Analyzing the effectiveness of parameters affecting the COVID-19 outbreak worldwide

Shahnawaz Ahmad[1], Shabana Mehfuz[1], and Javed Beg[2], Nadeem Ahmed Khan[3],
Neyara Radwan[4,5], K. Martin Sagayam[6]

[1]Department of Electrical Engineering, Jamia Millia Islamia (A Central University), New Delhi-110025, India
E-mail: shahnawaz98976@gmail.com
[2]Oracle, Noida, India.
[3]Department of Civil Engineering, Jamia Millia Islamia (A Central University), New Delhi-110025, India
[4]Faculty of Economics & Administration, King Abdulaziz University, Jeddah, Saudi Arabia
[5]Mechanical Department, Faculty of Engineering, Suez Canal University, Ismailia, Egypt
[6]Department of Electronic & Communication, Engineering Karunya Institute of Technology and Science, Coimbatore-641114, India

ABSTRACT

The COVID-19 pandemic has had a significant impact on spread of the coronavirus disease. Currently, valuation and analysis of COVID-19 is a trending topic in the research industry that has been explored deeply. Many researches have been done in the selected time frame so far that evaluated the damages within the given conditions through different approaches. In our Study by taking certain parameters, we have deduced the direct relationship between spread of COVID-19 and the probable factors. This will be going to help researchers and stakeholders of the healthcare system to understand the pandemic in a better way, which may lead to potential enhancements that can be done in this field to counter this pandemic in a smarter way.

Keywords: COVID-19, Pandemic, Factors, Stakeholders, Healthcare System.

1. INTRODUCTION

In scientific analysis, the key role is played by data visualization or information visualization. Since the analysis of Cholera epidemic by John Snow to till date, inherent trends in data are represented by a good visualization that are not possible to visualize from raw numbers alone. Recently, data scientists face new challenges, possessed by COVID-19 pandemic due to its vast and rapidly expansion and also due to its significant impact on economy. 126,389,672, the reported cases and 2,771,260, the global deaths of COVID-19 as on March 28, 2021 [9]. Figure 1. Shows the top most 20 countries of COVID-19 Worldwide. We have used graphs to represent the ongoing spread of the pandemic over various countries. The prime objective of our research work is to evaluate and analyze the data with the help of linear regression algorithm and using Python tools to delineate and bring out a result by comparing the COVID-19 outbreak in different countries around the world. The different important terminologies and parameters used in this work are explained below.

1.1 Normalization- The sole purpose of normalization is to alter the values of numeric columns available in the dataset to a common scale, without twisting the differences in the ranges of values. Not every data set in machine learning require normalization. It is only required when there are ranges in features. Normalize data means to transform dataset values into a common range like between 0 to 1. Results are better if all column values lie in the same range.

1.2 Body Mass Index (BMI)- BMI is basically a person's weight (kilograms) divided by the square of height (meters). Higher the value of BMI may indicate higher the value of body fatness. BMI can be used as a scale to categorized weight which may lead to critical health issues [9]. BMI is an easy-to-use scale based on tissue mass and height which is used to broadly categorize a person as underweight, normal weight, overweight, or obese. Commonly accepted ranges for BMI ranges are as: underweight (under 18.5 kg/m^2), normal weight (18.5 to 25), overweight (25 to 30), and

obese (over 30) [10]. The values of BMIs, other than the 20-25 range, have been associated with higher all-causes mortality, with the risk increasing with distance from this range [10].

1.3 Population density- The occupancy of number of people per square kilometer of area is termed as population density. It shows the population density per unit area or per unit volume. It is frequently applied to living organisms, especially to humans. It is a basic and important geographical term [11].

1.3 Death Percentage

Death Percentage = (No. of Death(s) / Total no. of Infected cases) * 100%

In simple terms, it is the likelihood of death after getting infected.

1.4 Infection Percentage

Infection Percentage = (Total infected cases/Total Population) *100 %

In simple terms, it is the percentage of people who got infected out of the total population.

