ISSN: 2634 - 680X

Journal of Clinical Case Studies Reviews & Reports

SCIENTIFIC Research and Community

Review Article

Open Access

Comparative Analysis and Forecasting on the Death Rate of COVID-19 Patients in Nigeria Using Random Forest and Multinomial Bayesian Epidemiological Models

Ozioma Collins Oguine[1]*, Kanyifeechukwu Jane Oguine[1], Chukwudindu Israel Okorie[1] and Munachimso Blessing Oguine[2]

[1]Department of Computer Science, University of Abuja, Abuja, Nigeria

[2]Department of Computer Science, National Open University of Nigeria

ABSTRACT

The novel COVID-19 (SARS-COV-2) is a disease currently ravaging the world, bringing unprecedented health and economic challenges to several nations. There are presently close to 175,000 reported cases in Nigeria with fatalities numbering over 2,163 persons. The main objective of this paper is to compare the analysis and predictive accuracy between the Random Forest and the Multinomial Bayesian Epidemiological model for a cumulative number of deaths for COVID-19 cases in Nigeria by identifying the underlying factors which may propagate future occurrences. It is worthy to note that the Random Forest algorithm is an ensemble learning approach for classification, regression, and other tasks that works by training a large number of decision trees G(t) while the Multinomial Bayesian algorithm provides an excellent theoretical framework for analyzing experimental data and the highlight of its success relies on its ability to integrate prior knowledge about the parameters of interest as a distribution function p(Ck|d).

***Corresponding author**

Ozioma Collins Oguine, Department of Computer Science, University of Abuja, Abuja, Nigeria. E-mail: oziomaoguine007@gmail.com

Received: August 13, 2021; **Accepted:** August 20, 2021; **Published:** August 25, 2021

Keywords: Covid-19, Random Forest, Multinomial Naïve Bayes, Covid-19 in Nigeria, Machine Learning, Pandemic

Abbreviations

WHO: World Health Organization
MNB: Multinomial Naïve Bayes
RF: Random Forest
MERS: Middle East Respiratory Syndrome
SARS: Severe Acute Respiratory Syndrome
ML: Machine Learning
NCDC: Nigeria Center for Disease Control.

Introduction

According to World Health Organization (WHO) 2020, Corona Viruses are a large family of viruses that are known to cause illnesses ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). These two diseases are spread by the coronaviruses named MERS-CoV and SARS-CoV. SARS was first seen in 2002 in China and MERS was first seen in 2012 in Saudi Arabia [1]. The latest virus seen in Wuhan, China is called SARS-COV-2 and it causes coronavirus.

Pneumonia of unknown cause detected in Wuhan; China was first reported to the World Health Organization (WHO) Country Office in China on 31 December 2019. Since then, the number of cases of coronavirus is increasing along with the high death toll. Coronavirus spread from one city to a whole country in just 30 days. On Feb 11, it was named COVID-19 by World Health Organization (WHO). As this COVID-19 is spread from person to person, Artificial intelligence-based electronic devices can play a pivotal role in preventing the spread of this virus.

As the role of healthcare epidemiologists has expanded, the pervasiveness of electronic health data has expanded too [2]. The increasing availability of electronic health data presents a major opportunity in healthcare for both discoveries and practical applications to improve healthcare [3].

This data can be used for training machine learning algorithms to improve their decision-making in terms of predicting diseases. As of March 21, 2021, a total of 123 million cases of COVID-19 have been registered and the total number of deaths was 2.7 million. COVID-19 has spread across the globe with around 213 countries and territories affected. As the rise in the number of cases of infected coronavirus quickly outnumbered the available medical resources in hospitals, resulting in a substantial burden on the health care systems. Due to the limited availability of resources at hospitals and the time delay for the results of the medical tests, it is a typical situation for health workers to give proper medical treatment to the patients. As the number of cases to test for coronavirus is increasing rapidly day by day, it is not

possible to test due to the time and cost factors, this study will try to predict potential COVID-19 patients to help manage the time and cost of testing.

