Clinical Case Reports International

Research Article
Published: 21 Nov, 2022

# Implementing Machine Learning Approach for the Prediction of Cumulative COVID-19 Cases in Nigeria

*Abdallahi J**

*Department of Civil Engineering, Baze University, Abuja, Nigeria*

## Abstract

Since December 2019, Coronavirus Diseases (COVID-19) has been a major topic of interest to the researchers, authorities and health officials due to its devastating nature and risk associated with its contraction and spread. Developing countries including Africa are at risks of getting more infected due to less awareness and ill-equipped health facilities. For such countries, accurate prediction of the COVID-19 cases could help the policy makers in decision making and proper utilization of the available resources. Machine Learning (ML) approaches are excellent tools for making predictions especially for an uncertain and stochastic phenomenon such as COVID-19. Therefore, Artificial Neural Network (ANN) as ML tool was employed to predict cumulative COVID-19 cases in Nigeria. In this way, data including daily and cumulative cases as well as daily and cumulative mortality rate were collected from February 2020 till December 2021. Five different models were developed using varied input combinations. Determination Coefficient (DC) and Root Mean Square Error (RMSE) were employed as performance metrics. The results showed all the developed models produced good prediction skill but M4 with DC=0.9996 and RMSE=0.0017 in the validation step led to better performance. The general results of this study emphasize the importance of using ANN for the prediction of COVID-19 pandemic in Nigeria.

**Keywords: COVID-19; Machine learning; Pandemic; Nigeria; Africa**

**OPEN ACCESS**

***Correspondence:***
*Jazuli Abdallahi, Department of Civil Engineering, Baze University, Abuja, Nigeria,*
*E-mail: jazuli.abdullahi@bazeuniversity.edu.ng*

**Received Date**: *26 Oct 2022*
**Accepted Date:** *17 Nov 2022*
**Published Date:** *21 Nov 2022*

***Citation:***
*Abdallahi J. Implementing Machine Learning Approach for the Prediction of Cumulative COVID-19 Cases in Nigeria. Clin Case Rep Int. 2022; 6: 1424.*



## Introduction

The total confirmed cases of COVID-19 have reached 396,558,014 as of February 9th, 2022 according to the World Health Organization. The pandemic has become a serious health challenge worldwide. The call for artificial intelligence integration in the fight against the deadly virus was made by the White House in collaboration with tech companies and research institute as well. Personalized preventive measures can significantly help from precise characterization of the population based on classified COVID-19 vulnerability. The initial examination that elderly populace has an elevated risk to coronavirus disease 2019 is challenged by a different study that more and more youthful adults experienced chronic COVID-19 signs, showing an emergency need of abroad risk assessment based on physiological features and personalized genetic.

Machine Learning (ML) algorithms have the capacity to integrate and analyze huge-scale data of coronavirus disease 2019 patients, which can help in understanding the viral spread pattern, develop new efficient therapeutics strategies, further improve testing/diagnostic accuracy and speed and potentially explore the most susceptible persons based on personalized genetic as well as physiological features [1]. Interestingly, within a limited time period since the pandemic started, ML methods have been utilized in taxonomic categorization of coronavirus disease 2019 genomes, survival projection of chronic COVID-19 patients [2]. Namasudra et al. [3] employed Artificial Neural Network (ANN) to predict COVID-19 cases. Nabi et al. [4] employed ANN and other ML tools for the prediction of COVID-19 diseases outbreak. De Araújo Morais and da Silva Gomes [5] applied ANN and other prediction model for the forecasting of daily COVID-19 cases in the world.

Despite been amongst the highest recorded number of cases and death in Africa, studies of COVID-19 cases in Nigeria are very limited. Therefore, owing to its robustness in dealing with the uncertain aspect of COVID-19, Artificial Neural Network (ANN) was employed to predict cumulative COVID-19 cases in Nigeria using several input combinations.

## Materials and Methods

### Study country

Nigeria is a tropical zone region located in west Africa with 923,770 km$^2$ of total area. A distance of 1,050 km separated southern Nigeria from northern Nigeria while 1,150 km is the west to east optimum distance. Benin, Cameroon, Niger and Chad are some of the African countries that share border with Nigeria [6]. Figure 1 shows the study country.

The data used in this study were obtained and can be downloaded from https://github.com/owid/covid-19-data/tree/master/public/data/. Table 1 shows statistical description of the data.

It can be seen from Table 1 that the COVID-19 cases in Nigeria within the selected period were up to a maximum 6,158 in a single day and cumulatively, the cases were up to 241,513 (as of February 28th, 2020 to December 31st, 2021). A maximum of 93 daily death can also be seen while cumulative death stands at 3,030.