*1.4 Median Age- "*Median age is referred the age that divides a given population into two numerically equally sized groups - that is, half the people are below this age and half are above. This single index is used to summarizes the age distribution of a population [12]".

1.5 Average Temperature of a Country Annually- "Average yearly temperature is calculated by taking the average of minimum and maximum daily temperatures in the country, averaged for the years 1961–1990, based on gridded climatologies from the Climatic Research Unit elaborated in 2011 [13]".

The target of the research study is-

- To analyse the dataset collected from WHO and from different Internet sources.
- To visualize the dataset using data visualization tools.
- To analyze the results obtained, to predict future outcomes relating to the dataset chosen.
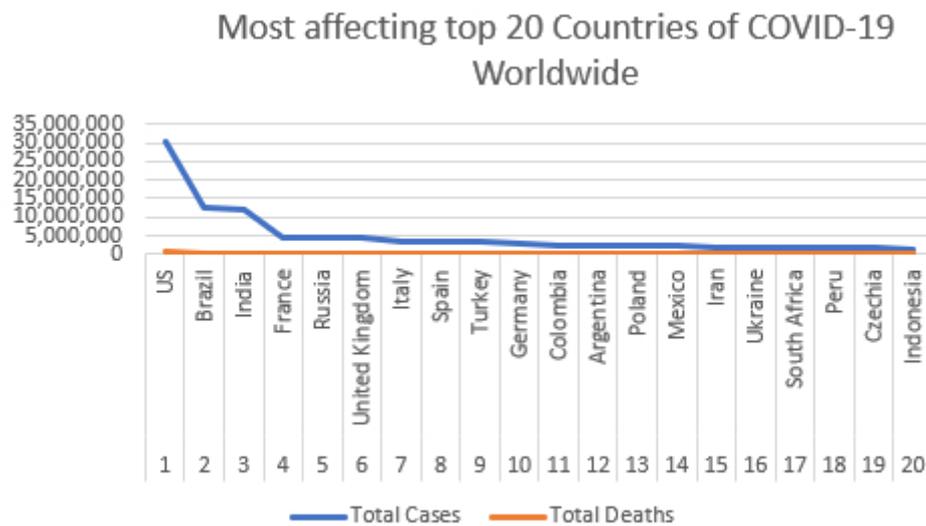

Figure 1. Most affecting top 20 countries of COVID-19 Worldwide [9]

The remaining part of this paper is systematized as follows. In section II, we have presented an insight into related works. Section III presents an understanding of gaps in literature review. Proposed methodology is given in Section IV. In section V results and interpretations are given. Finally, section VI shows the conclusion and future work.

2. RELATED WORK

People's courage and resilience can help to endure difficult situations like a global pandemic. This investigate teaches us that with world crises, comes the mindfulness among individuals through which it can be obstruct. Despite of many pandemics faced by the world; the COVID-19 extension is uncommon in history that produces it distinctive from others. Nowadays, estimation and analysis of COVID-19 is a much trending point that has been investigated profoundly within the investigate industryThere are a few who assessed the harms and conditions within the chosen time outline through diverse approaches. Anis Kourba [1], compared COVID-19 with other scourge episodes like "Ebola 2014", "MERS 2012", and "SARS 2003". In [1] comparison to January and February, "cumulative confirmed case", "death case", "recovered cases" rates are much higher in the month of March, 2020. He scrutinizes the responses, collected form internet search data, after the declaration of the primary COVID-19 case and appears that declarations had the greatest impact on looks for coronavirus, its indications, and hand sanitizer. Though, the searches for "coronavirus treatment", "testing", "isolation", "quarantine" and "coronavirus hoax" are not induced after the announcements.

