

## SURVEY ON ANALYSIS OF COVID-19 DATA USING MACHINE LEARNING ALGORITHMS

**Vinutha K<sup>\*1</sup>, Sushma T<sup>\*2</sup>, S Sriraksha<sup>\*3</sup>, Poornima RM<sup>\*4</sup>, Prajwal BR<sup>\*5</sup>**

<sup>\*1,2,4</sup>Asst. Prof., ACS College Of Engineering, Bangalore, India.

<sup>\*3,5</sup>Student, ACS College Of Engineering, Bangalore, India.

DOI : <https://www.doi.org/10.56726/IRJMETS34060>

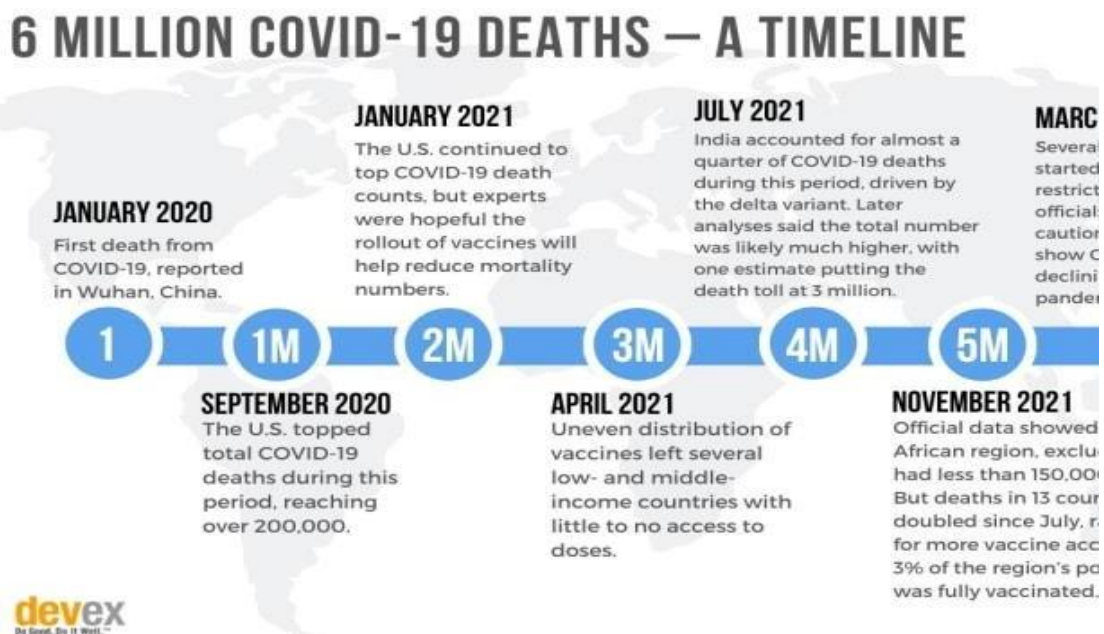
### ABSTRACT

With the spread of COVID-19 worldwide, people’s production and life have been significantly affected. It is very important to use data science and technology to help human in a timely manner to prevent and control the development of the pandemic, maintain social stability, and assess the impact of the pandemic. The outbreak of this novel corona virus has affected the world as a whole and caused millions of deaths. This remains the greatest pandemic in the world history. This paper aims to provide a better understanding of how various Machine Learning models can be implemented in real-world situations. With the help of datasets from the Ministry of Health and Family Welfare of India, this study puts forward various trends and patterns experienced in different parts of the world. The data to be studied has been obtained for 154 days i.e. from January 22, 2020, till July 30<sup>th</sup> 2020. For future references, the data can be further analyzed, and more results can be obtained.

**Keywords:** Covid-19, Machine Learning, Data Analysis, Trend Analysis.

### I. INTRODUCTION

According to the World Health Organization(WHO), viral and infectious diseases continue to appear and pose a serious threat to public health and well-being. Corona virus is a broad family of viruses which causes ailments ranging from common cold and flu to severe respiratory issues. According to NCBI, “In the last 20 years, there have been several viral epidemics that have been reported such as the Severe Acute Respiratory Syndrome Corona virus or better known as SARS- CoV which was declared a pandemic by WHO in 2002 - 2004 and H1N1 influenza in 2009.