### Data normalization and performance criteria

To ensure all variables have the same treatment and to eliminate their dimensional discrepancy, data Normalization is usually applied for AI-based modeling [7,8]. Therefore, the data were normalized to fall between 0 and 1 as:

$$D_{norm} = \frac{D_i - D_{\min}}{D_{\max} - D_{\min}} \qquad (1)$$

Where $D_{norm}$, $D_i$, $D_{min}$ and $D_{max}$ are the normalized data, ith data value, minimum and maximum data values, respectively.

To determine the performance of the developed models, Determination Coefficient (DC) and Root Mean Square Error (RMSE) were employed [9], given by:

$$DC = 1 - \frac{\sum_{i=1}^{N}\left(CC_i - \widehat{CC_i}\right)^2}{\sum_{i=1}^{N}\left(CC_i - \overline{CC_i}\right)^2} \qquad (2)$$

$$RMSE = \frac{\sum_{i=1}^{N}\left(CC_i - \widehat{CC_i}\right)^2}{N} \qquad (3)$$

Where $CC_i$ is cumulative case of the ith value, $N$, $\widehat{CC_i}$, and $\overline{CC_i}$ are the number of observations, predicted values and mean of the observed values, respectively. The DC ranges between - ∞ to 1 and RMSE between 0 to ∞ with DC towards 1 and RMSE close to 0 imply high efficiency.

### Artificial neural network (ANN)

ANN provides a determined approach in dealing with non-linear, noisy, and dynamic data, more specifically when the physical fundamental relationships are not completely known. Detail information regarding ANN can be found in study [10].

## Results and Discussion

In this study, the potential of ANN in predicting cumulative COVID-19 cases in Nigeria was investigated. For machine learning applications, selection of the most appropriate inputs are significant in determining the most efficient prediction performance, as incorrect inputs may results in error and performance deficiency. Recent studies including Ibrahim et al. [8] show that using COVID-19 cases at previous time step may lead to accurate prediction. However, the daily COVID-19 cases as well as the daily and cumulative death could have a significant impact on cumulative cases. Therefore, in this study, 5 different models were developed using different input combination to determine the highest performance. Table 2 shows the developed models and their training and validation percentages.

For the ANN model in both stations, FFBP was used for the model training using Levenberg Marquardt (LM) algorithm. Single hidden layer was used and via trial and error, the best number of hidden layer neurons was selected. Table 3 shows the results of the developed models. The structure x-y-z implies number of inputs, hidden layer neurons and output layer, respectively.

As shown in Table 3, different inputs combination led to different performance. These results show that using COVID-19 cumulative cases at previous time and other data could be employed sufficiently for COVID-19 case prediction. In the training phase, it can be seen that M2 ($CC_{(t-4)}$, $CC_{(t-5)}$, $CC_{(t-6)}$, $CC_{(t-7)}$), M4 (New case, new death, $CC_{(t-1)}$, $CC_{(t-2)}$) and M5 (New case, new death, $CC_{(t-3)}$, $CC_{(t-4)}$) provided the best prediction performance in terms of DC with 0.9999 value. However, with respect to RMSE, M4 (New case, new death, $CC_{(t-1)}$, $CC_{(t-2)}$) produced better skill than all other models with 0.0004 value. Therefore, this demonstrated the rationality of combining daily cases and death with previous cumulative cases to enhance prediction performance.

In the validation phase in Table 3, it can be seen that despite the nonlinear aspect of ANN and its ability to deal with uncertain stochastic phenomena, the performance of ANN decreases in the validation phase. Similar to the training phase, M4 in the validation phase led to best performance with lowest error RMSE=0.0017 and highest goodness of fit DC=0.9996. Comparing the results of this study with that of Ibrahim et al. [8] it can be seen that despite having

**Table 1:** Statistical description of the used data.

| Data | Period | Minimum | Maximum | St. deviation |
|---|---|---|---|---|
| New case | 28-02-2020 - 31-12-2021 | 0 | 6158 | 460.17 |
| Cumulative case | 28-02-2020 - 31-12-2021 | 1 | 241513 | 75734.58 |
| New death | 28-02-2020 - 31-12-2021 | 0 | 93 | 6.97 |
| Cumulative death | 28-02-2020 - 31-12-2021 | 0 | 3030 | 914.75 |

**Table 2:** Models development with their training and validation subsamples.

| Model | Inputs | Training subsample | Validation subsample |
|---|---|---|---|
| M1 | $CC_{(t-1)}$, $CC_{(t-2)}$, $CC_{(t-3)}$, $CC_{(t-4)}$ | 75% (471) | 25% (202) |
| M2 | $CC_{(t-4)}$, $CC_{(t-5)}$, $CC_{(t-6)}$, $CC_{(t-7)}$ | 75% (471) | 25% (202) |
| M3 | New case, new death, cum. death | 75% (471) | 25% (202) |
| M4 | New case, new death, $CC_{(t-1)}$, $CC_{(t-2)}$ | 75% (471) | 25% (202) |
| M5 | New case, new death, $CC_{(t-3)}$, $CC_{(t-4)}$ | 75% (471) | 25% (202) |