Beyond this, individuals have come up with their claim clarifications and so-called arrangements to overcome this widespread with the assistance of innovation to expect future results. Harshad Khadilkar et al., [2], Villalobos Arias et al., [3], and Md. Rezaul Karim et al., [4] have induced innovation like "Artificial Intelligence (AI)" and "Machine Learning (ML)" to use as the control measures to moderate the effects of the COVID-19. Harshad Khadilkar et al. [2] have shown the evolution of infection rates under lockdown as well as in the absence of lockdown with the help of reinforcement learning and Villalobos Arias et al., [3] has used machine learning model and curve fitting to anticipate an increase in people infected with COVID-19 in particular areas while Md. Rezaul Karim et al., [4] studied about predicting the COVID-19 patients for the screening in the hospital with the help of CXR (Chest Xray) images and propose an AI-assisted application which incorporates Deep Neural Network (DNN) based on automatic detection of COVID-19 symptoms followed by highlighting class-discriminating regions using gradient guiding method.

Shahnawaz et al. [14-15] have proposed FCB COVID-19 DA points to recognize the patients as affirmed, suspects, or suspicious of COVID-19 and FCB COVID-19 DA is to require control of this widespread COVID-19 and deaccelerate its rate of transmission among the society.

Nanning Zheng et al. [5] have proposed a hybrid artificial-intelligence (AI) model for COVID-19 prediction. Experimental findings on epidemic data from several traditional Chinese provinces and cities indicate that people with coronavirus have a higher rate of infection within the third to eighth day after infection, which is more in line with the epidemic's actual transmission rules.

The model's prediction results are strongly consistent with real epidemic events, which shows that the proposed hybrid model can analyze the propagation law and development pattern of the virus more accurately relative to previous models, and relevant news can further boost the prediction model's accuracy. Furthermore, they have deduced an efficient tool for forecasting the law of transmission as well as the trend of development of potential public health events. Similarly, Palash Ghosh et al. [6] have developed three development models to foresee tainted individuals over the next 30 days. They have taken in to the account the "exponential", "logistic", and "SIS models", along with the daily rate of infection (DIR). They have analyzed the results collectively from all simulations rather than individually. They have hypothesized the DIR to be zero or negative in order to establish the fact that COVID-19 spread has been throttled. Even a slight positive DIR (say 0.01) indicates that the virus proliferates in the community. In terms of being able to pronounce a conclusion to the widespread, DIR must end up zero or negative for 14 days in a push. Also, Muhammad et al. [7] investigates the inter-relation of COVID-19's rapid expansion and regional climate parameters globally. They conclude that almost all countries having relatively lower temperature show a rapid expansion of COVID-19 cases in comparison to the countries having warmer climate irrespective to their socio-economic conditions. An observation has been made on the correlation between meteorological parameters and COVID-19 cases. Normal sunshine hours and add up to COVID-19 cases are related with a coefficient of assurance

of 0.42, whereas a relationship of 0.59 and 0.42 of normal high-temperature with add up to COVID-19 cases and passing cases are calculated individually.

Philippe Monmousseau et al. [8] investigates effects of travel restriction measures implemented in the course of the COVID-19 passenger pandemic to the U.S. air transport system. Based on the obtained data from passengers and airlines on social networking sites, four metrics are proposed to determine the impact of travel bans on the relation among travelers and aircrafts in near to real-time.

The proposed measurements appear that from a traveler viewpoint, each aircraft has responded in an unexpected way to the COVID-19 travel restriction measures. Therefore, these proposed metrics can be used by airlines and passengers in order to improve their decision-making process.

The proposed passenger-centric measurements were made utilizing Twitter information, due to its major advantage of being accessible in real-time, and at whatever point required, it can hence be effortlessly overhauled on an hourly premise. The proposed measurements have the included advantage of empowering each traveler and carrier to effectively impact the scores. It ought to be that as it may be emphasized here that the measurements degree basically the communication quality particularly between carriers and travelers by means of Twitter, and ought to subsequently still be complemented with conventional flight-centric measures for completeness.

The talk between government organizations, aircrafts and travelers, included advance, ought to be taken under consideration to assist tune the proposed measurements in arrange to meet the desires of all concerned parties.

