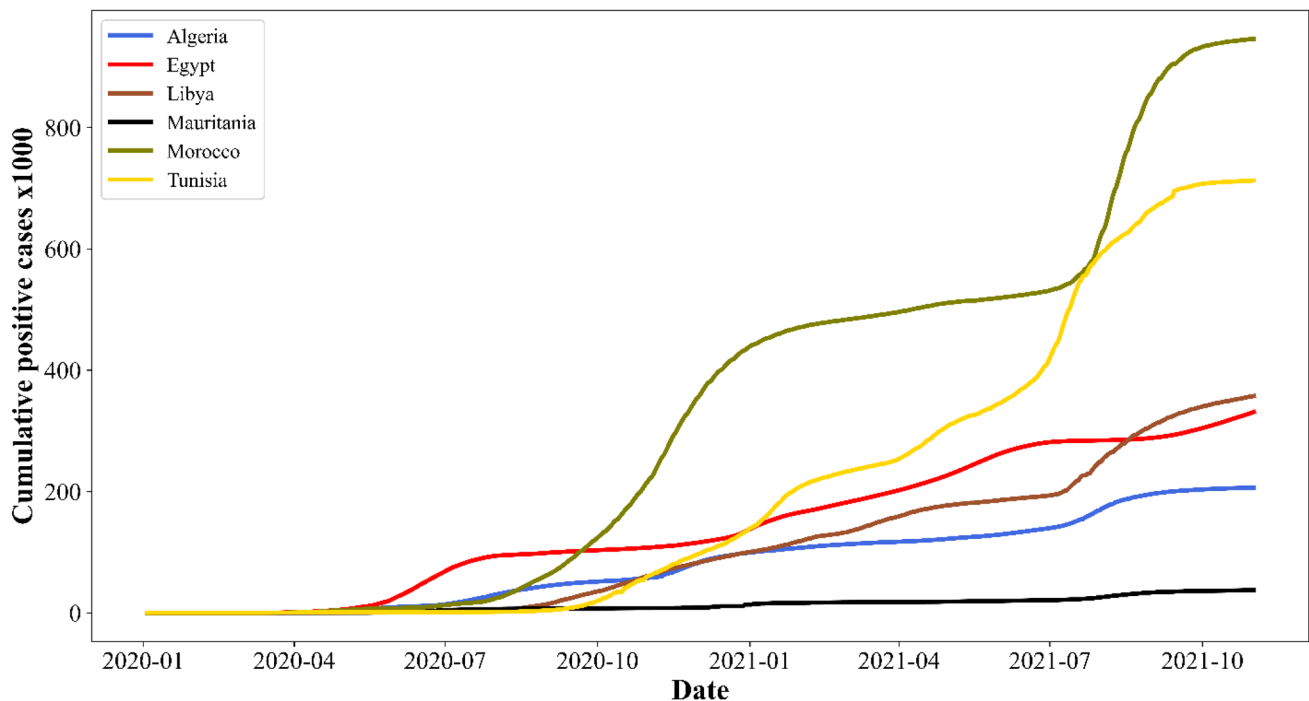


Fig. 5 Cumulative positive cases for Northern Africa

followed by Tunisia in this critical condition. On the other hand, Mauritania, on the other hand, has the lowest number of cases over time compared to other states in this region.

Libya has a relative increase in cases, with a gradual increase occurring between the months of October 2020 and July 2021. Beyond the month of July, a sharp increase that slowly reduces toward the month of October is observed. This clearly describes the first wave of COVID-19 cases in Libya. Algeria's trend is more similar to that of Libya's. However, it is observed that the cases reach a constant number, while in Libya there is an increase.

According to Fig. 6, it is observed that the LSTM model fits better than both the ARIMA and Prophet models. In Tunisia, it can be observed that the Prophet model performs the worst in predicting the test data. This is because while the test data flattens to a constant case value, the Prophet model predicts a sharp increase of over 800,000 cases. In countries like Egypt and Tunisia, the ARIMA and Prophet models predicted lower and higher cases, respectively, with respect to the actual data. Apart from these two countries, in the four other countries, both models predicted lower cumulative positive cases with regard to the actual data. This confirms the poor performance of these two models when compared to the LSTM model, which predicts better results close to the actual data in five countries except Egypt.

In Table 1, the best results in terms of the PSNR and R value can be observed with larger numbers, which implies

that the greater the number, the better is the model's relative performance.

Central Africa

In this region, five states were studied. At the time of this study, the most populated state in this group was Cameroon, with a population of 265,458,633 [3]. On the other hand, the least populated state is São Tomé and Príncipe, whose population is 219,159.

In Fig. 7 above, the COVID-19 cumulative cases from the five countries in this region have been given. According to this graph, COVID-19 cases in Cameroon are higher than in the rest of the countries, with more than two significant waves. Cameroon is followed by Gabon, which also has more than two waves. The rest of the countries maintain a slightly constant curve, with minor increases in COVID-19 cases. The lowest number of cases is seen in São Tomé and Príncipe. A positive correlation is observed between the population variable and the number of cases. This is because the highest number of cases is observed in Cameroon, which is also the most populated state in this region [3]. On the other hand, it can also be observed that the least number of cases are observed in São Tomé and Príncipe, a country with the smallest population. This makes Cameroon the member with the highest risk in terms of COVID-19 spread in this region.

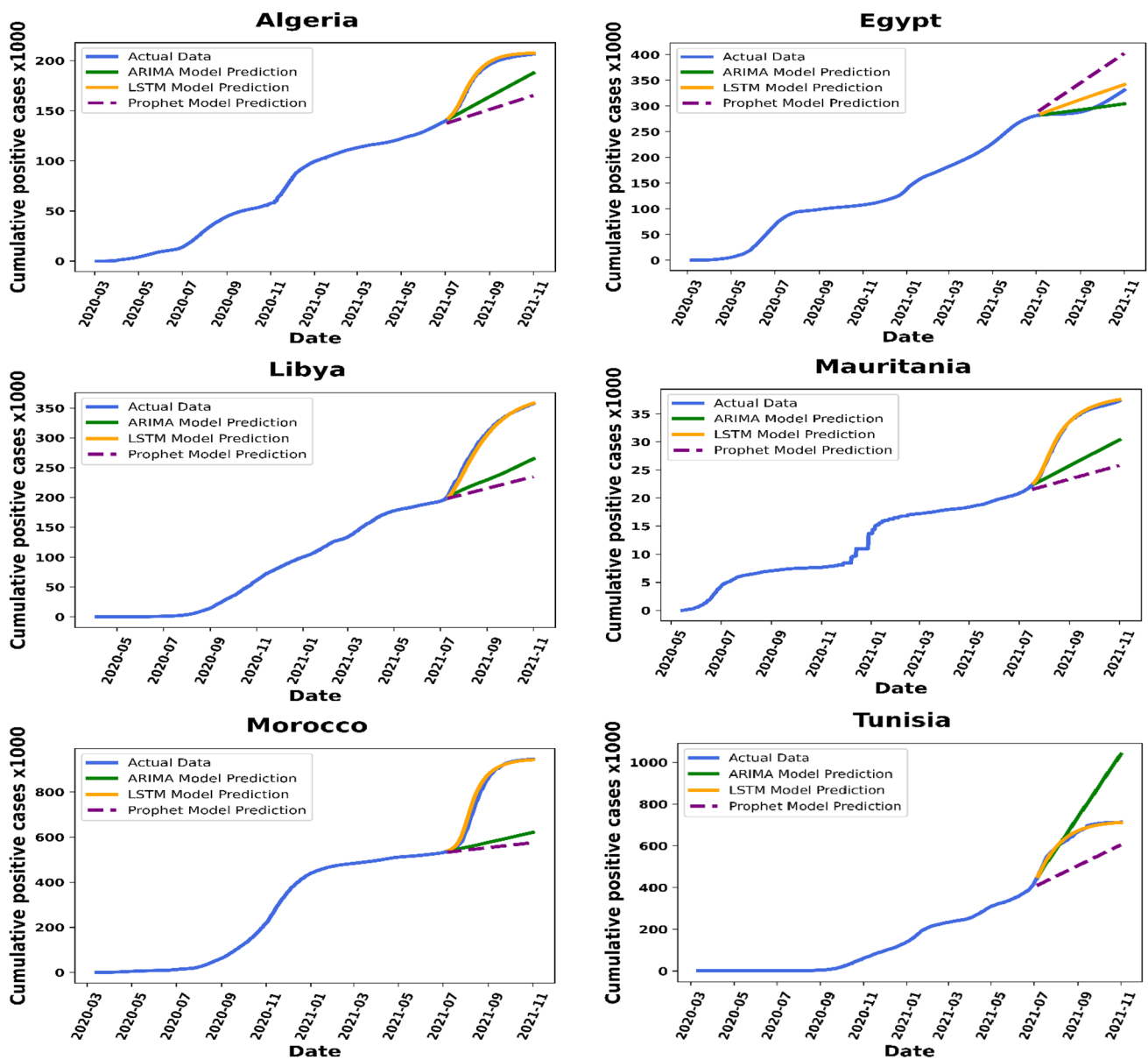


Fig. 6 Actual and predicted cumulative cases in Northern Africa

Figure 8 shows a plot of the model performance after prediction of the test data in various countries in the Central African region. In three countries, the LSTM model prediction generally matches well with the actual data. This implies that the best performance in this region was observed from the LSTM model. It is also observed that the worst model performance is given by the Prophet model, for example in Cameroon. In Chad, the ARIMA model performs relatively well in predicting the data, while in the rest of the countries, it comes immediately after the LSTM model.

Southern Africa

From this region, ten countries were used in this study. As shown in Fig. 3, the most densely populated country in this region is South Africa, with a population of 59308690. The least populated, on the other hand, is Eswatini, with a population of 1160164.

In Fig. 9, it is clearly observed that South Africa has the highest number of cases compared to other countries in the same region. This shows how fast the COVID-19 virus spreads in this country. This puts the other neighboring countries in the same region at a very high risk of having increased rates of spread of the virus. While the other

Table 1 Performance parameters of the models for Northern Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
Mauritania	ARIMA	-28.3210	-1.0387	0.4392	20.0932	6646.4399	44175163.3400	18.0172
	LSTM	<u>-1.0514</u>	<u>0.9962</u>	0.0190	0.7515	<u>287.8118</u>	82835.6322	<u>0.7551</u>
	Prophet	-31.3765	-3.1201	0.6244	30.2207	9448.5330	89274775.8500	25.8536
Algeria	ARIMA	-39.6473	-0.3443	0.3715	12.4041	24485.1867	599524367.7000	11.5554
	LSTM	-18.1055	0.9906	0.0311	1.0187	2050.2803	4203649.3090	1.0244
	Prophet	-43.4308	-2.2125	0.5744	20.2395	37851.3220	1432722577.0000	18.1621
Morocco	ARIMA	-59.9838	-1.7886	0.6184	29.8376	254525.4081	64783183368.0000	<u>25.0037</u>
	LSTM	-38.0839	0.9820	0.0497	1.8985	20452.0670	418287044.6000	1.9400
	Prophet	-60.8397	-2.3961	0.6824	33.7894	<u>280882.9106</u>	78895209467.0000	27.7835
Libya	ARIMA	-49.3128	-1.4477	0.4721	24.5338	74504.0709	5550856581.0000	21.5143
	LSTM	-25.4671	0.9899	0.0303	1.4573	4785.1562	22897719.8600	1.4395
	Prophet	-51.1119	-2.7040	0.5808	31.1302	91650.8128	8399871487.0000	<u>26.4318</u>
Egypt	ARIMA	-31.1439	0.5917	0.1873	2.0405	9198.8946	84619661.8600	<u>2.0052</u>
	LSTM	-37.1818	-0.6398	0.3753	5.5535	18434.4519	339829016.9000	<u>5.7408</u>
	Prophet	-46.5591	-13.2078	1.1048	15.3217	54261.6346	2944324989.0000	<u>16.7849</u>
Tunisia	ARIMA	-55.2048	-2.9832	0.5456	13.7406	146819.1010	21555848418.0000	15.5706
	LSTM	-28.8865	0.9907	0.0264	0.9365	7093.5494	50318443.0900	0.9371
	Prophet	-54.7576	-2.5934	0.5182	23.6105	139450.1419	19446342076.0000	21.0428

The highest PSNR and *R* values were obtained by the LSTM model in Mauritania. These values were -1.0514 and 0.9962, respectively. For the rest of the performance metrics other than PSNR and *R* values, the best results are observed with lower values. It is also evident in Mauritania that the lowest RMSE value of 287.8118 was obtained from the LSTM model. The ARIMA and Prophet models produced MAPE ranges of 2.0052–25.0037 and 16.7849–26.4318. On the other hand, it was observed that the LSTM model produced the lowest MAPE of 0.7551–5.7408. The highest MAPE value for the LSTM model is clearly observed to be lower than the lowest MAPE values for both the ARIMA and Prophet models. This makes the LSTM the best-performing model in predicting the COVID-19 cumulative in the Northern African region. Among the countries in this region, the best model performance was observed in Mauritania, while the worst model performance was observed in Morocco, with an RMSE value of 280882.9106 by the Prophet model

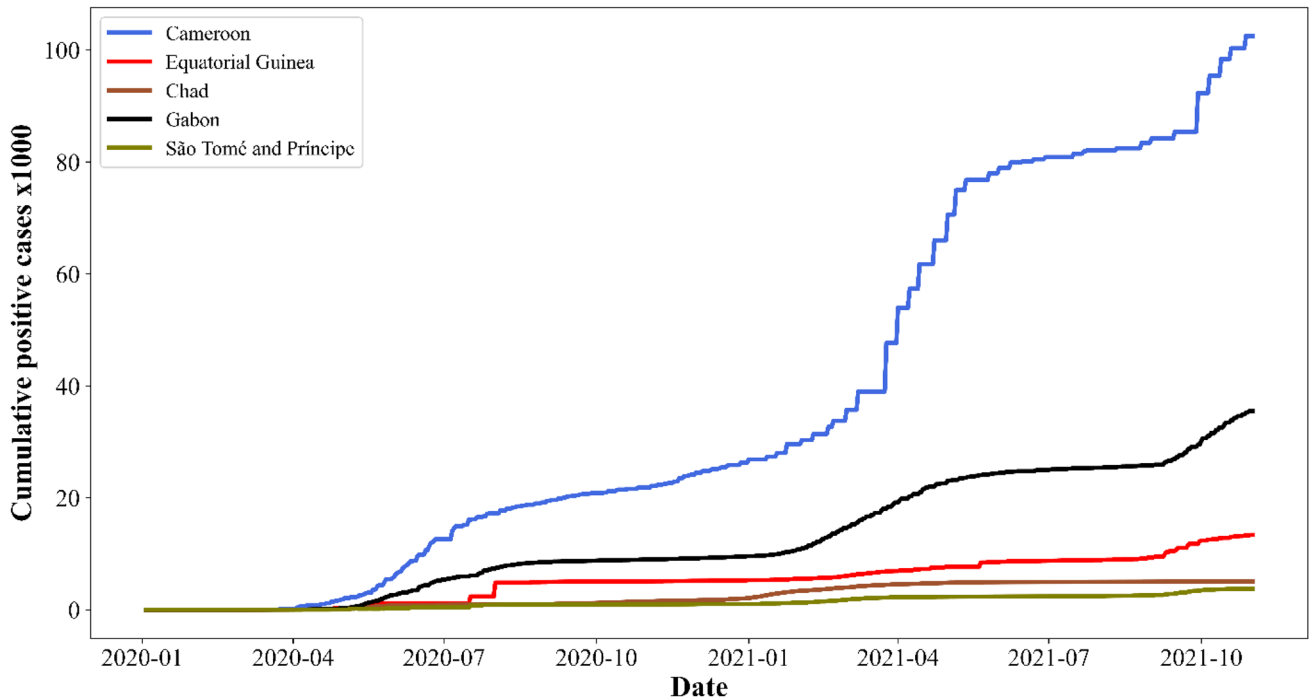


Fig. 7 Cumulative positive cases for Central Africa

Table 2 Performance parameters of the models for Central Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
Cameroon	ARIMA	-23.7735	0.6842	0.1819	3.8398	3937.4568	15503566.0519	3.8655
	LSTM	-29.9249	-0.3019	0.3694	7.0175	7994.4080	63910559.2705	7.4279
	Prophet	-40.3369	-13.3144	1.2249	24.9969	26508.4573	702698308.4259	28.8557
Gabon	ARIMA	-22.9030	-0.0850	0.3428	7.2343	3561.9664	12687604.6347	6.6077
	LSTM	-5.0523	0.9822	0.0439	1.0980	456.2018	208120.0823	1.1101
	Prophet	-19.1268	0.5452	0.2219	7.4839	2306.1026	5318109.2017	7.7485
Chad	ARIMA	14.9747	-0.3830	0.3921	0.7312	45.4783	2068.2758	0.7272
	LSTM	25.6706	0.8822	0.1144	0.2213	13.2742	176.2044	0.2212
	Prophet	0.0557	-41.9264	2.1842	4.6249	253.3710	64196.8636	4.7480
Equatorial Guinea	ARIMA	-19.4805	-0.9611	0.5230	16.0832	2401.9573	5769398.8710	13.9253
	LSTM	-10.6582	0.7428	0.1894	5.8417	869.8616	756659.2032	6.1844
	Prophet	-12.1101	0.6407	0.2238	8.8312	1028.1208	1057032.3790	8.8613
São Tomé and Príncipe	ARIMA	-7.0908	-0.2709	0.4364	12.6397	576.8751	332784.8810	11.1389
	LSTM	16.3165	0.9942	0.0295	0.8522	38.9686	1518.5518	0.8449
	Prophet	-8.3251	-0.6886	0.5030	15.7833	664.9582	442169.4077	13.7283

