**Abstract:**. Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. COVID-19 is caused by SARS- CoV-2. It has resulted in a worldwide pandemic that has put adults at risk of serious illness or death. The COVID-19 pandemic has challenged our worldwide healthcare resources. We have proposed analytical model to predict future COVID-19 cases using global COVID-19 data. The proposed model will work with pervious confirmed cases, deaths reported, recovered cases and daily cases. The proposed model can help us better prepare our future healthcare needs. (OUTCOMES)

**Key words:** COVID-19, predictive model, infectious disease, future cases

**Introduction:**

The World Health Organization declared COVID-19 outbreak an epidemic in March of 2020. COVID-19 is a respiratory disease caused by SARS-CoV-2. It causes illness similar to a cold, SARS(severe acute respiratory syndrome) and MERS(Middle East respiratory syndrome). The virus is thought to spread mainly from person to person through respiratory droplets produced when an infected person coughs, sneezes, or talks. Although some individuals will not have any symptoms, others may have mild to severe symptoms. Individuals with underlying conditions and adults over the age of 65 are at higher risk of severe illness.

The COVID-19 pandemic continues to pose threat to the wellbeing of individuals worldwide. As of May 21, 2021 there have been over 165,158,000 confirmed cases of COVID-19, including over 3,425,000 deaths, reported to WHO. As the number of COVID-19 cases grow, it continues to challenge our healthcare system. “International hospitals and healthcare facilities are facing catastrophic financial challenges related to the COVID-19 pandemic… Overall, a lack of preparedness was a major contributor to the struggles experienced by healthcare facilities around the world. Items such as personal protective equipment (PPE) for healthcare workers, hospital equipment, sanitizing supplies, toilet paper, and water were in short supply. These deficiencies were exposed by COVID-19 and have prompted healthcare organizations around the world to invent new essential plans for pandemic preparedness.”(Kaye, A. D., Okeagu)

In this paper, we propose an analytical model to predict future COVID-19 cases using global COVID-19 data. The proposed model will work with pervious confirmed cases, deaths reported, recovered cases and daily cases. The proposed model can help us better prepare our future healthcare needs. The unpredictive needs of the pandemic has resulted in a personal and economical loss. “The American Hospital Association estimates a financial impact of $202.6 billion in lost revenue for America's hospitals and healthcare systems, or an average of $50.7 billion per month.” (Kaye, A. D., Okeagu)

**Literature review:** Discuss how other researchers have addressed similar problems, what their achievements are, and what the advantage and drawbacks of each reviewed approach are. Explain how your investigation is similar or different to the state-of-the- art. Please cite the relevant papers where appropriate.

In Wuhan City, Hubei Province, China an outbreak of the coronavirus disease occurred. The World Health Organization(WHO) declared the outbreak as a public health emergency of international concern on January 30, 2020. Due to the recent nature of the pandemic, it must be noted that limited research is available. As the COVID-19 virus is still unknown to the world several studies have attempted to predict future COVID-19 outcomes.

On approach used in predicting COVID-19 cases is based on symptoms. Zoabi, Y., Deri-Rozov, S. & Shomron, N. established a machine-learning approach that model predicted COVID-19 test results with high accuracy using only eight binary features: sex, age ≥60 years, known contact with an infected individual, and the appearance of five initial clinical symptoms. The purpose of their study was to prioritize the use of testing resources on individuals who have a higher chance of testing positive for COVID-19 based on the model. The research found that “fever and cough were key to predicting contraction of the disease. As expected, close contact with an individual confirmed to have COVID-19 was also an important feature, thus corroborating the disease’s high transmissibility[15](https://www.nature.com/articles/s41746-020-00372-6#ref-CR15) and highlighting the importance of social distancing.” As noted in the research, there were several drawbacks. “We relied on the data reported by the Israeli Ministry of Health, which has limitations, biases and missing information regarding some of the features. For example, for patients labeled as having had contact with a person confirmed to have COVID-19, additional information such as the duration and location (indoors/outdoors) of the contact was not available.” A similar research done by Dr. Pallavi Mirajkar1, Dr. Rupali Dahake concluded that an analytical model can be employed to predict outbreak spreading trend are at high risk of developing Covid complications.

To better prepare our healthcare industry it would be beneficial to predict future COVID-19 cases. Hongwei Zhao et al. “describe a new approach that forecasts the number of incident cases in the near future given past occurrences using only a small number of assumptions” Using a Poisson distribution for the daily incidence number, and a gamma distribution for the series interval they modeled the observed incidence cases. They then estimated the effective reproduction number assuming its value stays constant during a short time interval and draw future incidence cases from their posterior distributions. Their model was focused on COVID-19 data available on Texas. Hongwei Zhao et al. stated “Our method produces reasonably accurate results when the effective reproduction number is distributed similarly in the future as in the past. Large deviations from the predicted results can imply that a change in policy or some other factors have occurred that have dramatically altered the disease transmission over time.” Some of the drawbacks of the study included complexity inherent in how data are collected. Hongwei Zhao et al. stated “Some major complexities of the data include: policies about testing algorithms (e.g. which suspect cases are tested); if screenings or surveillance is conducted, which diagnostic test is acceptable or required for reporting; accessibility and availability of testing; administrative issues such as reporting requirements, procedures, and infrastructure.” In our model we focus on global COVID-19 data. The proposed model will work with pervious confirmed cases, deaths reported, recovered cases and daily cases.

References:

Kaye, A. D., Okeagu, C. N., Pham, A. D., Silva, R. A., Hurley, J. J., Arron, B. L., Sarfraz, N., Lee, H. N., Ghali, G. E., Gamble, J. W., Liu, H., Urman, R. D., & Cornett, E. M. (2020). Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives. *Best Practice & Research. Clinical Anaesthesiology*, Advance online publication. <https://doi.org/10.1016/j.bpa.2020.11.009>

Pallavi Mirajkar, Rupali Dahake,, "Predictive System of COVID -19 Using Response Based Analytical Model", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456- 3307, Volume 7 Issue 2, pp. 05-10, March-April 2021.

Zhao H, Merchant NN, McNulty A, Radcliff TA, Cote MJ, et al. (2021) COVID-19: Short term prediction model using daily incidence data. PLOS ONE 16(4): e0250110.<https://doi.org/10.1371/journal.pone.0250110>

Zoabi, Y., Deri-Rozov, S. & Shomron, N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digit. Med.* **4,**3 (2021). https://doi.org/10.1038/s41746-020-00372-6

**Methodology:** Discuss the key aspects of your problem, data set and regression model(s). Given that you are working on real-world data, explain at a high-level your exploratory data analysis, how you prepared the data for regression modeling, your process for building regression models, and your model selection.

 **Experimentation and Results:** Describe the specifics of what you did (data exploration, data preparation, model building, model selection, model evaluation, etc.), and what you found out (statistical analyses, interpretation and discussion of the results, etc.).

 **Discussion and Conclusions:** Conclude your findings, limitations, and suggest areas for future work.

 **References:** Be sure to cite all references used in the report (APA format).

 **Appendices:**

 Supplemental tables and/or figures.

 R statistical programming code.