SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 4 | April 2022 - Peer Reviewed Journal

ANALYSIS OF COVID-19 IN INDIA

Sanchana.R[1], Nithya Devi.S[2], Mercy.P[3] Dhanush kumar.S[4]

[1]*Assistant Professor in Department of Information Technology, Sri Sairam Institute of Technology*
[2]*Department of Information Technology, Sri Sairam Institute of Technology*
[3]*Department of Information Technology, Sri Sairam Institute of Technology*
[4]*Department of Information Technology, Sri Sairam Institute of Technology*

ABSTRACT

COVID- 19 arose in the city of Wuhan in China. The situation started to be more critical when numerous number of person started to get infected which in turns increases the number of death cases. The novel coronavirus (COVID-19) that was first reported at the end of 2019 has impacted almost every aspect of life as we know it. This paper focuses on the incidence of the disease in India Using three simple machine learning algorithms—the linear regression model, polynomial regression and Support Vector machine, we model the daily and cumulative incidence of COVID-19 in the country during the early stage of the outbreak, and compute estimates for basic measures of the infectiousness of the disease including the active cases, death cases, growth rate, mortality rate and recovery rate. Estimates of the growth factor is calculated using the new confirmed, recovered and death cases divided using the previous confirmed, recovered and death cases. The growth factor above 1 indicates the increase in corresponding cases. The growth cases below 1 indicate the trending downward which indicates the sign of exponential growth. The predictive ability of the polynomial regression model was found to give a better fit and simple estimates of the daily incidence.

I. INTRODUCTION

COVID- 19 arose in the city of Wuhan in China. The situation started to be more critical when numerous number of person started to get infected which in turns increases the number of death cases. The government started documenting each and every case having symptoms like Pneumonia. The cases started to rapidly increases about 40 spans in 30 days. In the previous years the virus had a history in china known as SARS disease. The SARS disease took the lives of around 770 people in the year 2002 and 2003.

The corona virus is transmitted and spread in the following cases the first one is the person who in contact with the secretions of sneezing or coughing and float during the respiratory diseases. The second one is having physical contact with the affected person through the hands. Therefore in order to decrease the risk of the spreading corona virus doctors advised to wash the hand and avoid touching the infected person in order to eliminate the risk in case a person is exposed to it accidently. In order to overcome from the corona virus the vaccine has to be developed. But the history tells that it took decades to develop penicillin and polio vaccine. COVID-19 is one of the pandemic elements across the world. It is a disease spread by Coronavirus whose vaccine has not yet been discovered. Therefore, COVID -19 has become one of the pandemic elements in the leading Nordic countries. In this situation, the best approach every one needs to follow is stay sterile and maintains a social distance and stay enlightened about the situation.(WHO, 2020b).

About 87% of the population uses social media, and in the current situation, this percentage has increased. The general public prefers to stay at home the social media is a major source of information and info emic. Humans, in curiosity, try to stay updated and therefore are more gripped to be associated with social media. In the process, the data are encountered from various sources which could be appropriate or unsound in the same way.

Steven Taylor in his psychology book has predicted the world might experience a new pandemic in the upcoming years and people would prefer to stay at home as much as possible. The number of cases of the individual has increased rapidly which in turn created pressure on the medical services as healthcare. The healthcare providers test and diagnose the infected individuals. In addition to the medical services offered many cases trying to control covid 19 have led to the backlog for and deprivation of many medical procedures. The healthcare providers trying to balance the conflicts that may arise between the two notes which in turn change the nature of healthcare. The COVID 19 restrictions have led to much anxiety and distress which lead to an increase in psychiatric diseases. COVID 19 may have a long-term negative impact on the mental health of each and every individual.