$CC_{(t-n)}$ represents cumulative cases at previous time step

**Table 3:** Results of the developed models.

| Model | Structure | DC | RMSE | DC | RMSE |
|---|---|---|---|---|---|
| M1 | 4-10-1 | 0.9978 | 0.004 | 0.9637 | 0.0473 |
| M2 | 4-7-1 | 0.9999 | 0.0014 | 0.9993 | 0.0023 |
| M3 | 3-8-1 | 0.9998 | 0.0035 | 0.9836 | 0.0109 |
| M4 | 4-11-1 | 0.9999 | 0.0004 | 0.9996 | 0.0017 |
| M5 | 4-9-1 | 0.9999 | 0.0011 | 0.9991 | 0.0026 |

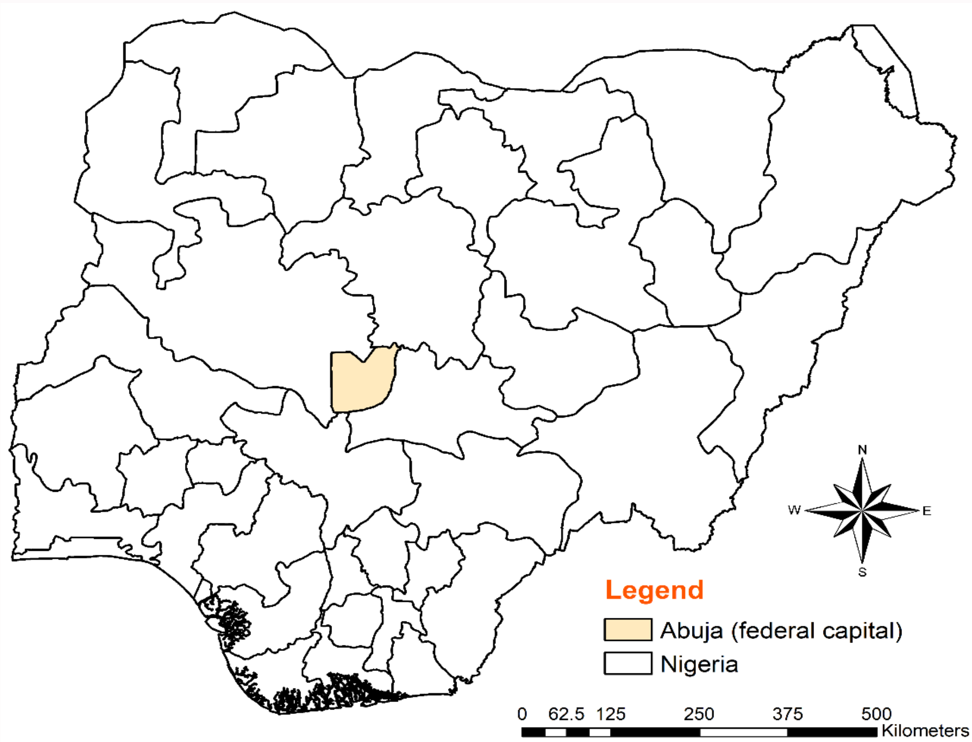


**Figure 1:** Study area.

same study location (Nigeria), performance of the ANN model in this study is superior. Many issues could lead to the discrepancy in the results. (i) The duration of the data could play a major role in the performance as this study covered entire period from the day of first cases in Nigeria (28-02-2020) to where decline in number of cases started. This implies that the study covered both the peak and least periods of the pandemic which might results in proper simulation of the uncertainty surrounding the system. (ii) The study by Ibrahim et al. [8] employed only the previous time step for their prediction while this study shows that addition of the cases and death input data could enhance prediction of the pandemic. Figure 2 shows the graphical comparison of the developed models.

The radar chart shown in Figure 2 compares the performance of the 5 developed models based on DC and RMSE. From the chart, as goodness of fit measure, the more open and wider the chart is, the better performance based on DC. While narrower chart signifies the most efficient performance with respect to RMSE. It can be seen from Figure 2 that all models led to performance close to 1. Similarly, the models performed excellent with a relatively small error for RMSE. This is an indication that once ANN model is well developed, the performance indicators irrespective of whether graphical or tabular approach is used the results would be accurate. Figure 3 compares the performances of the developed models against the observed data.