3. GAPS IN LITERATURE REVIEW

After doing considerable analysis on the research papers related to data analysis and data visualization, various methodologies and work related to it has been widely accepted. But there exist some drawbacks that need to be addressed for better accuracy in the results. The following observations has been made like:

- In earlier studies only fever countries' data were analyzed
- For more in-depth analysis we can add more countries
- Larger dataset will help us to yield more accurate results

4. PROPOSED METHODOLOGY

This section presents the proposed methodology analyzing the effectiveness of parameters affecting the Covid-19 outbreak worldwide. The steps involved in the proposed method are as follows: (as shown by Figure. 2):

1. A dataset has been constructed by collecting data related to various parameters.
2. After selection of the dataset, we have cleaned the data using NumPy.
3. After cleaning up the data we have filtered out the outliers.
4. After removal of outliers, data has been normalized.
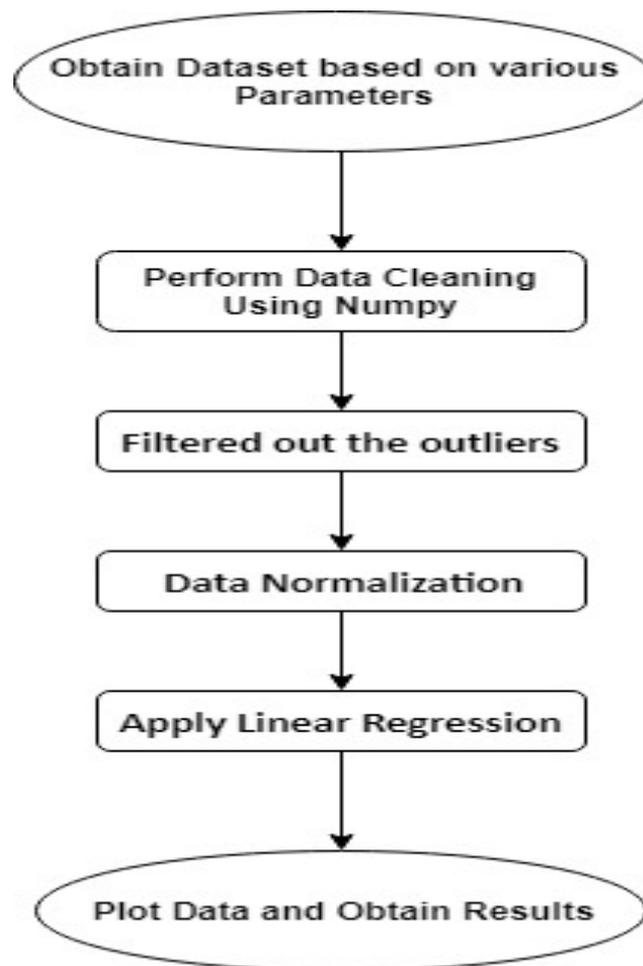5. And then finally we have applied the Linear Regression on the dataset, and have plotted the graph.

Figure 2. Proposed Methodology

5. RESULT ANALYSIS AND INTERPRETATION

In the proposed work, results have been obtained through meticulous process right from beginning to the end. Initially the dataset has been subjected to rigorous process of data cleaning. After obtaining the cleaned dataset, we removed the outliers and then normalized the data, to obtain more accurate results out of the dataset. The normalized dataset is then subjected to Linear Regression. Finally, the obtained results are plotted graphically. The experiment is carried out using Python language. Its libraries like Numpy are proved to be very useful in data cleaning.

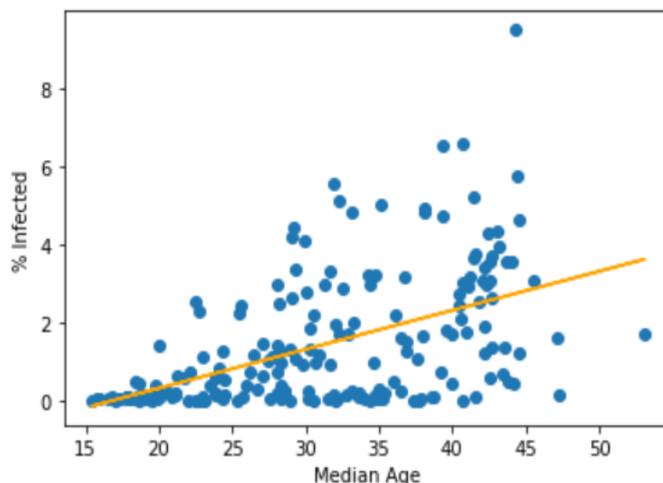The results obtained are shown in figures 3(a), 3(b), 3(c) and 3(d) respectively.

