Prediction of COVID 19 and Its Economic Impact in Africa

Habte Tadesse Likassa(PhD)

Department of Statistics, College of Natural and Computational Sciences, Addis Ababa University, Ethiopia,

Abstract

This paper presents a two way factor design incorporating both spatial and temporal variation in the prediction of COVID 19 in Africa. Inline to this, the impact of COVID-19 on the GDP in Africa is well scrutinized. In contrast to the existing works [1–3], this work also extends the two factor design into the one way factor design through incorporating covariates into spatial effects. The data which rely on the spatial and temporal are obtained from WHO datasets [4, 5]. The one factor design with more covariates are taken into consideration to identify the major potential predictor variables responsible for the deaths and confirmed cases due to COVID 19 in Africa. The MANCOVA considered population density, temperature, humidity; perception and wind are all considered as covariates. Simulations show that the two way analysis of variance has shown that there is statistical significant difference between the spatial ($F_{cal} = 8.2704$, $P_{value} = 3.099 * 10^{-6}$) and temporal ($F_{cal} = 48.7964$, $P_{value} = 9.147 * 10^{-16}$) effects. South Africa and Nigeria are highly influencing due to the pandemic where their GDP also relatively mostly declined. A significant economic change is observed before the pandemic and after the outbreak of the pandemic ($t_{cal} = 2.9548, P_{value} = 0.01805$). COVID 19 negatively influenced the economy of

most of African countries.The population density, temperature and the wind are found to be statistically significantly associated with COVID 19 cases and deaths.

1. Introduction

Prediction of Corona virus diseases is very essential to mitigate the spatial and temporal distribution of the pandemic [6–8]. The spatial and temporal variation has its own impacts on the COVID-19 [8–10]. There are several approaches proposed to identify the spatial and temporal effects on health [11–13].

Since the inception of the pioneering work of prediction algorithms by Likassa et al.[2], a myriad of algorithms has been addressed for prediction of COVID 19, e.g. [10, 14–16]. However, these methods do not work well in jointly identifying the impacts of covariates together.

To overcome this drawback, Bhati et al.[[17] considered a new model algorithm, which replaced the regression [18] with the regression to improve the performance of the approach. Thus, [2] proposed another predictive models where some potential predictor variables are considered. Additionally, One of the greatest challenges with any model is determining the current location and time on the epidemic curve[19]. To tackle this dilemma, [3, 20–22] addressed several predictive models to pinpoint the location. However, the impacts of predictor variables are totally ignored. Moreover, [23] proposed multiple models for prediction of COVID-19 in Africa, indicating the prediction rate is increasing. Thus, this paper advances the existing algorithms relying on the two factor design of experiments through considering more covariates. Despite the uncertainty in the prediction models, they are of central importance in directing many aspects of the response to COVID-19.

This paper presents a two way design of experiments considering the spatial and temporal effects into the model. To be more resilient with the outliers and heavy sparse noises and identify the potential impacts of covariates for the number of deaths and confirmed cases the MANCOVAis also considered including several predictor variables (population density, temperature, humidity, perception and wind). This work first, propose the two way design of experiments to determine the spatial (location) and temporal (time) on the COVID-19 spread. Following this, the one way design of experiments with more predictor variables considered to predict the number of confirmed cases and deaths due to the pandemic in Africa. Simulations show that the proposed method indicates as the spatial and temporal variation has an impact and spread of the pandemic at 0.05 level of significance statistically significantly impacting the both the COVID 19 cases and deaths. Following this, the population density, temperature and wind variation for the high spread of the pandemic at five level of significance.

The major contributions of this letter include:

Email address: `habte.tade@yahoo.com` (Habte Tadesse Likassa(PhD))

Table 1: The Data layout for Two factor Experimental Design

Temporal or Spatial	1	2	.....	k
1	Y_{11}	Y_{12}	....	Y_{1k}
2	Y_{21}	Y_{22}	....	Y_{2k}
....		....	....	
a	Y_{a1}	Y_{a1}	....	Y_{ak}

1) The spatial and temporal variation are incorporated in the two way factor design of an experiments to predict COVID 19 in Africa;

2) Following this, a number of covariates are considered to identify the major potential predictor variables related to and distribution of the pandemic.