The restriction due to COVID 19 has made non-essential business to get closed due to pandemic situations. This has impacted the national economics with much business being permanently closed. The permanently closing of business has significantly increased in unemployment. The travels have

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 4 | April 2022 - Peer Reviewed Journal

been severely affected the tourism and travel industries. The environment has limited on the business that have been able to continue operating in the pandemic time. This has led to possible improvements in the environment in terms of reduction of pollution. The two main countries affected due to COVID 19 are Europe, Spain and Italy. The majority of the literature review mainly focuses on the clinical aspect of the diseases. The review has only one limited number of exploring the prevalence of the diseases. The main contributions focus on modeling the incidence of the COVID 19 in several countries. The second one is provide the estimate of basic measure of the infectiousness and severity. The third one mainly focuses on predictive ability of the mathematical models. And finally forecast the incidence of COVID 19 in different countries.

II.LITERATURE REVIEW

The COVID 19 at present has affected over 49 million cases of infected individuals. The cases have been conformed in 180 countries with in excess of 1 million deaths [1]. The foundations of disease are very similar to the SARS. The SARS virus was initially found in Asia in 2003. SARS virus has spread much more easily and still there exist no vaccine.

The analyses of diseases in North and South America have similar classical methods. The model produces a progressive outbreak in the United States until the end of the 2021[13]. The countries like Brazil and Peru used a logistic growth model and machine learning techniques [14]. United States analysis completely used spatial log and error models. The spatial log and error models where able to estimate the number of deaths in the United States. The number of death was calculated using the modified logistic fault dependent detection methods [15]. The infected rate across the different states in United States was calculated using a sample selection model. The Sample selection model creates a relationship between the social media communication and the incidence in Colombia. The relationship between the social media and the incidence were calculated using the non-linear regression models [17].

The countries like Africa mainly focus on the method of stimulation and predict the spread of the diseases in different countries using a modified susceptible exposed infectious recovered model [19]. The West African mainly used predict the spread of the diseases using the deterministic susceptible exposed infectious recovered model [20]. The East African countries used to predict the spread of disease using each and every individual travel history and the personal contact in Nigeria. The spread of disease using the travel history is predicted using the ordinary least square regression method. The Auto regressive integrated moving average model is used to forecast the prevalence of COVID 19 in East Africa [22]. The real time forecasting of the daily confirmed cases in Saudi Arabia uses the logistic growth and susceptible infected recovered models. The above models were able to generate the real time forecasting of the confirmed cases [23].

The lead lag regression model is used to identify the relation between the cumulative numbers of daily cases of COVID 19 in various countries. In various countries the method of forecasting the future incidence is done using several machine learning techniques. This helps the countries in classifying the early middle and the later stage of the outbreak [24].

The number of cases increased has created a pressure on the medical services. In addition to the normal medical services that are offered by many health care providers. Trying to control COVID has led to backlog for and deprivation of medical procedures [25]. The health care providers trying to balance the conflicts that may arise between the two notes which in turn change the nature of healthcare. The COVID 19 restrictions have led to much anxiety and distress which lead to an increase in psychiatric diseases. COVID 19 may have a long-term negative impact on the mental health of each and every individual .

III. MODULE DESRIPTION

3.1 MORTALITY RATE

The equation [1] represents the mortality rate. Mortality rate is calculated based on the number of death cases divided by the number of confirmed cases.

Mortality rate = (death cases/confirmed cases)*100 [1]

3.2 RECOVERY RATE

The equation [2] represents the recovery cases. The recovery cases can be calculated using the formula total number of recovered cases divided by the total number confirmed cases.

Recovery rate= (recovered cases/confirmed cases)*100 [2]

3.3 FACTOR OF GROWTH

The growth factor is calculated using the new confirmed, recovered and death cases divided using the previous confirmed, recovered and death cases. The growth factor above 1 indicates the increase in corresponding cases. The growth cases below 1 indicate the trending downward which indicates the sign of exponential growth.

3.4 ACTIVE CASES

The equation [3] represents the active cases. The active cases can be calculated using the formula number of confirmed cases subtracted from the number of recovered cases subtracted from the number of death cases. Increase in the number of cases indicates the recovered cases or death cases number is dropping in comparison to number of confirmed cases drastically.