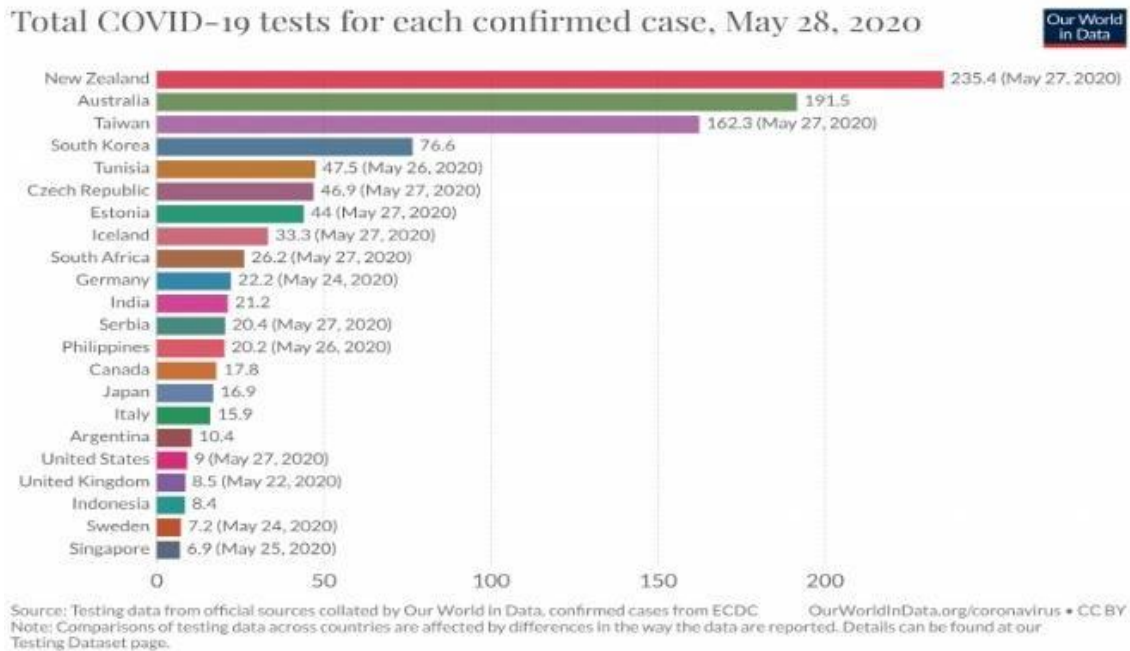


**Figure 1:** Timeline of the events of COVID-19 across different nations.

With most recently, Middle East Respiratory Syndrome Corona virus better known as MERS- CoV which hit its first outbreak in Saudi Arabia in 2012. Since this pandemic has affected the whole world not only in terms of health and hygiene but also in terms of the global economy. Apart from the adverse effects of COVID-19, there have been certain constructive influences around the world.

As the world was facing loses, our nature gained something from this pandemic, the harmful particulate matter was eliminated from the environment and most importantly the largest ever ozone hole detected was closed during this pandemic.

The timeline of the events of COVID-19 across different nations is shown in Figure 1. and the percentage of confirmed cases per country is shown by Figure 2.



**Figure 2:** The percentage of confirmed cases per country

## II. RELATED WORK

- It propose a mathematical model for the scenario of community spread of COVID19 (i.e., Stage-III). An approximate prediction for the spread of disease in coming days can be performed by using this model.
- The article describe modelling efforts for evaluating the current level of COVID-19 infections in India, using exponential model. The developed model is based on daily confirmed infections and not on cumulative infections and rationalization is carried out for the population of various regions, while predicting infections for various states.
- This Article suggests a simple and operable decision rule to quickly predict patients at the highest risk, allowing them to be prioritized and potentially reducing the mortality rate.
- This article proposes to utilize the machine learning and deep learning models with the aim for understanding its everyday exponential behavior along with the prediction of future reach ability of the COVID-2019 across the nations by utilizing the real-time information from the Johns Hopkins dashboard.
- This article focus the infected people in each state (restricting to only those states with enough data for prediction) and build three growth models to predict infected people for that state in the next 30 days. The impact of preventive measures on daily infected-rate is discussed for each state.

## III. PROBLEM STATEMENT

The worldwide pandemic of COVID-19 has led to a dangerous and rapidly spreading virus. People have been attacked because of this virus, and the virus is spreading rapidly in WHO designated countries.

To help simplify the daily activities related to COVID-19 data management, a project has been initiated to track cases, cured and death.

The main aim of this project is to make it easier for people to access and manage COVID-19 data automatically.

#### IV. METHODOLOGY

Methodology consists of four sections mainly A, B, C and D where A discusses on Data Acquisition, B describes the Data cleaning, C discusses on Data visualization and D discusses about the Data modeling.