3) The goodness of one way factor design after incorporating a set of covariates (population density, temperature, wind, perception and humidity) are well scrutinized;

4) The impacts of COVID-19 on GDP is also well scrutinized and experimentally justified;

This paper is structured as follows. Sec. 2 gives an overview of the problem formulation. Sec. 3 describes the materials and methods. Secs. 4 depicts the Simulation results to verify the effectiveness of the proposed algorithm. Sec. 5 draws some concluding remarks to summarize the paper.

2. Problem Formulation

Given i spatial locations $\{i\} \in \Re^{n \times 1}, and \{j\} \in \Re^{1 \times m}$, $i = 1, \cdots, a$ months which indicates the temporal effects, and $j = 1, \cdots, k$ denotes spatial effects respectively. Thus, the i spatial locations and the j temporal effects have response variable Y_{ij} [24]. While, we apply the ith location on the jth temporal denoted by months, then the overall mean denoted by μ and indicating the τ_i pointing out the spatial effects of the African countries and β_j the temporal effects categorized according to months. Then, the model of the two way design of experiments is given by

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \tag{1}$$

where i indicates the ith spatial effects, the jth effect denotes the jth temporal effects in predicting the COVID-19 and ε_{ij} indicates the random disturbance term which is $N(0, \sigma^2)$. The data layout corresponding to the Two factor design of experiments is given in Table 1. Following the same procedures as [25, 26], then from equation (1) can be partitioned as

$$SS_T = SS_{Treat} + SS_{Blocks} + SS_{Error} \tag{2}$$

where SS_T denotes the sum squares of total, SS_{Treat} denotes the spatial and SS_{Block} indicates the temporal effects. Each sum squares corresponding to source of variation are solved using [25, 26]. After we determined this, we also further extend model (2) to develop a new MANCOVA which incorporates three predictor variables.

$$Y_{ij} = \mu + \tau_i + \beta_j(X_{ij} - \bar{X}) + \varepsilon_{ij} \tag{3}$$

where Y_{ij} indicates the dependent variables (Cases and Deaths) and denotes all the other explanatory variables are denoted by X_{ij}.

3. Methods

The study is about novel coronavirus also known as COVID-19 predictions. The COVID-19 has proved a present potential threat to human life. It causes tens of thousands of deaths and the death rate is increasing day by day throughout the globe. To contribute to this pandemic situation control,this study attempts to perform future prediction of the number of confirmed cases and deaths for the upcoming several years.

3.1. Datasets

The aim of this study is the future prediction of COVID-19 spread focusing on the number of confirmed cases and deaths, and excluding the number of recoveries based on the WHO data [4, 5]. This dataset consists of all cases and deaths of COVID 19 in Africa reported by the WHO Region, as of 6th September 2020 to as of the 10th of January 2021. This data includes reporting country/territory/area along with cumulative cases, cumulative deaths and transmission classification. The aggregated cumulative cases per hundred thousands of population is 190.4 while the cumulative deaths per hundred thousands of population is 4.3.

3.2. Evaluation of Parameters

In this study, we evaluate the performance of each of the learning models in terms of R-squared R^2 score. The R^2 score is a statistical measure used to evaluate the performance of regression models [27–29]. The statistic shows the dependent variable's variance percentage that collectively determines the independent variable. It measures the relationship strength between the dependent variable and regression models on a convenient $0 ˘ 100$ scale. After training the regression model, we can check the goodness-of-fit of trained models by using the R^2 score. R^2 score finds the scatterings of data points around the regression line which can also be referred to as the coefficient of determination. 0 score implies the response variable has no variability around its mean explained by the model, and 100 implies that the response variable has all the variability around its mean.The high R^2 score shows the goodness of the trained model.