With the progress of the pandemic and the rising number of the confirmed cases and patients who experience severe respiratory failure and cardiovascular complications, there are solid reasons to be tremendously concerned about the consequences of this viral infection [4].

The need to develop innovative approaches to reach solutions for the COVID-19 related problems has received a great deal of attention. However, another huge problem that researchers and decision-makers have to deal with is the ever-increasing volume of the data, known as BIG DATA, that challenges them in the process of fighting against the virus. This justifies how and to what extent Machine Learning (ML) could be crucial in developing and upgrading health care systems on a global scale [5].

The urgency of this global menace has propelled the research on analyzing, modeling, and forecasting the novel COVID-19 pandemic both domestically and internationally. Some of the current researches conducted include; Modelling and Forecast the number of cases of the COVID-19 pandemic with the curve estimation models like; Box-Jenkins (ARIMA) and Brown/ Holt linear exponential smoothing to the number of COVID 19 epidemic cases in selected countries of G8 countries, Germany, United Kingdom, France, Italy, Russian, Canada, Japan, and Turkey [6]. In the AI practitioners applied ML to process internet activity, news reports, health organization reports, and media activity to predict the spread of the outbreak on the providence level in China [7].

Mathematical modeling was applied to the dynamics of a novel coronavirus (2019-nCoV) in Wuhan-China [8]. Mathematical modeling was applied to COVID-19 transmission mitigation strategies in Ontario Canada [9]. Some curve estimation statistical models and estimators were applied to the main factors affecting the spread of COVID-19 in Nigeria [10]. In the authors made use of the Bayesian approach to predict the number of deaths in Peru for 70 days in the future, using the empirical data from China [11]. Some parameters were used to calibrate the parameters of the SIRD model on the reported COVID-19 cases in the Hubei region, China, the selected model was used to forecast the evolution of the outbreak at the epicenter for three weeks ahead [12]. In the researchers implemented the random forest algorithm for severity analysis of COVID-19 patients using the Computed Tomography (CT) Scans [13]. A comprehensive comparison was carried out on COVID-19 cases using some mathematical models between Turkey and South Africa [14].

Given the above background, this study attempts to model the daily cumulative active, critical, and confirmed COVID-19 cases as it influences the number of reported deaths in Nigeria between January to February 2021, using two major mathematical models namely; Random Forest (RF) and Multinomial Naïve Bayes (MNB) models. This paper is organized in the following way. This paper is organized into four sections. In section 2, the methodology used for modeling is outlined. In section 3, the results and discussions for the study are presented while section 4 presents the conclusions.

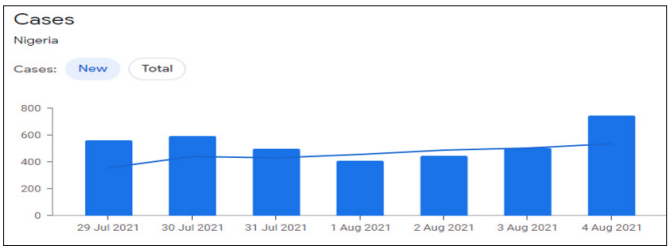


Figure 1: Infographics showing the Number of new covid-19 cases in Nigeria

Methodology

This study intends to create a COVID-19 prediction mechanism that first learns deep features of datasets of COVID-19 Patients records and then trains these learned features and uses it to predict the real-time COVID-19 occurrence using two major Machine learning models comparatively.

Data Collection and Description

Data collection was a necessary yet time-consuming activity. Regardless of the research subject, precision in data gathering is critical for maintaining cohesiveness. The Epidemiological dataset used to train the model to predict COVID-19 was obtained from NCDC which was made available on an open-source repository maintained by HDX. The data collection included information on COVID-19 cases recorded in Nigeria between a certain timeline. The original data set included information on the drugs used by doctors to treat the condition. However, those fields have been removed because our model does not require them. The dataset consists of multi-dimensional data that has been aggregated and integrated. It has textual data fields as well as fields with precise values. The dataset has one thousand and eighty-six (1,110) instances, the predictor variables are mainly gender of the recorded cases, number of confirmed cases, number of recovered cases, number of vaccinated people per day. The target variable is the number of deaths per day.