In Table 2, it is observed that the lowest RMSE value of 13.2742 was obtained by the LSTM model from Chad, as well as the highest RMSE of 26508.4573 was obtained by the Prophet model in Cameroon. It is evident that the lowest and highest PSNR values of 40.3369 and 16.3165 were observed in Cameroon and São Tomé and Príncipe by the Prophet and LSTM models, respectively. The best MAPE range of 0.2212–7.4279 was obtained by the LSTM model, followed by 0.7272–1.1389 and 4.7480–28.8557 by the ARIMA and Prophet models. In this region, the best model performance was obtained by the LSTM model, while the worst model performance was seen in the Prophet model

countries in the same region are experiencing their second wave of virus spread, South Africa is observed to have three waves. Since it has the largest population, there is a positive correlation between the large number of cases observed and the large population.

For clarity, in Fig. 10, South Africa was excluded to be able to perform a comparative analysis of the COVID-19 state in other countries in the same region. It can be observed that, apart from South Africa, Zambia has the largest number of cases compared to other countries. It is also the first country to have an earlier increase in the number of cases. It is also observed that all countries have had their second major wave of COVID-19 spread. It is worth noting that the lowest number of cases was observed in Lesotho. Beyond the month of October, it is clearly observed that in all countries, there is a constant number of cases with the curves flattened. This clearly signifies the effects of some form of control of the spread by a number of practices, such as quarantines and vaccinations.

In Fig. 11, in three countries (Botswana, Malawi, and Mozambique), the LSTM model provided the best-matching prediction results. In Lesotho, the ARIMA model performed better than the other two models. The Prophet model emerged as the worst performer, as clearly observed in four countries: Malawi, Mozambique, Eswatini, and Lesotho. In these countries, this model predicts a roughly constant number of cases, with slight increases in the predicted number of cases. In Angola, both the LSTM and Prophet models produced slightly matching predictions

close to the actual data, while the ARIMA model predicted a lower number of cases, quite different but also substantially close to the actual data. It is in this country that the three models show a significant uniformity in their predicted results. This can be generally attributed to the smooth rise in the number of cases in Angola, which makes it easier for all the models to capture the inherent data relationships and trends to be able to make better predictions.

In Fig. 12, it is observed that the ARIMA model performed the worst when compared to the other countries. This model made predictions that were generally higher than the actual data. In all four countries, the ARIMA model predicts a higher number of cases than the numbers predicted by the rest of the models. The LSTM model is also observed to provide the best performance with the best-matching predictions. The LSTM model is followed by the Prophet model, with the second-best prediction performance. In the South African region, the LSTM model is observed to provide the best overall prediction results compared to the ARIMA and Prophet models, as shown in both Figs. 11 and 12, while the worst prediction results are observed from the ARIMA model.

Table 3 displays the performance metrics used to determine the best prediction model in the Southern African region.

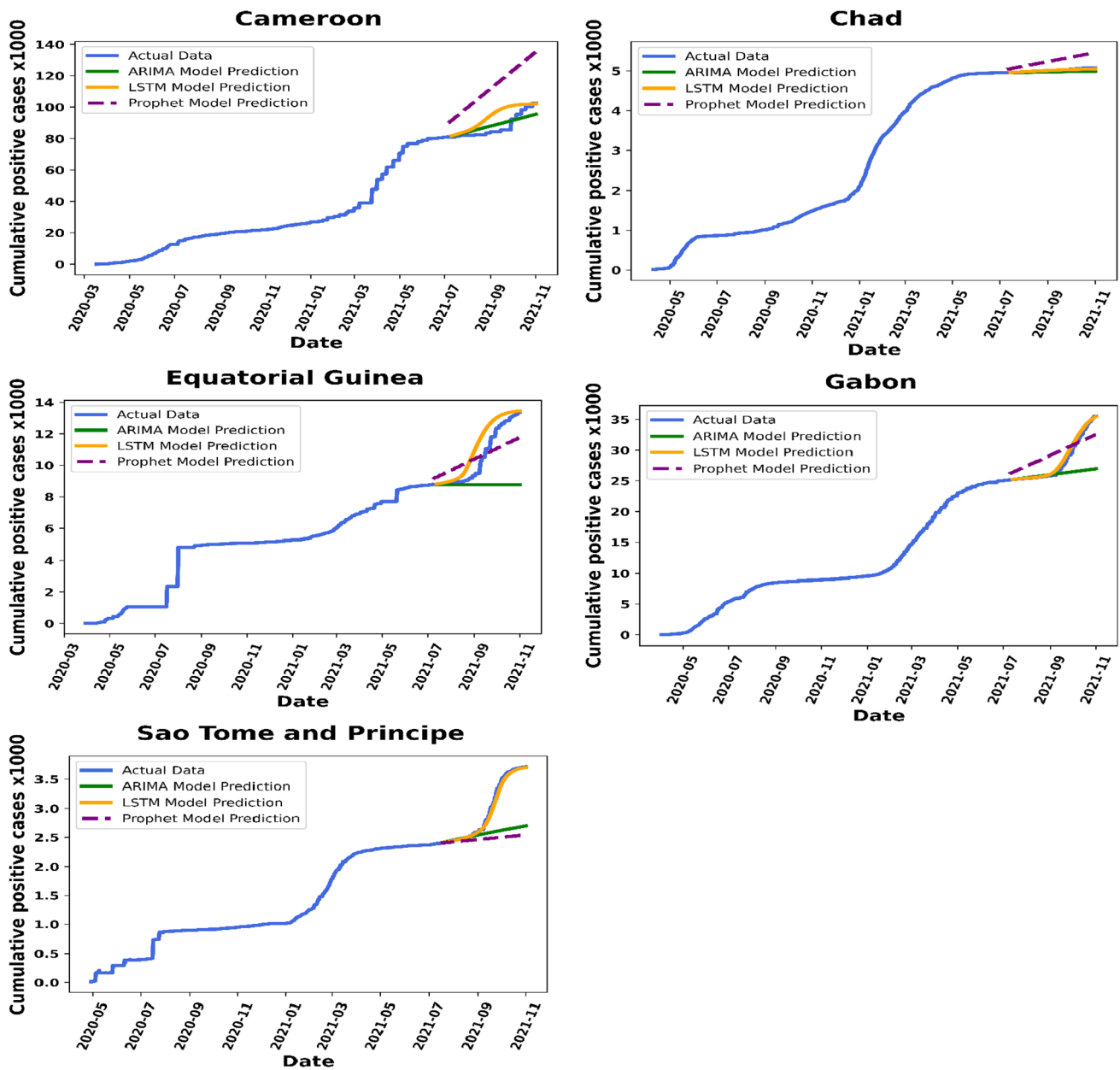


Fig. 8 Actual and predicted cumulative cases in Central Africa

Western Africa

In this research study, 12 countries from this region were used as case studies. In the Western region, Nigeria is the country with the largest population, with a total of 206139589 people. Guinea-Bissau, on the other hand, has the smallest population of 1968001 [3].

In Fig. 13, a comparative plot of the 12 countries used in this study from the Western region of Africa has been given. This shows the state of the COVID-19 pandemic in each of the 12 counties. It also displays the severity of the risk situation in terms of the COVID-19 spread given by

the cumulative positive cases. It is observed that between the months of January 2020 and April of the same year, no COVID-19 cases were reported in this region. However, beyond the month of April of the same year, the first cases have begun to be reported. Notably, after this, in about four countries, which include Nigeria, Ghana, Senegal, and Mali, there is a sharp increase in the number of cases, while in the other countries there is a gradual increase in the number of cases. Nigeria, followed by Ghana and Senegal, displays the highest number of cases over time. Nigeria, being the most populated country with over 200 million people and the highest number of cases, is the riskiest member in this

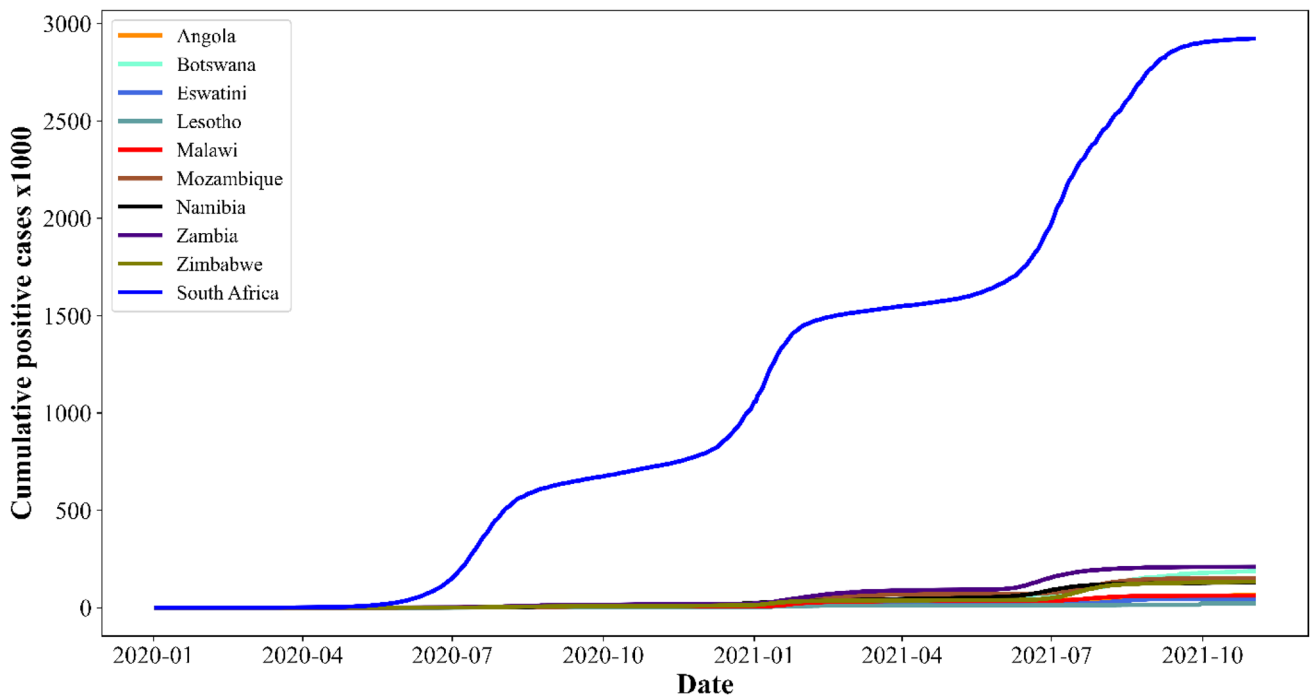


Fig. 9 Cumulative positive cases in the Southern African region including South Africa

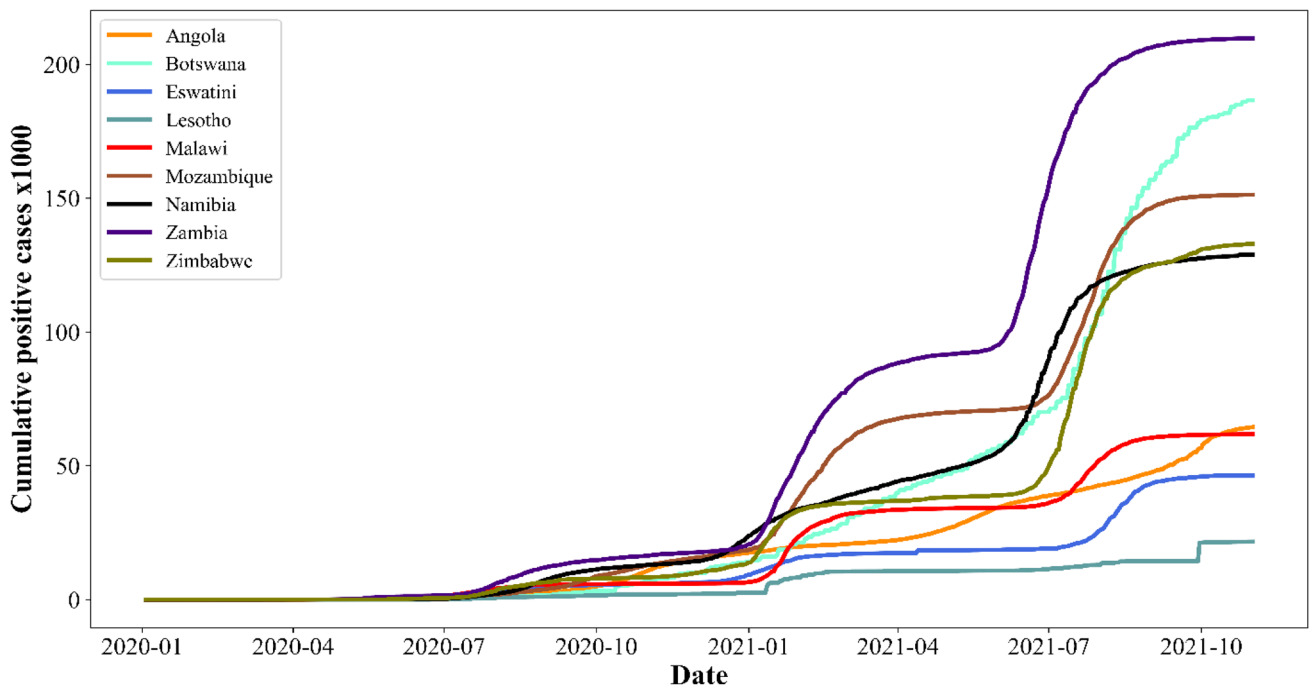


Fig. 10 Cumulative positive cases in the Southern African region excluding South Africa

region. If immediate measures are not taken, there are higher chances of a faster spread to other countries too.