For violin plots shown in Figure 3, the internal line represents the range of the values while the white central dot signifies mean of the data values. The more similarity and resemblance a model have with the observed data, the more efficient COVID-19 prediction, thus, as seen in Figure 3, all the models have a close resemblance with the observed data, which indicates the accurate performance by all models.

## Conclusion

In this study, Machine Learning (ML) approach was employed to predict cumulative COVID-19 cases in Nigeria. For this purpose, cumulative data at previous time step (up to 7-day period) as well as daily cases, daily death and cumulative death were used as inputs for 5 different models based on Artificial Neural Network (ANN).

The results showed that with unique inputs of the applied models, different results were achieved. The obtained results demonstrated that M4 (New case, new death, $CC_{(t-1)}$, $CC_{(t-2)}$) provided the most efficient performance in terms of DC and RMSE. This emphasizes the need of combining both cases at previous step and other COVID-19 data for successful prediction of cumulative cases in Nigeria. The results indicated that provided an efficient input selection approach and sufficient quality data are achieved, a successful prediction of COVID-19 cases in Nigeria could be achieved using ANN. The results

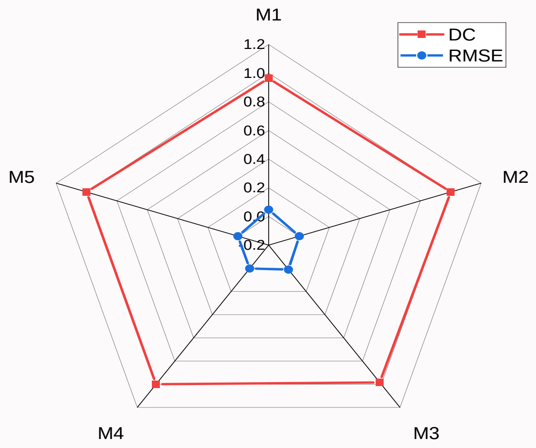


**Figure 2:** Performance comparison of all developed models.

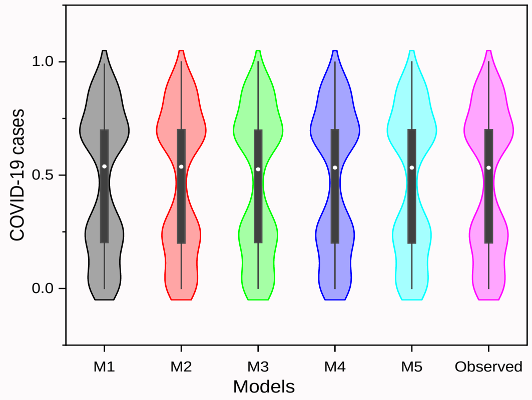


**Figure 3:** Models performance comparison using violin plots.

of this study could be useful in policy making and in prediction of other COVID related diseases. Further studies should involve other machine learning approaches and other countries in Africa.

## References

1. Usaini S, Hassan AS, Garba SM, Lubuma JS. Modeling the transmission dynamics of the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) with latent immigrants. J Interdiscip Math. 2019;22(6):903-30.
2. Lippi G, Henry BM. Chronic obstructive pulmonary disease is associated with severe coronavirus disease 2019 (COVID-19). Respir Med. 2020;167:105941.
3. Namasudra S, Dhamodharavadhani S, Rathipriya R. Nonlinear neural network-based forecasting model for predicting COVID-19 cases. Neural Process Lett. 2021;1-21.
4. De Araújo Morais LR, da Silva Gomes GS. Forecasting daily COVID-19 cases in the world with a hybrid ARIMA and neural network model. Appl Soft Comput. 2022;126:109315.
5. Nabi KN, Tahmid MT, Rafi A, Kader ME, Haider MA. Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks. Results Phys. 2021;24:104137.
6. Abdullahi J, Elkiran G, Aslanova F. Virtual water trade in the semi-arid regions of Nigeria. In IOP Conference Series: Earth and Environmental Science. IOP Publishing. 2020;614(1):012074.
7. Abdullahi J, Elkiran G, Nourani V. Application of artificial neural network to predict reference evapotranspiration in Famagusta, North Cyprus. In 11th International Scientific Conference on Production Engineering Development and Modernization of Production. 2017:549-54.
8. Ibrahim Z, Tulay P, Abdullahi J. Multi-region machine learning-based novel ensemble approaches for predicting COVID-19 pandemic in Africa. Environ Sci Pollut Res Int. 2022;1-23.
9. Nourani V, Elkiran G, Abdullahi J. Multi-step ahead modeling of reference evapotranspiration using a multi-model approach. J Hydrol. 2020;581:124434.
10. Nourani V, Elkiran G, Abdullahi J. Multi-station artificial intelligence-based ensemble modeling of reference evapotranspiration using pan evaporation measurements. J Hydrol. 2019;577:123958.