Active Cases = Number of Confirmed Cases - Number of Recovered Cases - Number of Death Cases [3]

3.5 CLOSED CASES

The equation [4] represents the closed cases. The closed cases are calculated using the formula number of recovered cases added with the number of death cases. Increase in number of cases indicates either more patients are getting recovered from the disease or more people are dying because of the covid disease.

Closed Cases = Number of Recovered Cases + Number of Death Cases [4]

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 4 | April 2022 - Peer Reviewed Journal

3.6 LINEAR REGRESSION MODEL

The linear regression model is the easiest and the popular machine learning algorithms. The linear regression model is a statistical method mainly used for predictive analysis. The model makes predictions based on continuous or numeric values. The linear regression model represents the relationship between the dependent and independent variables. The model finally provides a sloped straight line.

The sloped line represents the relationship between the variables. The linear regression uses the Mean Squared Error cost function. The cost function is calculated by taking the average of squared error occurred between the predicted values and the actual values. The cost function determines how the regression fits for the set of observations. The process of finding the best model from the various models is called optimization.

3.7 POLYNOMIAL REGRESSION MODEL

The Polynomial regression model represents the relationship between the dependent and independent variables. The polynomial model makes use of linear regression model to fit the complicated and nonlinear functions and the dataset. The initial step in the polynomial regression model is data pre-processing. The data preprocessing is very similar to the linear regression model except the model will not use the future scaling method and will also not split the dataset into training and test set. The training and the test data set is not splitted because the model contains very less information and the model will not be able to find the correlations between the variables. The second step involved is building a linear regression model and trying to fit the model into the dataset. The third step is building a polynomial regression model and trying to fit the model into the dataset. Finally, the result is visualized for linear regression and polynomial regression model.

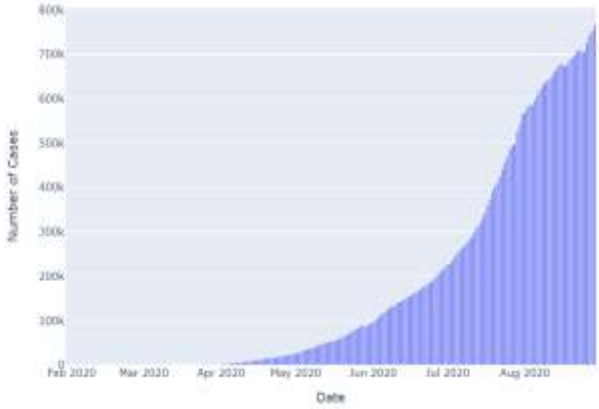
3.8 SUPPORT VECTOR MACHINE

Support vector machine is one of the most popular supervised learning algorithms. The support vector machine is mainly used for the classification and the regression model. The major goal of the model is to create a best line that can segregate the dimensional space into the classes. Once the data is segregated it can be easily put the new data points into the correct category of classes. The support vector machine algorithm chooses the extreme points that help in creating a hyper plane. The dimension of the hyper plane depends on the features. If there are only 2 features then the hyper plane will be a straight line. If there are three features then the hyper plane will be a 2 dimension plane. The hyper plane is always created with the maximum margin. Maximum margin means the maximum distance between the data points. Support vector machine is classified into two types they are linear and nonlinear svm. The linear svm model is mainly used for linearly separable data. Linearly separable data means the data set can be classified into two classes by using a single straight line. The nonlinear svm is mainly used for nonlinearly separable data. The dataset cannot be classified by using a straight line , then such data is termed as non-linear data and classifier used is called as Nonlinear svm Classifier.