Figure 14 displays the prediction results of the three models in the region of Western Africa. In this first group

of countries from this region, it can be observed that the LSTM model outperformed the other two models in producing the best-matching prediction results. This can be clearly observed in countries like Guinea, Guinea-Bissau, Gambia,

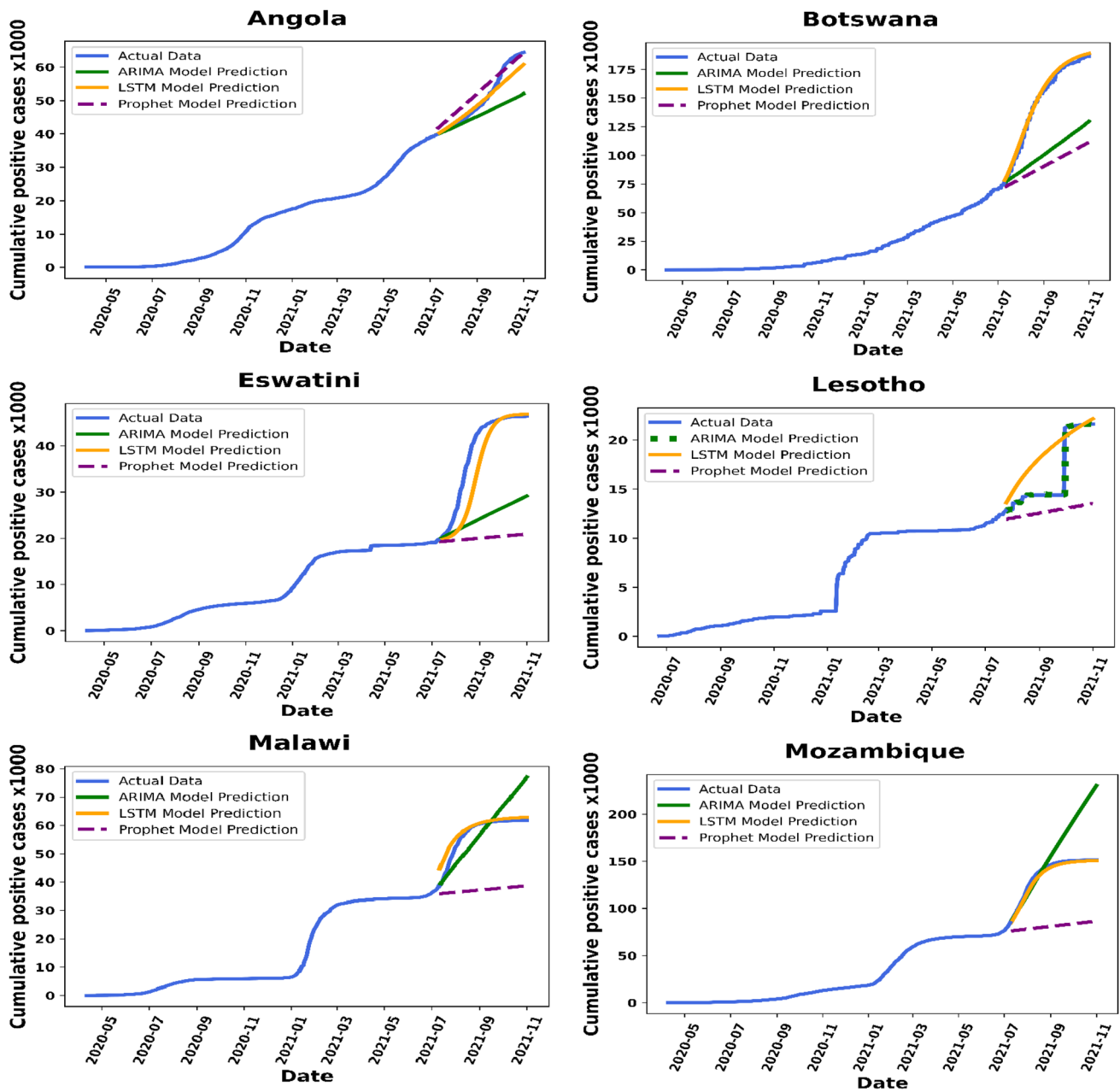


Fig. 11 Actual and predicted cumulative positive cases for Southern Africa (a)

Ghana, and Togo. In Burkina Faso, the Prophet model manages to make the most successful prediction. The ARIMA and Prophet are observed to make marching predictions in three countries: Guinea-Bissau, Ghana, and Togo. These predictions suggest a lower COVID-19 case number when compared to the actual data. This provides another proof of how these two models perform poorly when compared to the LSTM model. In Fig. 15, the second group of model predictions in the Western region of Africa is given. According to this figure, it can be observed that the best model prediction performance obtained in Niger is obtained from the Prophet

model. This is the only country where this model performs best when its performance is compared to the remaining countries. It can also be concluded from this figure that the ARIMA model did not display any top performance in any of the countries. In all the six countries in this group in the Western region of Africa, the LSTM model maintains the best-matching prediction results, which continues to affirm the LSTM model as the top performing model in this region. In Nigeria, both the ARIMA and Prophet models make matching predictions against each other, which is still lower and significantly different from the actual data. These results

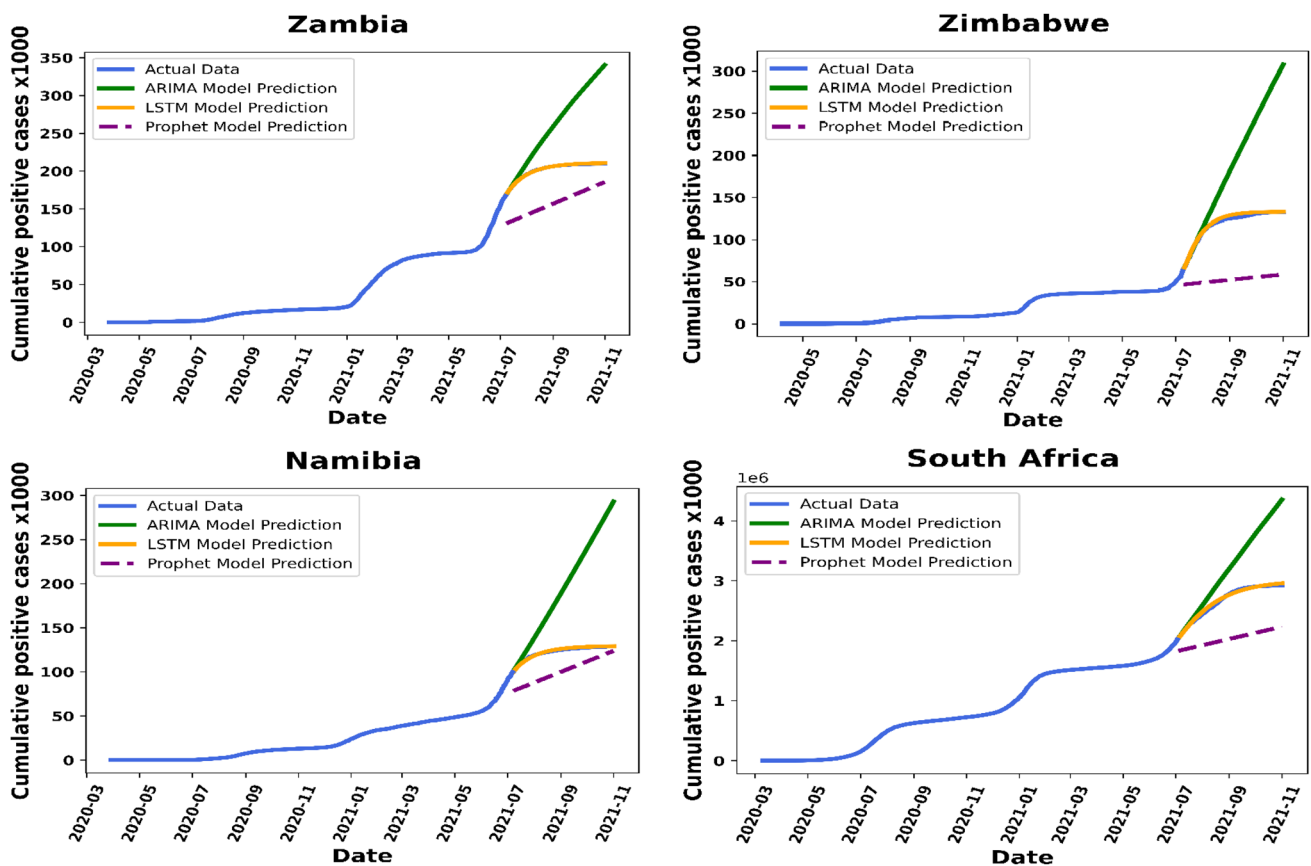


Fig. 12 Actual and predicted cumulative positive cases in Southern Africa (b)

prove the LSTM model to be the best prediction model in the West African region.

In Table 4, the prediction results based on the seven metrics used in this study for the three models are provided for the 12 countries from the Western region of Africa.

Eastern Africa

From this region, 12 countries were studied. Among these, the Comoros is observed to be the least populated country, with a population of 869601, while the most populated country is observed to be Ethiopia, with a population of 114963588 at the time of this study.

The cumulative positive COVID-19 cases for the countries in the Eastern region of Africa have been given in the plot in Fig. 16. It is notably clear that in this region, the highest number of cases is obtained in Ethiopia, which is followed by Kenya. It is worth noting that the population of Kenya, at 53771296 people immediately follows that of Ethiopia, while at the same time, its number of cumulative cases immediately follows that of Ethiopia, which means a roughly positive correlation between the population size and

the number of confirmed cases. If proactive measures are not applied, the Eastern region is at a higher risk of experiencing a surge in the spread of COVID-19. In the region, there was a relatively late occurrence of the first cases, which is observed from the fact that the significant numbers of cases started to be registered just after the month of July in 2020 in all countries. In this region, Kenya is observed to have the highest number of waves of the COVID-19 spread. Apart from Ethiopia, Kenya, Uganda, Rwanda, Madagascar, and Sudan, the rest of the countries are observed to have a relatively slow increase in the number of cases reported. This can be due to varying measures that might have been taken by the respective countries and also the general population. For example, in the Comoros, the least populated country in this region.

Both Figs. 16 and 17 display the prediction results from the LSTM, ARIMA, and Prophet models in the 12 countries used in this study from the Eastern region of Africa. These results display both the plots of the predicted data by the models and the expected actual data. It is observed from Fig. 16 that all three models performed relatively well in the Comoros, followed by Sudan, as displayed in Fig. 17. In the rest of the countries, in both figures, it can be observed

Table 3 Performance parameters of the models for Southern Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
South Africa	ARIMA	<u>-68.6754</u>	-5.7832	0.7905	16.5715	<u>692322.2002</u>	47931000000.0000	19.0097
	LSTM	-39.7648	0.9913	0.0283	0.7800	<u>24818.6368</u>	615964732.6000	0.7830
	Prophet	-68.2322	-5.1251	0.7512	26.7633	<u>657882.1667</u>	43280900000.0000	23.5047
Zambia	ARIMA	-48.9441	-48.6151	1.7532	23.3656	71407.6662	5099054792.0000	28.1140
	LSTM	<u>-6.7122</u>	<u>0.9970</u>	0.0136	0.2252	<u>552.2662</u>	304997.9557	<u>0.2252</u>
	Prophet	-44.8702	-18.4190	1.0968	24.6496	44673.6041	1995730903.0000	21.8100
Namibia	ARIMA	-50.7674	-168.8737	3.0293	40.7690	88086.5415	7759238793.0000	<u>57.0550</u>
	LSTM	-11.0236	0.9820	0.0312	0.6692	907.2334	823072.4421	0.6697
	Prophet	-39.0956	-10.5598	0.7902	19.9281	22978.5282	528012758.2000	17.8522
Eswatini	ARIMA	-35.7542	-1.6589	0.5871	40.7329	15640.5055	244625412.3000	32.6660
	LSTM	-25.7619	0.7336	0.1858	10.6050	4950.3420	24505885.9200	9.5222
	Prophet	-38.0908	-3.5536	0.7683	58.2823	20468.2129	418947739.3000	43.3898
Lesotho	ARIMA	-8.5919	0.9618	0.0775	0.6781	<u>685.6982</u>	470182.0215	0.6157
	LSTM	-21.4589	0.2614	0.3409	14.5044	3016.3651	9098458.4170	16.4016
	Prophet	-25.6834	-0.9537	0.5545	23.6798	4905.8203	24067072.8200	20.2240
Malawi	ARIMA	-28.4637	-0.0858	0.2957	9.3612	6756.5296	45650692.2400	9.5218
	LSTM	-20.0071	0.8451	0.1117	3.3325	2552.0723	6513073.0240	3.4858
	Prophet	-38.2399	-9.3125	0.9113	41.4722	20822.5839	433580000.3000	34.0048
Mozambique	ARIMA	-42.8290	-2.4604	0.5430	14.7160	35317.3358	1247314208.0000	16.9596
	LSTM	-21.2013	0.9762	0.0450	1.9387	2928.2214	8574480.5670	1.9128
	Prophet	-47.2140	-8.4979	0.8996	50.2045	58511.4488	3423589641.0000	39.7238
Botswana	ARIMA	-45.9388	-1.1362	0.4543	34.8145	50522.0752	2552480083.0000	29.1172
	LSTM	-21.9147	0.9915	0.0286	1.9982	3178.8901	10105342.2700	2.0284
	Prophet	-47.6825	-2.1915	0.5553	44.8292	61753.7964	3813531370.0000	35.9618
Angola	ARIMA	-28.5294	0.3054	0.2763	9.3038	6807.8362	46346633.7300	8.6089
	LSTM	-19.7646	0.9077	0.1007	3.0467	2481.8282	6159471.2140	2.9731
	Prophet	-21.5315	0.8613	0.1234	5.5594	<u>3041.7034</u>	9251959.5740	5.7561
Zimbabwe	ARIMA	-50.9603	-25.8469	1.3262	38.1986	90064.6700	8111644782.0000	54.1828
	LSTM	-19.8316	0.9793	0.0368	1.6663	2501.0312	6255157.0630	1.6868
	Prophet	-48.4614	-14.1010	0.9947	75.8921	67547.6906	4562690505.0000	54.7687

The highest *R* value of 0.9970 was obtained from Zambia by the LSTM model, while the smallest *R* value of -168.8737 was observed from Namibia by the ARIMA model. On the other hand, the smallest PSNR value -68.6754 was obtained from South Africa by the ARIMA model, while the highest PSNR value -6.7122 was obtained by the LSTM model from Zambia. Both the PSNR and *R* metrics suggest that the LSTM model is the best prediction model in this region, while the worst prediction model in this region is ARIMA. The RMSE metric ranges of 552.2662-24818.6368, 685.6982-692322.2002 and 3041.7034-657882.1667 were obtained by the LSTM, ARIMA and LSTM models, respectively. It is also evident from this metric that the best RMSE range was produced by the LSTM model compared to the rest of the models. With the smallest value of the MAPE metric of 0.2252 from Zambia, the overall best-performing model in the Southern Africa region is observed to be the LSTM model, while the worst-performing model, with the largest MAPE of 57.0550 from Namibia, is observed to be the ARIMA model.

that the three models show significant relative discrepancies in performance. In Fig. 16, both the LSTM and ARIMA models obtained better match prediction results when compared with the Prophet model in Madagascar. In Fig. 16, the worst model performance is observed in both Djibouti and Madagascar by the Prophet model. On the other hand, the best model performance is evidently obtained by the LSTM model in all countries represented by the same figure. In Fig. 17 too, the LSTM model is observed to have the overall best-matching prediction results when compared to the ARIMA and Prophet models. In both Mauritius and

Rwanda, the worst model performance can be observed from both the ARIMA and Prophet models. In this particular scenario, both models predicted extremely varied results from the actual data. These results conclude that the LSTM model outperformed the ARIMA and Prophet models in the Eastern region.