4. Simulations Results

In this sections, some simulations are conducted to assess the proposed methods. Two datasets are considered in the simulations, including the WHO data [4] taking country as a spatial and months as the temporal effects. The climatic related variables are obtained from the world climate guide [30], the population density from [31] and the GDP from https://www.imf.org/external/datamapper/NGDP$_{RP}CH@AFRREO/SSA/OEXP/OIMP$.
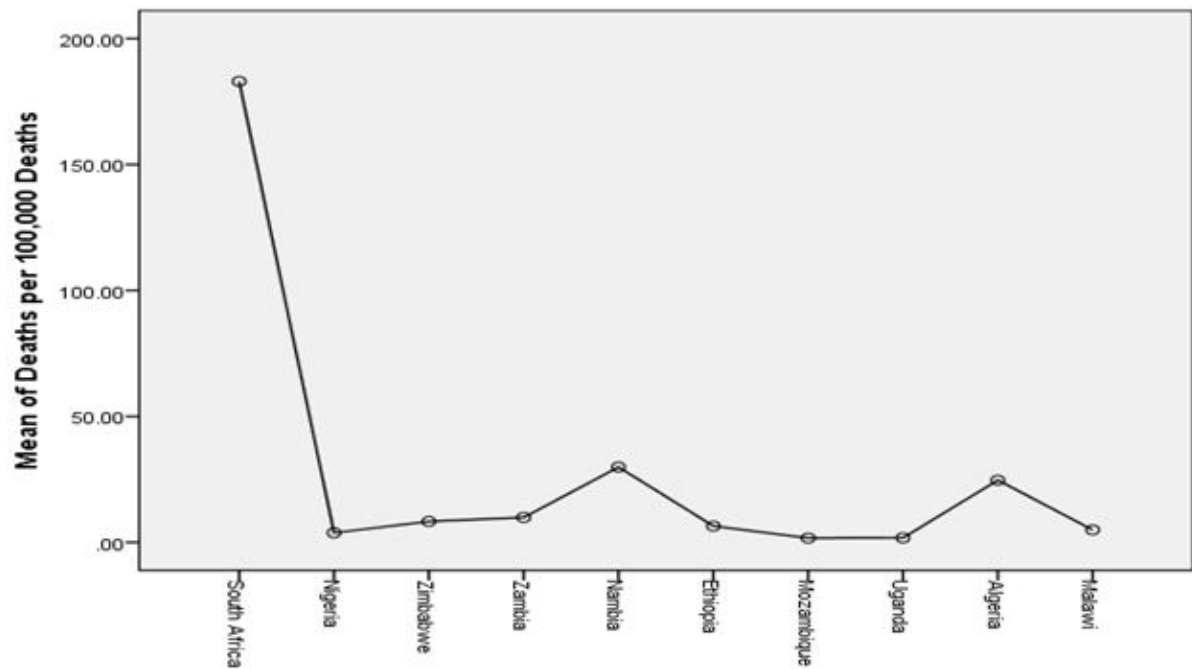
Figure 1: Top Ten Countries in Africa Highly Vulnerable due to COVID 19

Table 2: Two Way Factor Design

Source of Variation	DF	SS	Fcal	P_{value}
Spatial Effects (x_1)	9	91464	8.2704	$3.099 * 10^{-6}$
Temporal Effects (x_2)	8	479693	48.7964	$9.147 * 10^{-16}$
Error	0.0422	0.0382		

4.1. Spatial and Temporal Analysis

In this section, the experimental simulations which determine the effects of spatial and temporal variation in the prediction of COVID 19.The results of the study has shown that there is statistically significant mean difference deaths due to COVID 19 both spatially and temporally Table 2. This result is justified illustrated by the mean plots based on the spatial and temporal variations(Figure 1 and Figure 2).