Table 2.1: Data Instance Description

Features	Description	Role	Datatype
Reporting Date	Date of COVID-19 diagnosis	Predictor Variable	Date
State	State where individual originated from	Predictor Variable	String
Gender	Gender of COVID-19 Patient recorded	Predictor Variable	Int64
Age	Age of COVID-19 Patient recorded	Predictor Variable	Int64
From redzone	Shows if the COVID-19 Patient was picked from a red zone or came in from a red zone	Predictor Variable	Int64
Foreigner	Tells if the Patient is a Foreigner	Predictor Variable	Int64
Death	Did the recorded cases lead to the death of the patient or not	Target	Int64

Table 2.2: Sample of the instances of the dataset

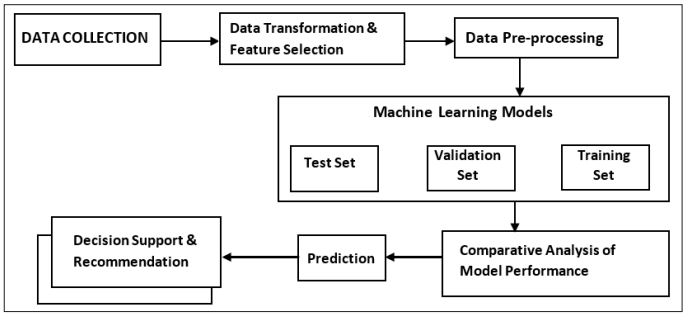
	Reporting date	state	gender	age	from_ redzone	foreigner	death	recovered
0	2/10/2021	Osun	NaN	28	0	0	0	1
1	2/5/2021	Plateau	Male	5	0	0	0	1
2	2/17/2021	Plateau	Male	31	0	0	0	1
3	1/25/2021	Lagos	Female	12	1	0	0	0
4	1/25/2021	Gombe	Male	47	0	0	0	1

Data Pre-processing and Data Splitting

The dataset also includes category variables. We performed label encoding of the category variables because the ML model wants all data supplied as input to be in numeric form. Every unique category value in the column is given a number. When supplied directly as an input, the dataset has several missing values, resulting in an error. As a result, we fill in the missing values with "NA." Because some patient data records have missing values for both the "death" and "recovered" columns, these were segregated from the main dataset and assembled into the test dataset, while the rest were built into the training dataset.

We divided the entire dataset into 60 percent training and 40 percent test sets for each experiment. We set a random state of tuning while splitting the data, ensuring the same data split every time the analysis was run. One of the most important tasks in machine learning is choosing the right features to utilize. Features that have little or a minor impact on results increase the size of the features vector unnecessarily. If their impact on performance is minor compared to their contribution. Generically, small datasets require models that have low complexity (or high bias) to avoid overfitting the model to the data.

2.3 Schematic Representation of Modeling Processes



Flow Diagram of the proposed Model for the comparative analysis of Covid-19 Predictive Algorithms

In the Figure 2.1 above, the process of COVID 19 detection is detailed, firstly the dataset is gathered from a reliable source with a real-time update, after which the dataset is cleaned and missing values are dropped. For a machine-learning algorithm to learn from the dataset and to avoid overfitting, feature engineering is done on the COVID 19 dataset to know which features have a great effect on the target variable.

The dataset is then split into test, train, and validation test, the training set is tuned using hyperparameters, which are used to control the learning process of the model. Then the tuned dataset is fed to the model to learn from after which the performance of the model is tested using the Test set.

Data Mining Techniques/Algorithms

The data collected in this research was evaluated using two major statistical algorithms (Random Forest and Multinomial Naïve Bayes), conclusions and recommendations were also drawn based on the comparative performance of the resulting model.