##### A. Data Acquisition

In simple words, Data Acquisition is composed of two words: Data and Acquisition, where data is the raw facts and figures, which could be structured and unstructured and acquisition means acquiring data for the given task at hand.

##### B. Data Cleaning

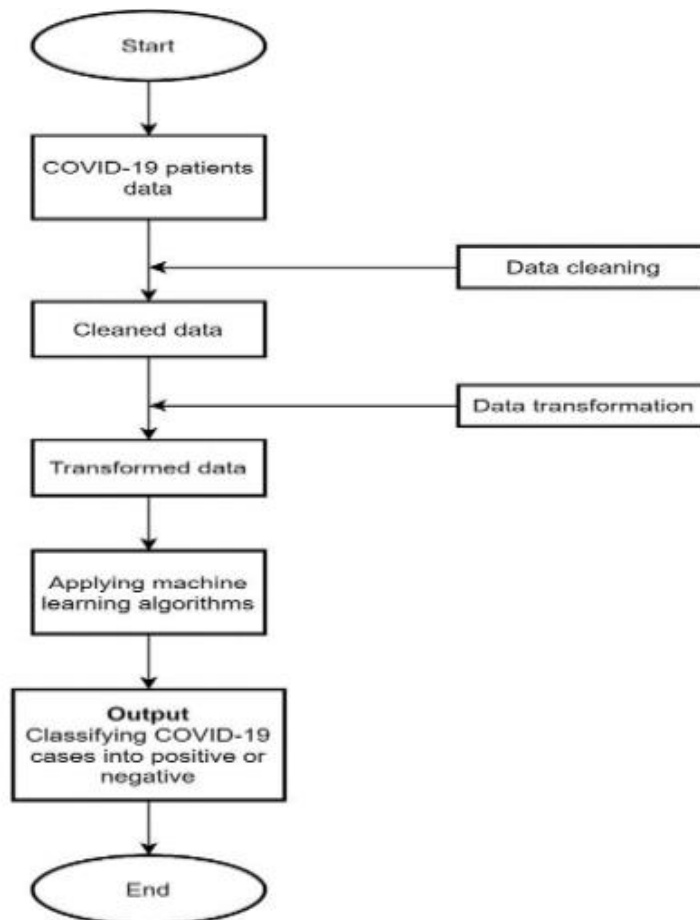
Data cleaning is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted. But, as we mentioned above, it isn't as simple as organizing some rows or erasing information to make space for new data. Data cleaning is a lot of muscle work.

##### C. Data Visualization

Data visualization is the graphical representation of information and data in a pictorial or graphical format (Example: charts, graphs, and maps). Data visualization tools provide an accessible way to see and understand trends, patterns in data, and outliers.

##### D. Data modeling

Data modeling is the process of creating a visual representation of either a whole information system or parts of it to communicate connections between data points and structures. this provides a common, consistent and predictable way of defining and managing data resources across an organization or even beyond.



**Figure 3:** data flow diagram for the covid-19 data analysis.

Data flow diagram (dfd) is a graphical representation of data flow in any system. it can illustrate incoming data flow, outgoing data flow and store data. data flow diagram describes anything about how data flows through the system.

The figure 3 represents the data flow diagram for the covid-19 data analysis.