4.2. Multivariate Analysis of Covariance

In this section, we considered more potential covariates that are responsible on the deaths and confirmed cases in Africa. As it is noted from Table 3, the climatic variables Temperature, Rainfall, Wind and Humidity are statistically significantly associated with the spread
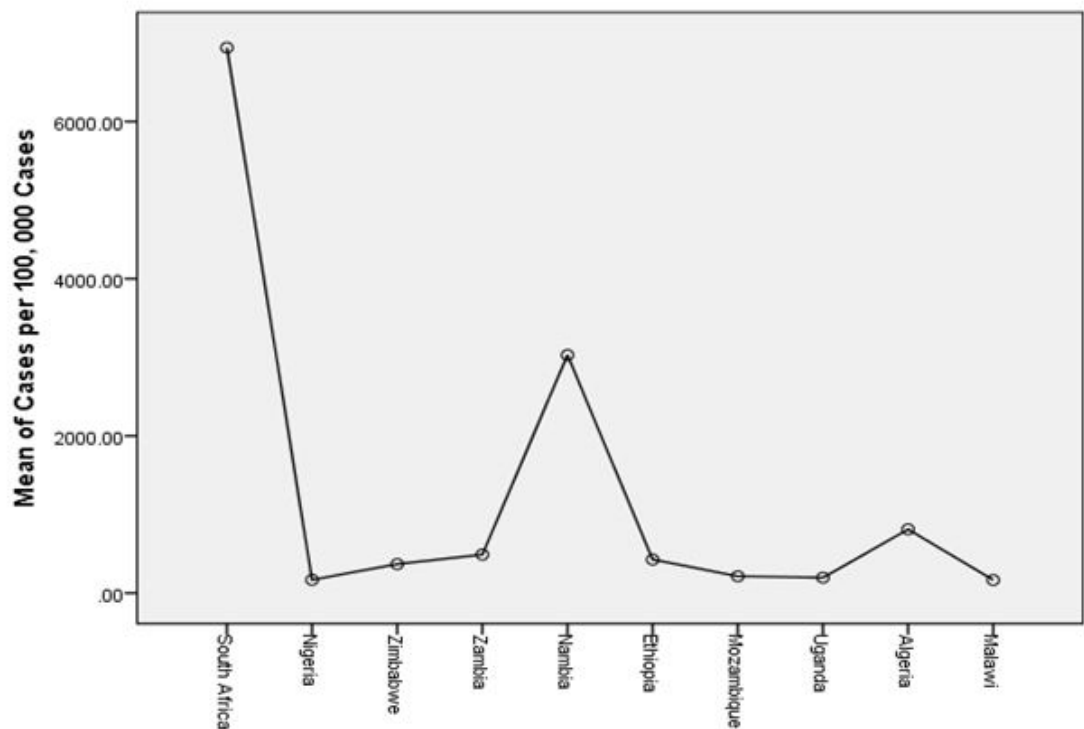
Figure 2: Top Ten Countries in Africa Highly Vulnerable due to COVID 19

of COVID 19 in Africa.

4.3. Evaluation of Parameters

Based on the fitted models, checking the goodness of the model is highly essential. Inline to this, the coefficient of determination which considers the joint impact of all covariates on the response variable is considered. The result of the study has shown as our proposed algorithm is best as compared to the state of the art of the works (Table 5).

5. Discussion and Conclusion

In this study, a two factor experimental design and MANCOVA through incorporating several covariates are considered to predict the number of COVID-19 cases and deaths in Africa.Unlike the predictive models done in Ethiopia [32–34], this study provides projected number of all active cases and all the deaths along with the number of population density and all other predictor variables that helps the government to choose and enforce better intervention mechanisms. This work illustrates the prediction of COVID 19 in Africa, where the impacts of temporal and spatial are both taken into consideration. Additionally, the population density, temperature variation and wind variation are all statistically significantly linked with the response variable (Table 2).

Thus, the impact of COVID 19 is also significantly influencing the GDP of most of African countries (Table 4). Moreover, Nageria and South Africa are the most highly influenced

6

Table 3: Multivariate Analysis of Covariance Table With More Covariates.

Source of Variation	Dependent Variable	SS	DF	MS	F	P_{value}
Corrected Model	Cases	40286869.71	5	8057373.941	16.195	0.009
	Deaths	26753.663	5	5350.733	22.458	0.005
Intercept	Cases	1791228.966	1	1791228.966	3.600	0.131
	Deaths	233.177	1	233.177	0.979	0.379
Population Density	Cases	12212981.4	1	12212981.4	24.547	0.008
	Deaths	5340.595	1	5340.595	22.415	0.009
Temperature	Cases	1078502.394	1	1078502.394	2.168	0.215
	Deaths	3369.294	1	3369.294	14.141	0.215
Perception	Cases	634991.567	1	634991.567	1.276	0.322
	Deaths	315.140	1	315.147	1.323	0.314
Humidity	Cases	28090.790	1	28090.790	0.056	0.824
	Deaths	8.628	1	8.628	0.036	0.858
Wind	Cases	20135705.590	1	20135705.590	40.471	0.003
	Deaths	16619.045	1	16619.045	69.713	0.001
Error	Cases	1990142.757	4	497535.689	— — —	— — —
	Deaths	953.029	4	248.257	— — —	— — —
Total	Cases	58695334.420	10	— — —	— — —	— — —
	Deaths	35248.068	10	— — —	— — —	— — —
Corrected Total	Cases	42277012.460	9	— — —	— — —	— — —
	Deaths	27706, 691	— — —	— — —	— — —	— — —

Table 4: GDP Comparison between 2019 and 2020

Source of Variation	DF	t_{cal}	P_{value}
Year 2019 versus 2020	8.0967	2.9548	0.01805

countries due to the outbreak of the pandemic.