Figure 3 (a). Graph showing Median Age[]] Vs Infected Population Percentage (Country wise)

Andorra, San Marino, and Belgium all having greater median age 44.3, 44.4 and 41.3 years, respectively, have been subjected to higher infection percentages 9.48, 5.73 and 5.22, respectively. Similarly, Solomon Island, Marshall Island and Papua New Guinea all having lesser median age 22.5, 22.9, and 23.1 years, respectively, have witnessed lower infection percentages 0.002, 0.006 and 0.008, respectively. From the above results, we can deduce that countries having higher median age, have witnessed higher infection rate compared to countries having lower median age. This will help us to strategize better to fight COVID 19.
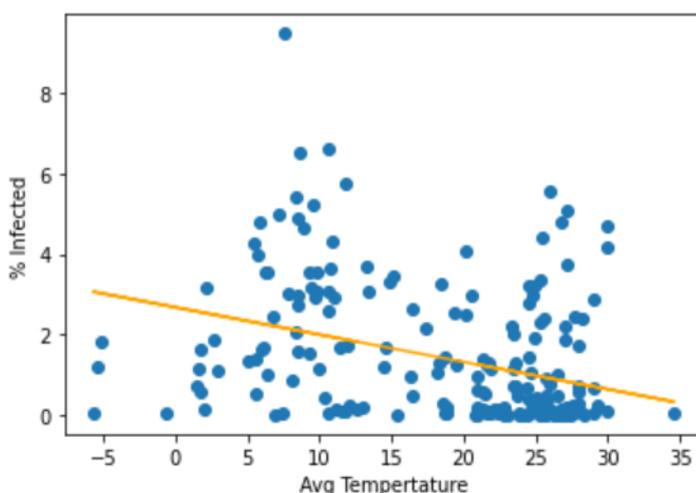

Figure 3 (b). Graph showing Average Yearly Temperature Vs Infected Population Percentage (country wise)

The infected population percentage in Bahrain which is 5.098, is way greater than Mongolia, which is 0.027. Bahrain average yearly temperature is very high (27.12 degree Celsius) as compared to Mongolia (-0.7 degree Celsius). Similarly, Puerto Rico whose yearly average Temperature (27.15 degree Celsius) is relatively very high as compared to Greenland (-5.7 degree Celsius), also witnessed a higher infection percentage (3.745) as compared to colder Greenland (0.033). It is very adamant from the above result that countries having average yearly temperature higher (hot weather countries), have suffered severely from COVID 19 and the infection percentage is much higher as

compared to countries having average temperature low (cold weather countries). This result will help us to formulate better plans to tackle COVID 19 in a more effective way.
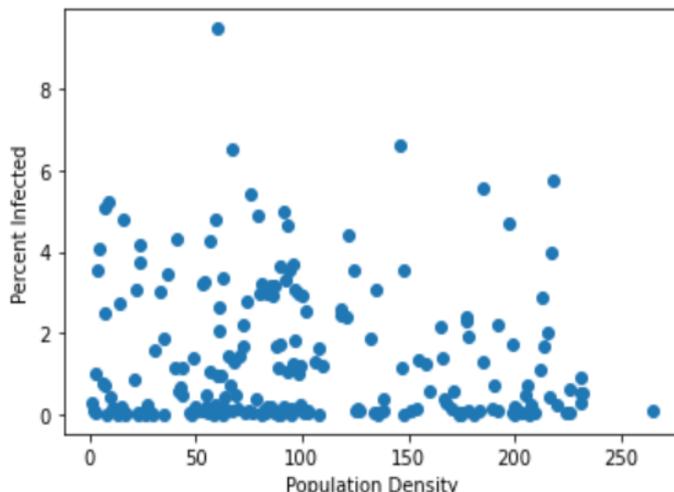