In Table 5, the three model performances have been given for the 12 countries from the Eastern region of Africa.

Figure 18 displays the overall combined model performance from all individual regions used in this study. It shows the percentage distributions both in the positive and

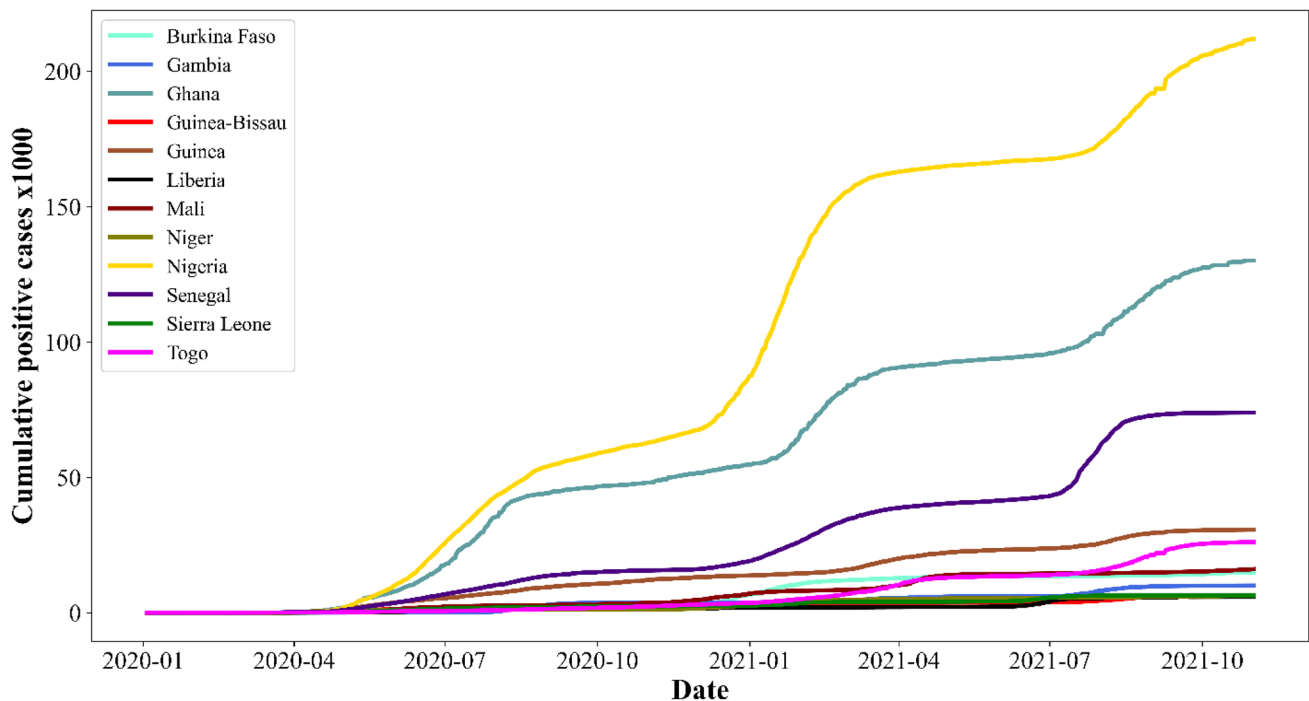


Fig. 13 Cumulative positive cases in Western Africa

negative directions to quantify each model's performance depending on its contribution to the total error value for the seven error metrics used in this study. In both PSNR and R, good performance is indicated by having more distribution toward the positive direction, just as bad performance can be observed by having a more negative percentage distribution. For RMSE, MAPE, NRMSE, SMAPE, and the MSE errors, good performance can be observed in having smaller percentage distributions tending in the positive direction. On the other hand, bad performance for the models can be observed in having a large positive percentage distribution. The RMSE, MAPE, NRMSE, SMAPE, and MSE metrics clearly state that the overall best performance in this study was obtained by the LSTM model, followed by the ARIMA model, and lastly, the Prophet model. This is because the LSTM model is observed to have obtained the smallest percentage distribution of the total error in all these five metrics. The ARIMA model follows, with relatively larger percentage distributions than the LSTM model, but smaller compared to the Prophet model. The PSNR and R values also clarify that the LSTM model is observed to outperform the other two models. Both the PSNR and R values for the LSTM model tend toward the positive direction, showing that it achieved the highest values for these two metrics compared to the ARIMA and Prophet models. It is again followed by the ARIMA and, lastly, the Prophet model, respectively. The LSTM model's performance is owed to the fact that it can process and handle sequential data of

all natures, while the other two models are affected by the quality of their inherent data properties. The ARIMA model works best with stationary data, and also requires a larger amount of data to fit well. With data that is not stationary, the ARIMA model performs poorly. The data used in this study was small in amount due to the fact that the COVID-19 pandemic is still a new ordeal with little data available. In most countries, the datasets were not significantly able to be made stationary, despite the differencing efforts to make them so during ARIMA model fitting. All of these factors contribute to its poor performance when compared to the LSTM model. On the other hand, in this study, it is observed that the overall worst-performing model is the Prophet model. Despite its ease of setup and not requiring data preprocessing, this Fourier series-based model failed to find and learn significant trends, seasonality, and holiday structures within the data to make best-matching predictions, which is because of the limited data available and given for training. The LSTM model's having several hyperparameter tuning points made it possible for it to be tuned until the best-matching results were reached. When compared to the other two models, the computational and time complexity of the LSTM model in order to achieve optimal results was the highest.

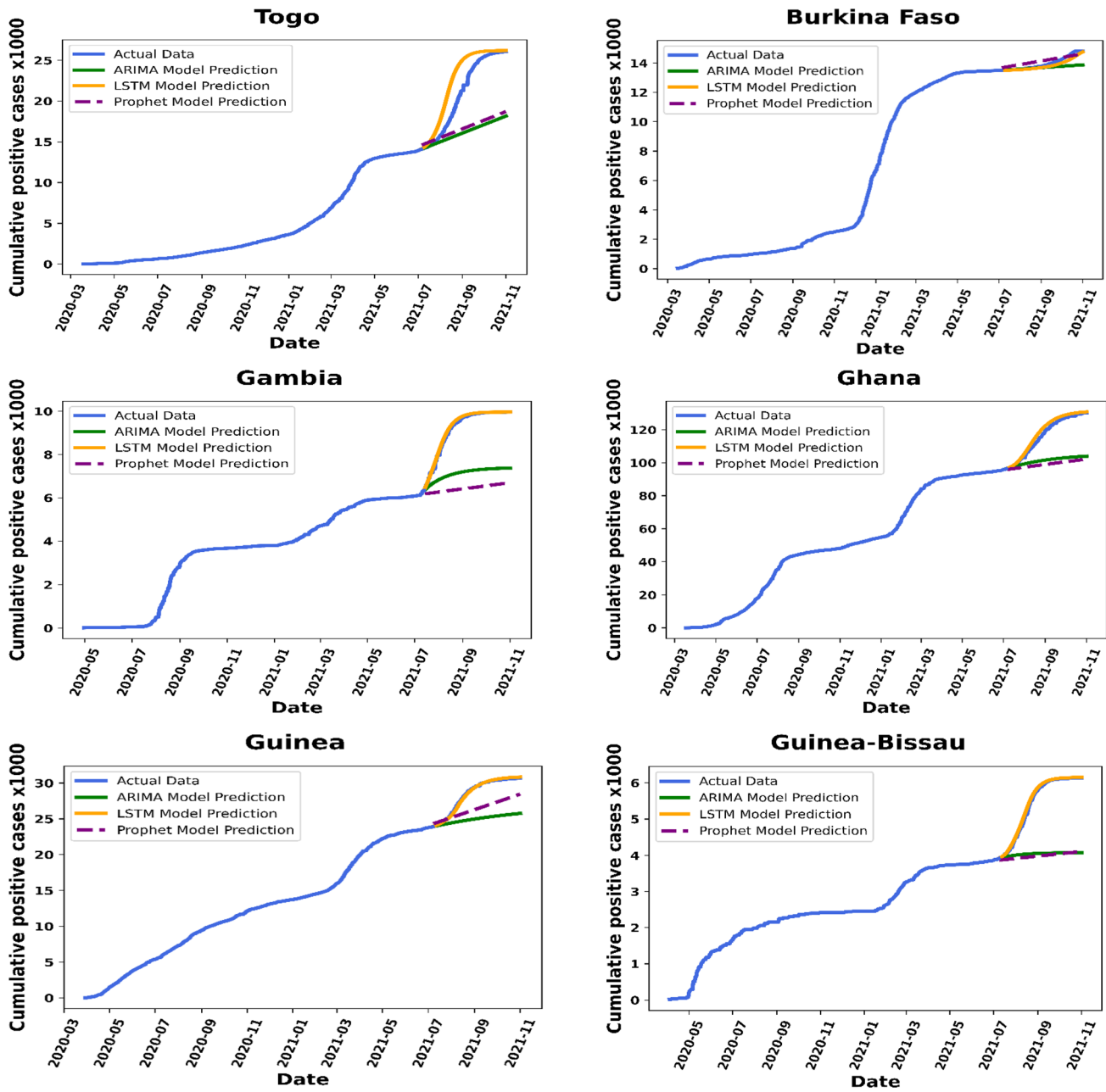


Fig. 14 Actual and predicted cumulative positive cases in Western Africa (a)

Forecasting for the Next 61 Days

In this study, after determining the best prediction model through the training and testing processes, the second major phase involved the forecasting of the cumulative positive cases by the best-performing model for each country for a period of 61 days. At the time of access to the main COVID-19 case dataset used in this study, the last date of the reported cases for each country in all regions was 2021-11-1. Cumulative positive cases were then forecasted from the last date of the original dataset up to the date of 2022-01-02

for each country in the five major regions of the African continent.

Northern Africa

As displayed in Fig. 19, the COVID-19 cumulative positive cases are expected to have a fast increasing rate in Egypt as well. While in countries like Tunisia, Algeria, and Mauritania, cases are expected to maintain a flat rate of increase, in Libya it is expected to show a gradual increase in the rate of

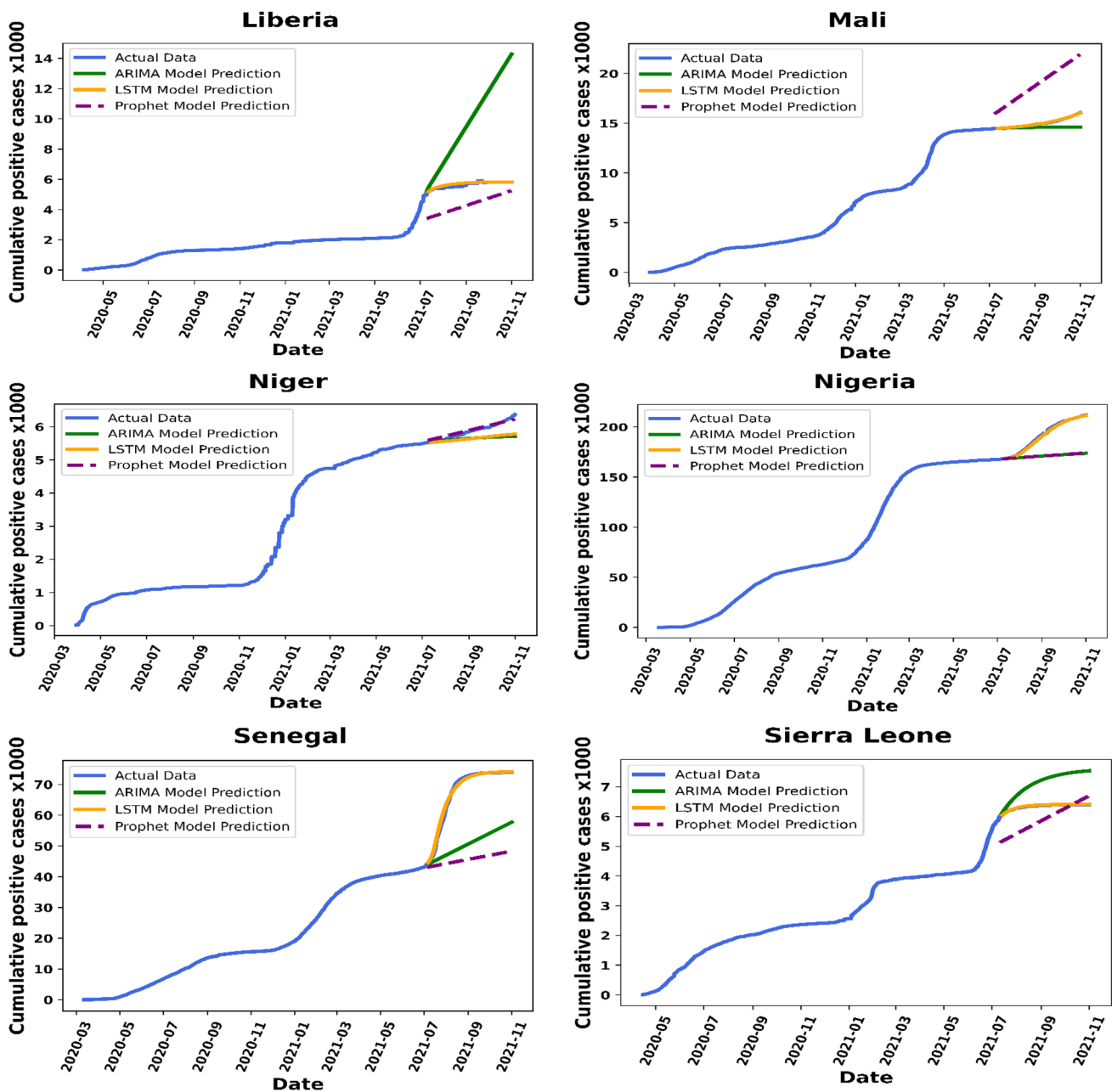


Fig. 15 Actual and predicted cumulative cases in Western African (b)

increase. In Morocco, a notable slight decrease is expected, after which a constant number of cases with a small increase at the end of the forecasting period is expected. At the end of the prediction period, all these countries in Northern Africa that reported cumulative cases are expected to show an increase. In Algeria, Mauritania, Tunisia, Egypt, Libya, and Morocco, cases are expected to increase from 206452 to 208009, 37320–38250, 712747–716835, 331017–370164, 357338–369986, and from 946145 to 947226, respectively. With an 11.83% increase in the number of cases at the end of the forecasting period, it is observed that Egypt is the

country in this region with the largest expected increase in the number of COVID-19 cumulative positive cases.