Random Forest (RF)

Random Forest (RF) algorithm is an ensemble learning technique for data mining classification and regression tasks. The algorithm constructs a multitude of decision trees at training time and outputting [15]. RF data mining algorithm is the best to be used for any decision tree with overfitting to its training dataset [16]. In comparison to other computer semester algorithms, the Random Forest (RF) needs fewer parameter adjustments. Below is the mathematical expression of the algorithm.

$$G(t) = 1 - \sum_{k=1}^{Q} P^2 (k|t) \quad (1)$$

RF Algorithm

Step 1: Randomly select "**k**" features from total "**m**" features k<<m.
Step 2: Among the "**k**" features, calculate the node "**d**" using the best split point.
Step 3: Split the node into daughter nodes using the best split.
Step 4: Repeat a to c steps until the "**l**" number of nodes has been reached
Step 5: Build a forest by repeating steps a to d for "**n**" number of times to create "**n**" number of trees.

Multinomial Naïve Bayes (MNB)

Naive Bayes is one kind of data mining classification algorithm and is used to discriminate dataset instances based on specified features or attributes. It is a probabilistic analysis-based method. For each k class, Multinomial Naïve Bayes computes the probability of an occurrence of COVID 19 presence d being of class *Ck*: $p(Ck|d)$.

No doubt, Naïve Bayes (NB) is a popular classifier that had been applied in several domains such as; weather forecasting, bioinformatics, image and pattern recognition, and medical diagnosis. Although such simplicity increases computational efficiency, it sometimes makes traditional NB insufficient with real-world conditions. Hence this study adopts the Multivariate Event model also referred to as Multinomial Naive Bayes

From the traditional Naïve Bayes mathematical equation in eqn. (2)

$$P\ (Ck|d) = \frac{P(Ck|d)\ P(Ck)}{P(d)} \quad (2)$$

d is given as;

$d = (d1, d2, d3, \ldots, dn)$

By substituting d in eqn.(2) and expanding using chain rule,we get

$$P(Ck|\, d_1, d_2, d_3, \ldots, d_n) = \frac{(P(d1\,|\,Ck)P(d2\,|\,Ck)\ldots P(dn\,|\,Ck)P(Ck))}{P(d1)\,P(d2)\,\ldots P(dn)} \quad (3)$$

Now,you can obtain the values for each by looking at the dataset and substitute them into the equation.

$$P(Ck|\, d_1,d_2,d_3,\ldots,d_n) \propto \pi\_(i=1)\text{^n } P(d1 \mid Ck) \quad (4)$$

There could be cases where the classification could be multivariate. Therefore,we need to find the class Ck with maximum probability. Hence,

$$Ck = argmax_{Ck}\, P(Ck)\, \pi_{i=1}^{n}\, P(d1 \mid Ck) \quad (5)$$

Using the above function in Eqn. (5), we can obtain the class, given the predictors. Hence the model development will follow the sequence highlighted below.

MNB Algorithm
Step 1: Separate by class
Step 2: Summarize dataset
Step 3: Summarize data by class
Step 4: Gaussian probability density function
Step 5: Class probabilities

Choice of Programming Language
Python programming language and Jupyter Notebook were used for data mining predictive tasks. Python is a well-known general-purpose and dynamic programming language that is being used for different fields such as data mining, machine learning, and the internet of things [17-21]. Data mining algorithms are being implemented using python with the help of special-purpose libraries. The models were developed using 5-fold cross-validation.

Experimental Result Presentation and Discussion
The two major data mining algorithms as stated earlier were applied directly to the dataset using python programming language. However, the model developed with Random Forest (RF) algorithm was found to be the most accurate with 97.61 % accuracy in contrast to the Multinomial Naïve Bayes algorithm with an accuracy of 75.98% as shown in Figure 7 below:

The model predicted a minimum and a maximum number of days for COVID-19 patients to recover from the virus. The model also predicted the age group of patients who are at high risk to die from the COVID-19 virus, as well as those who are likely to recover when diagnosed with the infection. From the comparative performance evaluation of the two models, the model developed with Random Forest (RF) algorithm is proven to be more efficient in predicting the possibility of "death or recovery" of infected patients from COVID-19 infection.