Figure 3 (c). Graph showing Population Density Vs Infected Population Percentage (country wise)

Indonesia Population Density is 151 person per sq km with a Population Density Rank 27, while Mauritania has a Population Density of 5 person per sq km and ranked 208th in the Population Density Country Wise list. These two countries have vastly different Population Densities but still their infection rates are almost similar at 0.224 and 0.226 percent respectively. Similarly, Nicaragua Population Density is 54 persons per square kilometer (Population Density Rank = 151) whereas Saint Vincent and The Grenadines has a Population Density of 284 persons per sq km (Population Density Rank = 265). These two countries have large differences in Population Densities but still their infection rates are almost similar at 0.087 and 0.088 percent respectively. This clearly shows that spread of infection is independent of Population Density. The graph does not show any clear-cut relationship between Population density of a country and the infection rate. On the contrary it has given mixed results. So, from the above result, we can say that the Infection Rate is not dependent on Population Density of a country.
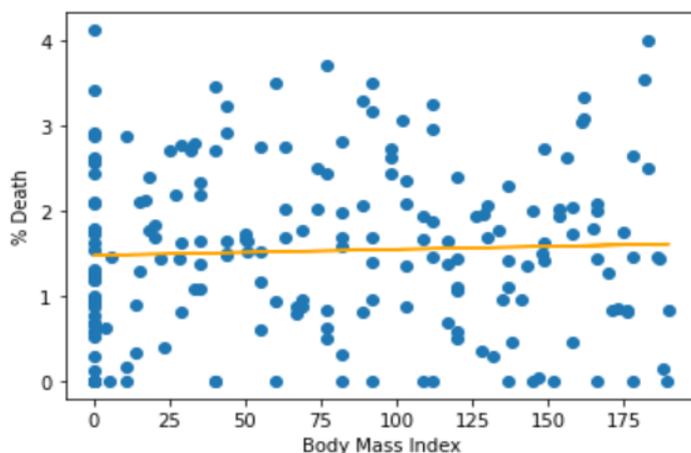

Figure 3 (d). Graph showing Body Mass Index Vs Death Percentage (country wise)

The death percent in Mozambique is 0.837 with a 171th rank in Body Mass Index Country wise list, whereas Israel has a much better Body Mass Index (Rank = 77) but they have also suffered with similar Death Percent 0.838. In the same way, Malaysia and Iceland have similar death percent 0.497 and 0.503, respectively. But former has BMI Rank 120th while latter one has 77th rank in BMI country wise ranking. There are large differences between the above countries in BMI ranking but they have shown a similar death percentage. The above graph shows random patterns. No relationship can be drawn on the basis of the above result between the Body Mass Index (BMI) of a country and the death rate. It has given ambiguous results. So, we can conclude that the Body Mass Index (BMI) of a person has no bearing on his likelihood of death from COVID-19.

5.1 Cases and Death counts include confirmed and probable- Recovered cases are estimates based on local media reports, and state and local reporting when available, and therefore may be substantially lower than the true number.
Incidence Rate = cases per 100,000 persons.
Case-Fatality Ratio (%) = Number recorded deaths / Number cases [9].

6. CONCLUSION AND FUTURE DIRECTION

This research paper analyzes and tries to deduce the relationship between probable parameters that may be affecting the Covid-19 outbreak worldwide. This will help researchers and stakeholders of the healthcare system to understand, plan, and counter the pandemic in a more efficacious way. We have chosen Median age, Average yearly temperature, Population density and Body Mass Index (BMI) as four parameters for the purpose of analysis. From the above results, we have observed that countries having colder weather and higher median age, have witnessed higher infection rate compared to countries having hotter weather and lower median age. On the other hand, the Body Mass Index (BMI) and Population density graph have shown random results. And do not establish any clear-cut relationship with the country's infection or death rate. So, we can conclude that the Infection Rate is not dependent on the country's Population Density and the Body Mass Index (BMI) of a person has no bearing on his likelihood of death from COVID-19. As a future work, we can analyze more probable parameters to obtain more holistic understanding of the COVID-19 worldwide spread.