6. Conclusions

Machine learning based prediction mechanisms have proved their significance to anticipate in perioperative outcomes to improve the decision making on the future course of actions. However, they still lack some robustness in fitting the data well. To tackle this dilemma, this paper propose first, a one way factor design through incorporating a set of covariates is considered. First, the effects of spatial and temporal on deaths and cases due to COVID-19 are well studied. Following this, to identify the potential covariates the one way factor design through considering a number of covariates is considered. The goodness of the fit is also well scrutinized. As the population density, temperature and wind are all increasing in the ith location and j^{th} time, then the spread of the pandemic is highly increasing which leads to high number of peoples infected and to the death in Africa. The simulation result has shown that both the spatial and temporal effects have significant impacts on COVID 19 in Africa. The simulation result has also shown that the impact of COVID 19 in both health and

Table 5: Models performance on Prediction for the Cases and Death rate

| Model	R^2	$R^2 Adjusted$
Proposed Method	0.9660	0.9450
Logistic Regression[35]	0.9422	0.9382
Shrinkage Regression[36]	0.8422	0.7382
Cubic Regression [2]	0.9545	0.9180

economy is big and draging the economy of African down. The simulation experiments have also shown that the population density variation, temperature variation and wind variation based on the location is significantly influencing the spatial and temporal distribution of COVID 19 in Africa. Social distancing where there is a high density, high fluctuation of temperature and heavy wind is highly recommended in Africa. Remedial action and imports are highly encouraged to boost the African GDP.

Acknowledgments

References

[1] Habte Tadesse Likassa, "The impacts of covariates on spatial distribution of corona virus 2019 (covid-19): What do the data show through ancova and mancova?," .

[2] Habte Tadesse Likassa, Wen Xain, Xuan Tang, and Gizachew Gobebo, "Predictive models on covid 19: What africans should do?," *Infectious Disease Modelling*, 2020.

[3] Antoine Bagula, Hloniphani Maluleke, Olasupo Ajayi, Amani Bagula, Nancy Bagula, and Moise Bagula, "Predictive models for mitigating covid-19 outbreak," in *2020 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2020, pp. 1–7.

[4] World Health Organization et al., "Coronavirus disease (covid-19): weekly epidemiological update," 2020.

[5] World Health Organization et al., "Covid-19 weekly epidemiological update, 10 november 2020," 2020.

[6] Mohammad Jamshidi, Ali Lalbakhsh, Jakub Talla, Zdeněk Peroutka, Farimah Hadjilooei, Pedram Lalbakhsh, Morteza Jamshidi, Luigi La Spada, Mirhamed Mirmozafari, Mojgan Dehghani, et al., "Artificial intelligence and covid-19: deep learning approaches for diagnosis and treatment," *IEEE Access*, vol. 8, pp. 109581–109595, 2020.

[7] Xiaolei Ma, Zhimin Tao, Yinhai Wang, Haiyang Yu, and Yunpeng Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[8] Parisa Rashidi and Azra Bihorac, "Artificial intelligence approaches to improve kidney care," *Nature Reviews Nephrology*, vol. 16, no. 2, pp. 71–72, 2020.

[9] Patricia Lustosa Brito, Monika Kuffer, Mila Koeva, Julio Cesar Pedrassoli, Jiong Wang, Federico Costa, and Anderson Dias de Freitas, "The spatial dimension of covid-19: The potential of earth observation data in support of slum communities with evidence from brazil," *ISPRS International Journal of Geo-Information*, vol. 9, no. 9, pp. 557, 2020.