Performance Evaluation of Models
For classification problems, the metrics used to evaluate an algorithm are accuracy, confusion matrix, and precision, recall, and F1 values. Since the project is a classification task, the comparative accuracy and the classification report of the two models were computed. Below is a little explanation of the evaluation techniques:

Accuracy - Accuracy is the percentage of correctly classified instances. It is one of the most widely used classification performance metrics.

$$\textbf{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total Number of predictions}} \quad (6)$$

In the case of binary classification models. The accuracy can be defined as:

$$\textbf{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (7)$$

Precision - Precision is the number of classified Positive or fraudulent instances that are positive instances.

$$\textbf{Precision} = \frac{\text{TP}}{\text{(TP+FP)}} \quad (8)$$

Recall - Recall is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made. Unlike precision that only comments on the correct positive predictions out of all positive predictions, recall indicates missed positive predictions. Recall is calculated as the number of true positives divided by the total number of true positives and false negatives.

$$\textbf{Recall} = \frac{\text{TP}}{\text{(TP+FN)}} \quad (9)$$

d. F1 Score - F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

$$\textbf{F1 Score} = \frac{2 * (\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (10)$$

Where:
TP = True Positive; **TN** = True Negative; **FP** = False Positive; **FN** = False Negative

Results and Discussion
The implementation of the experimental results is performed in Python. The results are computed based on Finding the Missing Values, Data Encoding and Feature Selection, Prediction, and Comparison of The Machine Learning Models. The discussion related to the results is summarized below.

Missing Values
The initial step is to find the missing values in the HDX dataset [22] and plot these missing values. The missing values in the dataset were determined, and as a substitute for these, we computed the mean and replaced the missing value with its mean. The default input is a numeric array with levels 0 and 1, where the minimum value is 0 and the maximum value is 1 as shown in Table 3.1 below.

Table 3.1: Showing the transformed description of the dataset after removing null values

	reporting_date	state	gender	age	foreigner	from_redzone	death	recovered
373	1/27/2021	17	0	38	0	1	0	0
751	2/23/2021	8	1	66	0	0	0	0
643	2/17/2021	8	1	56	0	0	0	0
89	2/1/2021	4	0	16	1	0	0	1
402	2/27/2021	19	1	41	0	0	0	0

Feature Selection

As shown in Table 3.2, we have selected 6 features among 8 features from the COVID-19 patient dataset. This selection is being made by analyzing the features after computing the feature importance score in the form of Extra Trees Classifier through the implementation of the decision tree method.

Table 3.2: Showing the features that were selected

	Specs	Score
0	state	17.245590
1	gender	0.188206
2	age	167.282592
3	foreigner	0.369398
4	from_redzone	0.109826
5	death	14.872310

Prediction

The model predicted a minimum and a maximum number of days for COVID-19 patients to recover from the virus. The model also predicted the age group of patients who are at high risk not to recover from the COVID-19 pandemic, those who are likely to recover, and those who might be likely to recover quickly from the COVID-19 pandemic.

Analysis with Random Forest

The Accuracy of the Random Forest model after applying principal component analysis was 97.61%. Along with Accuracy, other performance metrics; Precision, Recall, and F1 score raise after the introduction of PCA which can be seen from Table 3.3.

	precision	recall	f1-score	accuracy
Random Forest	75.0	72.0	73.0	97.61

Table 3.3: Performance metrics for the Random Forest Model

Analysis with Naïve Bayes

The Accuracy of the Naïve Bayes model after applying principal component analysis was 75.98%. Along with Accuracy, other performance metrics; Precision, Recall, and F1 score raise after the introduction of PCA which can be seen from Table 3.4.

	precision	recall	f1-score	accuracy
M. Naive Bayes	70.0	71.0	70.0	75.98

Table 3.4: Performance metrics for the Multinomial Naïve Bayes Model

Comparative Analysis

Data mining models are evaluated using evaluation techniques to determine their accuracy. The techniques determine the quality and efficiency of the model using the data mining algorithm or machine learning algorithms. These main performance evaluation techniques for the data mining model include specificity, sensitivity, and accuracy. However, in this study, the only accuracy is considered to evaluate the developed models.