Central Africa

In Fig. 20, the forecasted cases for the five Central African countries have been plotted. In Cameroon, the cases are expected to slightly drop to a constant rate of increase. In Gabon and Equatorial Guinea, a gradual increase is expected, while in Chad and São Tomé and Príncipe, a constant rate of change in the cases is expected. At the end of

Table 4 Performance parameters of the models for Western Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
Niger	ARIMA	-1.1740	-0.6132	0.3475	3.9254	291.9044	85208.1787	3.8076
	LSTM	-0.5325	-0.3917	0.3228	3.8508	271.1221	73507.1931	3.7488
	Prophet	12.2196	0.9262	0.0743	0.9621	62.4539	3900.4896	<u>0.9659</u>
Mali	ARIMA	-7.2888	-0.7804	0.3645	2.8408	590.1751	348306.6487	2.7670
	LSTM	19.7407	<u>0.9965</u>	0.0162	0.1374	26.2729	690.2653	<u>0.1375</u>
	Prophet	-24.2073	-86.5715	2.5566	22.7287	4139.1138	17132263.0500	25.9383
Liberia	ARIMA	-25.5340	<u>-558.9784</u>	5.2358	49.7815	4822.1454	23253086.2600	<u>72.6324</u>
	LSTM	7.4258	0.7167	0.1178	1.4477	108.4557	11762.6389	1.4647
	Prophet	-14.4240	-42.3685	1.4571	26.6279	1341.9666	1800874.3560	23.2167
Guinea	ARIMA	-23.7940	-1.8990	0.5884	12.7953	3946.7398	15576755.0500	11.8515
	LSTM	3.7225	0.9949	0.0248	0.5042	166.1170	27594.8577	0.5034
	Prophet	-19.6152	-0.1076	0.3637	7.5991	2439.4969	5951145.1250	7.2484
Guinea-Bissau	ARIMA	-15.9092	-2.9237	0.7195	28.2155	1592.2175	2535156.5670	23.9589
	LSTM	13.1228	0.9951	0.0254	0.8095	56.2861	3168.1251	0.8156
	Prophet	-16.0789	-3.0801	0.7337	29.4110	1623.6256	2636160.0890	24.9430
Ghana	ARIMA	-36.9592	-1.3960	0.5336	13.5331	17967.9858	322848513.7000	12.4038
	LSTM	-18.7581	0.9637	0.0656	1.5789	2210.2812	4885342.9830	1.5967
	Prophet	-37.8953	-1.9724	0.5943	15.6611	20012.7714	400511019.1000	14.2355
Gambia	ARIMA	-18.8193	-3.6199	0.6265	24.9010	2225.9044	4954650.3980	21.8765
	LSTM	9.2716	0.9928	0.0247	0.7308	87.6925	7689.9746	0.7358
	Prophet	-21.1720	-6.9414	0.8214	34.6748	2918.3552	8516797.0730	29.1977
Burkina Faso	ARIMA	-4.7163	-0.0108	0.3379	2.0304	438.8913	192625.5732	<u>1.9840</u>
	LSTM	3.2095	0.8370	0.1357	0.9683	176.2237	31054.7924	0.9607
	Prophet	-0.2520	0.6384	0.2021	1.7649	262.5062	68909.5050	1.7794
Togo	ARIMA	-27.1499	-0.6747	0.4841	23.3990	5808.0929	33733943.1400	20.0786
	LSTM	-18.9326	0.7475	0.1880	8.1182	2255.1162	5085549.0760	<u>8.7542</u>
	Prophet	-26.4842	-0.4367	0.4484	21.2872	5379.6047	28940146.7300	18.5120
Sierra Leone	ARIMA	-10.4028	-88.3031	2.1384	11.3549	844.6563	713444.2651	12.1692
	LSTM	<u>22.5264</u>	0.9545	0.0483	0.2727	<u>19.0642</u>	363.4437	0.2728
	Prophet	-6.9960	-39.7560	1.4446	8.0593	570.6140	325600.3370	7.6311
Nigeria	ARIMA	<u>-39.5807</u>	-1.5257	0.5510	10.7816	<u>24298.1514</u>	590400161.5000	10.0128
	LSTM	-12.3440	0.9952	0.0239	0.4389	1056.1850	1115526.7540	0.4374
	Prophet	-39.5353	-1.4995	0.5481	10.7166	24171.6715	584269703.1000	9.9560
Senegal	ARIMA	-36.9374	-2.5463	0.5945	27.3320	17923.0581	321236011.7000	23.6709
	LSTM	-9.9932	0.9928	0.0267	0.9606	805.7498	649232.7402	0.9669
	Prophet	-39.2209	-4.9995	0.7732	37.3071	23312.2755	543462189.0000	<u>30.8797</u>

As annotated in the table based on the PSNR metric, the best model performance was obtained from the LSTM model from Sierra Leone, with a PSNR value of 22.5264. On the other hand, the worst performance was observed in Nigeria, with a PSNR value of -39.5807, which was obtained by the ARIMA model for Nigeria. From the *R* metric, it is also observed that the highest value of 0.9965 is produced by the LSTM model in Mali, as well as the lowest value of -558.9784 was obtained by the ARIMA model in Liberia. It is worth noting that a higher PSNR and *R* value imply better results. Values obtained with this metric show that the LSTM models outperform the other two models in the Western African region as well. The lowest RMSE value of 19.0642 obtained by the same model in Sierra Leone further reinforces this observation, while the worst performance with regard to the same metric can be notably seen in Nigeria by the ARIMA model with the highest value of 24,298.1514. When the MAPE metric ranges from all the three models are taken into consideration, it is observed that the LSTM model provides the best range, reflecting the best performance of 0.1375–8.7542, which is followed by the ARIMA model with a range of 1.9840–72.6324 and the Prophet model with a range of 0.9659–30.8797. These performance metric results affirm that the best-performing model is the LSTM model in the Western region.

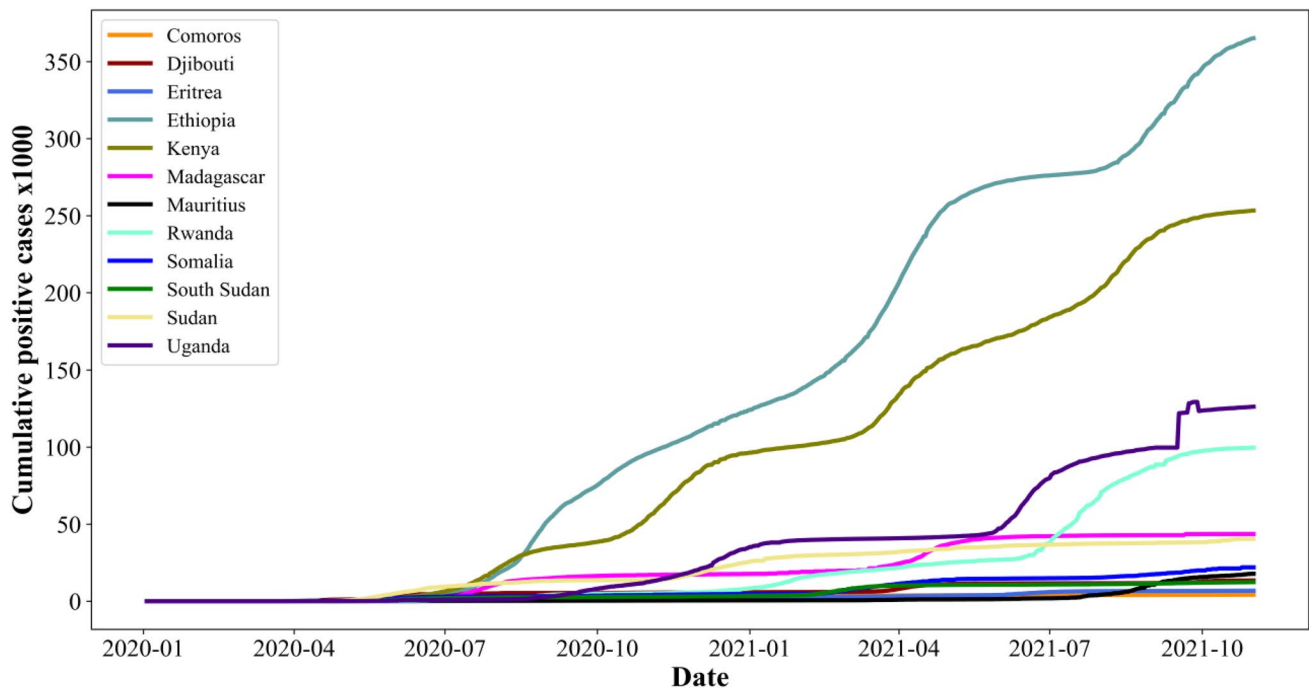


Fig. 16 Cumulative positive cases for Eastern Africa

the forecasting period in Cameroon, a decrease in the number of cases is expected to occur from 102,499 to 102,129. In the Central African region, Cameroon is the only country with an expected decrease in the number of cases.

The rest of the countries are expected to experience an increase in the number of cases as well. Cases are expected to increase from 35525 to 36522, 5069–5072, 13368–13508 and 3714–3717 in Gabon, Chad, Equatorial Guinea and São Tomé and Príncipe respectively. The largest increase in the number of cases in this region is expected to occur in Gabon, with an expected percentage increase of 2.81%.

Southern Africa

For the sake of clarity, countries from the Southern African region were separated into two plots showing the forecasted cumulative cases. This is because the number of cases in South Africa is so much bigger than in the rest of the countries in this region. This would result in plots for other countries being stacked together and not being able to be examined. In Fig. 21, a plot for the actual and forecasted cumulative cases for seven countries in the Southern African region is provided. According to this figure, it is observed that in Angola, the expected rate of increase in the cumulative positive cases is higher than in the rest of the countries. Angola is followed by Lesotho, with a moderate rate of increase in the number of cumulative cases. Lesotho is also followed by Botswana, with a small but notable increase in the cumulative cases. The rest of the countries, apart from

these three, are observed to maintain a constant number of cases with insignificant increases.

Figure 22 is a continuation of Fig. 21, which also shows a plot of the forecasted cases and actual cases for three countries in the Southern African region. In both Zambia and Mozambique, the number of cumulative cases is expected to maintain a constant course while a significant gradual increase in the number of cumulative cases is expected to occur. At the end of the forecasting period among the countries of this region, it is only in Mozambique that the number of COVID-19 cumulative cases is expected to decrease from 151292 to 151051. In the rest of the countries, the cases are expected to increase. In Angola, Botswana, Malawi, Namibia, South-Africa, Zambia, Eswatini, Lesotho, and Zimbabwe, the number of cases is expected to increase from 64433 to 76655, 186594 to 193024, 61796 to 63201, 128886 to 129401, 209734 to 210955, 46421 to 46874, 21635 to 24334, and 132977 to 133267, respectively. In this region, the highest percentage increase is observed to be 18.97% from Angola.

Eastern Africa

Forecasted cases in the Eastern region of Africa have been plotted in two separate graphs (Figs. 23 and 24). This made it possible to analyze and observe clearly the forecasted cases in all countries studied in this region.

In Fig. 23, a plot of the actual and forecasted cases for seven countries from the Eastern African region has been

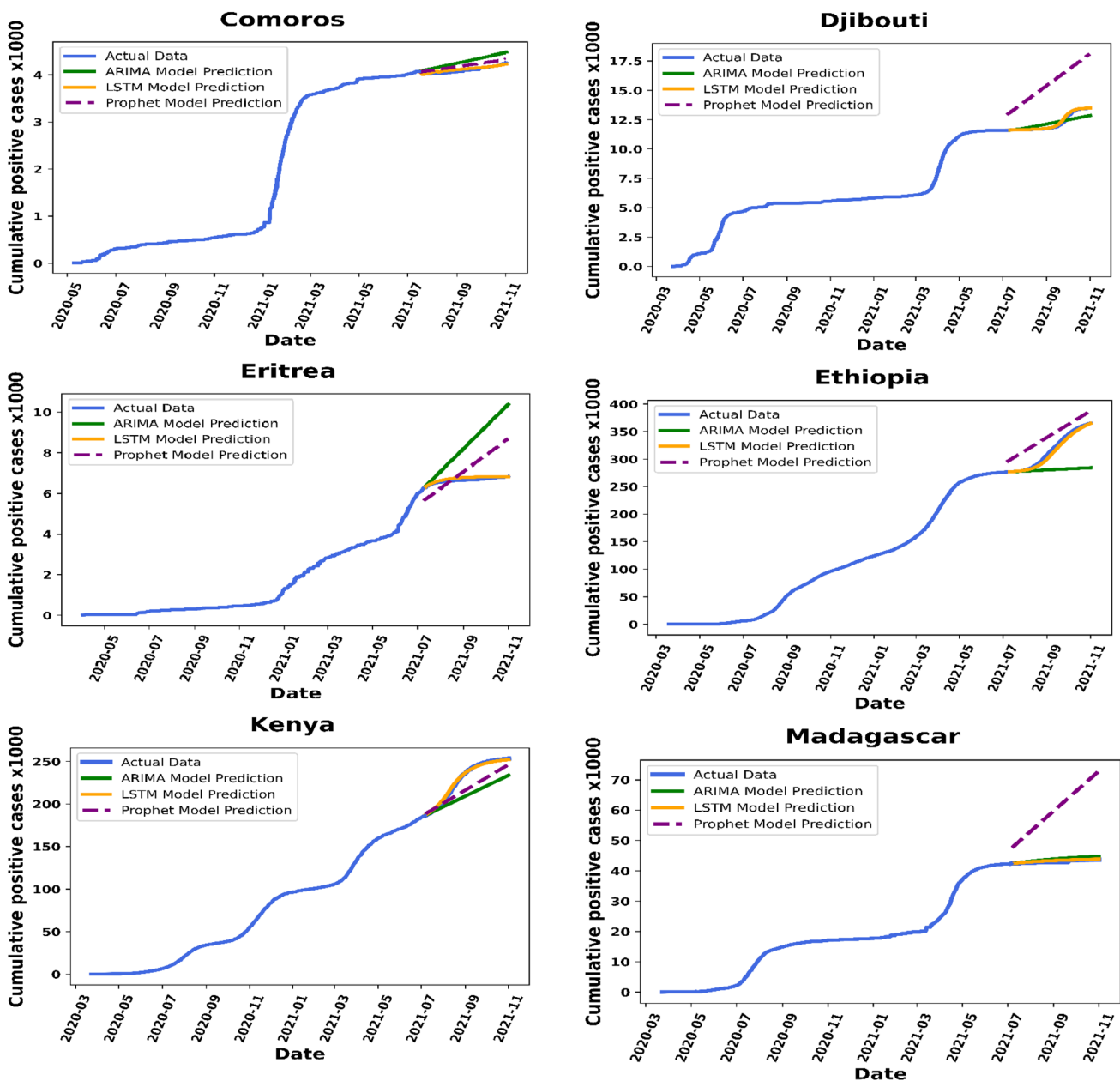


Fig. 17 Actual and predicted cumulative positive cases for Eastern Africa (a). Actual and predicted cumulative positive cases for Eastern Africa (b)

given. This forecast has been produced by the top performing model, which is the LSTM in most countries. According to this forecast, it is observed that in two countries, Rwanda and Mauritius, there is an expected gradual increase in the rate of increase of cumulative positive cases. Apart from these two countries and Djibouti, which are expected to have the same number of cases, the rest of the countries are expected to have small fluctuations in the number of cases.

In Fig. 24, five countries in the Eastern region of Africa have been shown with their respective COVID-19 cumulative positive cases. In Kenya, a constant number of cases is

expected, while in Ethiopia and Somalia, a notable increase is expected to occur. In both Uganda and Sudan, a small increase, which will be followed by a small but significant decrease, is expected to take place.