[10] Rui Huang, Miao Liu, and Yongmei Ding, "Spatial-temporal distribution of covid-19 in china and its prediction: A data-driven modeling analysis," *The Journal of Infection in Developing Countries*, vol. 14, no. 03, pp. 246–253, 2020.
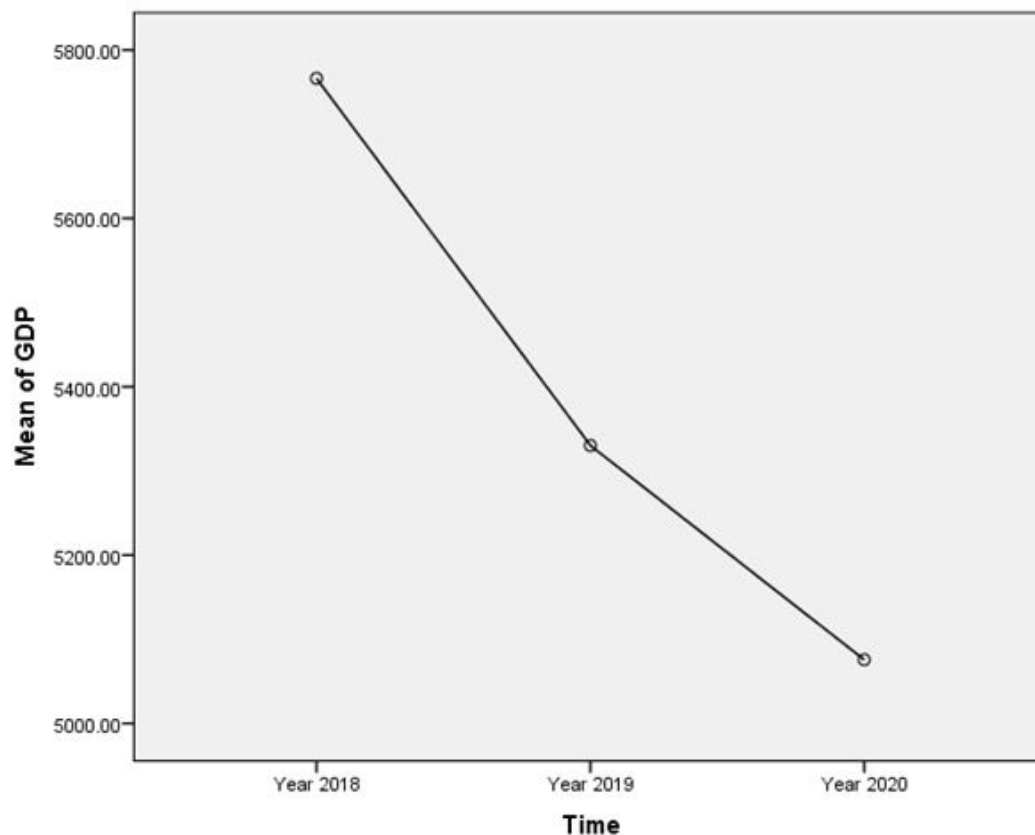
[11] Diego Giuliani, Maria Michela Dickson, Giuseppe Espa, and Flavio Santi, "Modelling and predicting the spatio-temporal spread of coronavirus disease 2019 (covid-19) in italy," *Available at SSRN 3559569*, 2020.

[12] Anne Weiss, Mads Jellingsø, and Morten Otto Alexander Sommer, "Spatial and temporal dynamics of sars-cov-2 in covid-19 patients: A systematic review and meta-analysis," *EBioMedicine*, vol. 58, pp. 102916, 2020.

[13] Zhanshan Ma, "Spatiotemporal fluctuation scaling law and metapopulation modeling of the novel coronavirus (covid-19) and sars outbreaks," *arXiv preprint arXiv:2003.03714*, 2020.

[14] Patrick Flandrin, Charles G Lucas, Pierre Borgnat, et al., "Spatial and temporal regularization to estimate covid-19 reproduction number r (t): Promoting piecewise smoothness via convex optimization," *medRxiv*, 2020.

[15] Ashish U Mandayam, AC Rakshith, S Siddesha, and SK Niranjan, "Prediction of covid-19 pandemic based on regression," in *2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, 2020, pp. 1–5.

[16] D Haritha, N Swaroop, and M Mounika, "Prediction of covid-19 cases using cnn with x-rays," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*. IEEE, 2020, pp. 1–6.

[17] Amit Bhati and Anurag Jagetiya, "Prediction of covid-19 outbreak in india adopting bhilwara model of containment," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2020, pp. 951–956.