REFERENCES

1. Anis Koubaa, "Understanding the COVID19 Outbreak: A Comparative Data Analytics and Study", https://www.riotu-lab.org/resources/covid19-analytics.pdf, March 2020.
2. Harshad Khadilkar, Tanuja Ganu, and Deva P. Seetharam, "Optimising Lockdown Policies for Epidemic Control using Reinforcement Learning", US National Library of Medicine, National Institutes of Health, PMCID: PMC7311597, June 2020.
3. Villalobos Arias and Mario Alberto, "Using generalized logistic regression to forecast population infected by COVID-19", https://arxiv.org/abs/2004.02406, April 2020.
4. Md. Rezaul Karim, et al., "DeepCOVIDExplainer: Explainable COVID-19 Predictions Based on Chest X-ray Images", https://creativecommons.org, April 2020.
5. Nanning Zheng, Shaoyi Du, Jianji Wang, He Zhang, Wenting Cui, Zijian Kang, Tao Yang, Bin Lou, Yuting Chi, Hong Long, Mei Ma, Qi Yuan, Shupei Zhang, Dong Zhang, Feng Ye, and Jingmin Xin "Predicting COVID-19 in China Using Hybrid AI Model", IEEE TRANSACTIONS ON CYBERNETICS, VOL. 50, NO. 7, JULY 2020.
6. Palash Ghosh, Rik Ghosh, and Bibhas Chakraborty, "COVID-19 in India State-wise Analysis and Prediction', JMIR Public Health and Surveillance doi: 10.2196/20341, April 2020.
7. Muhammad Mazhar Iqbal, Irfan Abid, Saddam Hussain, Naeem Shahzad, Muhammad Sohail Waqas, and Muhammad Jawed Iqbal, "The effects of regional climatic condition on the spread of COVID-19 at global scale", Sci Total Environ, Oct 2020.
8. Philippe Monmousseau, Aude Marzuoli, Eric Feron, Daniel Delahaye, "Impact of Covid-19 on passengers and airlines from passenger measurements: Managing customer satisfaction while putting the US Air Transportation System to sleep, Transportation Research Interdisciplinary Perspectives", Volume 7, 2020, 100179, ISSN 2590-1982, https://doi.org/10.1016/j.trip.2020.100179.
9. (https://www.sciencedirect.com/science/article/pii/S2590198220300907)https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

10. Zhe Xu, Lei Shi, Yijin Wang, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir Med. 2020. DOI: 10.1016/s2213-2600(20)30076-x.

11. Bernheim A, Mei X, Huang M, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. Radiology. 2020, 0(0): 200463.

12. Internet of Things-aided diagnosis and treatment of COVID-19 Chinese experts' group of Clinical of eHealth. Chinese experts' consensus on the Internet of Things-aided diagnosis and treatment of COVID-19. Fudan Univ J Med. 2020, 47(2): 151-160.

13. Li D. 5G and intelligence medicine-how the next generation of wireless technology will reconstruct healthcare? Precis Clin Med 2019; 2(4): 205-8.

14. Shanghai Respiratory Clinical Quality Control Center. Shanghai expert consensus for respiratory clinic quality control during epidemic 2019-nCoV time. Fudan Univ J Med. 2020, 47(2): 143-150.

15. Shahnawaz Ahmad, Shabana Mehfuz, Javed Beg, Nadeem Ahmad Khan, Afzal Husain Khan, Fuzzy cloud based COVID-19 diagnosis assistant for identifying affected cases globally using MCDM, Materials Today: Proceedings, 2021, ISSN 2214-7853, doi: https://doi.org/10.1016/j.matpr.2021.01.240.