At the end of the forecasting period in Djibouti, the cases are expected to remain constant. The previous number in the original dataset was 13478 cases, which was expected to remain the same at the end of the forecast for Djibouti. In Eritrea, a small decrease is expected to happen from 6834 to 6820 cases. On the other hand, in the

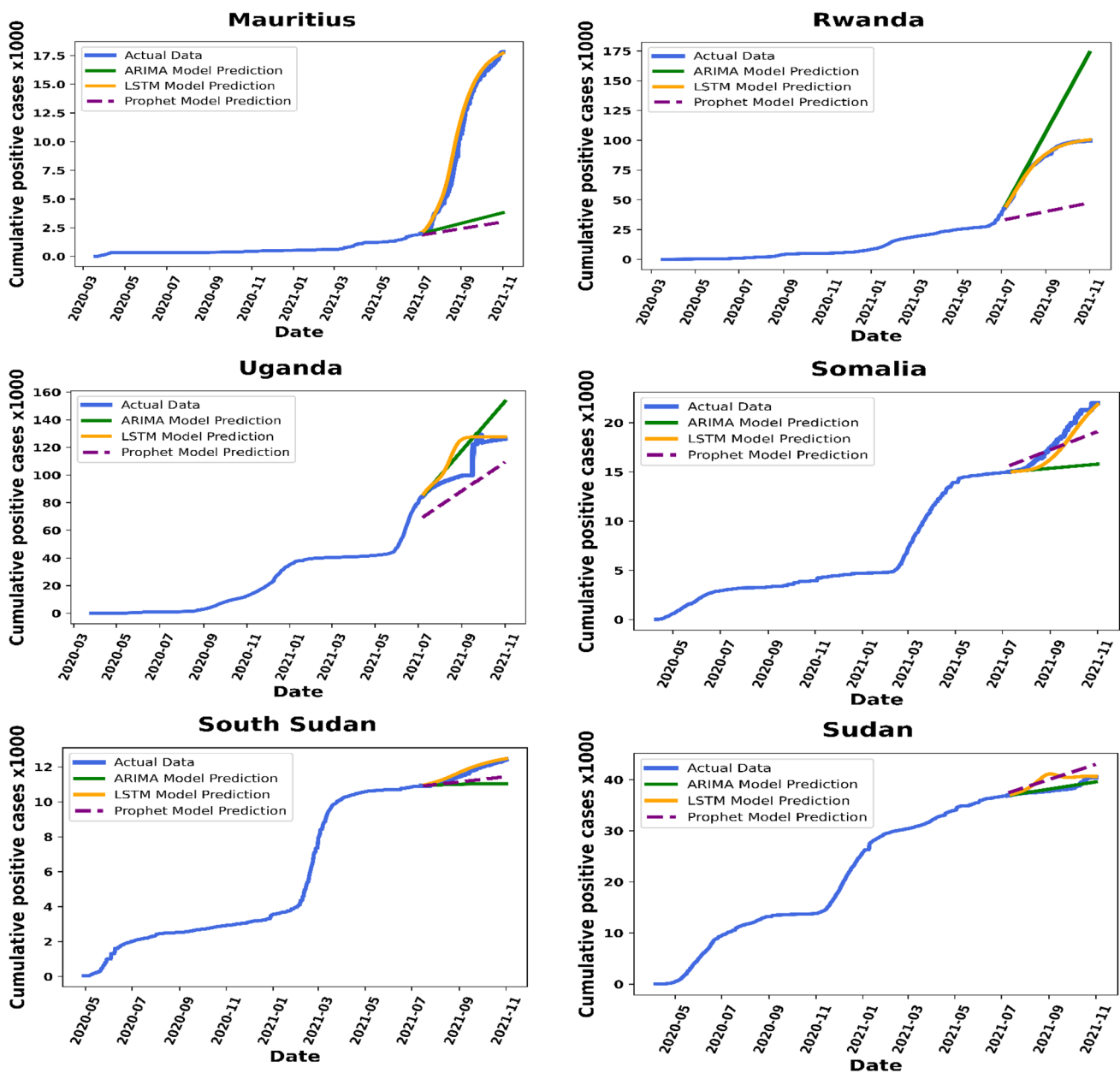


Fig. 17 (continued)

rest of the countries, an increase is expected by the end of the forecasting period. In these countries, Uganda, Sudan, Madagascar, Kenya, South Sudan, Somalia, Rwanda, Mauritius, Ethiopia, and Comoros, cases are expected to increase from 126236 to 127628, 40433 to 40598, 43626 to 44150, 253310 to 253901, 12410 to 12761, 21998 to 24356, 99698 to 102205, 17812 to 18297, 365167 to 377935, and 4259 to 4472, respectively. The highest expected increase in the cumulative number of cases is

observed to take place in Somalia, with a 10.72% expected percentage increase.

Western Africa

The forecasted cases from the Western African countries were grouped into two groups. As shown in Figs. 25 and 26, six countries were plotted together in each group. This was done in order to separate countries that have closer numbers of cumulative cases for a clear analysis of the results from the forecasting stage.

Table 5 Performance parameters of the models for Eastern Africa

Country	Model	PSNR	R value	NRMSE	SMAPE	RMSE	MSE	MAPE
Mauritius	ARIMA	-30.9283	-1.5722	0.5692	93.1564	8973.3930	80521781.9300	59.9527
	LSTM	-9.8988	0.9797	0.0506	7.7913	797.0358	635266.0665	8.3734
	Prophet	-31.3969	-1.8653	0.6007	104.3478	9470.7650	89695389.6900	65.2334
Madagascar	ARIMA	-10.3860	-2.8964	0.6832	1.7484	843.0209	710684.2378	1.7672
	LSTM	-2.1830	0.4107	0.2657	0.6019	327.8600	107492.1796	0.6048
	Prophet	-37.2296	<u>-1882.7825</u>	15.0212	32.4367	18536.1595	343589209.0000	39.7624
Kenya	ARIMA	-38.5762	0.1094	0.3237	8.2885	21644.5874	468488163.7000	7.8676
	LSTM	-16.2976	0.9947	0.0249	0.6621	1665.0265	2772313.2460	0.6640
	Prophet	-34.9861	0.6104	0.2141	5.2961	14316.8267	204971526.8000	5.1147
Rwanda	ARIMA	-42.6049	-3.1725	0.6113	22.6311	34418.0092	1184599357.0000	27.5678
	LSTM	-11.4368	<u>0.9968</u>	0.0169	0.9610	951.4349	905228.3689	0.9569
	Prophet	-44.8293	-5.9637	0.7897	66.9946	<u>44463.6562</u>	1977016723.0000	49.7882
Eritrea	ARIMA	-17.9013	-234.2806	3.3829	21.4673	2002.6522	4010615.8340	25.1679
	LSTM	8.3570	0.4431	0.1646	1.2999	97.4299	9492.5854	1.3106
	Prophet	-11.2635	-50.0280	1.5754	10.6952	932.6455	869827.6287	11.4391
Ethiopia	ARIMA	<u>-45.0386</u>	-1.0228	0.5137	11.0389	<u>45548.2859</u>	2074646348.0000	10.0965
	LSTM	-25.4272	0.9779	0.0537	1.1764	4763.2242	22688304.7800	1.1653
	Prophet	-40.5723	0.2767	0.3072	8.1788	27236.8169	741844194.8000	8.5723
Comoros	ARIMA	2.7683	-6.6139	0.7476	4.2035	<u>185.4063</u>	34375.4961	4.3021
	LSTM	<u>18.6599</u>	0.8039	0.1200	0.5967	<u>29.7540</u>	885.3005	0.5992
	Prophet	8.3879	-1.0876	0.3915	2.2780	<u>97.0829</u>	9425.0895	2.3056
Djibouti	ARIMA	-3.8103	0.6871	0.2116	2.6073	395.4184	156355.7111	2.5933
	LSTM	7.8871	0.9788	0.0550	0.4601	102.8451	10577.1146	<u>0.4634</u>
	Prophet	-22.6304	-22.8474	1.8469	23.6379	3451.9329	11915840.7500	27.0253
Uganda	ARIMA	-35.2449	0.0511	0.3294	10.4174	14749.7766	217555909.7000	11.2426
	LSTM	-34.4800	0.2044	0.3017	8.4224	<u>13506.4698</u>	182424726.5000	9.2481
	Prophet	-37.2039	-0.4897	0.4128	17.6531	18481.4583	341564300.9000	16.1231
South Sudan	ARIMA	-9.3188	-1.3764	0.4994	5.1883	745.5587	555857.7751	4.9893
	LSTM	3.9456	0.8879	0.1084	1.2819	161.9053	26213.3262	1.2915
	Prophet	-6.4872	-0.2381	0.3604	3.6874	538.1478	289603.0546	3.5850
Somalia	ARIMA	-22.5747	-1.0788	0.4898	15.0670	3429.8690	11764001.3600	13.4804
	LSTM	-10.2801	0.8774	0.1189	3.9871	832.8056	693565.1674	3.8825
	Prophet	-15.7176	0.5714	0.2224	6.7348	1557.4683	2425707.5060	6.5082
Sudan	ARIMA	-4.6451	0.8122	0.1243	0.8412	435.3047	189490.1818	0.8396
	LSTM	-17.3794	-2.5243	0.5385	3.9144	1885.8611	3556472.0890	4.0335
	Prophet	-18.5771	-3.6434	0.6181	5.0666	2164.6931	4685896.2170	5.2195

Seven metrics have been used to measure the model’s performance. Of these metrics, PSNR and R value, show the best performance for higher numbers. The rest of the metrics, apart from these, display the best performance for lower values. Basing on the PSNR metric, it is observed that the LSTM model obtained the highest value of 18.6599 from the Comoros. Using the same metrics, it is evident that in Ethiopia the worst performance, with the smallest PSNR value of -45.0386, was obtained by the ARIMA model. This implies that there is a significantly greater amount of noise in the predicted data by the ARIMA model, which directly reflects poor performance. This is also evident in Fig. 17, whereby there is a greater deviation of the predictions from the ARIMA with regard to the actual data. In Madagascar, the Prophet is observed to have the worst R value of -1882.7825. This implies that the predicted values by the Prophet model in Madagascar have the poorest correlation with the actual data in this country when compared with the other two model predictions and their respective actual values in both Madagascar and the other countries. Using the same metric, it is observed that the LSTM model in Rwanda obtained the best and highest value of 0.9968. This shows a strong correlation between the LSTM predicted cases and the actual cases, as further represented in Fig. R in Rwanda. When ranked by the RMSE ranges, it is observed that the best range is 29.7540–13506.4698, which is obtained by the LSTM, followed by the Prophet model with a range of 97.0829–44463.6562 and lastly, the ARIMA model with a range of 185.4063–45548.2859. It is clear from this metric that the most accurate model is the LSTM model in the Eastern region

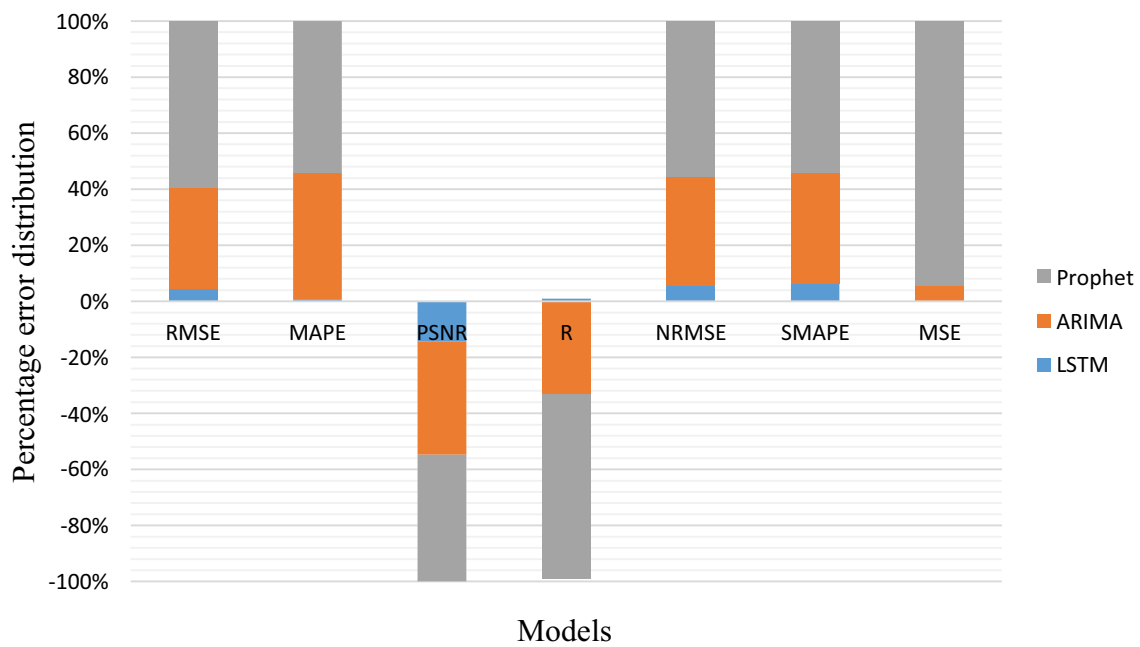


Fig. 18 Total error distribution of the models

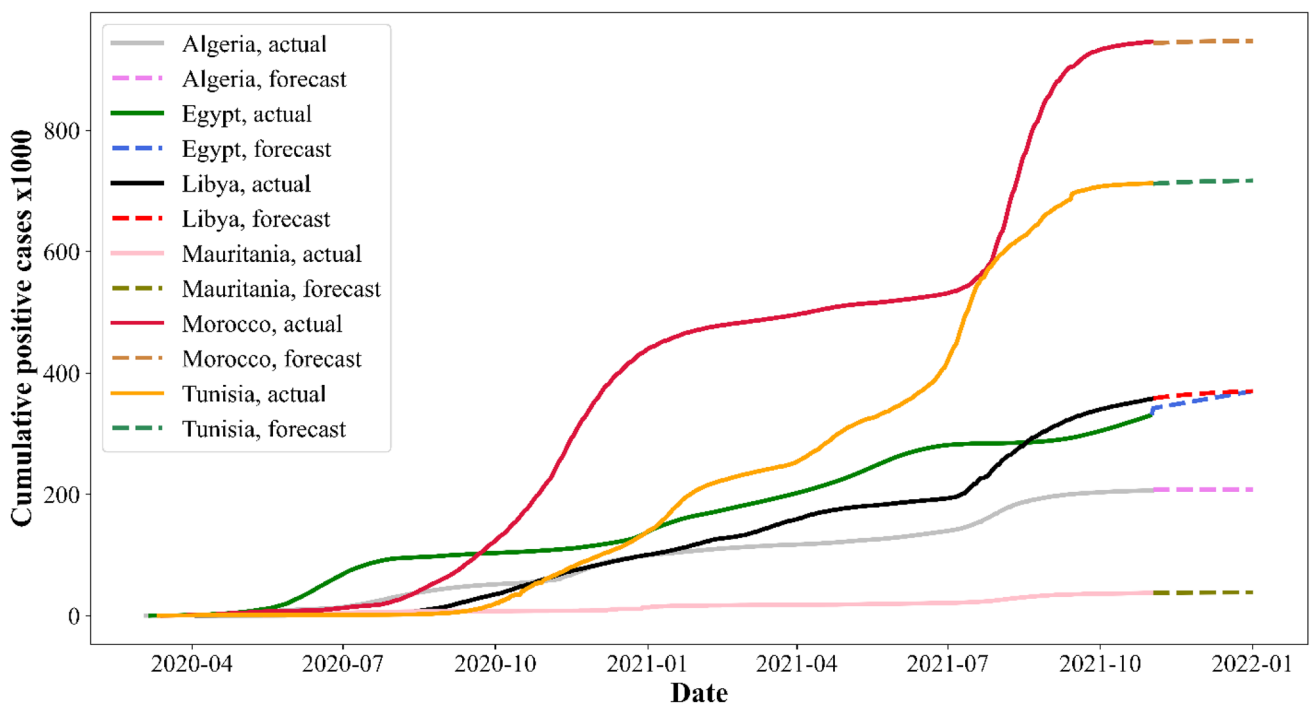


Fig. 19 Actual and forecasted COVID-19 cumulative positive cases for Northern Africa