RESULT AND DISCUSSION

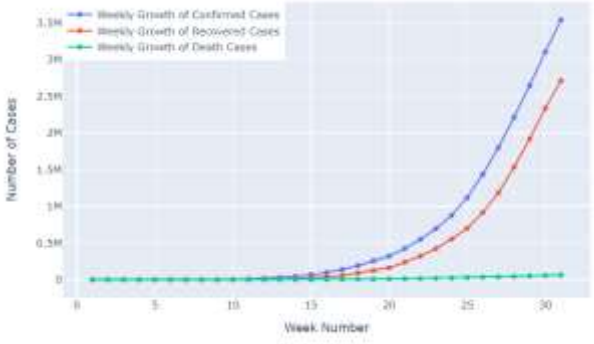
The following figure 4.1 represents the distribution of active cases in India. The active cases are calculated using the number of confirmed cases subtracted from the number of recovered cases subtracted from the number of death cases.

Figure 4.1 Active Cases in India



The following figure 4.2 represents the closed cases. The closed cases are calculated using the number of recovered cases added with the number of death cases. The graph is plotted against the number of cases vs week number.

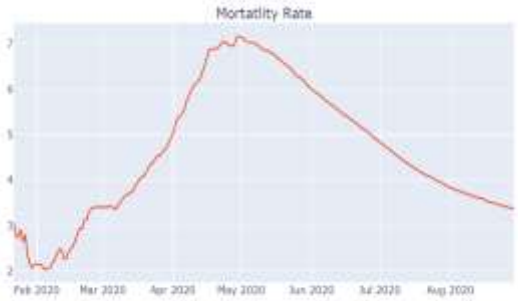
Figure 4.2 Closed Cases



The following figure 4.3 represents the mortality rate. The mortality rate is calculated using the number of death cases divided by the number of confirmed cases. The mortality rate is considerable for a long time which is a positive sign. The graph is plotted against the mortality rate vs the dates.

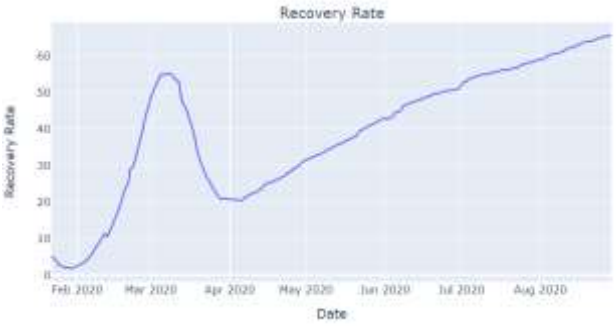
SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 4 | April 2022 - Peer Reviewed Journal

Figure 4.3 Mortality Rate



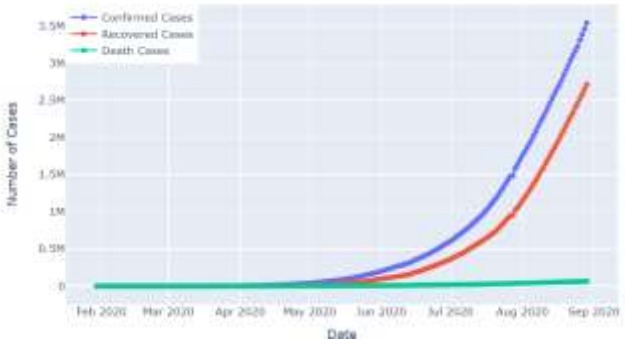
The following figure 4.4 represents the recovery rate. The recovery rate in the graph is plotted against the recovery rate and the dates in which the person has recovered. The recovery rate has picked up with a good sign.

Figure 4.4 Recovery Cases



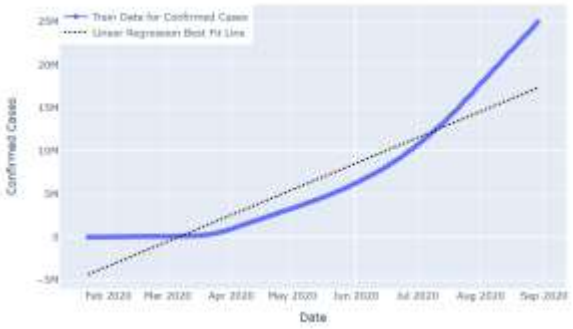
The following figure 4.5 represents the growth factor against the confirmed cases, recovered cases and death cases. The growth factor is calculated using the formula new confirmed, recovered and death cases divided using the previous confirmed, recovered and death cases.