In Table 3.5 and Figure 3.1 below, we compared the two algorithms used in the analysis of the dataset. The Random Forest performs better in all metrics than the Multinomial Naïve Bayes. Below are visual representations of the comparative result;

Table 3.5: Comparative Performance metrics for both Random Forest and the Multinomial Naïve Bayes Model

	precision	recall	f1-score	accuracy	average
Random Forest	75.0	72.0	73.0	97.61	79.4025
M. Naive Bayes	70.0	71.0	70.0	75.98	71.7450

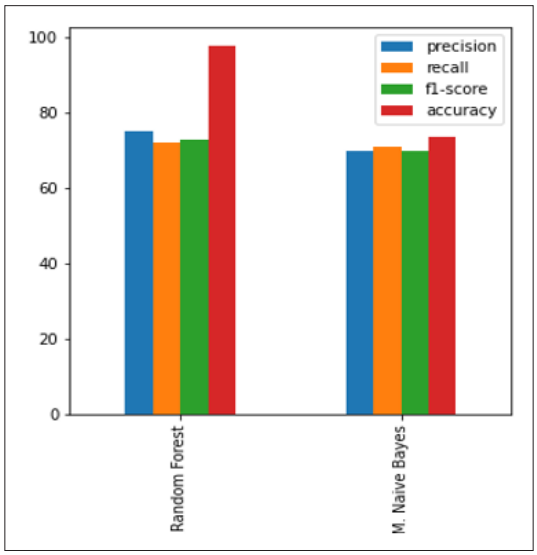


Figure 3.1: Visualization of the Comparative Performance metrics

Summary

The need for an accurate predictor for the prediction of COVID 19 cannot be overemphasized. Many researchers have employed the techniques of machine learning and artificial intelligence for the prediction and classification of COVID 19 disease. These techniques take data as input, learn from the data and next time will be able to make predictions on any new data that has the same dimension with that which they learn from. In this research, two machine learning techniques employed are Random Forest and Multinomial Naïve Bayes. For the classification task, the data is split into two sets, which are the training set (80% of data) and the test set (20% of data). The training set is used to train the models and subsequently, the test set is used to test the trained model. The performance metrics used for our performance evaluation are precision score, recall score, and f1-score.

Conclusion

The machine learning models used in this study do not necessarily require data feature scaling of data, neither is it greatly affected by unbalanced data nor dependency among data set features. Hence, for medium-size data, they are good probabilistic prediction models to employ for a binary classification problem, because of its simplicity and less time complexity; therefore, it can be used for the prediction of COVID 19, which greatly help physicians to make a proper and early diagnosis. In addition to that, Increased collaboration in the development of the AI prediction models can enhance their applicability in the clinical practice and assist healthcare providers and developers in the fight against this pandemic and other public health crises will go a long way in increasing the survivability rate of patients.

Recommendation

An even bigger dataset should be provided and a similar analysis performed and see if the results are identical. Furthermore, the dataset used is lacking in useful information that can help the prediction, more criteria for prediction and improved Machine Learning Models must have been available to attain more accurate numerical metrics. It would also be groundbreaking if the right parameters can be identified from our current and future datasets to generate ROC curves and a possible confusion matrix. Additionally, besides the models that have been implemented, future research should look at the possibility of comparative analysis on sophisticated models in a bid to determine the best prediction model. The idea of applying other feature selection such as the Recursive Feature Elimination and the Correlation Heat Map on the currently used models should also be considered as well.

Acknowledgments

The authors gratefully acknowledge the Nigeria Center for Disease Control (NCDC) and the Humanitarian Data Exchange (HDX) for publicly releasing updated datasets on the number of confirmed, recovered, and death COVID-19 cases in Nigeria [22].

Author Contributions

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work.

Funding

The authors received no funding for this work.

Disclosure

The authors reported no conflicts of interest for this work.