In Fig. 25, six countries from the Western region of Africa, including their respective forecasted and actual cumulative cases, are shown. According to this figure, it is clear that the expected cases in Guinea will have a small increase, which is immediately followed by a generally constant number of

cases. In the rest of the five countries, a constant number of cases is expected, with small fluctuations by the end of the forecasting period. Since all countries in this figure maintained their respective fluctuation courses in the number of cases, it is evident that countries with a higher number of cases before

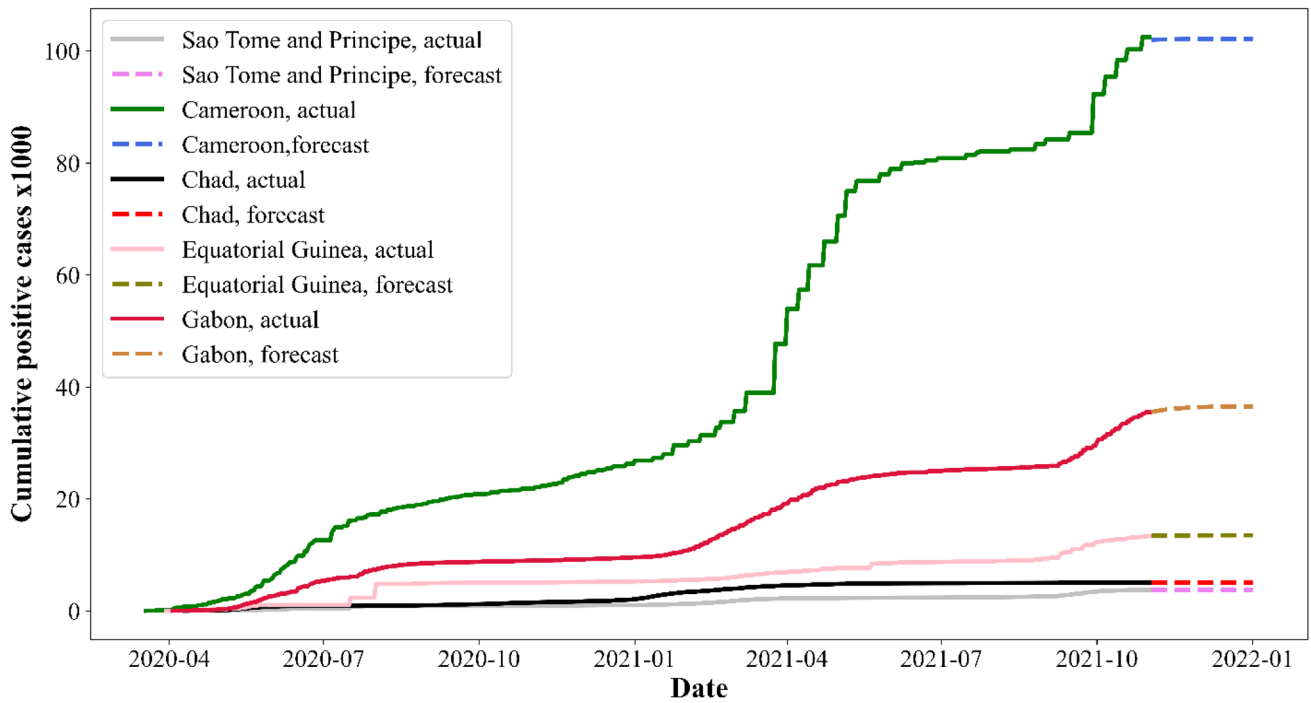


Fig. 20 Actual and forecasted COVID-19 cumulative positive cases for Central Africa

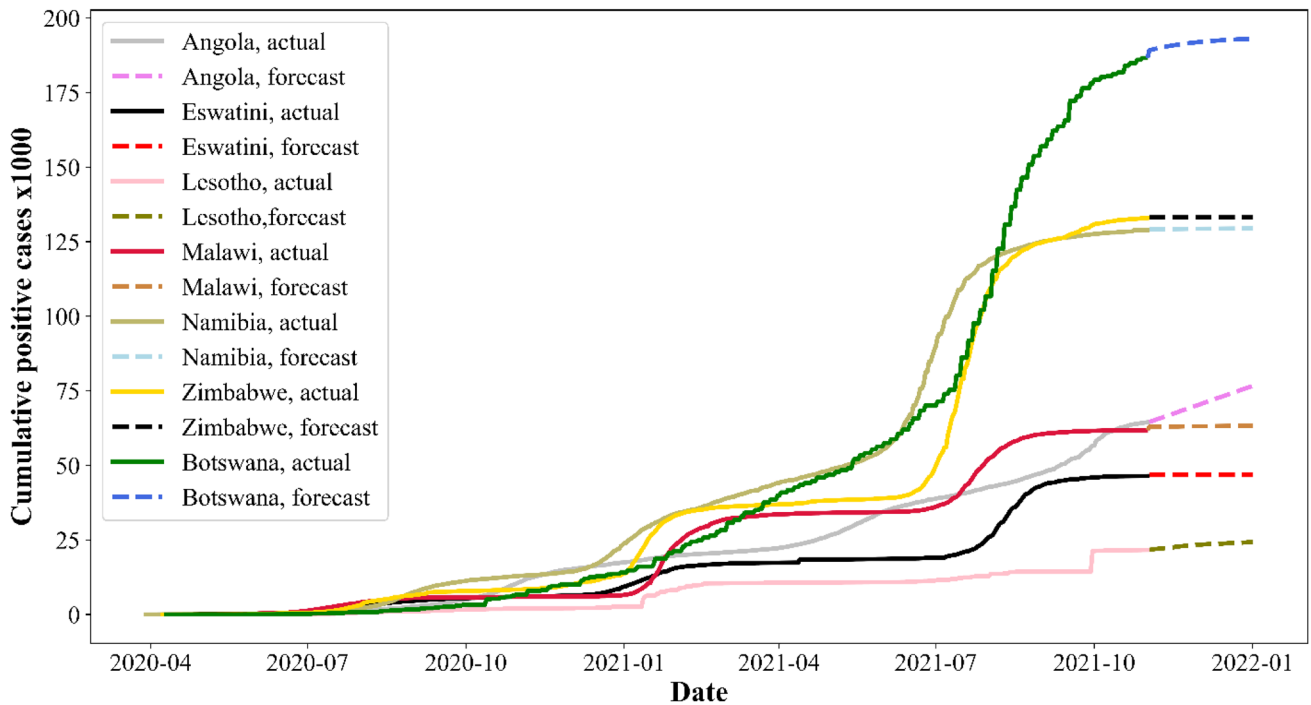


Fig. 21 Actual and forecasted COVID-19 cumulative positive cases for Southern Africa (a)

the forecasting processes maintained these higher numbers after forecasting. Countries such as Guinea, with the highest number of actual cases, are still expected to have the highest

number of forecasted cases, as depicted in Fig. 25. Since there is no expected significant decrease in the forecasted cases, this

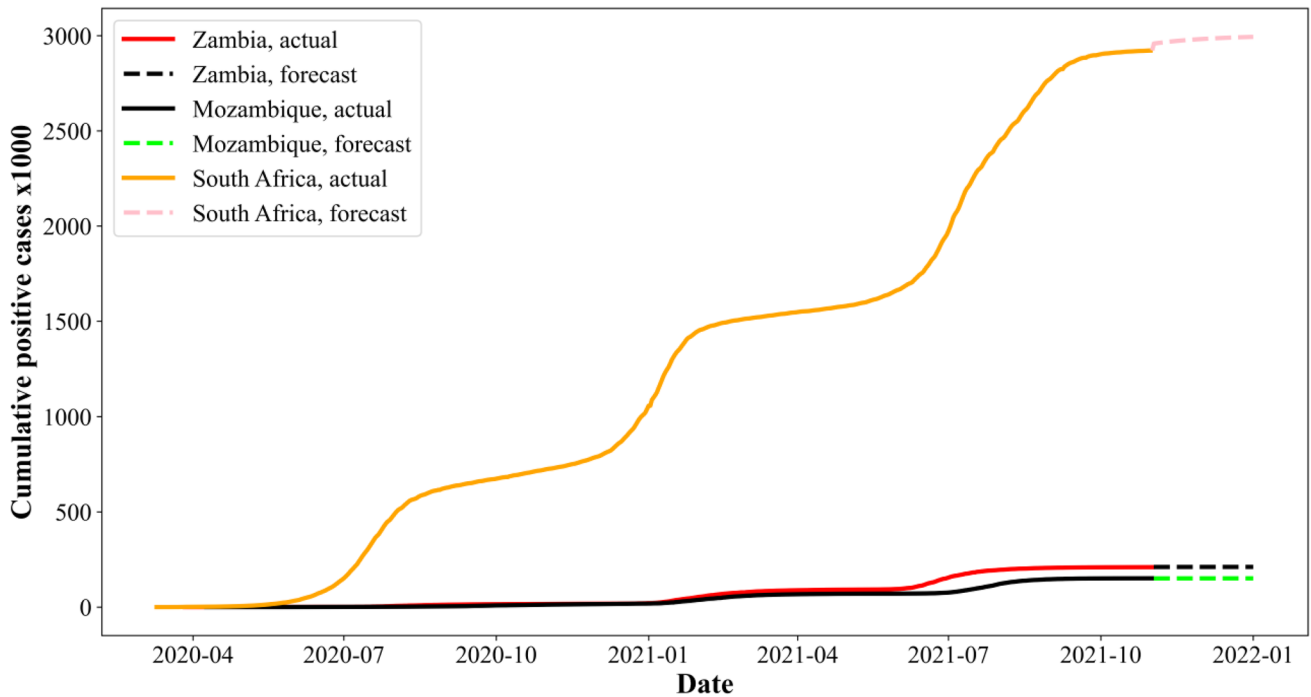


Fig. 22 Actual and forecasted COVID-19 cumulative positive cases for Southern Africa (b)

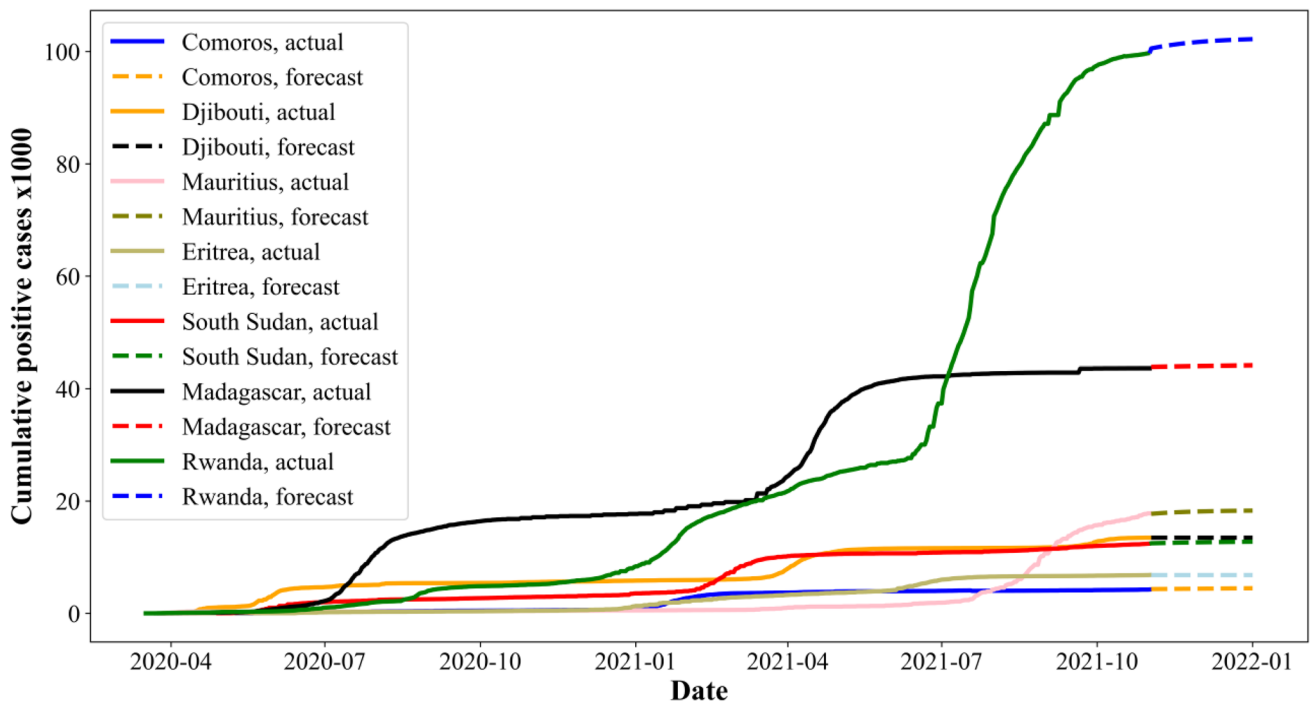


Fig. 23 Actual and forecasted COVID-19 cumulative positive cases for Eastern Africa (a)

still presents a great risk for the region if preemptive measures are not taken.

In Fig. 26, the rest of the six countries from the Western region of Africa are given, including the forecasted and actual cases in each state. A significant increase in the expected cases

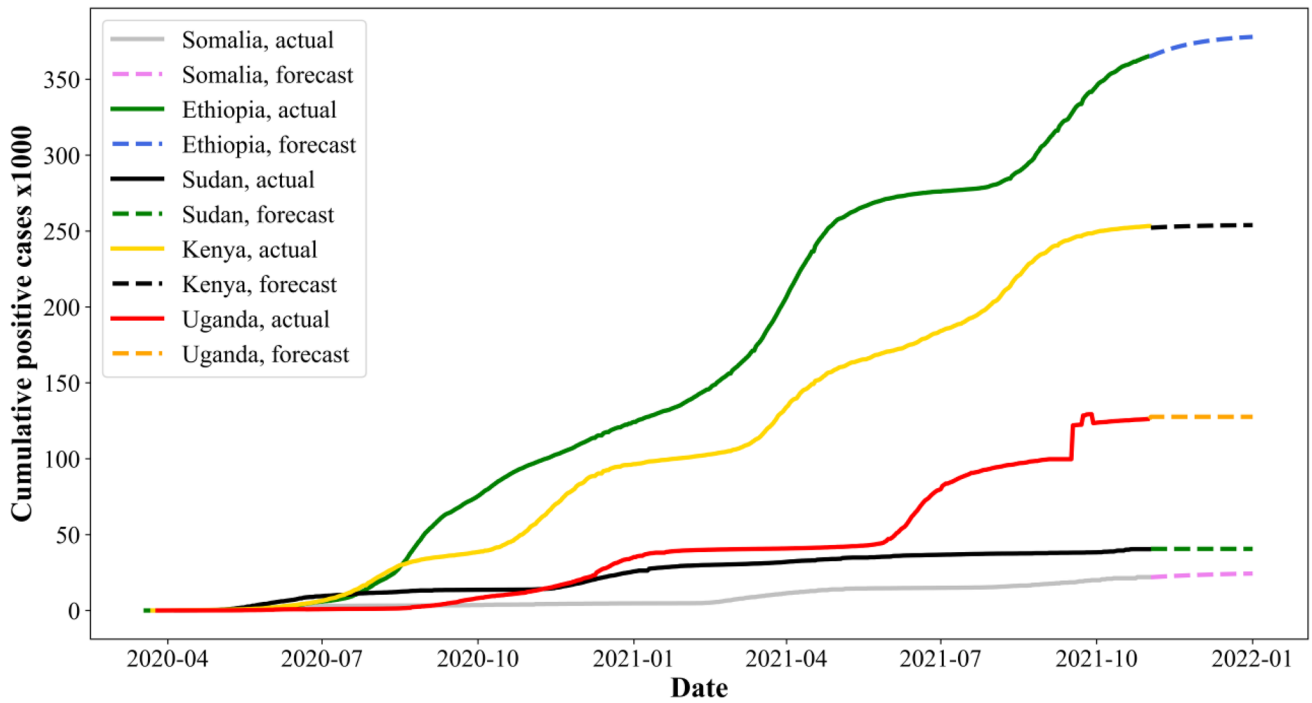


Fig. 24 Actual and forecasted COVID-19 cumulative positive cases for Eastern Africa (b)