Figure 4.5 Growth factor



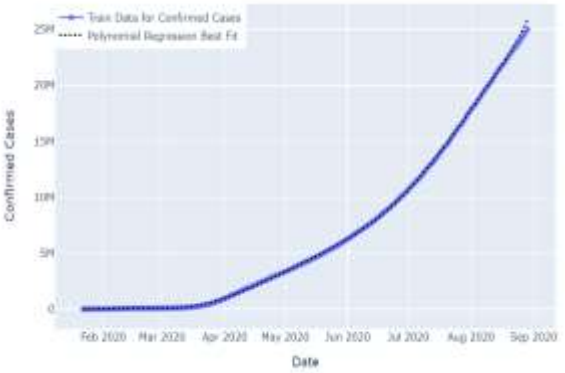
The following figure 4.6 represents the confirmed cases using the linear regression model. The linear regression model which predicted the output is falling apart. The graph is plotted against the confirmed cases and the dates. The linear regression model shows a clear view that the trend in the confirmed cases is absolutely not a linear model.

Figure 4.6 Confirmed Cases using Linear Regression



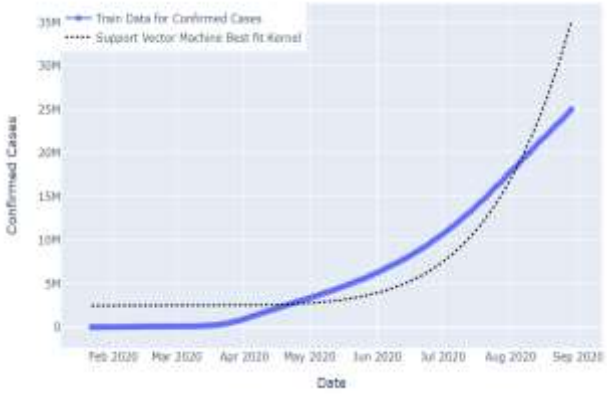
The figure 4.7 represents the confirmed cases using the polynomial regression model. The final output is compared using the linear regression model. Predict () method is used for the comparison of the linear and polynomial regression model. The predicted output for the polynomial regression is much closer to the real value.

Figure 4.7 Confirmed Cases using the Polynomial Regression



The following figure 4.8 represents the confirmed cases using the support vector prediction model. The support vector model has not provided a better result. The output that is predicted is either very high or very low than what expected.

Figure 4.8 Confirmed cases using the Support vector Machine



SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 4 | April 2022 - Peer Reviewed Journal

V. CONCLUSION

In this paper, we have provided a simple statistical analysis of the novel Coronavirus (COVID-19) outbreak in India. The novel coronavirus (COVID-19) that was first reported at the end of 2019 has impacted almost every aspect of life as we know it. This paper focuses on the incidence of the disease in India Using three simple machine learning algorithms—the linear regression model, polynomial regression and Support Vector machine, we model the daily and cumulative incidence of COVID-19 in the country during the early stage of the outbreak, and compute estimates for basic measures of the infectiousness of the disease including the active cases, death cases, growth rate, mortality rate and recovery rate. The predictive ability of the polynomial regression model was found to give a better fit and simple estimates of the daily incidence.