References

1. Ali A (2016). Challenges presented by MERS coronavirus and SARS coronavirus to global health. Saudi journal of biological sciences, Publisher: Elsevier 23: 507-511.
2. Bates DW, Suchi S, Lucila O, Anand S, Gabriel E (2014) Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Affairs 33: 1123-1131.
3. Jenna Wiens, Erica S Shenoy (2018) Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology, Clinical Infectious Diseases 66: 149-153.
4. Guo J, Shuai W, Bo K, Jinlu M, Xianjun Z, et al. (2020) A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). Pre-*/8print, Infectious Diseases (except HIV/AIDS) https://doi.org/10.1101/2020.02.14.20023028.
5. Hamet P, Tremblay J (2017) ‘‘Artificial intelligence in medicine,’’ Metabolism 69: S36-S40.
6. Yonar H, Yonar A, Tekindal MA, Tekindal M (2020) “Modeling and Forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the Box-Jenkins and exponential smoothing methods.” EJMO 4: 160-165.
7. Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, et al. (2020) A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19outbreak using Internet searches, news alerts, and estimates from mechanistic models. arXiv 2004.04019. Available online at https://arxiv.org/abs/2004.04019.
8. Khan MA, Atangana A (2020) “Modeling the dynamics of novel coronavirus (2019-n-Cov) with fractional derivative.”

Alexandria Engineering Journal 59: 2379-2389.

9. Tuite AR, Fisman DN, Greer LA (2020) "Mathematical modeling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada." CMAJ 192: E497-E505.
10. Ayinde K, Lukman AF, Rauf IR, Alabi OO, Okon CE, et al. (2020) Modeling Nigerian Covid-19 cases: A comparative analysis of models and estimators. Chaos: Solitons and Fractals 138:109911.
11. Bayes C, Valdivieso L (2020) Modelling death rates due to COVID-19: A Bayesian approach. arXiv. (2020) 2004.02386. Available online at https://arxiv.org/abs/2004.02386.
12. Anastassopoulou C, Russo L, Tsakris A, Siettos C (2020) "Data-based analysis, modeling and forecasting of the COVID-19 outbreak." PLoS ONE 15: 0230405.
13. Tang Z, Zhao W, Xie X, Zhong Z, Shi F, et al. (2020) Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images arXiv 2003.11988. Available online at https://arxiv.org/abs/2003.11988.
14. Atangana A, Araz, SI (2020) "Mathematical Model of COVID-19 spread in Turkey and South Africa: Theory, Methods, and Applications 2020: 659.
15. Haque MR, Islam MM, Iqbal H, Reza MS, Hasan MK (2018) Performance Evaluation of Random Forests and Artificial Neural Networks for the Classification of Liver Disorder. In: 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2). IEEE 1-5.
16. Muhammad LJ, Ahmed Abba Haruna, Ibrahim Alh Mohammed, Mansir Abubakar, Bature Garba B, et al (2019) Performance Evaluation of Classification Data Mining Algorithms on Coronary Artery Disease Dataset: IEEE 9th International Conference on Computer and Knowledge Engineering (ICCKE 2019), Ferdowsi University of Mashhad DOI: 10.1109/ICCKE48569.2019.8964703.
17. Hasan MK, Islam MM, Hashem MMA (2016) Mathematical model development to detect breast cancer using multigene genetic programming. In: 2016 5th International Conference on Informatics, Electronics, and Vision (ICIEV). IEEE 574-579.
18. Islam Ayon S, Milon Islam M (2019) Diabetes Prediction: A Deep Learning Approach. Int J Inf Eng Electron Bus 11: 21-27.
19. Hasan M, Islam MM, Zarif MII, Hashem MMA (2019) Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet of Things 7: 100059.
20. Islam M, Neom N, Imtiaz M, Nooruddin S, Islam M, Islam M (2019) A Review on Fall Detection Systems Using Data from Smartphone Sensors. Ingénierie des systèmes d Inf 24: 569-576.
21. Nooruddin S, Islam MM, Sharna FA (2020) An IoT-based device-type invariant fall detection system. Internet of Things 9: 100130.
22. Coronavirus dataset of Nigeria Center for Disease Control (NCDC) retrieved from the Humanitarian Data Exchange (HDX) repository. https://data.humdata.org/dataset/nigeria_covid19_subnational.