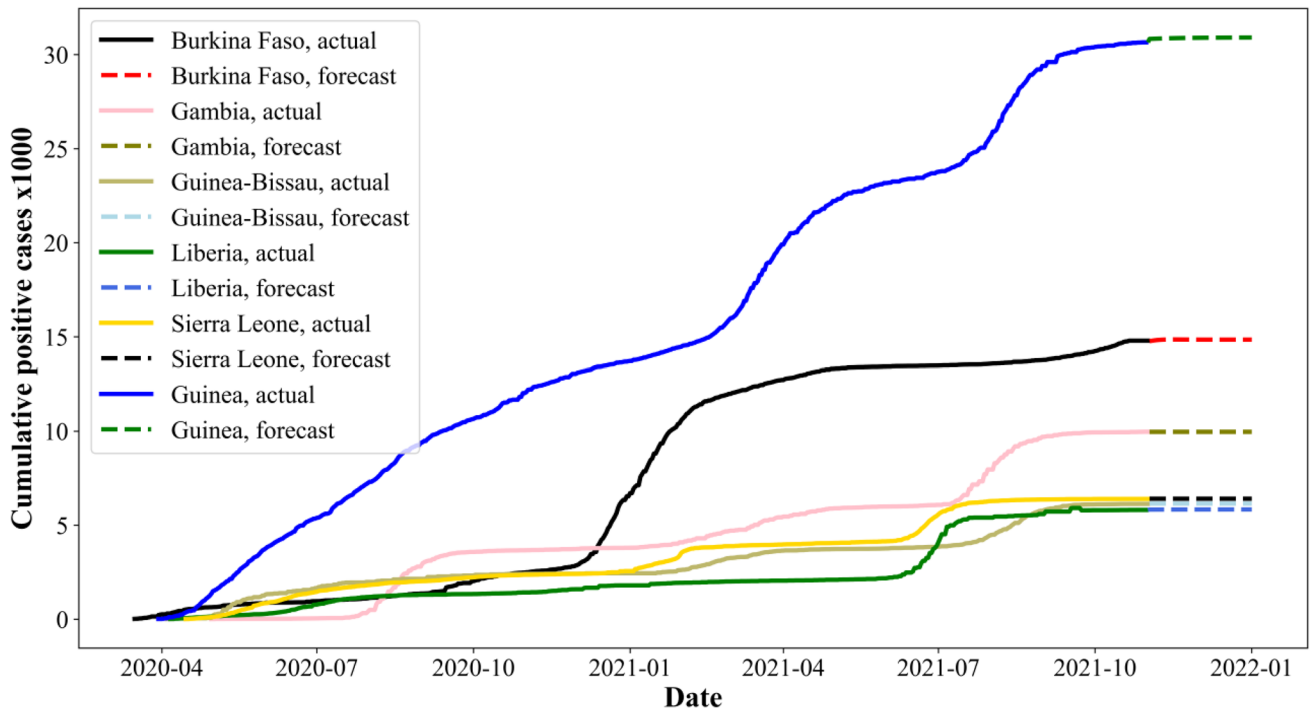


Fig. 25 Actual and forecasted COVID-19 cumulative positive cases for Western Africa (a)

in Mali is observed, while in the rest of the countries, a constant number of cases with minor fluctuations is observed.

In Gambia, a very small decrease is expected to occur in the forecasted number of cumulative cases at the end of the forecasting period. In this country, cases are forecasted to decrease

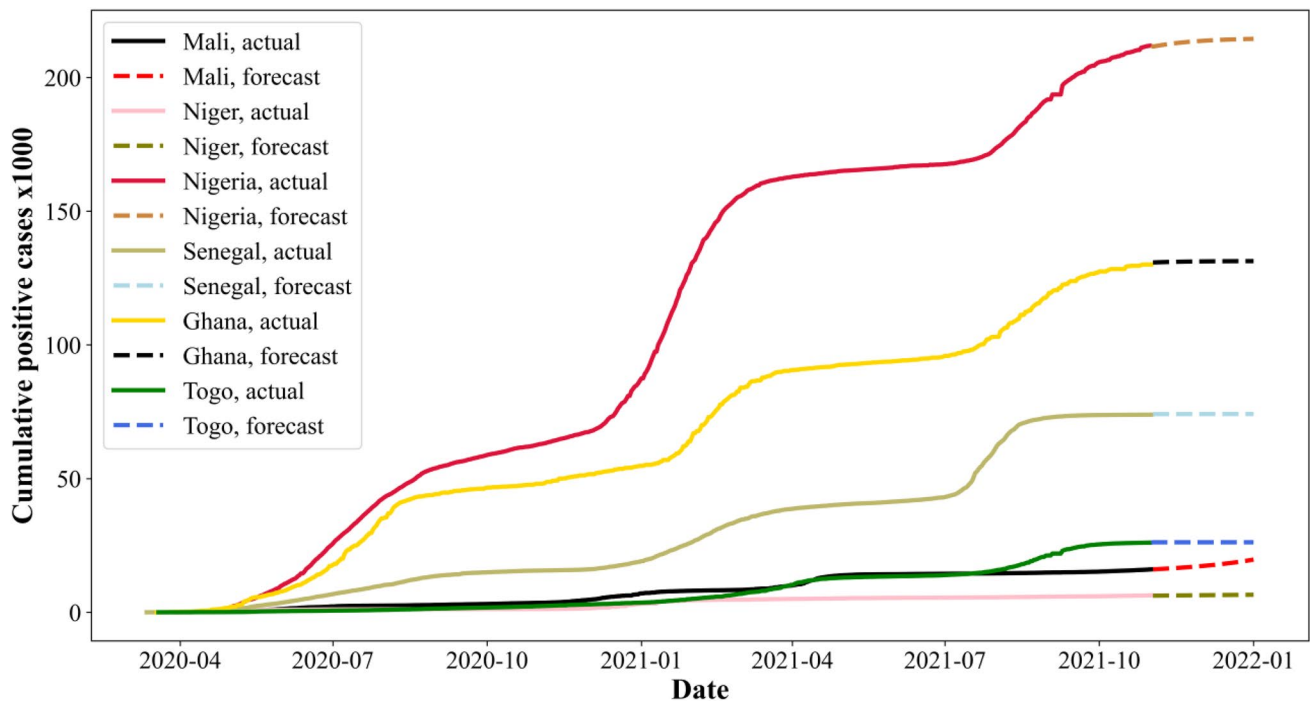


Fig. 26 Actual and forecasted COVID-19 cumulative positive cases for Western Africa (b)

from 9967 to 9964. In other countries in the Western region, apart from the Gambia, there is an expected increase in the number of cases. The COVID-19 cumulative positive cases are expected to increase from 6366 to 6565, 6134 to 6151, 30653 to 30909, 14793 to 14848, 26079 to 26195, 6398 to 6408, 73917 to 74171, 211961 to 214460, 16074 to 19734, 5815 to 5838, and 130077 to 131347 in countries such as Niger, Guinea-Bissau, Guinea, Burkina Faso, Togo, Sierra Leone, Senegal, Nigeria, Mali, Liberia, and Ghana, respectively. According to these results, it is observed that the highest expected percentage increase of 22.77% is expected to occur in Mali.

Conclusions and Suggestions

This study involves the forecasting of COVID-19 cumulative positive cases in countries from the five major regions of the African continent, which include the Northern, Eastern, Western, Central, and Southern regions. To contain and control the spread of the COVID-19 pandemic, there is a great need for strategies that can predict the

future course that the pandemic might take beforehand. This is because it would enable authorities to plan ahead of time and eventually allocate resources effectively and efficiently to more critical areas. There is a significant gap in the literature for studies that consider a continent's perspective, especially in Africa, when dealing with the forecasting of COVID-19. This study aimed at closing this gap by focusing on the forecasting and investigation of the expected future COVID-19 cumulative positive cases for a period of sixty-one days. From the forecasted values, this study aims to also identify the most critical states in each of the five major regions that have the highest expected percentage increase in the number of cases.

To achieve these objectives, this study employed both statistical and deep learning approaches, which consisted of three prediction models that were composed of the ARIMA, Prophet, and LSTM models. In a comparative analysis of the performance of these three models, seven performance metrics were used. These included the MSE, RMSE, MAPE, SMAPE, R2 score, NRMSE, and PSNR. The best-performing model was then selected to perform the forecasting of the future COVID-19 cumulative positive cases for a 61-day

perspective. In this study, the best-performing model was the LSTM model, while the worst-performing model was the Prophet model. The highest expected increase in the number of cases from the Western African region is expected to be 22.77% from Mali. On the other hand, in Angola, a country from the Southern region, the overall highest expected increase is 18.97%. The highest expected increase from the Northern region is expected to take place in Egypt, at 11.83%. In the Eastern region, the highest increase of 10.72% is expected to occur in Somalia. Lastly, from the Central African region, the highest expected increase is 2.81% in Gabon. There is a need for studies that consider the influence of population demographics on the spread of COVID-19

Author Contributions All authors have participated in (a) the conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. This manuscript has not been submitted to, nor is under review for, another journal or other publishing venue.

Funding No funding was granted to this research work.

Data availability The dataset used as case study was obtained and retrieved on October 1, 2021, from <https://data.humdata.org/dataset/africa-covid19-infected>

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

- Abdulmajeed K, Adeleke M, Popoola L. Online forecasting of COVID-19 cases in nigeria using limited data. *Data Brief*. 2020;30:105683. <https://doi.org/10.1016/j.dib.2020.105683>.
- Africa: COVID-19 Infections (National)—Humanitarian Data Exchange. (2021). Humanitarian Data Exchange. Retrieved October 1, 2021, from <https://data.humdata.org/dataset/africa-covid19-infected>
- African Countries by Population (2022)—Worldometer. (2021). Worldometer. Retrieved November 1, 2021, from <https://www.worldometers.info/population/countries-in-africa-by-population/>
- Archived: WHO Timeline—COVID-19. (2020, April 28). WHO Timeline. <https://www.who.int/news/item/27-04-2020-who-timeline---covid-19>
- Gebretensae YA, Asmelash D. Trend analysis and forecasting the spread of COVID-19 pandemic in Ethiopia using Box-Jenkins modeling procedure. *Int J Gen Med*. 2021;14:1485–98. <https://doi.org/10.2147/ijgm.s306250>.
- Hssayeni MD, Chala A, Dev R, Xu L, Shaw J, Furht B, Ghoraani B. The forecast of COVID-19 spread risk at the county level. *J Big Data*. 2021. <https://doi.org/10.1186/s40537-021-00491-1>.
- Hu, Z. (2020, February 17). Artificial Intelligence Forecasting of COVID-19 in China. *arXiv.Org*. <https://arxiv.org/abs/2002.07112>
- Huang CJ, Chen YH, Ma Y, Kuo PH. Multiple-input deep convolutional neural network model for COVID-19 forecasting in China. *medRxiv*. 2020. <https://doi.org/10.1101/2020.03.23.20041608>.
- Huang Y, Xu C, Ji M, Xiang W, He D. Medical service demand forecasting using a hybrid model based on ARIMA and self-adaptive filtering method. *BMC Med Inform Decis Mak*. 2020. <https://doi.org/10.1186/s12911-020-01256-1>.
- Kırbaş S, Sözen A, Tuncer AD, Kazancıoğlu FI. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. *Chaos, Solitons Fractals*. 2020;138:110015. <https://doi.org/10.1016/j.chaos.2020.110015>.
- Le H, Lee J. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water*. 2019;11(7):1387. <https://doi.org/10.3390/w11071387>.
- Marzouk M, Elshaboury N, Abdel-Latif A, Azab S. Deep learning model for forecasting COVID-19 outbreak in Egypt. *Process Saf Environ Prot*. 2021;153:363–75. <https://doi.org/10.1016/j.psep.2021.07.034>.
- Noureen S, Atique S, Roy V, Bayne S. Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: a case study of small scale agricultural load. In: 2019 IEEE 62nd International midwest symposium on circuits and systems (MWSCAS), Dallas, TX, USA, 2019, p. 521–524. <https://doi.org/10.1109/MWSCAS.2019.8885349>
- Pal R, Sekh AA, Kar S, Prasad DK. Neural network based country wise risk prediction of COVID-19. *Appl Sci*. 2020;10(18):6448. <https://doi.org/10.3390/app10186448>
- Ribeiro MH, da Silva RG, Mariani VC, dos Santos Coelho L. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos, Solitons Fractals*. 2020;135:109853. <https://doi.org/10.1016/j.chaos.2020.109853>
- Shastri S, Singh K, Kumar S, Kour P, Mansotra V. Time series forecasting of COVID-19 using deep learning models: India-USA comparative case study. *Chaos, Solitons Fractals*. 2020;140:110227. <https://doi.org/10.1016/j.chaos.2020.110227>.
- Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, Nanda C, Sharma S, Sharma YD, Rabaan AA, Rahmani J, Kumar P. Prediction of the COVID-19 Pandemic for the Top 15 affected countries: advanced autoregressive integrated moving average (ARIMA) Model. *JMIR Pub Health Surveil*. 2020;6(2):e19115. <https://doi.org/10.2196/19115>.
- Taylor SJ, Letham B. Forecasting at scale. *Am Stat*. 2017. <https://doi.org/10.7287/peerj.preprints.3190v2>.
- Wang P, Zheng X, Li J, Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons Fractals*. 2020;139:110058. <https://doi.org/10.1016/j.chaos.2020.110058>.
- Wang YW, Shen ZZ, Jiang Y. Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. *PLoS ONE*. 2018;13(9):e0201987. <https://doi.org/10.1371/journal.pone.0201987>.
- Wang Y, Wang J, Zhao G, Dong Y. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China. *Energy Policy*. 2012;48:284–94. <https://doi.org/10.1016/j.enpol.2012.05.026>.
- W.H.O. (2020, January 10). Coronavirus. WHO. https://www.who.int/health-topics/coronavirus#tab=tab_1
- Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, Liu P, Cao X, Gao Z, Mai Z, Liang J, Liu X, Li S, Li Y, Ye F, Guan W, Yang Y, Li F, Luo S, He J. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health

- interventions. *J Thorac Dis.* 2020;12(3):165–74. <https://doi.org/10.21037/jtd.2020.02.64>.
24. Yu CS, Chang SS, Chang TH, Wu JL, Lin YJ, Chien HF, Chen RJ. A COVID-19 pandemic artificial intelligence-based system with deep learning forecasting and automatic statistical data acquisition: development and implementation study. *J Med Internet Res.* 2021;23(5):e27806. <https://doi.org/10.2196/27806>.
 25. Zeroual A, Harrou F, Dairi A, Sun Y. Deep learning methods for forecasting COVID-19 time-series data: a comparative study. *Chaos, Solitons Fractals.* 2020;140:110121. <https://doi.org/10.1016/j.chaos.2020.110121>.
 26. Zhang X, Yu Y, Xiong F, Luo L. Prediction of daily blood sampling room visits based on ARIMA and SES model. *Comput Math Methods Med.* 2020;2020:1–11. <https://doi.org/10.1155/2020/1720134>.
 27. Zoabi Y, Deri-Rozov S, Shomron N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *Npj Digital Med.* 2021. <https://doi.org/10.1038/s41746-020-00372-6>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.