REFERENCES

1. *Akele R., 2020. Stochastic modelling for predicting COVID-19 prevalence in East Africa Countries. Infectious Disease Modelling, 5, pp. 598–607. pmid:32838091.*
2. *Alboaneen D., Pranggono B., Alshammari D., Alqahtani N. and Alyaffer R., 2020. Predicting the Epidemiological Outbreak of the Coronavirus Disease 2019 (COVID-19) in Saudi Arabia. International Journal of Environmental Research and Public Health, 17, 4568. pmid:32630363.*
3. *Arora P., Kumar H. and Panigrahi B.K., 2020. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. Chaos, Solitons & Fractals, 139, 110017. pmid:32572310*
4. *Atkeson, A., 2020. What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. National Bureau of Economic Research, Working Paper 26867.*
5. *Benatia, D., Godefroy, R. and Lewis, J., 2020. Estimating COVID-19 Prevalence in the United States: A Sample Selection Model.*
6. *Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), 2020. Coronavirus COVID-19 (2019-nCoV).*
7. *Garba S.M., Lubuma J.M.-S. and Tsanou B., 2020. Modeling the transmission dynamics of the COVID-19 Pandemic in South Africa. Mathematical Biosciences, 328, 108441. pmid:32763338*
8. *Ghysels E., 2016. Macroeconomics and the reality of mixed frequency data. Journal of Econometrics, 193, pp. 294–314.*
9. *Kuzin V., Marcellino M. and Schumacher C., 2011. MIDAS vs. mixed-frequency VAR—Nowcasting GDP in the euro area. International Journal of Forecasting, 27, pp. 529–542.*
10. *Li Q., Guan X., Wu P., Wang X., Zhou L., Tong Y., et al. 2020. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. The New England Journal of Medicine, 382, pp. 1199–1207. pmid:31995857*
11. *Mizumoto K., Kagaya K., Zarebski A. and Chowell G., 2020. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Eurosurveillance, 25, 2000180. pmid:32183930*
12. *Mollalo A., Vahedi B. and Rivera K.M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Science of The Total Environment, 728, 138884. pmid:32335404*
13. *Ogundokun R.O., Lukman A.F., Kibria G.B.M., Awotunde J.B. and Aladeitan B.B., 2020. Predictive modelling of COVID-19 confirmed cases in Nigeria. Infectious Disease Modelling, 5, pp. 543–548. pmid:32835145.*
14. *Omori R., Mizumoto K. and Nishiura H., 2020. Ascertainment rate of novel coronavirus disease (COVID-19) in Japan. International Journal of Infectious Diseases, 96, pp. 673–675. pmid:32389846*
15. *Park H. and Kim S.H., 2020. A Study on Herd Immunity of COVID-19 in South Korea: Using a Stochastic Economic-Epidemiological Model. Environmental and Resource Economics, 76, pp. 665–670.*
16. *Pham H., 2020. On Estimating the Number of Deaths Related to Covid-19. Mathematics, 8, 655.*
17. *Sarkar K., Khajanchi S. and Nieto J.J., 2020. Modeling and forecasting the COVID-19 pandemic in India. Chaos, Solitons & Fractals, 139, 110049.*
18. *Shim E., Tariq A., Choi W., Lee Y. and Chowell G., 2020. Transmission potential and severity of COVID-19 in South Korea. International Journal of Infectious Diseases, 93, pp. 339–344. pmid:32198088*
19. *Taboe H.B., Salako K.V., Tison J.M., Ngonghala C.N. and Kakaï R.G., 2020. Predicting COVID-19 spread in the face of control measures in West Africa. Mathematical Biosciences, 328, 108431. pmid:32738248*
20. *Wang P., Zheng X., Li J. and Zhu B., 2020. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. Chaos, Solitons & Fractals, 139, 110058. pmid:32834611*
21. *Wu J.T., Leung K. and Leung G.M., 2020c. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet, 395, pp. 689–697.*
22. *Wu J.T., Leung K., Bushman M., Kishore N., Niehus R., de Salazar P.M., et al. 2020a. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine, 26, pp. 506–510.*
23. *Zhao S., Lin Q., Ran J., Musa S.S., Yang G., Wang W., et al. 2020a. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases, 92, pp. 214–217.*
24. *Zhao Z., Li X., Liu F., Zhu G., Ma C. and Wang L., 2020b. Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. Science of the Total Environment, 729, 138959.*
25. *Zhu N., Zhang D., Wang W., Li X., Yang B., Song J., et al. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. The New England Journal of Medicine, 382, pp. 727–733. pmid:31978945.*