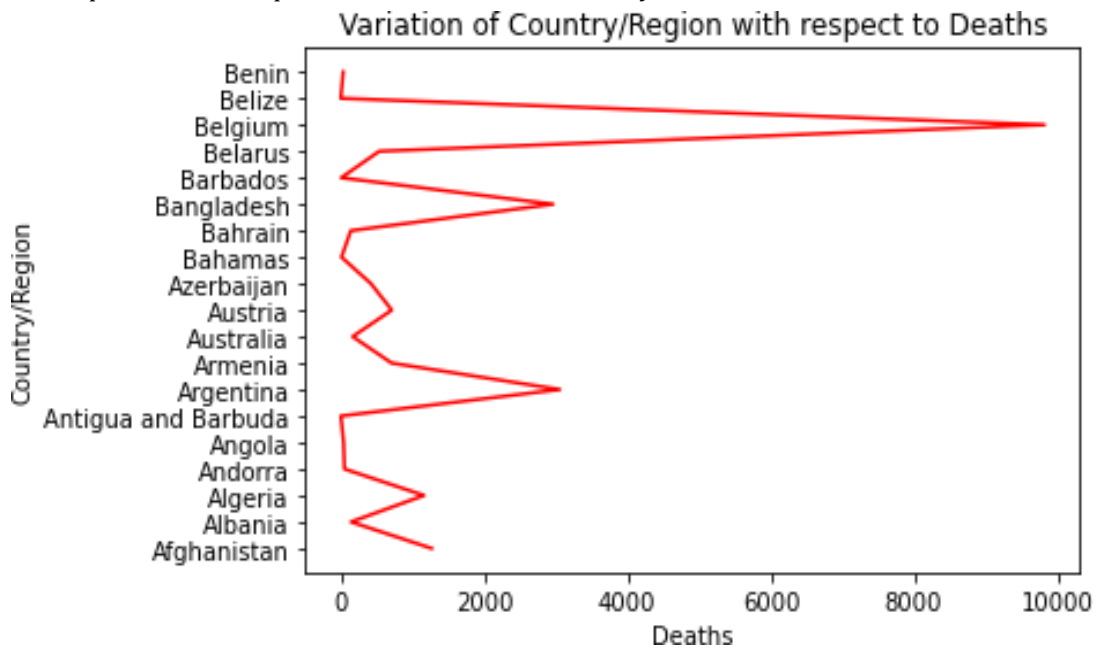
### V. RESULTS AND DISCUSSION

Through this paper the analysis on COVID-19 data has been performed successfully. This analysis on this pandemic spread has been done and compared between different countries. The analysis of confirmed cases, active cases, recovered cases, and deaths are done separately to give a clear look on how the virus is spreading, which countries are getting affected mostly and how different countries are recovering.

The above analysis has been for vaccines part also. Finally, the accurate check using different metrics is performed all over the analysis done in this project. The purpose of this study is to analyze the dataset and obtain important insights from it.

Coronavirus disease, (COVID-19) pandemic has quickly spread to various countries, with many cases having been reported worldwide. As of May 8th, 2020, 56,342 positive cases have been reported in India. India, with a population of more than 1.34 billion—the second largest population in the world—will have difficulty controlling the transmission of severe acute respiratory syndrome coronavirus 2 among its population.

The Figure 4 represents the snapshots of COVID-19 cases country-wise scenario.



**Figure 4:** COVID-19 country-wise scenario.

COVID-19 has had a global impact on every aspect of life, from living and interacting with each other to work, travel and communication. The repercussions of this pandemic are expected to last for years. Governments need access to the best information when making decisions, and the international statistics community is collaborating with national statistical offices and systems to provide high quality data and statistics during and after the crisis.

The Figure 5 represents the snapshots of COVID-19 cases world-wide scenario.