[18] Anupam Prakash, Piyush Sharma, Indrajeet Kumar Sinha, and Upendra Pratap Singh, "Spread & peak prediction of covid-19 using ann and regression (workshop paper)," in *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*. IEEE, 2020, pp. 356–365.

[19] Junling Ma, "Estimating epidemic exponential growth rate and basic reproduction number," *Infectious Disease Modelling*, vol. 5, pp. 129–141, 2020.

[20] Swati Arora, Rishabh Jain, and Harendra Pal Singh, "Epidemiological models of sars-cov-2 (covid-19) to control the transmission based on current evidence: A systematic review," 2020.

[21] Manoj Kumar, Jatin Bareja, Manjot Singh, and Rupanshu Sharma, "An intelligent prediction model of covid-19 in india using hybrid epidemic model," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE, 2020, pp. 389–396.

[22] Saad Awadh Alanazi, MM Kamruzzaman, Madallah Alruwaili, Nasser Alshammari, Salman Ali Alqahtani, and Ali Karime, "Measuring and preventing covid-19 using the sir model and machine learning in smart health care," *Journal of healthcare engineering*, vol. 2020, 2020.

[23] Wandong Zhang, WG Will Zhao, Dana Wu, and Yimin Yang, "Predicting covid-19 trends in canada: a tale of four models," *Cognitive Computation and Systems*, vol. 2, no. 3, pp. 112–118, 2020.

[24] Bram Burger, Marc Vaudel, and Harald Barsnes, "Importance of block randomization when designing proteomics experiments," *Journal of Proteome Research*, 2020.

[25] Oscar Kempthorne, "The design and analysis of experiments.," 1952.

[26] Douglas C Montgomery, *Design and analysis of experiments*, John wiley & sons, 2017.

[27] Furqan Rustam, Aijaz Ahmad Reshi, Arif Mehmood, Saleem Ullah, Byungwon On, Waqar Aslam, and Gyu Sang Choi, "Covid-19 future forecasting using supervised machine learning models," *IEEE Access*, 2020.

[28] Josep Lupón, Hanna K Gaggin, Marta De Antonio, Mar Domingo, Amparo Galán, Elisabet Zamora, Joan Vila, Judith Peñafiel, Agustín Urrutia, Elena Ferrer, et al., "Biomarker-assist score for reverse remodeling prediction in heart failure: the st2-r2 score," *International journal of cardiology*, vol. 184, pp. 337–343, 2015.

[29] Ji-Hyeong Han and Su-Young Chi, "Consideration of manufacturing data to apply machine learning methods for predictive manufacturing," in *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2016, pp. 109–113.

[30] LJB McArthur, "World climate research programme-baseline surface radiation network (bsrn)-operations manual version 2.1," 2005.

[31] Simon Chamaillé-Jammes, Hervé Fritz, Marion Valeix, Felix Murindagomo, and Jean Clobert, "Resource variability, aggregation and direct density dependence in an open context: the local regulation of an african elephant population," *Journal of Animal Ecology*, vol. 77, no. 1, pp. 135–144, 2008.

[32] Joseph Waogodo Cabore, Humphrey Karamagi, Hillary Kipruto, James Avoka Asamani, Benson Droti, Aminata Binetou-Wahebine Seydi, Regina Titi-Ofei, Benido Impouma, Michel Yao, Zabulon Yoti, et al., "The potential effects of widespread community transmission of sars-cov-2 infection in the who african region: a predictive model," .

[33] Rediat Takele Figa, "Stochastic modelling for predicting covid-19 prevalence in east africa countries," *Infectious Disease Modelling*.

[34] N Erasmus, "Age discrimination in critical care triage in south africa: The law and the allocation of scarce health resources in the covid-19 pandemic," *SAMJ: South African Medical Journal*, vol. 110, no. 12, pp. 1172–1175, 2020.

[35] Hsiao-Lin Hwa, Wen-Hong Kuo, Li-Yun Chang, Ming-Yang Wang, Tao-Hsin Tung, King-Jen Chang, and Fon-Jou Hsieh, "Prediction of breast cancer and lymph node metastatic status with tumour markers using logistic regression models," *Journal of evaluation in clinical practice*, vol. 14, no. 2, pp. 275–280, 2008.

[36] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.