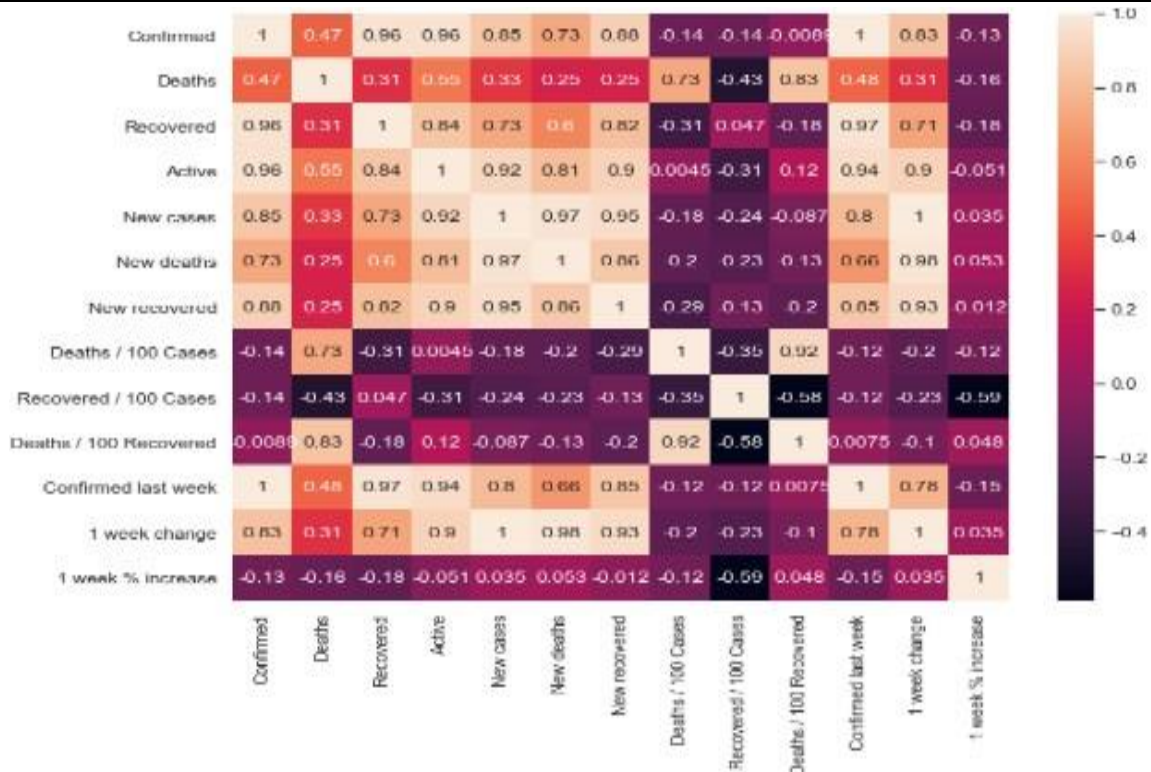


Figure 5: COVID-19 world-wide scenario.

## VI. CONCLUSION

This paper successfully analyzed the current trend of the transmission of Covid-19 in the world. Anticipating the further spread of the Covid-19 or better known as Novel Corona virus will help in taking necessary actions to control the spread. The paper also presented a comprehensive study of the spread of the virus outbreak situation in India which will further help in taking necessary steps to manage the huge population of India.

## VII. FUTURE ENHANCEMENTS

Further, more attributes can be included in the study in order to add more accuracy during the process. Further investigation may involve the analysis on India dataset by predicting the number of cases in future and how the mortality rate varies with the rise in the number of cases. Hope this article contributes to the world's response to this epidemic and puts forward some references for further research in future.

## VIII. REFERENCES

- [1] Arti. M. K, K. Bhatnagar, "Modeling And Predictions For Covid 19 Spread In India", Available: DOI: 10.13140/RG.2.2.11427.81444, PUBLISHED: APRIL (2020).
- [2] H. Shekhar, "Prediction Of Spreads Of Covid-19 In India From Current Trend", Medrxiv, Available: DOI: HTTPS://DOI.ORG/10.1101/2020.05.01.20087460, MAY 06, (2020).
- [3] Yan, L., Zhang, H., Goncalves, J. Et Al. "An Interpretable Mortality Prediction Model For Covid-19 Patients". Nature Machine Intelligence 2, Available: HTTPS://DOI.ORG/10.1038/S42256-020-0180-7, PP. 283-288, PUBLISHED: 14 MAY (2020)
- [4] N. S Punn, S. K. Sonbhadra, S. Agarwal, "Covid-19 Epidemic Analysis Using Machine learning And Deep Learning Algorithms", Medrxiv, Available: DOI: HTTPS://DOI.ORG/10.1101/2020.04.08.20057679, JUNE 1 (2020).
- [5] P. Ghosh, R. Ghosh, B. Chakraborty, "Covid-19 In India: State-Wise Analysis And Prediction", Medrxiv, Available: DOI: HTTPS://DOI.ORG/10.1101/2020.04.24.20077792, MAY 19 (2020).
- [6] S. F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. R. Varkonyikoczy, U.Reuter, T. Rabczuk, P. M. Atkinson, "Covid-19 Outbreak Prediction With Machine Learning", Available At SSRN: HTTPS://SSRN.COM/ABSTRACT=3580188 OR HTTP://DX.DOI.ORG/10.2139/SSRN.3580188, APRIL 19 (2020).

- 
- [7] F. Petropoulos, S. Makridakis, "Forecasting The Novel Coronavirus Covid-19", Available: [HTTPS://DOI.ORG/10.1371/JOURNAL.PONE.0231236](https://doi.org/10.1371/journal.pone.0231236), Published: March 31 (2020). Proceedings Of The International Conference On Smart Electronics And Communication (Icosec 2020) Ieee Xplore Part Number: Cfp20v90-Art; Isbn: 978-1-7281-5461-9 70 Authorized Licensed Use Limited To: Ieee Xplore. Downloaded On December 21,2021 At 07:53:12 Utc From Ieee Xplore. Restrictions Apply.
- [8] Features, Evaluation And Treatment Coronavirus (Covid-19) . Available: [HTTPS://WWW.NCBI.NLM.NIH.GOV/BOOKS/NBK554776/](https://www.ncbi.nlm.nih.gov/books/NBK554776/), MAY 18 (2020).